Laura Machado Lara Carvalho

Investigação genética da obesidade sindrômica e desenvolvimento de modelos biológicos para estudo da síndrome de Xia-Gibbs

Genetic investigation of syndromic obesity and development of biological models to study the Xia-Gibbs syndrome

São Paulo - SP
Janeiro de 2023

EXEMPLAR CORRIGIDO

Laura Machado Lara Carvalho

Investigação genética da obesidade sindrômica e desenvolvimento de modelos biológicos para estudo da síndrome de Xia-Gibbs

Genetic investigation of syndromic obesity and development of biological models to study the Xia-Gibbs syndrome

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo como requisito parcial para a obtenção de Título de Doutora em Ciências Biológicas (Biologia Genética).

Orientadora: Dra. Carla Rosenberg

Coorientadores: Dra. Ana Cristina V. Krepischi
Dra. Celia P. Koiffmann
Dr. Tiago Antônio O. Mendes

São Paulo - SP
Janeiro de 2023

Ficha catalográfica elaborada pelo Serviço de Biblioteca do Instituto de Biociências da USP, com os dados fornecidos pela autora.

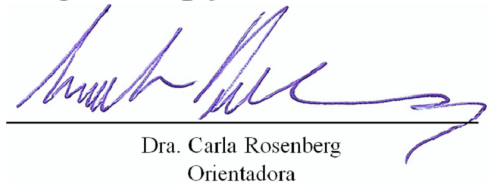
Machado Lara Carvalho, Laura.
Investigação genética da obesidade sindrômica e desenvolvimento de modelos biológicos para estudo da síndrome de Xia-Gibbs / Laura Machado Lara Carvalho; orientadora Carla Rosenberg; coorientadores Ana Cristina Victorino Krepischi, Celia Priszkulnik Koiffmann, Tiago Antônio de Oliveira Mendes -- São Paulo, 2023.

256 f.

Tese (Doutorado) – Instituto de Biociências da Universidade de São Paulo. Ciências Biológicas (Biologia Genética).

1. Obesidade Sindrômica. 2. Deficiência Intelectual. 3. Diagnóstico Genético; 4. *AHDC1*; 5. Síndrome de Xia-Gibbs; I. Rosenberg, Carla, orient. II. Cristina Victorino Krepischi, Ana, coorient. III. Koiffmann, Celia Priszkulnik, coorient. IV. Antônio de Oliveira Mendes, Tiago, coorient. V. Título.

Bibliotecária responsável pela catalogação: Elisabete da Cruz Neves. CRB - 8/6228.



Dra. Carla Rosenberg
Orientadora

Janus

Universidade de São Paulo

ATA DE DEFESA

Ata de defesa de Tese do(a) Senhor(a) Laura Machado Lara Carvalho no Programa: Ciências Biológicas (Biologia Genética), do(a) Instituto de Biociências da Universidade de São Paulo.

Aos 17 dias do mês de janeiro de 2023, no(a) Sala da Congregação, realizou-se a Defesa da Tese do(a) Senhor(a) Laura Machado Lara Carvalho, apresentada para a obtenção do título de Doutora intitulada:

"Investigação genética da obesidade sindrômica e desenvolvimento de modelos biológicos para estudo da síndrome de Xia-Gibbs"

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Nome dos Participantes da Banca	Função	Sigla da CPG	Resultado
Carla Rosenberg	Presidente	IB - USP	Não Votante
Maria Rita dos Santos e Passos Bueno	Titular	IB - USP	APROVADA
Alexander Augusto de Lima Jorge	Titular	FM - USP	Aprovado
Juliana Ferreira de Oliveira	Titular	Externo	Aprovada

Resultado Final: APROVADA

Parecer da Comissão Julgadora *

Eu, Helder Rossi Santos Souza _____, lavrei a presente ata, que assino juntamente com os(as) Senhores(as). São Paulo, aos 17 dias do mês de janeiro de 2023.

Maria Rita dos Santos e Passos Bueno

Alexander Augusto de Lima Jorge

participou remotamente
Juliana Ferreira de Oliveira

Carla Rosenberg
Presidente da Comissão Julgadora

* Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

A defesa foi homologada pela Comissão de Pós-Graduação em 19/01/23 e, portanto, o(a) aluno(a) _____ jus ao título de Doutora em Ciências obtido no Programa Ciências Biológicas (Biologia Genética) - Área de concentração: Biologia (Genética).

Presidente da Comissão de Pós-Graduação

Antonio Carlos Marques
Presidente da CPG do IB

Ao Felipe, meu amado companheiro;
ao tio Waldir (*in memorian*), meu pai do coração
que dizia que eu ia ser “doutora de verdade”;
e a todos os pacientes que participam das pesquisas do LGH.
Esta tese é dedicada a vocês, com todo o meu carinho.

Agradecimentos

Expresso toda a minha gratidão à Universidade de São Paulo, em especial ao Instituto de Biociências, ao Centro de Estudos do Genoma Humano e Células-tronco (CEGH-CEL) e ao Laboratório de Genética Humana (LGH) pela formação e experiências construtivas que me proporcionaram. Sinto-me honrada em poder conviver com seus pesquisadores excepcionalmente perspicazes e que desenvolvem trabalhos tão inspiradores.

Meu especial agradecimento às professoras Carla Rosenberg e Ana Krepischi, que me orientaram em todas as atividades do doutorado. Agradeço pela confiança e incentivo, por me oferecerem ótimas condições de trabalho e por terem me dado liberdade para definir alternativas metodológicas e procurar ajuda de outros pesquisadores para atender a questões específicas que o projeto demandou. Destaco ainda toda a elegância, serenidade e disposição das duas nas vezes em que eu me frustrei porque algo deu errado. Agradeço a oportunidade que me deram de poder batalhar pelo meu sonho de ser professora e pesquisadora e por, nesse sentido, serem para mim uma importante inspiração.

Agradeço aos meus coorientadores Celia Koiffmann e Tiago Mendes. À Profa. Celia por incluir seus pacientes neste projeto e me dar os primeiros ensinamentos sobre aconselhamento genético. E ao Prof. Tiago por contribuir com seus conhecimentos de ferramentas moleculares que até então não utilizávamos no LGH.

Foram valiosos os ensinamentos que os professores do Departamento de Genética e Biologia Evolutiva me proporcionaram, dentro e fora das salas de aula. A interlocução próxima e constante, sobretudo quando o contato presencial era possível, foi indispensável para me nortear. Sou grata por terem compartilhado comigo seu conhecimento e experiência. Em especial, destaco a Profa. Regina Celia Mingroni Netto, sempre com comentários muito pertinentes, e que também cedeu a estrutura do seu laboratório para os ensaios de Biologia Molecular.

Estendo os agradecimentos a todos os professores que participaram da minha formação e despertaram minha curiosidade. Pelos seus exemplos, minha pretensão é de ser uma eterna estudante, mas também ter o ensino e a investigação como ofício. Gostaria de expressar meu especial carinho pela Profa. Karina Braga, que foi minha

orientadora no mestrado e é uma constante incentivadora, pessoa inspiradora e que sempre me emociona profundamente.

Não posso deixar de agradecer à Natália Linhares, quem me sugeriu fazer o doutorado com a Carla Rosenberg e também me recomendou a ela por carta.

Agradeço aos colegas do LGH (Giovanna, Gustavo, Juliana S., Luiza e Sara) pela companhia e aprendizado coletivo, mas principalmente pelas conversas regadas a café no final da tarde. Nós convivemos em um período político e sanitário complicados e essa rede de apoio certamente foi muito importante para nos fortalecer.

No LGH tive a imensa sorte de poder participar de discussões sobre os projetos de pesquisa dos colegas, que têm metodologias e temas diversos, o que foi bastante relevante para a minha formação. Também agradeço ao Davi, bolsista de iniciação científica do LGH, que tem me ajudado a cuidar do nosso modelo animal.

De certo, é inevitável que um trabalho de mais de quatro anos, que envolve sequenciamento de nova geração, modelo animal e cultivo celular, tenha muitos colaboradores. Os procedimentos envolvidos nesta tese foram realizados por muitas mãos talentosas e mentes curiosas. Sou muito grata a todos os que generosamente contribuíram e, por isso, agradeço a cada um dos coautores dos capítulos que compõem esta tese. Agradeço em especial ao grupo da Profa. Maria Rita Passos-Bueno pela colaboração, apoio e acolhimento, mas principalmente à Elisa e ao Gerson, que compartilharam comigo conhecimentos sobre o trabalho com células-tronco e se empenharam em me treinar. Agradeço também à equipe do laboratório da Profa. Lygia da Veiga Pereira, sobretudo à Raquel e ao Fabiano, por toda a ajuda para concluir a caracterização das iPSC (*induced pluripotent stem cells*). Também agradeço ao pessoal do laboratório da Profa. Mariz Vainzof, onde eu provavelmente passei mais tempo que no LGH.

O Programa de Pós-graduação em Biologia Genética tem trabalhado arduamente por alcançar excelência nas várias atividades que desempenha. Também valorizo que esse programa dá uma ampla liberdade na escolha das disciplinas, o que me permitiu procurar em outros institutos por conteúdos relacionados aos meus temas de estudo ou mesmo aos motivos que despertavam minha curiosidade. Gostaria de mencionar também os seminários do CEPID CEGH-CEL e do Programa de Pós-Graduação, que sempre renderam discussões construtivas, e aproveito para agradecer aos colegas e professores que fizeram parte das comissões organizadoras do Curso de

Férias em Genética e Evolução, pois de 2020 até agora nós aprendemos, trabalhamos e nos divertimos muito juntos.

Tive a honra de participar de duas comissões da Universidade (Comissão de Pesquisa do IB e Comissão Coordenadora do Programa de Pós-graduação em Biologia Genética), que contribuíram no meu entendimento sobre questões institucionais, de fomento e de avaliação. Pude ver de perto o impacto negativo dos cortes em educação, ciência e tecnologia e acompanhei a aflição de colegas sem bolsa. Entendo que essas participações como representante discente poderão ser um diferencial na minha formação, visto que pretendo seguir carreira acadêmica e atuar na defesa do ensino público, gratuito e inclusivo, bem como defender os investimentos em ciência e tecnologia.

Durante os anos em que meu doutorado foi desenvolvido, assistimos no governo federal um projeto eleito democraticamente, porém avesso à democracia, que fez constantes ataques às universidades e à ciência, com cortes, constrangimentos e desqualificação ao trabalho dos cientistas. Também enfrentamos uma pandemia que levou ao adoecimento e até à morte de conhecidos, com números diários de óbitos absurdos no mundo inteiro. Havia ainda o agravante de um movimento antivacina em curso. O próprio presidente da república negava a necessidade de isolamento social e outras medidas sanitárias, minimizava a gravidade da situação e demonstrava insensibilidade ao adoecimento e luto de tantos brasileiros. Foram anos de angústia e incerteza, mas não esmorecemos, seguimos desempenhando nosso trabalho, nos manifestando e tentando convencer as pessoas que se deve confiar na ciência. Agradeço, portanto, a todos profissionais de saúde que trabalharam durante a pandemia e aos cientistas que desenvolveram tecnologias fundamentais para o controle da doença. Aproveito agradecer ainda ao grupo de trabalho para diagnóstico da Covid-19 do CEGH-CEL, que nos atendeu em momentos de suspeita de infecção, tendo sido fundamental para nos adequarmos na retomada do trabalho presencial.

Preciso reconhecer meu contexto privilegiado em termos de financiamento durante o doutorado, por ter sido agraciada com uma bolsa FAPESP - que conta ainda com uma reserva técnica que faz muita diferença - e estar em um laboratório de pesquisadoras vinculadas a um CEPID da mesma agência de fomento. O contexto da maioria dos laboratórios de pesquisa do país hoje em dia é bem mais complicado e por isso eu gostaria de cumprimentar a todos os cientistas brasileiros que seguem

trabalhando com afinco mesmo em tempo de “vacas magras”. Severo Ochoa, que recebeu em 1959 o Prêmio Nobel de Fisiologia e Medicina e que teve sua carreira algumas vezes impactada por guerras, crises e outras questões políticas, disse que “em princípio, a pesquisa precisa de mais cabeças do que meios” e, com tudo o que a ciência brasileira tem passado e mesmo assim resistido e produzido, acredito que ele estava certo. Ainda assim, por uma questão de sobrevivência, qualidade de vida e soberania, precisamos seguir na defesa de mais investimentos na ciência e valorização dos cientistas, pela construção um futuro melhor. Investimentos em ciência precisam ser encarados como uma das questões prioritárias do país.

Toda a minha gratidão pelo carinho e incentivo dos familiares que apoiam minha decisão de seguir carreira acadêmica. Vovó Neyde, tia Helena, tia Carminha, tia Dadá, tio Laurinho e tia Diana, vocês têm para mim importância existencial. E agradeço à vovó Stelita (*in memorian*) e à tia Dadá porque soube que custearam a maior parte dos meus estudos antes de eu ingressar na graduação. Agradeço ao tio Waldir (*in memorian*) por ter sido meu principal incentivador e maior inspiração, exemplo de honestidade intelectual e profissionalismo, uma alma inquieta e metódica que ainda me influencia.

Ao Felipe agradeço por tudo isso e mais um pouco. Pelo amor e pela compreensão, mesmo quando as células tinham que ser cuidadas diariamente ou os peixes precisavam de mim aos fins de semana. Por ter paciência quando eu não conseguia dormir pensando em algum experimento. Por colocar a “mão na massa” comigo e me ajudar a fazer experimentos funcionarem. Por me ajudar a pensar no plano A, no B e em todas as alternativas que sucederam. Por ser uma pessoa tranquila, tão diferente de mim, por ser o equilíbrio em contraste com a minha ansiedade.

Finalmente, agradeço àqueles que são a razão de todos os projetos desenvolvidos no LGH: os pacientes e suas famílias. No modelo que se costuma adotar no LGH, os pacientes muitas vezes são beneficiados com os resultados das pesquisas. Esses resultados também integram trabalhos acadêmicos que levam a benefícios coletivos, além do aprendizado dos alunos envolvidos. Pela experiência que me proporcionaram e por permitirem o uso dos dados nas publicações contidas nesta tese, eu os agradeço.

Ao rememorar as discussões e experimentos envolvidos nesta tese, percebo que são muitas as pessoas que participaram e que tive muita sorte em poder contar com elas. Em complemento a esta manifestação, dediquei ainda mais alguns agradecimentos ao final de artigos que integram esta tese. Adicionalmente, eu gostaria de agradecer à Mara Pinheiro e à Helenice Hirata, funcionárias de importância central e contato constante. Pelos eventuais esquecimentos de alguns nomes, eu peço antecipadas desculpas e asseguro não são intencionais.

Por óbvio, eximo as pessoas citadas de qualquer relação com as eventuais falhas e deficiências contidas nesta tese, que certamente são de responsabilidade exclusivamente minha.

Como o doutorado me rendeu muitas alegrias e realizações, eu preciso agradecer também por me proporcionarem o prazer que foi trabalhar com todos vocês. Espero que possamos continuar nos divertindo juntos em projetos futuros.

Apoio financeiro

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fundamental apoio financeiro:

- FAPESP, 2018/08486-3 – bolsa de doutorado (01/03/2019 a 31/12/2022);
- CAPES, 1805008 – bolsa de doutorado (01/08/2018 a 31/02/2019);
- FAPESP, 2013/08028-1 – CEPID Centro de Estudos do Genoma Humano e de Células-Tronco (01/07/2013 a 30/06/2024);
- CAPES, PROAP (Programa de Apoio à Pós-Graduação).

Resumo

INTRODUÇÃO: A obesidade sindrômica é definida pela associação de obesidade com uma ou mais das seguintes características: atraso de desenvolvimento, deficiência intelectual, traços dismórficos e malformações congênitas. A maior parte dos casos não tem sua origem genética elucidada, o que impacta o aconselhamento genético da família e o desenvolvimento de métodos terapêuticos havendo, portanto, a necessidade de se ampliar estudos clínicos e genéticos. Nesta tese, a obesidade sindrômica foi o tema de estudo, inicialmente com a investigação genética de uma casuística dessa condição. Posteriormente, identificamos nessa casuística, entre outros, um paciente afetado pela síndrome de Xia-Gibbs, uma forma sindrômica de deficiência intelectual causada por mutações em heterozigose em *AHDC1* que tem mecanismos fisiopatológicos pouco conhecidos e para a qual ainda não há associação reconhecida com obesidade. Decidimos então desenvolver modelos biológicos para seu estudo. **OBJETIVOS: (1)** Identificar variantes genéticas causais em uma coorte de 20 pacientes afetados por obesidade sindrômica que testaram negativo para as síndromes de Prader-Willi e X-Frágil, além de para os testes MLPA SALSA P064, P036 e P070; **(2)** desenvolver modelos biológicos (celular e animal) para estudo da síndrome de Xia-Gibbs. **MÉTODOS: (1)** Para casos com resultado negativo prévio em análise cromossômica por microarranjos (CMA), realizamos o sequenciamento de exomas (Agilent V5) dos pacientes e seus genitores (cinco trios). Para os demais, foi realizado sequenciamento de exomas dos trios (15 no total) com captura por sondas tanto para regiões codificadoras quanto para um *backbone* otimizado para avaliação de número de cópias de DNA (CNVs - *copy number variation*; OneSeq – Agilent). **(2)** Além do paciente com mutação de perda de função em *AHDC1* identificado na coorte de obesidade sindrômica, outros dois pacientes com síndrome de Xia-Gibbs foram recrutados. Para desenvolvimento de linhagens de células-tronco pluripotentes induzidas (iPSC - *induced pluripotent stem cells*), reprogramamos células dos três pacientes. Para desenvolvimento do modelo animal (*zebrafish*), editamos o gene *ahdc1* utilizando CRISPR-Cas9. **RESULTADOS: (1)** No grupo em que analisamos apenas SNVs e indels (por já ter diagnóstico negativo de CMA), tivemos uma taxa diagnóstica de 40% (2/5 – genes *MYT1L* e *GRIA3*); no grupo em que combinamos as análises de CNVs, SNVs e indels, a taxa diagnóstica foi de 47% (7/15 – genes

MED13L, *AHDC1*, *EHTM1*, *STAG2*, *KIAA0442*, *MEIS2 e SETD1A*). Dois desses pacientes (com variantes causais em *MYT1L* e *MED13L*) foram reportados também como relatos de casos, além do artigo referente à casuística total. **(2)** Obtivemos três linhagens de iPSC com diferentes mutações de perda de função em *AHDC1*. Essas células têm morfologia típica de iPSC humanas, apresentaram expressão de fatores de pluripotência (*SOX2*, *SSEA-4*, *OCT3/4* e *NANOG*) e capacidade de diferenciação nos três folhetos germinativos evidenciada pela expressão de *PAX6*, *HAND1* e *SOX17* nos corpos embrioides derivados *in vitro*. A linhagem de *zebrafish* obtida tem uma inserção de quatro pares de bases na região codificadora de *ahdc1*, é fértil e as frequências genotípicas de sua prole seguem proporções mendelianas. **CONCLUSÕES: (1)** As nove variantes causais identificadas em pacientes da coorte de obesidade sindrômica envolvem genes conhecidos de deficiência intelectual sindrômica; apenas dois desses genes (*MYT1L* e *EHMT1*) estão relacionados com síndromes em que a associação com obesidade é bem definida. Em investigações de casos de obesidade sindrômica, não se deve descartar a possibilidade de que a etiologia do quadro sindrômico possa ser independente da obesidade. Métodos de investigação abrangente do genoma (como microarranjos e sequenciamento de exoma/genoma) devem ser priorizados em detrimento de painéis de genes de obesidade sindrômica. **(2)** Os modelos biológicos que desenvolvemos são os primeiros para a síndrome de Xia-Gibbs e pretendemos futuramente utilizá-los para estudar sua ainda mal compreendida fisiopatologia, bem como para avaliar a nossa hipótese de que mecanismos neurais de fome e saciedade possam estar alterados nos pacientes. **CONTRIBUIÇÕES DESTA TESE:** Por meio do estudo genético de uma casuística de obesidade sindrômica, da apresentação de relatos de casos e de uma revisão abrangente sobre esse tema, esta tese contribuiu para a o diagnóstico e aconselhamento dos pacientes e suas famílias. Com o desenvolvimento de modelos para a síndrome de Xia-Gibbs, pretendemos investigar os mecanismos subjacentes às mutações em *AHDC1*, inclusive vislumbrando a proposição de alvos terapêuticos. Por fim, esta tese também envolve produções de divulgação científica que poderão ser utilizadas em ensino de genética, contribuindo para levar a ciência para além dos muros da nossa universidade.

Palavras-chave: obesidade sindrômica; deficiência intelectual; diagnóstico genético; *AHDC1*; síndrome de Xia-Gibbs; iPSC; *zebrafish*

Abstract

INTRODUCTION: Syndromic obesity is defined by the association of obesity with one or more of the following features: developmental delay, intellectual disability, dysmorphic traits, and congenital malformations. Most cases of syndromic obesity do not have their genetic origin elucidated, which impairs genetic counseling of the families and precludes the development of therapeutic methods; therefore, further clinical and genetic studies are required. In this thesis, syndromic obesity was the topic of study, with the initial genetic investigation of a cohort with this condition. We identify in this cohort a case of Xia-Gibbs syndrome, a syndromic form of intellectual disability caused by *AHDC1* heterozygous mutations, that has poorly understood pathophysiological mechanisms, and for which there is still no recognized association with obesity; we then decided to develop biological models for its study. **OBJECTIVES: (1)** To identify causal genetic variants in a cohort of 20 patients with syndromic obesity who were negative for Prader-Willi, Fragile X syndromes, and MLPA SALSA P064, P036 and P070 tests; **(2)** to develop biological models (cellular and animal) to study the Xia-Gibbs syndrome. **METHODS: (1)** For those cases with a previous negative result in chromosome microarray analyses (CMA), we performed exome sequencing (Agilent V5) of the patients and their parents (five trios). For the others, exome sequencing of the trios (total of 15) was performed with probe capture for coding regions and for an optimized backbone for evaluation of copy number variation (CNVs; OneSeq – Agilent). **(2)** In addition to the patient with the loss-of-function *AHDC1* variant identified in the syndromic obesity cohort, two other Xia-Gibbs patients were recruited. For the development of induced pluripotent stem cells (iPSC) lines, we used the reprogramming method of cells from the three patients. To develop the animal model (zebrafish), we edited the *ahdc1* gene using CRISPR-Cas9. **RESULTS: (1)** In the group in which we analyzed only SNVs and indels (because they were already negative for CMA), we had a diagnostic rate of 40% (2/5 – *MYT1L* and *GRIA3* genes); in the group in which we combined the analysis of CNVs, SNVs and indels, the diagnostic rate was 47% (7/15 – *MED13L*, *AHDC1*, *EHTM1*, *STAG2*, *KIAA0442*, *MEIS2* and *SETD1A* genes). Two of these patients (with *MYT1L* and *MED13L* causal variants) were also described as case reports, in addition to the paper referring to the complete cohort. **(2)** We obtained three iPSC lines with different loss-

of-function *AHDC1* variants. These cells have a human iPSC typical morphology, show expression of pluripotency factors (*SOX2*, *SSEA-4*, *OCT3/4* and *NANOG*) and have the ability to differentiate into the three germ layers – which is evidenced by the expression of *PAX6*, *HAND1* and *SOX17* by the *in vitro* derived embryoid bodies. The zebrafish line obtained has an insertion of four base pairs in the *ahdc1* coding region, it is fertile, and the genotypic frequencies of its offspring follow Mendelian proportions. **CONCLUSIONS: (1)** The nine causal variants identified involve known syndromic intellectual disability genes; only two of these genes (*MYT1L* and *EHMT1*) are related to syndromes in which the association with obesity is well defined. In syndromic obesity investigations, the possibility that the etiology of the syndromic condition may be independent of obesity should not be ruled out. Genome-wide investigation approaches (such as microarray and exome/genome sequencing) should be prioritized rather than gene panels focused on syndromic obesity. **(2)** We have developed the first biological models for Xia-Gibbs syndrome, which we will use to study its yet poorly understood pathophysiology, as well as to evaluate the hypothesis that the neural mechanisms for hunger and satiety are altered in patients. **CONTRIBUTIONS OF THIS THESIS:** The genetic study of a syndromic obesity cohort, the presentation of case reports and a comprehensive review about this topic presented in this thesis contributed to the diagnosis and genetic counseling of patients and their families. The development of biological models of Xia-Gibbs syndrome, should lead to the understanding of the mechanisms underlying *AHDC1* mutations and emerging therapeutic targets. Finally, this thesis also comprises articles meant to support genetic education, helping to take science beyond the walls of our university.

Keywords: syndromic obesity; intellectual disability; genetic diagnosis; *AHDC1*; Xia-Gibbs syndrome; iPSC; zebrafish

Índice

Introdução Geral.....16

Capítulo 1. Fundamentação teórica.....26

Capítulo 1.1 *A comprehensive review of syndromic forms of obesity: genetic etiology, clinical features and molecular diagnosis*.....27

Capítulo 1.2 Tecnologias de sequenciamento e microarranjos para diagnóstico de doenças congênitas.....74

Capítulo 1.3 Reprogramação celular para estudo de doenças genéticas.....96

Capítulo 1.4 *Zebrafish* como modelo para estudo de doenças genéticas.....105

Capítulo 2. Investigação genética de casos de obesidade sindrômica: artigo original e relatos de casos.....114

Capítulo 2.1. *Genetic investigation of syndromic forms of obesity* - https://doi.org/10.1038/s41366-022-01149-5.....115

Capítulo 2.2. *A novel MYT1L mutation in a boy with syndromic obesity: Case report and literature review* - https://doi.org/10.1016/j.orcp.2021.01.001.....131

Capítulo 2.3. *Two novel pathogenic variants in MED13L: one familial and one isolated case* - https://doi.org/10.1111/jir.12891.....141

Capítulo 2.4. *SCAF4-related syndromic intellectual disability* - https://doi.org/10.1002/ajmg.a.63032.....151

Capítulo 3. *Establishment of iPSC lines and zebrafish with loss-of-function AHDC1 variants: models for Xia-Gibbs syndrome*.....157

Capítulo 4. Artigos de educação em genética e divulgação científica.....190

Capítulo 4.1. Inteligência e deficiência intelectual: bases genéticas e fatores ambientais - https://doi.org/10.55838/1980-3540.ge.2020.332.....191

Capítulo 4.2. Modelos animais no estudo de doenças genéticas humanas - https://doi.org/10.55838/1980-3540.ge.2022.420.....200

Capítulo 4.3 CRISPR-Cas9: conceitos fundamentais e aplicações em saúde humana.....216

Discussão geral.....236

Anexos.....248

Anexo 1 - *SnapShot: Directed Differentiation of Pluripotent Stem Cells*.....249

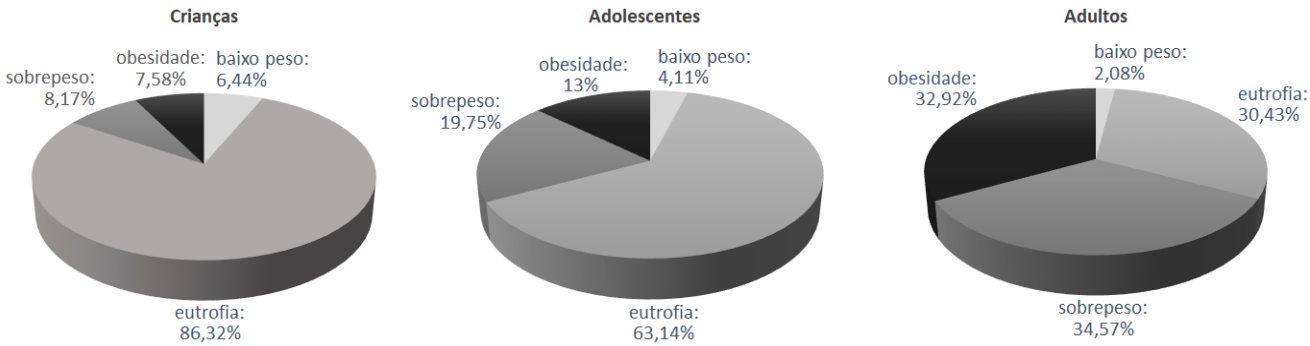
Anexo 2 – Dados dos pacientes com síndrome de Xia-Gibbs que tiveram células utilizadas para desenvolvimento de modelos de iPSC.....250

Anexo 3 - Outras produções durante o período do doutorado e distinções.....253

Introdução geral

A obesidade é um problema de saúde pública relevante em nível mundial (SOUSA; GUEDES; BENCHIMOL, 2022; WILLIAMS et al., 2015), associada a uma importante morbimortalidade por seus impactos no bem-estar físico (JUNG; LEE; SONG, 2017) e psíquico (PAN et al., 2018). No Brasil, de acordo com os dados de 2021 do SISVAN (Sistema de Vigilância Alimentar e Nutricional – Governo Federal), quase 33% dos adultos, 13% dos adolescentes e 7,6% das crianças (0 a 5 anos) atendidos nos serviços de atenção básica do Sistema Único de Saúde (SUS) têm obesidade (**Figura 1**).

Figura 1: Dados epidemiológicos de 2021 do SISVAN.



Tamanho amostral: 4.529.930 crianças (0 a 5 anos apenas), 4.241.617 adolescentes e 13.903.316 adultos. Para adultos, as categorias consideram o IMC. Para crianças e adolescentes considerou-se IMC x idade. Dados obtidos em sisaps.saude.gov.br/sisvan/

O índice de massa corporal (IMC) é amplamente utilizado para diagnóstico de obesidade. Se calcula utilizando a equação *peso (Kg) / altura (m)²* (OKORODUDU et al., 2010; SOUSA; GUEDES; BENCHIMOL, 2022). O IMC não avalia quantidade e distribuição de gordura corporal, pois um IMC elevado não necessariamente reflete um excesso de massa gorda (KIM; ALBANO; BERTOLA, 2022), mas ainda assim é o preditor mais utilizado internacionalmente (SOUSA; GUEDES; BENCHIMOL, 2022). A classificação da Organização Mundial de Saúde (OMS) de sobrepeso e níveis de obesidade é apresentada no **Quadro 1** e foi determinada com base no risco de complicações associadas à obesidade (WHO, 2000).

Quadro 1 – Classificação de risco de complicações com base no IMC de acordo com a OMS.

Classificação	IMC	Risco de complicações
Baixo peso	<18,5	Baixo
Peso normal	18,5-24,99	Médio
Sobrepeso	25,0-29,99	Levemente aumentado
Obesidade de classe I	30,0-34,99	Moderado
Obesidade de casse II	35,0-39,99	Alto
Obesidade de classe III	≥40	Muito alto

Fonte: WHO, 2000

No entanto, a classificação acima é considerada mais arbitrária para crianças e adolescentes por não se correlacionar robustamente com probabilidades de morbidade e mortalidade. Para crianças e adolescentes, o limite de normalidade é estabelecido por curvas de percentil do IMC, de acordo com idade e sexo. O diagnóstico de obesidade se dá quando o IMC está acima do percentil 95 (p95) ou z-IMC > +2 desvios padrão para idade e sexo. Considera-se sobrepeso quando o IMC está entre percentis 85 e 95 (P85-P95) ou z-IMC entre +1 e +2 (SOUSA; GUEDES; BENCHIMOL, 2022).

A obesidade se dá por um desequilíbrio positivo crônico entre ingestão calórica e gasto energético e, nesse sentido, há uma série de fatores comportamentais, sociais e genéticos com importante influência (ANGELUCCI; MANCINI, 2022). No entanto, nem todos os indivíduos ganham a mesma quantidade de peso quando expostos a dietas hipercalóricas. A obesidade tem, em geral, alta herdabilidade (CAMPBELL AM, 2017; SERRA-JUHÉ et al., 2017) e origem multifatorial, estando portanto relacionada a uma série de componentes genéticos e ambientais (ALBUQUERQUE et al., 2015; APALASAMY; MOHAMED, 2015; PETTERSSON et al., 2017; SERRA-JUHÉ et al., 2017). Estudos de gêmeos sugerem uma herdabilidade de massa corporal entre 40 e 70% (ANGELUCCI; MANCINI, 2022).

Além da obesidade comum (multifatorial), existem duas outras classes de obesidade: a monogênica não sindrômica e a sindrômica (ALBUQUERQUE et al., 2015). As variantes genéticas que determinam casos monogênicos não sindrômicos incluem principalmente mutações no gene da leptina - *LEP* (OMIM*164160) -, do receptor de leptina - *LEPR* (OMIM*601007) - e também do receptor de melanocortinas tipo 4 - *MC4R* (OMIM*155541) (PETTERSSON et al., 2017; SERRA-JUHÉ et al., 2017), que levam ao fenótipo de hiperfagia (KAUR et al.,

2017). Esses e outros genes relacionados a formas monogênicas de obesidade são apresentados no **Quadro 2**.

Quadro 2: Genes relacionados a formas monogênicas não sindrômicas de obesidade.

Gene	Proteína	Função
LEP	Leptina	Hormônio anorexígeno produzido por adipócitos
LEPR	Receptor da leptina	Receptor de membrana expresso em neurônios
POMC	Pró-opiomelanocortina	Neuropeptídio anorexígeno hipotalâmico
MC4R	Receptor de melanocortinas tipo 4	Receptor para produtos da POMC, tais quais: MSH-α e MSH-β
PCSK1	Proproteína convertase 1	Processa propeptidases (incluindo a POMC)

Fontes: Mancini et al., 2022; Nordang et al., 2017; Velloso & Araújo, 2022

A obesidade sindrômica – matéria deste estudo - é a associação de obesidade com fenótipos adicionais como deficiência intelectual, atraso no desenvolvimento, dismorfismos e malformações congênitas (D'ANGELO et al., 2018; HUVENNE et al., 2016; VUILLAUME et al., 2014). A ocorrência da obesidade sindrômica é rara (KAUR et al., 2017; MILANI et al., 2014) e há uma elevada variabilidade fenotípica (GEETS; MEUWISSEN; VAN HUL, 2019; KAUR et al., 2017). Prader-Willi e Bardet-Biedl são as mais conhecidas síndromes genéticas em que a obesidade é uma característica central (KAUR et al., 2017). Essas duas síndromes e outras são apresentadas no **Capítulo 1.1**, que tem como finalidade apresentar uma revisão narrativa ampla da etiologia genética, descrição clínica e diagnóstico molecular de formas sindrômicas de obesidade. A revista *Current Obesity Reports* nos convidou para um artigo de revisão sobre obesidade sindrômica e o **Capítulo 1.1** está sendo adaptado para submissão a essa revista.

A obesidade sindrômica costuma seguir padrões mendelianos de herança (KAUR et al., 2017). No entanto, a maioria dos casos não tem a etiologia genética desvendada (KAUR et al., 2017; VUILLAUME et al., 2014). Em uma revisão sistemática, Kaur e colaboradores (2017) identificaram publicações que descrevem 79 formas sindrômicas de obesidade. De acordo com a sua análise desses documentos, das 79 formas apenas 19 tiveram a origem genética elucidada satisfatoriamente. Assim sendo, esses autores defenderam a necessidade de se ampliar os estudos clínicos e genéticos envolvendo a obesidade sindrômica. Essas lacunas no conhecimento relativo às causas da obesidade sindrômica impactam no

aconselhamento genético e atrasam o desenvolvimento de possíveis métodos terapêuticos (KAUR et al., 2017).

A investigação sobre a etiologia da obesidade sindrômica é uma linha de pesquisa estabelecida há anos em nosso departamento e algumas publicações, especialmente relacionadas à investigação de alterações de número de cópias (CNVs - *Copy Number Variations*), são produtos desses estudos (D'ANGELO et al., 2010, 2014, 2013, 2018; D'ANGELO; KOIFFMANN, 2012). A Dra. Celia Koiffmann dedicou grande parte da sua carreira aos pacientes do ambulatório de Prader-Willi e Angelman do nosso instituto, coordenado por ela. Para os indivíduos com resultados negativos para essas hipóteses diagnósticas, a investigação genética podia continuar, em contexto de projetos de pesquisa, com o intuito de se avaliar diagnósticos diferenciais.

Como parte desta tese, foi investigada uma casuística de pacientes da Dra. Celia Koiffman com hipótese diagnóstica inicial de síndrome de Prader-Willi e resultados negativos para os seguintes testes genéticos: (1) teste de metilação da região de Prader-Willi; (2) cariótipo (bandas G); (3) MLPA SALSA P064, P036 e P070 (MRC-Holland); e (4) PCR para síndrome do X-Frágil (AmplideX PCR/CE FMR1–Asuragen). Para esses pacientes, realizamos sequenciamento de nova geração (NGS - *Next-Generation Sequencing*) com captura de exoma, para análise de *single nucleotide variants* (SNVs) e indels, com o intuito de identificar variantes causais. Para pacientes que ainda não tinham análise de CNVs por microarranjos, utilizamos também os dados de exoma para essa avaliação. Uma fundamentação teórica sobre tecnologias de sequenciamento e de microarranjos é apresentada no **Capítulo 1.2.** A metodologia em mais detalhes da investigação genética da casuística do ambulatório da Dra. Celia e respectivos resultados são apresentados no **Capítulo 2.1**, um artigo que publicamos recentemente na revista *International Journal of Obesity*.

Nos **Capítulos 2.2** e **2.3** são apresentados relatos de casos específicos da casuística do **Capítulo 2.1**. Decidimos publicar o relato de um caso envolvendo o gene *MYT1L* juntamente com uma revisão dos casos já descritos na literatura (**Capítulo 2.2**) pois, naquele momento, havia a necessidade de sumarizar as informações disponíveis e definir frequências aproximadas das características clínicas da síndrome causada por mutações nesse gene. Esse artigo foi publicado na revista *Obesity Research & Clinical Practice*.

Já o artigo sobre casos envolvendo o gene *MED13L* (**Capítulo 2.3**) foi motivado não apenas pela identificação de um paciente em nossa coorte, mas também pelo fato de o nosso laboratório ter previamente identificado em outra casuística uma família com dois afetados. Até então, todas as descrições na literatura eram de casos *de novo*, de maneira que consideramos interessante publicar o primeiro caso familial. Dessa forma, reunimos os dois casos diagnosticados em nosso laboratório (o familial da casuística de doutorado da Dra. Silvia Costa e o isolado da nossa casuística de obesidade sindrômica) em um artigo que foi publicado no *Journal of Intellectual Disability Research*.

O artigo envolvendo o gene *SCAF4* (**Capítulo 2.4**), publicado recentemente pela revista *American Journal of Medical Genetics: Parte A*, é a segunda publicação relacionando esse gene a uma doença mendeliana do neurodesenvolvimento. Embora o paciente não faça parte da casuística apresentada no **Capítulo 2.1**, pois foi investigado após o encerramento do estudo dessa coorte, esse paciente atendido pelo Prof. Paulo Otto apresenta obesidade sindrômica. Decidimos, portanto, apresentar sua descrição nesta tese.

Os avanços das técnicas moleculares e suas aplicações têm contribuído significativamente para entender a funcionalidade dos genes tanto em estado normal quanto patológico. O aprofundamento desses conhecimentos é relevante no aconselhamento genético, um campo de trabalho que busca auxiliar indivíduos e famílias a compreender as implicações médicas e reprodutivas decorrentes de patologias de cunho genético. São objetivos centrais do aconselhamento genético: definição e/ou confirmação do diagnóstico, comunicação à família, prognóstico, identificação de possibilidades terapêuticas atuais e informação sobre risco de ocorrência e recorrência (BERTOLLO et al., 2013).

Estudos genéticos de coorte podem levar à identificação de variantes patogênicas já descritas, revelar variantes novas ou mesmo em genes que ainda não foram relacionados a doenças. Quando há indícios que levam um gene a ser considerado candidato para determinada condição clínica, muitas vezes é preciso realizar estudos complementares para avaliar essa hipótese. Estudos funcionais envolvendo modelos celulares e animais são ferramentas importantes nesse tipo de avaliação, além de poderem ser úteis para elucidar mecanismos fisiopatológicos relacionados.

Em um dos pacientes da casuística apresentada no **Capítulo 2.1** foi identificada uma variante *nonsense* no gene *AHDC1*, o que levou ao diagnóstico molecular de síndrome de Xia-Gibbs, uma forma sindrômica de deficiência intelectual descrita pela primeira vez em 2014 (XIA et al., 2014). Por não existirem modelos celulares e animais bem estabelecidos para estudo dessa síndrome e por seus mecanismos fisiopatológicos ainda serem mal compreendidos, nos empenhamos no desenvolvimento de modelos biológicos. Recrutamos mais dois pacientes com mutações de perda de função em *AHDC1* e reprogramamos células adultas desses três indivíduos com a síndrome de Xia-Gibbs para obter iPSC (*induced pluripotent stem cells* – células pluripotentes induzidas); essas células são passíveis de serem diferenciadas em outros tipos celulares, como por exemplo neurônios, para investigação de mecanismos subjacentes às variantes causais. Nós também editamos o gene ortólogo em *zebrafish* (*ahdc1*) e obtivemos uma linhagem com inserção de quatro pares de bases, que é fértil e utilizaremos nos estudos da síndrome de Xia-Gibbs. O desenvolvimento desses dois modelos (celular e animal) é apresentado no **Capítulo 3**, que corresponde a um artigo submetido em dezembro de 2022 à revista *Gene*. Uma fundamentação teórica sobre esses dois tipos de modelos (iPSC e *zebrafish*) para estudo de doenças genéticas humanas é apresentada nos **Capítulos 1.3** e **1.4**, respectivamente.

A síndrome de Xia-Gibbs, no entanto, não é considerada uma forma sindrômica de obesidade e sim de deficiência intelectual. Nos chamou a atenção que os três pacientes que temos com essa síndrome apresentam uma ansiedade importante por comida. Dois deles têm obesidade. O paciente que não tem obesidade é submetido, segundo relato da genitora, a controle alimentar para não comer em excesso. Cabe ainda mencionar que o site da *Xia-Gibbs Society* (organização de pais - xia-gibbs.org/) tem uma parte dedicada a apresentar os gostos de alguns pacientes, chamada de "*our favorite things*" e, para muitos deles, "comer" - de maneira geral ou pratos específicos – está entre as três atividades favoritas descritas.

Como não há na literatura científica referências de que hiperfagia e obesidade poderiam ser características da síndrome de Xia-Gibbs e as fotos de pacientes presentes em relatos de casos publicados nos indicam que muitos não têm obesidade, nós entramos em contato com a *Xia-Gibbs Society* e com o Dr. Richard Gibbs (*Baylor College of Medicine*), quem descreveu essa síndrome. Ambos relataram que alguns

pacientes têm obsessão por comida e problemas de saciedade. Dessa forma, trabalhamos com a hipótese de que variantes em *AHDC1* poderiam interferir em vias de fome e saciedade, podendo causar quadros de obesidade sindrômica.

Alguns dos temas explorados no desenvolvimento do projeto que levou a esta tese foram abordados em três artigos de educação e divulgação científica. O primeiro desses artigos trata de fatores genéticos e ambientais que se relacionam com o neurodesenvolvimento, a inteligência e a deficiência intelectual (**Capítulo 4.1**). O segundo artigo discorre sobre o uso de modelos animais para estudo de doenças genéticas humanas (**Capítulo 4.2**). Já o terceiro artigo trata da técnica CRISPR-Cas9 (**Capítulo 4.3**), que tem sido uma revolução em diversas áreas da ciência e foi utilizada também no projeto que gerou esta tese.

Em suma, esta tese pretende contribuir com (1) novos conhecimentos sobre a investigação genética de formas sindrômicas de obesidade, (2) dois modelos biológicos para estudo da fisiopatologia da síndrome de Xia-Gibbs (iPSC e *zebrafish*) e (3) produções que poderão ser utilizadas em ensino de genética e divulgação científica. A metodologia dos trabalhos originais é descrita nos capítulos correspondentes e a bibliografia é apresentada ao final de cada capítulo.

Referências

ALBUQUERQUE, D.; STICE, E.; RODRÍGUEZ-LÓPEZ, R.; MANCO, L.; NÓBREGA, C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. **Molecular Genetics and Genomics**, v. 290, n. 4, p. 1191–1221, 2015. ISSN: 1617-4615. DOI: 10.1007/s00438-015-1015-9

ANGELUCCI, A. P.; MANCINI, M. C. Fisiopatologia da obesidade e da ciclicidade do peso. *In*: MANCINI, M. (org.). **Tratado de obesidade**. 3. ed. Rio de Janeiro: Guanabara Koogan, 2022. p. 51–62

APALASAMY, Y. D.; MOHAMED, Z. Obesity and genomics: role of technology in unraveling the complex genetic architecture of obesity. **Human Genetics**, v. 134, n. 4, p. 361–374, 2015. ISSN: 0340-6717. DOI: 10.1007/s00439-015-1533-x

BERTOLLO, E. M. G.; CASTRO, R.; CINTRA, M. T. R.; PAVARINO, É. C. O processo de Aconselhamento Genético. **Arq Ciênc Saúde** , v. 20, n. 1, p. 30–36, 2013

CAMPBELL, A. M.; Lesley V. Genetics of obesity. **Australian family physician**, v. 46, n. 7, p. 456–459, 2017. ISSN: 0300-8495. Disponível em:

https://www.racgp.org.au/afp/2017/july/genetics-of-obesity

D'ANGELO, C. S.; KOHL, I.; VARELA, M. C.; DE CASTRO, C. I. E.; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; KOIFFMANN, C. P. Extending the phenotype of monosomy 1p36 syndrome and mapping of a critical region for obesity and hyperphagia. **American Journal of Medical Genetics Part A**, v. 152A, n. 1, p. 102–110, 2010. ISSN: 15524825. DOI: 10.1002/ajmg.a.33160

D'ANGELO, C. S.; VARELA, M. C.; DE CASTRO, C. I. E.; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; PEREZ, A. B. A.; KOIFFMANN, C. P. Investigation of selected genomic deletions and duplications in a cohort of 338 patients presenting with syndromic obesity by multiplex ligation-dependent probe amplification using synthetic probes. **Molecular Cytogenetics**, v. 7, n. 1, p. 75, 2014. ISSN: 1755-8166. DOI: 10.1186/s13039-014-0075-6

D'ANGELO, C. S. et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. **Molecular Cytogenetics**, v. 11, n. 1, 2018. ISSN: 17558166. DOI: 10.1186/s13039-018-0363-7

D'ANGELO, C. S.; KOHL, I.; VARELA, M. C.; DE CASTRO, C. I. E.; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; PEREZ, A. B. A.; KOIFFMANN, C. P. Obesity with associated developmental delay and/or learning disability in patients exhibiting additional features: Report of novel pathogenic copy number variants. **American Journal of Medical Genetics Part A**, v. 161, n. 3, p. 479–486, 2013. ISSN: 15524825. DOI: 10.1002/ajmg.a.35761

D'ANGELO, C. S.; KOIFFMANN, C. P. Copy Number Variants in Obesity-Related Syndromes: Review and Perspectives on Novel Molecular Approaches. **Journal of Obesity**, v. 2012, p. 1–15, 2012. ISSN: 2090-0708. DOI: 10.1155/2012/845480

GEETS, E.; MEUWISSEN, M. E. C.; VAN HUL, W. Clinical, molecular genetics and therapeutic aspects of syndromic obesity. **Clinical Genetics**, v. 95, n. 1, p. 23–40, 2019. ISSN: 0009-9163. DOI: 10.1111/cge.13367

HUVENNE, H.; DUBERN, B.; CLÉMENT, K.; POITOU, C. Rare Genetic Forms of Obesity: Clinical Approach and Current Treatments in 2016. **Obesity Facts**, v. 9, n. 3, p. 158–173, 2016. ISSN: 1662-4025. DOI: 10.1159/000445061

JUNG, C. H.; LEE, W. J.; SONG, K. Metabolically healthy obesity: a friend or foe? **The Korean Journal of Internal Medicine**, v. 32, n. 4, p. 611–621, 2017. ISSN: 1226-3303. DOI: 10.3904/kjim.2016.259

KAUR, Y.; DE SOUZA, R. J.; GIBSON, W. T.; MEYRE, D. A systematic review of genetic syndromes with obesity. **Obesity Reviews**, v. 18, n. 6, p. 603–634, 2017. ISSN: 14677881. DOI: 10.1111/obr.12531

KIM, C. A.; ALBANO, L. M. J.; BERTOLA, D. R. Síndromes genéticas associadas à obesidade. *In*: MANCINI, M. C. (org.). **Tratado de obesidade**. 3. ed. Rio de Janeiro – RJ: Guanabara Koogan, 2022. p. 253–261

MANCINI, M.; FERNANDES, A.; MELO, M. E. Genética molecular da obesidade.

In: MANCINI, M. C. (org.). **Tratado de obesidade**. 3. ed. Rio de Janeiro: Guanabara Koogan, 2022. p. 94–101

MILANI, D.; CERUTTI, M.; PEZZANI, L.; MAFFEI, P.; MILAN, G.; ESPOSITO, Susanna. Syndromic obesity: clinical implications of a correct diagnosis. **Italian Journal of Pediatrics**, v. 40, n. 1, p. 33, 2014. ISSN: 1824-7288. DOI: 10.1186/1824-7288-40-33

NORDANG, G. B. N.; BUSK, O. L.; TVETEN, K.; HANEVIK, H. I.; FELL, A. K. M.; HJELMESAETH, J.; HOLLA, O. L.; HERTEL, J. K. Next-generation sequencing of the monogenic obesity genes LEP , LEPR , MC4R , PCSK1 and POMC in a Norwegian cohort of patients with morbid obesity and normal weight controls. **Molecular Genetics and Metabolism**, v. 121, n. 1, p. 51–56, 2017. ISSN: 10967192. DOI: 10.1016/j.ymgme.2017.03.007

OKORODUDU, D. O.; JUMEAN, M. F.; MONTORI, V. M.; ROMERO-CORRAL, A.; SOMERS, V. K.; ERWIN, P. J.; LOPEZ-JIMENEZ, F. Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis. **International Journal of Obesity**, v. 34, n. 5, p. 791–799, 2010. ISSN: 0307-0565. DOI: 10.1038/ijo.2010.5

PAN, L.; LI, X.; FENG, Y.; HONG, L.. Psychological assessment of children and adolescents with obesity. **Journal of International Medical Research**, v. 46, n. 1, p. 89–97, 2018. ISSN: 0300-0605. DOI: 10.1177/0300060517718733

PETTERSSON, M.; VILJAKAINEN, H.; LOID, P.; MUSTILA, T.; PEKKINEN, M.; ARMENIO, M.; ANDERSSON-ASSARSSON, J. C.; MÄKITIE, O.; LINDSTRAND, A. Copy Number Variants Are Enriched in Individuals With Early-Onset Obesity and Highlight Novel Pathogenic Pathways. **The Journal of Clinical Endocrinology & Metabolism**, v. 102, n. 8, p. 3029–3039, 2017. ISSN: 0021-972X. DOI: 10.1210/jc.2017-00565

SERRA-JUHÉ, C.; MARTOS-MORENO, G. Á.; BOU DE PIERI, F.; FLORES, R.; GONZÁLEZ, J. R.; RODRÍGUEZ-SANTIAGO, B.; ARGENTE, J.; PÉREZ-JURADO, L. A. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. **PLOS Genetics**, v. 13, n. 5, p. e1006657, 2017. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006657

SOUSA, L.; GUEDES, E. P.; BENCHIMOL, A. Definições antropométricas da obesidade. *In*: MANCINI, M. C. (org.). **Tratado de obesidade**. 3ª ed. Rio de Janeiro: Guanabara Koogan, 2022. p. 5–10

VELLOSO, L.; ARAÚJO, E. Regulação Central do Balanço Energético. *In*: MANCINI, M. (org.). **Tratado de obesidade**. 3. ed. Rio de Janeiro – RJ: Guanabara Koogan, 2022. p. 63–67

VUILLAUME, M. et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. **American Journal of Medical Genetics Part A**, v. 164, n. 8, p. 1965–1975, 2014. ISSN: 15524825. DOI: 10.1002/ajmg.a.36587

WHO. Obesity: preventing and managing the global epidemic. **WHO consultation 2000**, v. 0512-3054 (Print), n. 0512–3054, 2000. Disponível em: https://apps.who.int/iris/handle/10665/42330

WILLIAMS, E. P.; MESIDOR, M.; WINTERS, K.; DUBBERT, P. M.; WYATT, S. B. Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem. **Current Obesity Reports**, v. 4, n. 3, p. 363–370, 2015. ISSN: 2162-4968. DOI: 10.1007/s13679-015-0169-4

XIA, F. et al. De Novo Truncating Mutations in AHDC1 in Individuals with Syndromic Expressive Language Delay, Hypotonia, and Sleep Apnea. **The American Journal of Human Genetics**, v. 94, n. 5, p. 784–789, 2014. ISSN: 00029297. DOI: 10.1016/j.ajhg.2014.04.006

Capítulo 1
Fundamentação teórica

A finalidade deste primeiro capítulo é apresentar uma fundamentação teórica com base na literatura referente aos temas e técnicas centrais desta tese.

O **Capítulo 1.1** apresenta uma revisão narrativa a respeito da etiologia genética, descrição clínica e diagnóstico molecular de formas sindrômicas de obesidade. Esse manuscrito está em vias de submissão à revista *Current Obesity Reports*.

O **Capítulo 1.2** discorre sobre as tecnologias de sequenciamento e as baseadas em microarranjos, que foram utilizadas em produções contidas nesta tese.

No **Capítulo 1.3** é apresentada a técnica de reprogramação celular para obtenção de células pluripotentes induzidas (iPSC - *induced pluripotent stem cells*) e sua importância no desenvolvimento de modelos para estudo de doenças genéticas humanas.

O **Capítulo 1.4** trata do *zebrafish* como modelo para estudos funcionais de doenças genéticas humanas.

Capítulo 1.1

A comprehensive review of syndromic forms of obesity: genetic etiology, clinical features and molecular diagnosis

CARVALHO, L. M. L; JORGE, A. A. L.; BERTOLA, D. R.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. A comprehensive review of syndromic forms of obesity: genetic etiology, clinical features and molecular diagnosis.

(artigo a ser submetido à revista *Current Obesity Reports*)

A comprehensive review of syndromic forms of obesity: genetic etiology, clinical features and molecular diagnosis

Running title: Syndromic obesity review

Laura Machado Lara Carvalho[1], Alexander Augusto de Lima Jorge[2], Débora Romeo Bertola[1,3], Ana Cristina Victorino Krepischi[1], Carla Rosenberg[1*]

1 – Human Genome and Stem Cell Research Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo (USP), São Paulo, SP, Brazil.
2 – Genetic Endocrinology Unit, Cellular and Molecular Endocrinology Laboratory (LIM/25), Faculty of Medicine, University of São Paulo (USP), São Paulo, SP, Brazil.
3 - Genetics Unit of Instituto da Criança, Faculty of Medicine, University of São Paulo, São Paulo, SP, Brazil.

* Corresponding Author: Carla Rosenberg, Department of Genetics and Evolutionary Biology, Laboratory of Human Genetics – LGH, Institute of Biosciences, University of Sao Paulo, Matão Street 277 – Room 350, São Paulo, SP, Brazil. Zip code: 05508-090. Email: carlarosenberg@ib.usp.br

STATEMENTS AND DECLARATION

Competing interests statement: The authors declare that there are no conflicts of interest.

Funding: This study was supported by São Paulo Research Foundation - FAPESP (2018/08486-3 and 2013/08028-1), National Council for Scientific and Technological Development - CNPq (305806/2019-0 and 303294/2020-5) and Coordination for the Improvement of Higher Education Personnel - CAPES (1805008).

Author contributions: LMLC prepared the first version of the manuscript, and all other authors made relevant subsequent contributions.

ABSTRACT

Syndromic obesity is a generic clinical term that refers to obesity occurring with additional phenotypes, such as intellectual disability/developmental delay, dysmorphic features, and/or organ-specific developmental abnormalities. **Purpose of review:** To present a narrative review regarding the genetic etiology, clinical description, and molecular diagnosis of syndromic obesity, which is a rare condition with high phenotypic variability and genetic heterogeneity. Prader-Willi, Bardet-Biedl, Pseudohypoparathyroidism, Kleefstra, Börjeson‐Forssman‐Lehmann, Cohen, Alström, Carpenter (types 1 and 2), MORM, Temple, 16p11.2 and 1p36 deletions, *MYT1L*-related, and Luscan-Lumish syndromes are described in this review. **Summary:** The etiology of syndromic obesity tends to be linked to neurodevelopment (central) and is related to a diversity of genes and biological pathways. There are three main groups of mechanisms: imprinting disorders, related to genes direcly involved in regulation of transcriptional activity, and ciliopathies. In the genetic investigation of patients with syndromic obesity, the possibility that the etiology of the syndromic condition is independent of obesity should be considered. Methods of genome-wide investigation should be prioritized over the target sequencing of well-characterized syndromic obesity gene panels. The correct genetic diagnosis impacts medical management, treatment, and prognosis, and allows for proper genetic counseling.

Keywords: syndromic obesity; genetic etiology; molecular diagnosis; neurodevelopment; polyphagia

INTRODUCTION

Obesity is a growing public health problem with high prevalence [1], affecting an increasing number of countries worldwide [2]. The etiology is mostly multifactorial [3] and heritability is high. Evidence from twins, families and adoption studies has shown heritability estimates ranging from 40 to 70% [4]. Cases of major underlying medical causes (monogenic or endocrine disorders, cerebral or medication-induced obesities) are rare and a relatively low diagnostic yield has been reported, but quantitative evidence is limited [5].

Syndromic obesity is a generic clinical term that refers to cases of obesity occurring with additional phenotypes [6–10]. The majority of cases present with intellectual disability/developmental delay [8,11,12] and may also exhibit dysmorphic features or organ-specific developmental abnormalities [6–10] (**Figure 1**). Syndromic forms of obesity are rare [8,13–15] and present high phenotypic variability [8,10,16,17]. Frequently, syndromic obesity has a monogenic origin (autosomal or X-linked) [8] and occurs as a consequence of mutations in genes with pleiotropic effects [8,10].

Prader-Willi and Bardet-Biedl are the best-known genetic syndromes in which obesity is a central feature [8]. These and other examples of genetic disorders in which obesity is often a relevant feature are briefly presented in **Table 1** and will be discussed in detail in this narrative review. Our focus is on the syndromes of greatest interest in clinical practice, their genetic etiology, clinical description, and molecular diagnosis.

Classically, obesity cases related to deleterious variants in leptin-melanocortin pathway genes such as *LEP, LEPR, POMC, MC4R, MC3R, PCSK1*, and *SH2B1* are classified as nonsyndromic monogenic forms [8]. Kleinendorst et al. (2018) [18] investigated 1230 patients with obesity; 48 (~4%) were diagnosed with genetic obesity, 29 of whom had pathogenic variants in these genes. Another study conducted by the same research group [5] focused only on cases of pediatric obesity (younger than 18 years); among 282 individuals, 37 (~13%) were diagnosed with genetic obesity and 17 (~6%) had causal mutations in these leptin-melanocortin pathway genes.

Genetic changes in the leptin-melanocortin pathway result in hypogonadotropic hypogonadism, low blood pressure, behavioral problems, cholestasis and seizures, in addition to obesity, and these pleiotropic clinical characteristics could fit the definition of syndromic obesity. In this manuscript, we will discuss forms of obesity that are more

commonly classified as syndromic. We disregard the forms related to the leptin-melanocortin pathway, which are most commonly described in the literature as nonsyndromic monogenic forms, although we recognize that there is a gray area in the distinction between syndromic and nonsyndromic forms of obesity, as discussed by Kaur and colleagues (2017) [8].

MAIN SYNDROMIC FORMS OF OBESITY

Prader-Willi syndrome

The prevalence of Prader-Willi syndrome is estimated at 1:52,000-8,333 live births [19–25]**.** Frequent features of Prader-Willi syndrome patients include profound hypotonia in early childhood (with feeding difficulties up to approximately six months of age), hypogonadotropic hypogonadism (in both men and women), developmental delay, intellectual disability, and short stature [26,27]. Although facial dysmorphisms are not prominent and highly variable [27]**,** patients often have a narrow forehead, almond-shaped eyes, and a narrow nasal bridge [26]. Hands are small [28], and the fingers can be tapered at the tips; hypopigmentation of the hair, eyes and skin is common due to the loss of a copy of the *OCA2* gene (**Figure 2A)**, relevant for melanocytes. Features such as strabismus and scoliosis are often also present in Prader-Willi syndrome [26]**.** In addition, some patients have a high pain threshold [27]**.**

There is an evident delay in motor development**.** Regarding intellectual development, they can learn to read, write and do simple arithmetic calculations. They also have a special skill in assembling puzzles. The average intelligence quotient tends to be between 60 and 70 [29]. Behavioral problems are common in Prader-Willi syndrome [27]. Anger, stubbornness, compulsiveness and difficulty in changing routine occur in early childhood in 70-90% of cases. Psychosis occurs in adulthood in 10 to 20% of cases [26]. Some individuals have depression, possessiveness, and a tendency to steal and lie. The habit of picking wounds is also frequent [27,29].

In Prader-Willi patients, there is marked polyphagia and development of severe obesity [26], starting between 1 and 6 years old [29], and primarily central (abdomen, buttocks, and thighs) in both sexes [26]. Satiety and cognition problems often cause patients to ingest food in unconventional conditions (frozen, raw and even spoiled and found in the trash) [29]. Obesity is the main cause of metabolic complications, poor quality of life [30], and death [31]. Early dietary and lifestyle interventions, with

restricted access to food and control food intake (food security approach), associated with the use of growth hormone therapy, have profound impacts in preventing the development of obesity in these patients [32].

The phenotype results from a lack of the expression of paternally inherited alleles of genes in the 15q11-q13 imprinted region [10,26,33], usually due to paternal microdeletion (approximately 65-75%) or maternal uniparental disomy 15 (20-30%), but imprinting defects also occur in a minority of cases (1-3%) [26,34]. Chromosomal rearrangements such as translocations can also cause Prader-Willi syndrome but are rarer [35]. Angelman syndrome is another genetic disorder related to changes in the same region of chromosome 15, but caused by the absence of expression of alleles on the maternal chromosome and with markedly distinct phenotypes [36].

The Prader-Willi region is crucial for hypothalamus development and function [35]. **Figure 2A** shows the structure of the 15q11-q13 region and specifies the genes that are expressed exclusively from the paternal chromosome (maternal imprinting) and those expressed from the maternal chromosome (paternal imprinting). Hypothalamic dysfunction is recognized as the basis of Prader-Willi syndrome pathophysiology, which includes a deficiency in growth hormone and reproductive hormones, circadian rhythm abnormalities, and a lack of satiety, leading to extreme obesity. Structural hypothalamic changes are described in patients with Prader-Willi syndrome: significantly smaller hypothalamic paraventricular nucleus volume and reduction in cell number [37]. High-level of ghrelin, a hormone produced by the gastric mucosa that has an orexigenic action, is another molecular feature of Prader Willi syndrome [38].

Obesity is a recognized cause of morbidity and mortality in Prader-Willi syndrome, but not in Angelman syndrome [39]. In an evaluation of 110 adolescents and adults with Angelman syndrome conducted by Larson et al. (2015) [40], 32% were overweight or obese, which is not higher than the prevalence in the general population. However, specifically Angelman patients with uniparental disomy or imprinting center defects have increased body mass in infancy/early toddlerhood [39].

Angelman/Prader-Willi syndrome imprinting centers (ICs – **Figure 2A**) regulate imprinting in the SNURF-SNRPN gene cluster. Under normal conditions, Prader-Willi IC is methylated on the maternal allele, while it is unmethylated on the paternal allele. Silencing of the paternal allele or maternal uniparental disomy of the SNURF/SNRPN gene cluster results in loss of expression of several genes in the Prader-Willi region and leads to the development of the syndrome [41].

DNA methylation analysis is the only technique that can diagnose Prader-Willi syndrome in all three molecular genetic classes (microdeletion, maternal uniparental disomy and mutations in the imprinting center) and differentiate Prader-Willi from Angelman syndrome [34]. In the 15q11-q13 region, there are sequences more prone to rearrangements known as BP1, BP2 and BP3 (breakpoints 1, 2 and 3, respectively – **Figure 2A**). Rearrangements between these breakpoints resulting in microdeletions are classified as class I (B1-B3; approximately 6.6 Mb) and II (B2-B3; approximately 5.3 Mb); both are almost always *de novo* events, but class 1 is usually associated with a more severe clinical condition. There are also cases involving minor or major microdeletions, but these are considered atypical [42,43].

Bardet-Biedl syndrome

Bardet-Biedl syndrome is a multisystem genetic syndrome [44,45] that occurs in 1:160,000 to 1:13,000 [46–50] live births and has considerable interfamilial and intrafamilial clinical variation [29,44,51]. It is characterized by six main features: retinal degeneration (>90%), obesity (72–92% of patients), polydactyly (63–81%), ID (50–61%), kidney disorders (20–53%), and hypogonadotrophic hypogonadism in men (59–98%). Severe intellectual disability is uncommon, accounting for only approximately 9% of the cases [29]. Other features are also frequently reported, such as speech development delay, hearing loss, olfactory problems [44,52], cataracts, astigmatism [52], craniofacial dysmorphisms (e.g., brachycephaly, macrocephaly, dental changes [44,52], arched palate, and strabismus), coordination problems [52], ataxia [44,52]**,** brachydactyly, syndactyly [52]**,** metabolic defects (e.g., type 2 diabetes mellitus), hypertension, heart and liver disease [44,52], uterine and/or vaginal malformations, delayed puberty in girls, amenorrhea [51], polyuria, polydipsia [52], skin dermatoses [45,53], and behavioral problems (e.g., autism and psychosis) [44]. Most signs may not be present at birth, but they appear and progressively worsen during the first and second decades of life [51].

Bardet-Biedl syndrome should always be considered in patients with obesity and progressive visual impairment. Retinal dystrophy occurs in virtually all cases. It usually begins between four and nine years of age, with progressive night blindness. First, central vision and, later, peripheral vision are lost. For this reason, while peripheral vision is still preserved, it is common for the child to bump into objects

placed in front of him or even tilt his head to look out of the corner of his eye at objects on the floor. Blindness occurs at 15.5 years of age on average [29].

Differential diagnosis should be made, especially considering McKusick-Kaufman syndrome (OMIM # 236700) and other diseases in which there are visual difficulties and/or polydactyly. Previously, Bardet-Biedl syndrome was known as Laurence-Moon-Bardet-Biedl. However, there is now a separation: (1) Laurence-Moon syndrome presents with ataxia and spastic paralysis, which are absent in Bardet-Biedl syndrome; (2) Bardet-Biedl syndrome includes polydactyly and obesity, which are absent in Laurence-Moon syndrome [29].

Hippocampal volume tends to be reduced in patients with Bardet-Biedl syndrome [44]. Bardet-Biedl syndrome is a ciliopathy [45,54]; most related genes encode proteins involved in the formation and/or functioning of the BBSome, a multiprotein complex important in ciliary function [44,45]. It is known that the primary cilium is an important structure in neurogenesis and development of the hippocampus. Its impairment is probably related to the neurological problems of Bardet-Biedl syndrome [44]. As shown in **Table 2**, many genes present on different chromosomes have been identified as related to this syndrome. *BBS1* and *BBS10* are the most frequently mutated genes among Bardet-Biedl cases, followed by *BBS12* [55]. A total of 20 to 30% of individuals with Bardet-Biedl syndrome do not have identifiable mutations in known related genes, which means that as other yet unidentified genes may be involved [29]. Bardet-Biedl syndrome is usually caused by individual *locus*, with an autosomal recessive pattern of inheritance, and therefore is more common in inbred populations [45,56]**,** but there are reports of the interaction of two or more loci [44]**.** In some families, three mutated alleles were observed for the manifestation of Bardet-Biedl syndrome (and not just two, as for autosomal recessive inheritance), suggesting a triallelic inheritance model (involving two *loci*), with an estimated frequency of less than 10% of bulletin board cases [29]**.**

Knockout mouse models for the *BBS2* [57], *BBS4* [58] and *BBS6* [59] genes were developed. All of them exhibited obesity, retinopathy and male infertility, in addition to other signs particular to a single model, such as problems of social interaction, deficits in smell and polydactyly [57–59]. These three models had hyperleptinemia secondary to leptin (anorexigenic hormone) resistance and showed reduced hypothalamic POMC (anorexigenic neuropeptide) expression, which suggests that Bardet-Biedl syndrome genes play an important role in maintaining neuronal

sensitivity to leptin [60]; this is the probable mechanism of origin of obesity in patients with Bardet-Biedl syndrome [45].

Pseudohypoparathyroidism

Hypoparathyroidism is characterized by low or undetectable parathyroid hormone and consequent reduction in calcium and elevation in phosphorus levels [61]. Pseudohypoparathyroidism (or inactivating PTH/PTHrP signaling disorder [62]) is characterized by target organ resistance or unresponsiveness to parathyroid hormone, hypocalcemia, hyperphosphatemia, and increased circulating parathyroid hormone [63].

Pseudohypoparathyroidism prevalence was estimated in a Danish study as 1.1/100,000 inhabitants [64]. The three disorders, pseudohypoparathyroidism type 1a, 1c, and pseudopseudohypoparathyroidism, are caused by alterations in the *GNAS* gene [61], a gene subjected to imprinting that produces four different transcripts (**Figure 2B**) using alternative promoters, including G protein subunit alpha (Gαs). The five protein encoded are Gαs (expressed by both alleles in most tissues), A/B, XLαS and GNαS-AS1, which are paternally expressed, and *NESP55*, which is maternally expressed [65,66]. Since exons 2–13 of *GNAS* are shared by XLαs, NESP55 and A/B proteins, mutations in those exons lead to deficiency of not only Gsα but also these other products; the phenotypes resulting from the deficiency of each of these *GNAS* products are unclear [67]. Gsα is an ubiquitously expressed signaling protein and plays important roles in the regulation of various hormones; resistance to additional hormones, such as thyroid stimulating hormone (TSH), growth hormone releasing hormone (GHRH) and gonadotropins, can also be observed in patients with pseudohypoparathyroidism 1a and 1c [67]. The obesity phenotype occurs primarily in patients who also have multiple hormone resistance [68].

Typical pseudohypoparathyroidism physical findings are called Albright hereditary osteodystrophy (OMIM #300800), in which patients have round facies (approximately 92%), short stature (80%), thickened calvaria (62%), ectopic soft tissue or dermal ossification(s) (osteoma cutis – 56%), and obesity (50%). Subcutaneous calcifications (55%), dental hypoplasia (51%), cataracts and band keratopathy (44%) are also common and are likely the consequences of longstanding hypocalcemia. Approximately 75% of patients with pseudohypoparathyroidism and Albright hereditary osteodystrophy have cognitive impairment [68], but diagnosed cases of

intellectual disability are usually mild [29]. Many patients have brachydactyly, which may be asymmetric or not and may involve one or both hands or feet; 68% have brachymetacarpia, 50% brachyphalangia, and 43% brachymetatarsia [68].

Pseudohypoparathyroidism 1a and 1c forms are caused by germline or somatic mutations in the maternal allele of the *GNAS* gene (Gsα coding region). The obesity mechanism is likely related to Gsα signaling reduction at the α-MSH (α-melanocyte stimulating hormone) receptors in the central nervous system [69]. α-MSH is a neuropeptide that is part of the central melanocortin system and is related to anorexigenic effects [70].

GNAS mutations in the paternal allele lead to pseudopseudohypoparathyroidism, which also manifests as Albright hereditary osteodystrophy, but without resistance to parathyroid hormone or other hormones [67].

Regarding the difference between pseudohypoparathyroidism 1a and 1c, the 1c form presents normal Gsα activity measured in patient-derived erythrocytes [67]. In patients with pseudohypoparathyroidism 1a, there is a deficiency of approximately 50% of Gsα activity [29].

Patients with pseudohypoparathyroidism type 1b show imprinting abnormalities at the complex *GNAS* locus, with parathyroid hormone and other hormone resistance, usually in the absence of Albright hereditary osteodystrophy. However, there are cases of pseudohypoparathyroidism type 1b with an Albright hereditary osteodystrophy phenotype, including obesity. The molecular cause of autosomal dominant familial pseudohypoparathyroidism type 1b has been well defined with the identification of microdeletions within the *GNAS* locus or the nearby *STX1* (approximately 200 kb of distance). This results in isolated loss of imprinting at exon *A/B* or deletions involving the *NESP55* differentially methylated region, resulting in loss of all maternal *GNAS* imprints [71]. In some cases of sporadic pseudohypoparathyroidism type 1b, an isodysomy of paternal chromosome 20 (entire [72,73] or just the q arm [71,73]) was identified.

In the cohort of 282 patients with pediatric obesity evaluated by Kleinendorst et al. (2020) [5], 5 (~1.8%) were positive for *GNAS* mutations. In the analysis of exome data from 2548 children with severe obesity, de Oliveira et al. (2021) [74] identified 22 (~0.86%) patients with *GNAS* mutations and without suspected pseudohypoparathyroidism. Traditionally, *GNAS* sequencing has been performed only in patients with clinical features suggestive of pseudohypoparathyroidism, but

pathogenic mutations in this gene may manifest only with obesity [74]. The importance of the *GNAS* gene has been recognized in the investigation of childhood obesity.

Temple syndrome

Temple syndrome is characterized by hypotonia (93%, with feeding difficulties frequently reported), small feet (96%) and hands (87%), low weight at birth (87% <5th centile), precocious puberty (86%), motor development delay (83%), short stature (79%), hyperextensible joints (63%), speech delay (59%), obesity (49%), intellectual disability (39%), premature birth (30%), scoliosis (23%), and low birth head circumference (27% <5th centile) [75]. Obesity is of the truncal type and develops as early as 4–6 years [76]. Early-onset weight gain is associated with decreased resting energy expenditure and hyperphagic behavior similar similar to that observed common obesity. Regarding facial features, there are facial similarities between many of the patients, including a broad, tall forehead nose often short with a wide, fleshy nasal tip, and a short philtrum [75]. Facial features become distinguishable with age. It is worth mentioning that frontal bossing and micrognathia are common in infants [76]. Often in early childhood, patients with Tempe syndrome have a phenotype that overlaps those of patients with Silver-Russell syndrome.

Temple syndrome maps to the 14q32.2 region [77] and, like Prader-Willi syndrome, is also an imprinting disorder [76]. Of 51 Temple syndrome cases reviewed by Ioannides (2014) [75], 40 had maternal chromosome 14 disomy as an etiology, six had imprinting defects, and five had a microdeletion on the paternal chromosome. Cases of Robertsonian translocations [78,79] or mosaicism [80–82] are rare. All these mechanisms have the absence paternal 14q32 expression in common [75]. The same mechanisms involving absence of expression of the maternal alleles are the origin of Kagami-Ogata syndrome (OMIM #608149) [76], which has a very different phenotype characterized by facial abnormality, a small, bell-shaped thorax, abdominal wall defects and polyhydramnios [83]. To date, no SNV (single nucleotide variant) or indel mutations that cause Temple syndrome have been identified [84,85].

We present in **Figure 2C** the 14q32.2 region highlighting the imprinting profile of the genes contained therein. Imprinting is controlled by an IC intergenic differentially methylated region or IG-DMR, which is normally methylated only on the paternal chromosome. IG-DMR methylation is acquired in the male germline and subsequently directs the acquisition of methylation on the paternal allele of a somatic

DMR (differentially methylated region) within *GTL2/MEG3*. The unmethylated IG-DMR on the maternal allele is associated with the expression of *GTL2/MEG3* and *RTL1as*, which represses the expression of *DLK1* and *RTL1* in cis. Genetic testing based on ratiometric measurement of methylated/unmethylated DNA within the IG-DMR can be used as a screening Temple syndrome test [75].

Loss-of-function *DLK1* mutations cause central precocious puberty and are also associated with metabolic disorders. DLK1, which inhibits adipogenesis, is likely involved in the Temple syndrome mechanisms of obesity, precocious puberty, and metabolic problems such as type 2 diabetes mellitus and hyperlipidemia [86]. Furthermore, *DLK1* expression in the mouse hypothalamus and in kisspeptin neuron-derived cell lines suggests a neuroendocrine function [84].

16p11.2 deletion syndromes

Deletions of different segments in the 16p11.2 region, which do not necessarily have overlapping gene content, cause syndromic obesity forms. In a cohort study of 680 unrelated children with syndromic obesity and without Fragile-X or Prader-Will syndromes, Gimeno-Ferrer and colleagues (2018) [87] identified 11 cases (1.6%) with 16p11.2 copy number variations (CNVs) by microarray analysis. In this region, there are recurrent CNVs originating from nonhomologous allelic recombination between five breakpoints (BPs) (**Figure 3A**) [88].

The 16p11.2 deletion between BP4 and BP5 (~600 kb, OMIM #611913) is one of the most frequent neurodevelopmental and autism spectrum disorders [88,89], with a prevalence of 1:2000 in the general population [89]. It is most typically diagnosed with microarray testing for genetic etiologies of developmental delay, intellectual disability, or autism spectrum disorders [90]. Approximately 0.5% of the cases of autism spectrum disorders present this syndrome [89].

Obesity is associated with polyphagia and evolves throughout childhood. In adulthood, 75% of patients have obesity (approximately 45% of them with grade III), although birth weight is frequently below average. Although head circumference is smaller at birth, there is an overall increase by 2 years of age, and macrocephaly (z score ≥2) is present in 17% of patients. Frequent features are speech impairment (70%), motor coordination difficulties (60%), seizures (24%), and autism (20–25%). The intelligence quotient is shifted 1.8 standard deviations lower than that of 16p11.2 deletion-related noncarriers. Facial dysmorphisms are present in half of the carriers

[89], but there are no characteristic facial features [90]. Congenital anomalies are more frequent than in the general population, with the most common being vertebral abnormalities (hemivertebrae or kyphoscoliosis affect ~20% of carriers). Altered brain phenotypes include volume increases, changes in white matter microstructural properties, and early electrophysiological cortical responses from the auditory cortex [89].

The deletion between BP4 and BP5 is most often *de novo*, but is inherited in approximately 7% of probands [89]. Reciprocal 16p11.2 duplication syndrome (OMIM #614671) is a rare disorder often associated with intellectual disability, attention deficit, hyperactivity disorder, and a predisposition to epilepsy and schizophrenia; there are no specific dysmorphic features [91]. In contrast to 16p11.2 deletion, patients with 16p11.2 duplications are underweight, with a typical body mass index lower than that of relatives [90].

The 16p11.2 deletion between BP2 and BP3 has 220 Kb (OMIM #613444) [88] (**Figure 3A**) and causes severe early-onset obesity with cognitive deficits, developmental delay, and autism [92]. This region encompasses nine genes, including *SH2B1*, which was postulated to be principally responsible for obesity and intellectual disability [93,94]. *SH2B1* is a critical regulator of body weight and metabolism in both animals and humans. Mice with hypothalamus-specific ablation of *Sh2b1* develop obesity, insulin resistance, and liver steatosis. In contrast, *SH2B1* hypothalamic overexpression protects against high-fat diet-induced obesity and metabolic syndrome [95]. Other 16p11.2 deletions related to syndromic obesity phenotypes involve genes of the BP4-BP5 and/or BP2-BP3 portions. Some of them are represented in **Figure 3A.**

1p36 deletion syndrome

1p36 deletions are the most common terminal deletions in humans [96] and are detected in approximately 1 in 5,000 newborns [96,97]. In a cohort of 279 syndromic obesity patients analyzed by genomic microarrays and negative for Prader-Willi, five (1.8%) were positive for 1p36 deletions [6]. The major clinical features are developmental delay, hypotonia, short stature and craniofacial dysmorphisms such as large anterior fontanel, prominent forehead and chin, deep eyes, ear asymmetry, a flat nasal bridge, and maxillary hypoplasia. Other possible clinical manifestations are skeletal malformations, cardiac, gastrointestinal, and visual abnormalities, seizures, and

behavioral problems [97]. There is also an association of 1p36 deletions with obesity and polyphagia [98].

There are two critical portions related to the phenotype, one proximal/interstitial and one distal/terminal (also called classic) [96,97] (**Figure 3B**); the distal portion is associated with obesity [98]. A majority of the cases (52-67%) have a terminal deletion on chromosome 1. There is evidence of the participation of the *MMP23B*, *GABRD, SKI, PRDM16, KCNAB2, RERE, UBE4B, CASZ1, PDPN, SPEN, ECE1, HSPG2*, and *LUZP1* genes in the phenotype [98], and the deleted region and genomic content are related to phenotypic variations [97]. Regions related to some features are presented in **Figure 3B**, according to a review published by Jordan et al. [96]; we also represent a critical region for the manifestation of obesity and polyphagia (1p36.33–36.32) as suggested by D'Angelo et al (2009) [98].

Kleefstra syndrome type 1

Kleefstra syndrome type 1 is characterized by moderate to severe intellectual disability, severe delayed or absent speech, neonatal hypotonia, delayed motor development [99–101] and sleep disorder [99]. Facial features include microcephaly, brachycephaly, hypertelorism, synophrys, midface hypoplasia, everted lower lip, prognathism, and macroglossia. Patients with Kleefstra syndrome type 1 ara often large for gestational age and childhood obesity is very common. Congenital heart disease and seizures are also frequent [101]. With the exception of individuals with somatic mosaicism, patients with Kleefstra syndrome type 1 do not reproduce [100]. The prevalence remains unprecise, but it is estimated to occur in 1:200,000 intellectual disability cases [102]. Gastroesophageal reflux and tracheobronchomalacia, and genitourinary abnormalities such as micropenis, cryptorchidism and vesicourethral reflux are rarely reported in Kleefstra syndrome type 1. Brain imaging tests usually show results within the normal range, but in some cases there are dilated ventricles, myelination problems and other changes [101].

Behavioral problems are common in Kleefstra syndrome type 1 [99,101]. Apathy, aggression, psychosis, autistic features, catatonia, bipolar disorder, and regression in daily functions and cognitive abilities are possible manifestations. In adulthood, they may also manifest severe psychiatric pathologies [101].

Kleefstra syndrome type 1 is caused by *EHMT1* haploinsufficiency [103]. There are cases of deletions in 9q34.3, including at least part of *EHMT1* and intragenic

heterozygous variants. Almost all cases are due to *de novo* mutations. There are few reports of cases inherited, for example, from somatic mosaicism or balanced translocation involving 9q34.3 from one parent [100].

EHMT1 is a major methyltransferase that acts for the mono- and dimethylation of lysine 9 on histone 3 (H3K9me1/2) and can also monomethylate both H3K27 and H3K56 in some contexts [104]. Its predominant histone mark (H3K9me2) is related to transcriptional repression. EHMT1 also interacts with other transcription factors [103] and is essential for repressing gene transcription in a highly tissue- and temporal-specific manner [105]. Pyramidal CA1 hippocampal neurons of *Ehmt1*$^{+/-}$ mice had significant reductions in spine density, number of mature spines, and dendritic arborization, indicating a postsynaptic deficit [106], which may be related to the pathophysiology of human Kleefstra syndrome type 1.

Börjeson-Forssman-Lehmann syndrome

Börjeson-Forssman-Lehmann syndrome is a rare syndrome in which mild to moderate intellectual disability is the main feature [29,107]. In this syndrome, there is a delay in speech and in the development of motor skills. In adolescence, patients have generalized learning difficulties that make them dependent on the care of others [29].

Börjeson-Forssman-Lehmann syndrome is an X-linked recessive syndrome [108] that is caused by mutations involving the *PHF6* gene [109] that encodes an epigenetic regulator [110]. It manifests mainly in males; female carriers are usually unaffected or have only mild clinical signs [109], but recently *de novo* mutations in *PHF6* have been identified in women with a distinct and severe intellectual disability disorder [109,111–113]. Obesity, which is common in male individuals, is not a striking feature in these female patients [113].

Some dysmorphisms with narrow palpebral fissures and large ears are also prevalent, especially with long and flaccid lobes, gynecomastia (97%), characteristic digit anomalies (96% - such as short, flat and hyperflexible fingers, short and separated toes, big and short hallux and/or hammer toe), hypogonadism (86%), and truncal obesity (76%) [29]. Patients may also have microcephaly (6%) or macrocephaly (15%). Epileptic seizures were observed in approximately 8% of patients. At birth, weight and head circumference are usually normal [107]. However, hypotonia occurs in the neonatal period and in infancy, which causes growth and developmental problems. Börjeson-Forssman-Lehmann syndrome has great interfamilial and intrafamilial

phenotypic variability [29], and some clinical features only become more evident with age [107].

A transcriptomic analysis of the cerebral cortex of mice with a CRISPR-generated *PHF6* mutation demonstrated that this gene promotes the expression of neurogenic genes and represses synaptic genes [114]. It was also shown in mice that the *PHF6* gene encodes a critical protein for the migration of neurons in the cerebral cortex. PHF6 dysregulation generates white matter changes and neuronal hyperexcitability, which may be relevant to the pathogenesis of intellectual disability and seizures in Börjeson-Forssman-Lehmann syndrome [107,115].

Cohen syndrome

Cohen syndrome is an autosomal recessive disease with multisystem clinical features [116,117] that is characterized by obesity, hypotonia (usually detected during breastfeeding and may persist until adolescence), intellectual disability, and narrow hands and feet. Craniofacial features can include microcephaly [117], thick hair and eyebrows, long eyelashes [117,118], hypertelorism [117], arched or wave-shaped eyelids [117,118], high nasal bridge, bulbous nasal tip, short philtrum, high and narrow palate, prominent maxillary central incisors, maxillary hypoplasia, and micrognathia [117]. However, clinical facial recognition is difficult before the age of six, which makes the diagnosis of Cohen syndrome in younger children difficult [117,119]. Patients may also have leukopenia (mainly neutropenia) [117,120,121], ophthalmologic changes (e.g., severe myopia and early-onset and progressive retinal dystrophy) [117,120,121], high-pitched crying and voice, possibly secondary to laryngeal abnormalities [117], and joint hypermobility [117,120,121]. There are also case reports of congenital heart disease [117]. Increased waist circumference becomes evident between 5 and 12 years [29], but this abnormal distribution of fat does not always lead to an obesity BMI.

All Cohen syndrome patients have some degree of intellectual disability, speech development is often impaired, and motor developmental milestones are significantly delayed. Patients also often rely on help for some daily activities. Antisocial, violent, rebellious, or untrustworthy behavior is rare; a friendly personality is more common [117].

Cohen syndrome is caused by mutations involving the *VPS13B* gene (also known as *COH1*) [117,122]. This syndrome is mainly described in the Finnish

population [29,123], in which bottleneck events such as geographic and cultural isolation significantly impacted the enrichment of some pathogenic variants [124]. Cohen syndrome tends to be more heterogeneous with wide phenotypic variability in non-Finnish patients [117].

The *VPS13B* gene appears to be important in the development and function of the eyes, hematopoietic system, and central nervous system [117]. It encodes a membrane protein involved in Golgi integrity [116] and vesicle-mediated transport [117]. Deleterious *VPS13B* variants disturb the Golgi, leading to an unusual pattern of glycosylation of several proteins and an impairment in endosomal-lysosomal trafficking. Functional studies indicate that preadipocytes without *VPS13B* differentiate more easily, which justifies the accumulation of fat [117]. These biochemical and cellular changes are believed to be involved in the pathophysiology of Cohen syndrome [116].

Alström syndrome

Alström syndrome is related to mutations involving the *ALMS1* gene and has an autosomal recessive pattern of inheritance [29,125,126]. Obesity, retinal degeneration, type 2 diabetes mellitus and sensorineural deafness are the most important features of Alström syndrome [29,126,127]. Intellectual disability is uncommon, but patients can have early learning difficulties and speech impairment; delay in early motor developmental milestones is seen in 27% of patients [126]. The prevalence of Alström syndrome is not precisely known, but according to a review published in 2021, approximately 1,200 cases of Alström syndrome have been identified worldwide [128].

Obesity is always present [127]. Polyphagia and excessive weight gain begin during the first years, resulting in childhood obesity [129]. In general, birth weight is normal, but excessive gain occurs after the first year of life [29,129]. The BMI of most young children is >95th centile [126], and obesity is centripetal in childhood [29]. In some individuals, body weight tends to normalize, decreasing into the high-normal to normal range after adolescence [129]. Height is normal in early childhood, but growth slows in adolescence, and final adult height is usually <5th centile [126].

They exhibit early onset retinal degeneration of rods and cones, leading to progressive vision problems, nystagmus, and photophobia [29,129]. Many people lose all perception of light by the end of their second decade of life, albeit a minority retains

the ability to read large prints by the third decade. Hearing loss is also progressive and severe (40-70 decibels), observed in the first decade in up to 70% of individuals [129].

In addition, patients have dyslipidemia, especially markedly increased levels of triglycerides and VLDL cholesterol. They are also often diagnosed with arterial hypertension around the age of two [127]. They may present with hypothyroidism, alterations in the insulin-like growth factor (IGF) system, low testosterone in males and hyperandrogenism in females; hyperinsulinemia and type 2 diabetes develop in childhood [126].

Dilated cardiomyopathy can occur suddenly in infancy (first months of life), and restrictive cardiomyopathy develops slowly in adolescents and adults. They may develop absence seizures and febrile epilepsy. Respiratory problems such as pulmonary fibrosis, chronic obstructive respiratory syndrome, and acute respiratory distress syndrome are also common. Liver dysfunction begins with hepatic steatosis and can progress to hepatic fibrosis and cirrhosis. Renal failure develops slowly in late adolescence and adulthood [126] and is the main cause of death in patients with Alström syndrome [29]. Birth complications include neonatal hypoxia and hypotonia [130].

The ALMS1 protein localizes to centrosomes and basal bodies of ciliated cells and operates in cell cycle regulation and intraciliary transport. Like Bardet-Biedl syndrome, Alström syndrome is also a ciliopathy [126] and there is an important phenotypic overlap between the two syndromes; therefore, Bardet-Biedl syndrome should be considered in the differential diagnosis of Alström syndrome [29]. However, patients with AS do not have polydactyly or syndactyly and rarely have intellectual disability [29,127], but 30% of them have some learning disability [29].

Carpenter syndromes type 1 and type 2

In humans, biallelic mutations involving the *RAB23* gene lead to Carpenter syndromes type 1 [131,132]. This syndrome is characterized by multiple suture craniosynostosis (sagittal, lambdoid, and coronal), facial dysmorphisms, broad or bifid thumbs, absent or small middle phalanges, and polysyndactyly of hands and feet. Additionally, the characteristics of Carpenter syndrome type 1 are intellectual disability, high birth weight, late obesity, congenital heart disease, umbilical hernia, cryptorchidism, hypoplastic testes, arched femur and tibia, and malformations of the nervous system [131]. The patients' obesity is of the central type. Overgrowth manifestations and advanced bone age have also been described [133,134].

RAB23 is a GTPase that acts as a regulator in *sonic hedgehog* signaling [131], an important pathway in animal development [135]. Nonsense *RAB23* mutations in mice lead to neural tube defects and embryonic lethality [136]. Notably, RAB23 is implicated in ciliary trafficking [137,138], including this syndrome in the group of ciliopathies.

As Carpenter syndrome type 1, Carpenter syndrome type 2 presents multiple congenital malformations, such as multisuture craniosynostosis, polysyndactyly of hands and feet, umbilical hernia, cryptorchidism and heart disease. Obesity is also a feature of Carpenter syndrome type 2. However, in Carpenter syndrome type 2, which is caused by biallelic mutations in *MEGF8*, there may also be defective lateralization ranging from transposition of the great arteries, dextrocardia, to complete *situs inversus*. In addition, thoracic skeletal abnormalities and spaced, hypoplastic, and/or supernumerary nipples appear to be more frequent features in patients with Carpenter syndrome type 2 [139].

Megf8$^{-/-}$ mice have a disruption in axon guidance in the peripheral nervous system and exhibit defects in development of the limb, heart, and left-right patterning. Furthermore, *MEGF8* appears to be involved in BMP4 signaling [140], a protein that regulates a number of developmental processes, including the dorsal-ventral axis, neural patterning [141], and orofacial development [142].

MORM syndrome

Homozygous deleterious variants in the *INPP5E* gene lead to a syndrome characterized by mental retardation, truncal obesity, retinal dystrophy, and micropenis, or MORM syndrome [143,144]. In MORM syndrome, retinal dystrophy is usually not progressive, except for one reported case that was progressive in the first decade of life. Patients can also have type 2 diabetes mellitus and dyslipidemia [145]. However, few case reports are available in the literature, which limits the establishment of the phenotypic spectrum.

The phenotype of MORM syndrome overlaps those of Bardet-Biedl and Cohen syndrome, but MORM can be distinguished by the earlier age of onset and the absence of dysmorphic facies, microcephaly, polydactyly, and testicular anomalies [144].

INPP5E plays an important role in the development and maintenance of ciliary homeostasis. Therefore, MORM is also a ciliopathy [143,146].

MYT1L-related syndrome

According to the most comprehensive study of this syndrome, which involved the analysis of 62 cases, the main phenotypic features of *MYT1L*-related syndrome are developmental delay with behavioral disorders (98%), language impairment (95%), intellectual disability (70%), overweight/obesity (58%), and epilepsy (23%) [147]. Polyphagia is an overarching mechanism leading to excessive weight gain in these patients. Strabismus and macrocephaly are also common features [148], but overall, a facial phenotype does not seem to be recognizable [148,149]. Among the behavioral features, it is worth mentioning autism, aggressiveness [147,148], attention deficit hyperactivity disorder, anxiety, nervous tics, and obsessive–compulsive disorder [147].

CNVs, indels and SNVs involving *MYT1L* are the underlying cause of this autosomal dominant form of syndromic obesity [147–151]; microdeletions are more frequent [149,152]. As functional evidence, *Myt1l* haploinsufficient mice recapitulate several clinical phenotypes observed in humans carrying heterozygous *MYT1L* mutations, such as obesity and behavioral alterations. MYT1L is a zinc finger neural family of transcription factors [153]. It is an important regulator of neuronal differentiation [154–157], expressed in different parts of the brain, with peak expression in the fetal period [152]. There is a hypothesis that intellectual disability and obesity observed in patients with *MYT1L* mutations could be caused by dysregulation of gene expression and development of the neuroendocrine hypothalamus [148].

Luscan-Lumish syndrome

Heterozygous mutations in the *SETD2* gene cause Luscan-Lumish syndrome, whose main features are macrocephaly, intellectual disability, speech delay, low sociability, and behavioral problems (most typically an autism spectrum disorder but also attention-deficit disorder, aggressive outbursts, self-mutilating behaviors, frustration intolerance, anxiety, hyperphagia, and stereotypies) [158]. Patients with this syndrome may also have postnatal overgrowth, obesity, advanced carpal ossification, developmental delay, seizures, hypotonia [158–160], cryptorchidism, nasal polyps, large tongue with creases, high pain threshold, and constipation [161]. SETD2 is a histone methyltransferase responsible for trimethylation of lysine 36 of histone H3 [159].

DISCUSSION

In this article, we summarized the main forms of syndromic obesity. As can be observed, the diagnosis of syndromic obesity patients is challenging due to genetic heterogeneity. Kabuki and Rubinstein-Taybi syndromes may also be associated with obesity [162,163]; however, the current evidence is insufficient to establish them as forms of syndromic obesity.

Genetic investigation of syndromic obesity

Here, we focus on genetic syndromes in which the occurrence of obesity is frequent and well characterized, but studies investigating the genetic etiology of syndromic obesity cases also identified genetic variants mapped to intellectual disability genes, in which obesity is not a central feature [6,8,164–166]. It should be noted that multifactorial obesity is quite prevalent [1], and therefore, the obesity presented by a syndromic patient may not be directly related to intellectual disability syndrome [166], although several publications have reported a higher prevalence of obesity among people with cognitive impairment [167–169]. Many factors may contribute to this higher prevalence, such as low levels of physical activity, poor nutritional habits [170], and the use of psychiatric drugs [171] (i.e., antiepileptic, antidepressant and antipsychotic drugs) [5]. Several publications also reported a relationship between obesity and psychiatric disorders [172–175], which are present in many syndromic patients and can also contribute to an increased risk of obesity [176]. Thus, in the genetic investigation of patients with syndromic obesity, one should not rule out the possibility that the etiology of the syndromic condition can be independent of obesity.

Even before the manifestation of obesity, some clinical features may be indicative of the need for genetic investigation. Prader-Willi syndrome, for example, can be diagnosed during the neonatal period due to typical hypotonia and feeding problems. Another relevant fact is that early-onset obesity (especially <5 years) is more likely to have a genetic origin [5].

The genetic investigation of syndromic obesity requires the evaluation of multiple genetic mechanisms, considering that the etiology can be related to CNVs, SNVs, indels and alterations in the methylation pattern, in addition to presenting high genetic heterogeneity that poses substantial challenges for diagnosis.

Methods of genome-wide investigation (DNA microarray and whole exome/genome sequencing) should be prioritized over the target sequencing of well-characterized syndromic obesity gene panels. The simultaneous analysis of CNVs, SNVs and indels using NGS data is a cost-effective option for genetic investigation of syndromic obesity [166]. Additionally, epigenetic platforms that can collectively analyze and distinguish between genomic imprinting disorders, such as Prader-Willi and Temple syndrome, are useful to aid in clinical diagnosis [76].

Biological mechanisms

Although the biological pathways related to syndromic obesity genes are quite diversified, it is clear that the etiology tends to be central (linked to neurodevelopment), with less contribution of causal peripheral mechanisms. There are three main groups of mechanisms (Supplementary Figure):

(1) imprinting disorders (e.g., Prader-Willi syndrome, pseudoparathyroidism, and Temple syndrome). Various imprinted genes are critical regulators of growth and development; thus, related disorders have clinical features impacting growth, development, behavior and metabolism [177]. Prader-Willi, Temple and syndromes related to *GNAS* mutations have obesity and other clinical features possibly related to hypothalamic disturbances.

(2) caused by mutations in genes directly involved in transcriptional activity (e.g., Kleefstra type 1, Börjeson‐Forssman‐Lehmann, *MYT1L*-related, and Luscan-Lumish syndromes). Neurodevelopmental disorders can involve mutations in several genes; however, there is a significant overrepresentation of genes encoding proteins that alter chromatin structure [178].

(3) ciliopathies (e.g., Bardet-Biedl, Alström, MORM, and Carpenter type 1 syndrome). An increasing number of genetic disorders are termed ciliopathies, which are associated with obesity. Neuronal cilia likely play diverse signaling roles throughout the brain and impact several behaviors including feeding. Additionally, cilia modulate peripheral tissue signaling and homeostasis (pancreas, skeletal muscle, adipocyte differentiation), but the mechanisms associated with cilia dysfunction and obesity are complex and not fully understood [138].

Conclusions

A correct genetic diagnosis impacts medical management, treatment, and prognosis, in addition to enabling proper genetic counseling. Additionally, it may benefit the family from the psychological point of view by ending the diagnostic odyssey [179]. However, the etiology is not identified in most cases of syndromic obesity [8,165]. Kaur and colleagues (2017) [8] reported that the genetic etiology has been elucidated for only a fraction of syndromic forms of obesity described in the literature (19 out of the 79). There are gaps in knowledge about syndromic obesity and a low genetic diagnostic rate, which impairs genetic counseling and hinders the development of therapeutic approaches. Knowledge of pathophysiology pathways could reveal therapeutic opportunities for syndromic forms of obesity.

REFERENCES

1. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism. 2019;92:6–10. https://doi.org/10.1016/j.metabol.2018.09.005

2. Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health. Int J Surg Oncol. 2017;2:17. https://doi.org/10.1097/IJ9.0000000000000017

3. Reddon H, Guéant J-L, Meyre D. The importance of gene–environment interactions in human obesity. Clin Sci. 2016;130:1571–97. https://doi.org/10.1042/CS20160221

4. Sandholt CH, Hansen T, Pedersen O. Beyond the fourth wave of genome-wide obesity association studies. Nutr Diabetes. 2012;2:e37–e37.

5. Kleinendorst L, Abawi O, van der Voorn B, Jongejan MHTM, Brandsma AE, Visser JA, et al. Identifying underlying medical causes of pediatric obesity: Results of a systematic diagnostic approach in a pediatric obesity center. Buchner DA, editor. PLoS One. 2020;15:e0232990. https://doi.org/10.1371/journal.pone.0232990

6. D'Angelo CS, Varela MC, De Castro CIE, Otto PA, Perez ABA, Lourenço CM, et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. Mol Cytogenet. BioMed Central Ltd.; 2018;11.

7. Huvenne H, Dubern B, Clément K, Poitou C. Rare Genetic Forms of Obesity: Clinical Approach and Current Treatments in 2016. Obes Facts. 2016;9:158–73. https://doi.org/10.1159/000445061

8. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. Obes Rev. 2017;18:603–34.

https://doi.org/10.1111/obr.12531
9. Clément K, Mosbah H, Poitou C. Rare genetic forms of obesity: From gene to therapy. Physiol Behav. 2020;227:113134. https://doi.org/10.1016/j.physbeh.2020.113134
10. Geets E, Meuwissen MEC, Van Hul W. Clinical, molecular genetics and therapeutic aspects of syndromic obesity. Clin Genet. 2019;95:23–40. https://doi.org/10.1111/cge.13367
11. Delrue M-A, Michaud J. Fat chance: genetic syndromes with obesity. Clin Genet. 2004;66:83–93. https://doi.org/10.1111/j.0009-9163.2004.00300.x
12. Micleaa D, Al-Khzouza C, Osan S, Bucerzan S, Cret V, Popp RA, et al. Genomic study via chromosomal microarray analysis in a group of Romanian patients with obesity and developmental disability/intellectual disability. J Pediatr Endocrinol Metab. 2019;32:667–74. https://doi.org/10.1515/jpem-2018-0439
13. Campbell LVC. Genetics of obesity focus. Aust Fam Physician. 2017;46:456–9.
14. Milani D, Cerutti M, Pezzani L, Maffei P, Milan G, Esposito S. Syndromic obesity: clinical implications of a correct diagnosis. Ital J Pediatr. 2014;40:33. https://doi.org/10.1186/1824-7288-40-33
15. Wiegand S, Krude H. Monogene und syndromale Krankheitsbilder bei morbider Adipositas. Internist (Berl). 2015;56:111–20. https://doi.org/10.1007/s00108-014-3532-8
16. Wiegand S, Krude H. Monogene und syndromale Krankheitsbilder bei morbider Adipositas: Selten aber wichtig. Internist. Springer Verlag; 2015;56:111–20.
17. Geets E, Meuwissen MEC, Van Hul W. Clinical, molecular genetics and therapeutic aspects of syndromic obesity. Clin Genet. 2019;95:23–40. https://doi.org/10.1111/cge.13367
18. Kleinendorst L, Massink M, Cooiman MI, Savas M, van der Baan-Slootweg OH, Roelants RJ, et al. Genetic obesity: next-generation sequencing results of 1230 patients with obesityNo Title. J Med Genet. 2018;55:578–86.
19. Lionti T, Reid SM, White SM, Rowell MM. A population-based profile of 160 Australians with Prader-Willi syndrome: Trends in diagnosis, birth prevalence and birth characteristics. Am J Med Genet Part A. 2015;167:371–8. https://doi.org/10.1002/ajmg.a.36845
20. Thomson A, Glasson E, Bittles A. A long-term population-based clinical and morbidity review of Prader-Willi syndrome in Western Australia. J Intellect Disabil

Res. 2006;50:69–78.
21. Vogels A, Van Den Ende J, Keymolen K, Mmortier G, Devriendt K, Legius E, et al. Minimum prevalence, birth incidence and cause of death for Prader-Willi syndrome in Flanders. . Eur J Hum Genet. 2004;2012:238–40.
22. Whittington J, Holland A, Webb T, Butler J, Clarke D, Boer H. Population prevalence and estimated birth incidence and mortality rate for people with Prader-Willi syndrome in one UK Health Region. J Med Genet. 2001;38:792–8.
23. Ehara H, Kousaku O, Kenzo T. Frequency of the Prader –Willi syndrome in the San-in district. . Japan Brain Dev. 1995;17:324–6.
24. Burd L, Besely B, Martsolf J, Korbeshian J. Prevalence study of Prader-Willi syndrome in North Dakota. Am J Med Genet. 1990;37:97–9.
25. Akefeldt A, Gillberg C, Larsson C. Prader-Willi syndrome in a Swedish rural county: Epidemiological aspects. . Dev Med Child Neurol. 1991;33:715–21.
26. Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. Prader-Willi syndrome. Genet Med. 2012;14:10–26. https://doi.org/10.1038/gim.0b013e31822bead0
27. Fridman C, Koiffmann CP. Genomic imprinting: genetic mechanisms and phenotypic consequences in Prader-Willi and Angelman syndromes. Genet Mol Biol. 2000;23:715–24. https://doi.org/10.1590/S1415-47572000000400004
28. Butler MG, Miller JL, Forster JL. Prader-Willi Syndrome - Clinical Genetics, Diagnosis and Treatment Approaches: An Update. Curr Pediatr Rev. 2019;15:207–44. https://doi.org/10.2174/1573396315666190716120925
29. Kim CA, Albano LMJ, Bertola DR. Síndromes genéticas associadas à obesidade. Tratado de obesidade. 2nd ed. Rio de Janeiro – RJ: Guanabara Koogan; 2022. p. 253–61.
30. Khan MJ, Gerasimidis K, Edwards CA, Shaikh MG. Mechanisms of obesity in Prader–Willi syndrome. Pediatr. Obes. John Wiley and Sons Ltd; 2016. p. 3–13.
31. Einfeld SL, Kavanagh SJ, Smith A, Evans EJ, Tonge BJ, Taffe J. Mortality in Prader-Willi syndrome. Am J Ment Retard. 2006;111:193–8. https://doi.org/10.1352/0895-8017(2006)111[193:MIPS]2.0.CO;2
32. Deal CL, Tony M, Höybye C, Allen DB, Tauber M, Christiansen JS, et al. GrowthHormone Research Society workshop summary: consensus guidelines for recombinant human growth hormone therapy in Prader-Willi syndrome. J Clin Endocrinol Metab. 2013;98:E1072-87. https://doi.org/10.1210/jc.2012-3888
33. Alves C, Franco RR. Prader-Willi syndrome: endocrine manifestations and

management. Arch Endocrinol Metab. 2020;64:223–34. https://doi.org/10.20945/2359-3997000000248

34. Angulo MA, Butler MG, Cataletto ME. Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings. J Endocrinol Invest. 2015;38:1249–63. https://doi.org/10.1007/s40618-015-0312-9

35. Tauber M, Hoybye C. Endocrine disorders in Prader-Willi syndrome: a model to understand and treat hypothalamic dysfunction. Lancet Diabetes Endocrinol. 2021;9:235–46. https://doi.org/10.1016/S2213-8587(21)00002-4

36. Ehrhart F, Janssen KJM, Coort SL, Evelo CT, Curfs LMG. Prader-Willi syndrome and Angelman syndrome: Visualisation of the molecular pathways for two chromosomal disorders. World J Biol Psychiatry. 2019;20:670–82. https://doi.org/10.1080/15622975.2018.1439594

37. Bochukova EG. Transcriptomics of the Prader–Willi syndrome hypothalamus. 2021. p. 369–79. https://doi.org/10.1016/B978-0-12-820683-6.00027-0

38. Tauber M, Coupaye M, Diene G, Molinas C, Valette M, Beauloye V. Prader-Willi syndrome: A model for understanding the ghrelin system. J Neuroendocrinol. 2019;31. https://doi.org/10.1111/jne.12728

39. Brennan M-L, Adam MP, Seaver LH, Myers A, Schelley S, Zadeh N, et al. Increased body mass in infancy and early toddlerhood in Angelman syndrome patients with uniparental disomy and imprinting center defects. Am J Med Genet Part A. 2015;167:142–6. https://doi.org/10.1002/ajmg.a.36831

40. Larson AM, Shinnick JE, Shaaya EA, Thiele EA, Thibert RL. Angelman syndrome in adulthood. Am J Med Genet Part A. 2015;167:331–44. https://doi.org/10.1002/ajmg.a.36864

41. Elhamamsy AR. Role of DNA methylation in imprinting disorders: an updated review. J Assist Reprod Genet. 2017;34:549–62. https://doi.org/10.1007/s10815-017-0895-5

42. Kim S-J, Miller JL, Kuipers PJ, German JR, Beaudet AL, Sahoo T, et al. Unique and atypical deletions in Prader–Willi syndrome reveal distinct phenotypes. Eur J Hum Genet. 2012;20:283–90. https://doi.org/10.1038/ejhg.2011.187

43. Butler MG. Magnesium Supplement and the 15q11.2 BP1–BP2 Microdeletion (Burnside–Butler) Syndrome: A Potential Treatment? Int J Mol Sci. 2019;20:2914. https://doi.org/10.3390/ijms20122914

44. Khan SA, Muhammad N, Khan MA, Kamal A, Rehman ZU, Khan S. Genetics of

human Bardet-Biedl syndrome, an updates. Clin Genet. 2016;90:3–15. https://doi.org/10.1111/cge.12737
45. Florea L, Caba L, Gorduza EV. Bardet–Biedl Syndrome—Multiple Kaleidoscope Images: Insight into Mechanisms of Genotype–Phenotype Correlations. Genes (Basel). 2021;12:1353. https://doi.org/10.3390/genes12091353
46. Beales PL, Warner AM, Hitman GA, Thakker R, Flinter FA. Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. J Med Genet. 1997;34:92–8. https://doi.org/10.1136/jmg.34.2.92
47. Farag TI, Teebi AS. High incidence of Bardet Biedl syndrome among the Bedouin. Clin Genet. 2008;36:463–4. https://doi.org/10.1111/j.1399-0004.1989.tb03378.x
48. Klein D, Ammann F. The syndrome of Laurence-Moon-Bardet-Biedl and allied diseases in Switzerland. J Neurol Sci. 1969;9:479–513. https://doi.org/10.1016/0022-510X(69)90091-4
49. M'hamdi O, Ouertani I, Maazoul F, Chaabouni-Bouhamed H. Prevalence of Bardet–Biedl syndrome in Tunisia. J Community Genet. 2011;2:97–9. https://doi.org/10.1007/s12687-011-0040-6
50. Moore S, Green J, Fan Y et al. Clinical and genetic epidemiology of Bardet–Biedl syndrome in Newfoundland: a 22-year prospective, population-based cohort study. . Am J Med Genet A. 2005;132:352–60.
51. Suspitsin EN, Imyanitov EN. Bardet-Biedl Syndrome. Mol Syndromol. 2016;7:62–71. https://doi.org/10.1159/000445491
52. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet . 1999;36:347–446.
53. Haws RM, McIntee TJ, Green CB. Cutaneous findings in Bardet-Biedl syndrome. Int J Dermatol. 2019;58:1160–4. https://doi.org/10.1111/ijd.14412
54. Tsang SH, Aycinena ARP, Sharma T. Ciliopathy: Bardet-Biedl Syndrome. 2018. p. 171–4. https://doi.org/10.1007/978-3-319-95046-4_33
55. Muller J, Stoetzel C, Vincent MC, Leitch CC, Laurier V, Danse JM, et al. Identification of 28 novel mutations in the Bardet–Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. Hum Genet. 2010;127:583–93. https://doi.org/10.1007/s00439-010-0804-9
56. Priya S, Nampoothiri S, Sen P, Sripriya S. Bardet–Biedl syndrome: Genetics,

molecular pathophysiology, and disease management. Indian J Ophthalmol. 2016;64:620. https://doi.org/10.4103/0301-4738.194328
57. Nishimura DY, Fath M, Mullins RF, Searby C, Andrews M, Davis R, et al. Bbs2-null mice have neurosensory deficits, a defect in social dominance, and retinopathy associated with mislocalization of rhodopsin. Proc Natl Acad Sci. 2004;101:16588–93. https://doi.org/10.1073/pnas.0405496101
58. Mykytyn K, Mullins RF, Andrews M, Chiang AP, Swiderski RE, Yang B, et al. Bardet-Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly. Proc Natl Acad Sci. 2004;101:8664–9. https://doi.org/10.1073/pnas.0402354101
59. Fath MA, Mullins RF, Searby C, Nishimura DY, Wei J, Rahmouni K, et al. Mkks-null mice have a phenotype resembling Bardet–Biedl syndrome. Hum Mol Genet. 2005;14:1109–18. https://doi.org/10.1093/hmg/ddi123
60. Rahmouni K, Fath MA, Seo S, Thedens DR, Berry CJ, Weiss R, et al. Leptin resistance contributes to obesity and hypertension in mouse models of Bardet-Biedl syndrome. J Clin Invest. 2008;118:1458–67. https://doi.org/10.1172/JCI32357
61. Bilezikian JP. Hypoparathyroidism. J Clin Endocrinol Metab. 2020;105:1722–36. https://doi.org/10.1210/clinem/dgaa113
62. Thiele S, Mantovani G, Barlier A, Boldrin V, Bordogna P, De Sanctis L, et al. From pseudohypoparathyroidism to inactivating PTH/PTHrP signalling disorder (iPPSD), a novel classification proposed by the EuroPHP network. Eur J Endocrinol. 2016;175:P1–17. https://doi.org/10.1530/EJE-16-0107
63. Mantovani G, Bastepe M, Monk D, de Sanctis L, Thiele S, Usardi A, et al. Diagnosis and management of pseudohypoparathyroidism and related disorders: first international Consensus Statement. Nat Rev Endocrinol. 2018;14:476–500. https://doi.org/10.1038/s41574-018-0042-0
64. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. Pseudohypoparathyroidism - epidemiology, mortality and risk of complications. Clin Endocrinol (Oxf). 2016;84:904–11. https://doi.org/10.1111/cen.12948
65. Butler MG. Imprinting disorders in humans: a review. Curr Opin Pediatr. 2020;32:719–29. https://doi.org/10.1097/MOP.0000000000000965
66. Jüppner H. Molecular Definition of Pseudohypoparathyroidism Variants. J Clin Endocrinol Metab. 2021;106:1541–52. https://doi.org/10.1210/clinem/dgab060
67. Turan S, Bastepe M. GNAS Spectrum of Disorders. Curr Osteoporos Rep.

2015;13:146–58. https://doi.org/10.1007/s11914-015-0268-x
68. Hendy GN, Cole DE, Bastepe M. Hypoparathyroidism and Pseudohypoparathyroidism [Internet]. Endotext. 2000. https://doi.org/25905388
69. Linglart A, Levine MA, Jüppner H. Pseudohypoparathyroidism. Endocrinol Metab Clin North Am. 2018;47:865–88. https://doi.org/10.1016/j.ecl.2018.07.011
70. Lu X-Y, Barsh GS, Akil H, Watson SJ. Interaction between alpha-melanocyte-stimulating hormone and corticotropin-releasing hormone in the regulation of feeding and hypothalamo-pituitary-adrenal responses. J Neurosci. 2003;23:7863–72. https://doi.org/10.1523/JNEUROSCI.23-21-07863.2003
71. Dixit A, Chandler KE, Lever M, Poole RL, Bullman H, Mughal MZ, et al. Pseudohypoparathyroidism Type 1b due to Paternal Uniparental Disomy of Chromosome 20q. J Clin Endocrinol Metab. 2013;98:E103–8. https://doi.org/10.1210/jc.2012-2639
72. Bastepe M, Altug-Teber Ö, Agarwal C, Oberfield SE, Bonin M, Jüppner H. Paternal uniparental isodisomy of the entire chromosome 20 as a molecular cause of pseudohypoparathyroidism type Ib (PHP-Ib). Bone. 2011;48:659–62. https://doi.org/10.1016/j.bone.2010.10.168
73. Fernández-Rebollo E, Lecumberri B, Garin I, Arroyo J, Bernal-Chico A, Goñi F, et al. New mechanisms involved in paternal 20q disomy associated with pseudohypoparathyroidism. Eur J Endocrinol. 2010;163:953–62. https://doi.org/10.1530/EJE-10-0435
74. Mendes de Oliveira E, Keogh JM, Talbot F, Henning E, Ahmed R, Perdikari A, et al. Obesity-Associated GNAS Mutations and the Melanocortin Pathway. N Engl J Med. 2021;385:1581–92. https://doi.org/10.1056/NEJMoa2103329
75. Ioannides Y, Lokulo-Sodipe K, Mackay DJG, Davies JH, Temple IK. Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. J Med Genet. 2014;51:495–501. https://doi.org/10.1136/jmedgenet-2014-102396
76. Prasasya R, Grotheer K V, Siracusa LD, Bartolomei MS. Temple syndrome and Kagami-Ogata syndrome: clinical presentations, genotypes, models and mechanisms. Hum Mol Genet. 2020;29:R107–16. https://doi.org/10.1093/hmg/ddaa133
77. Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, et al. Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. Nat Genet. 2008;40:237–42.

https://doi.org/10.1038/ng.2007.56
78. Bertini V, Fogli A, Bruno R, Azzarà A, Michelucci A, Mattina T, et al. Maternal Uniparental Disomy 14 (Temple Syndrome) as a Result of a Robertsonian Translocation. Mol Syndromol. 2017;8:131–8. https://doi.org/10.1159/000456062
79. Wang X, Pang H, Shah BA, Gu H, Zhang L, Wang H. A Male Case of Kagami-Ogata Syndrome Caused by Paternal Unipaternal Disomy 14 as a Result of a Robertsonian Translocation. Front Pediatr. 2020;8. https://doi.org/10.3389/fped.2020.00088
80. Suzumori N, Kagami M, Kumagai K, Goto S, Matsubara K, Sano S, et al. Clinical and molecular findings in a patient with 46,XX/47,XX,+14 mosaicism caused by postzygotic duplication of a paternally derived chromosome 14. Am J Med Genet Part A. 2015;167:2474–7. https://doi.org/10.1002/ajmg.a.37194
81. Haug MG, Brendehaug A, Houge G, Kagami M, Ogata T. Mosaic upd(14)pat in a patient with mild features of Kagami-Ogata syndrome. Clin Case Reports. 2018;6:91–5. https://doi.org/10.1002/ccr3.1300
82. Yakoreva M, Kahre T, Pajusalu S, Ilisson P, Žilina O, Tillmann V, et al. A New Case of a Rare Combination of Temple Syndrome and Mosaic Trisomy 14 and a Literature Review. Mol Syndromol. 2018;9:182–9. https://doi.org/10.1159/000489446
83. Kagami M, Nishimura G, Okuyama T, Hayashidani M, Takeuchi T, Tanaka S, et al. Segmental and full paternal isodisomy for chromosome 14 in three patients: Narrowing the critical region and implication for the clinical features. Am J Med Genet Part A. 2005;138A:127–32. https://doi.org/10.1002/ajmg.a.30941
84. Dauber A, Cunha-Silva M, Macedo DB, Brito VN, Abreu AP, Roberts SA, et al. Paternally Inherited DLK1 Deletion Associated With Familial Central Precocious Puberty. J Clin Endocrinol Metab. 2017;102:1557–67. https://doi.org/10.1210/jc.2016-3677
85. Macedo DB, Kaiser UB. DLK1, Notch Signaling and the Timing of Puberty. Semin Reprod Med. 2019;37:174–81. https://doi.org/10.1055/s-0039-3400963
86. Gomes LG, Cunha-Silva M, Crespo RP, Ramos CO, Montenegro LR, Canton A, et al. DLK1 Is a Novel Link Between Reproduction and Metabolism. J Clin Endocrinol Metab. 2019;104:2112–20. https://doi.org/10.1210/jc.2018-02010
87. Gimeno-Ferrer F, Albuquerque D, Guzmán Luján C, Marcaida Benito G, Torreira Banzas C, Repáraz-Andrade A, et al. The effect of copy number variations in chromosome 16p on body weight in patients with intellectual disability. J Hum Genet.

2019;64:221–31. https://doi.org/10.1038/s10038-018-0545-5
88. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, et al. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet. 2012;49:660–8. https://doi.org/10.1136/jmedgenet-2012-101203
89. Chung WK, Roberts TP, Sherr EH, Snyder LG, Spiro JE. 16p11.2 deletion syndrome. Curr Opin Genet Dev. 2021;68:49–56. https://doi.org/10.1016/j.gde.2021.01.011
90. Goldenberg P. An Update on Common Chromosome Microdeletion and Microduplication Syndromes. Pediatr Ann. United States; 2018;47:e198–203.
91. Levkova M, Stoyanova M, Staneva R, Hachmeriyan M, Angelova L. 16p11.2 Duplication Syndrome - a Case Report. Folia Med (Plovdiv). 2021;63:138–41. https://doi.org/10.3897/folmed.63.e52763
92. Liangyou R. SH2B1 regulation of energy balance, body weight, and glucose metabolism. World J Diabetes. 2014;5:511–526.
93. Doche ME, Bochukova EG, Su H-W, Pearce LR, Keogh JM, Henning E, et al. Human SH2B1 mutations are associated with maladaptive behaviors and obesity. J Clin Invest. 2012;122:4732–6. https://doi.org/10.1172/JCI62696
94. Gimeno-Ferrer F, Albuquerque D, Guzmán Luján C, Marcaida Benito G, Torreira Banzas C, Repáraz-Andrade A, et al. The effect of copy number variations in chromosome 16p on body weight in patients with intellectual disability. J Hum Genet. England; 2019;64:221–31.
95. Jiang L, Su H, Wu X, Shen H, Kim M-H, Li Y, et al. Leptin receptor-expressing neuron Sh2b1 supports sympathetic nervous system and protects against obesity and metabolic disease. Nat Commun. 2020;11:1517. https://doi.org/10.1038/s41467-020-15328-3
96. Scott D, Jordan V, Zaveri H. 1p36 deletion syndrome: an update. Appl Clin Genet. 2015;189. https://doi.org/10.2147/TACG.S65698
97. Rocha CF, Vasques RB, Santos SR, Paiva CLA. Mini-Review Monosomy 1p36 syndrome: reviewing the correlation between deletion sizes and phenotypes. Genet Mol Res. 2016;15. https://doi.org/10.4238/gmr.15017942
98. D'Angelo CS, Kohl I, Varela MC, de Castro CIE, Kim CA, Bertola DR, et al. Extending the phenotype of monosomy 1p36 syndrome and mapping of a critical region for obesity and hyperphagia. Am J Med Genet Part A. 2010;152A:102–10.

https://doi.org/10.1002/ajmg.a.33160
99. Verhoeven WMA, Kleefstra T, Egger JIM. Kleefstra syndrome: neuropsychiatric sequelae. Eur Psychiatry. 2011;26:819–819. https://doi.org/10.1016/S0924-9338(11)72524-X
100. Kleefstra T, de Leeuw N. Kleefstra Syndrome. GeneReviews. 2010.
101. Kleefstra T et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. . J Med Genet . 2009;46:598–606.
102. Kleefstra Syndrome. J Coll Physicians Surg Pakistan. 2022;32:S76–8. https://doi.org/10.29271/jcpsp.2022.Supp1.S76
103. Koemans TS, Kleefstra T, Chubak MC, Stone MH, Reijnders MRF, de Munnik S, et al. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLOS Genet. 2017;13:e1006864. https://doi.org/10.1371/journal.pgen.1006864
104. Fiszbein A, Giono LE, Quaglino A, Berardino BG, Sigaut L, von Bilderling C, et al. Alternative Splicing of G9a Regulates Neuronal Differentiation. Cell Rep. 2016;14:2797–808. https://doi.org/10.1016/j.celrep.2016.02.063
105. Deimling SJ, Olsen JB, Tropepe V. The expanding role of the Ehmt2/G9a complex in neurodevelopment. Neurogenesis. 2017;4:e1316888. https://doi.org/10.1080/23262133.2017.1316888
106. Balemans MCM, Nadif Kasri N, Kopanitsa M V., Afinowi NO, Ramakers G, Peters TA, et al. Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. Hum Mol Genet. 2013;22:852–66. https://doi.org/10.1093/hmg/dds490
107. Jahani-Asl A, Cheng C, Zhang C, Bonni A. Pathogenesis of Börjeson-Forssman-Lehmann syndrome: Insights from PHF6 function. Neurobiol Dis. 2016;96:227–35. https://doi.org/10.1016/j.nbd.2016.09.011
108. Bellad A, Bandari AK, Pandey A, Girimaji SC, Muthusamy B. A Novel Missense Variant in PHF6 Gene Causing Börjeson-Forssman-Lehman Syndrome. J Mol Neurosci. 2020;70:1403–9. https://doi.org/10.1007/s12031-020-01560-5
109. Kasper BS, Dörfler A, Di Donato N, Kasper EM, Wieczorek D, Hoyer J, et al. Central nervous system anomalies in two females with Borjeson-Forssman-Lehmann syndrome. Epilepsy Behav. 2017;69:104–9. https://doi.org/10.1016/j.yebeh.2017.01.022

110. Loontiens S, Dolens A-C, Strubbe S, Van de Walle I, Moore FE, Depestel L, et al. PHF6 Expression Levels Impact Human Hematopoietic Stem Cell Differentiation. Front Cell Dev Biol. 2020;8. https://doi.org/10.3389/fcell.2020.599472
111. Di Donato N, Isidor B, Lopez Cazaux S, Le Caignec C, Klink B, Kraus C, et al. Distinct phenotype of PHF6 deletions in females. Eur J Med Genet. 2014;57:85–9. https://doi.org/10.1016/j.ejmg.2013.12.003
112. Wieczorek D, Bögershausen N, Beleggia F, Steiner-Haldenstätt S, Pohl E, Li Y, et al. A comprehensive molecular study on Coffin–Siris and Nicolaides–Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum Mol Genet. 2013;22:5121–35. https://doi.org/10.1093/hmg/ddt366
113. Zweier C, Kraus C, Brueton L, Cole T, Degenhardt F, Engels H, et al. A new face of Borjeson–Forssman–Lehmann syndrome? De novo mutations in PHF6 in seven females with a distinct phenotype. J Med Genet. 2013;50:838–47. https://doi.org/10.1136/jmedgenet-2013-101918
114. Cheng C, Deng P-Y, Ikeuchi Y, Yuede C, Li D, Rensing N, et al. Characterization of a Mouse Model of Börjeson-Forssman-Lehmann Syndrome. Cell Rep. 2018;25:1404-1414.e6. https://doi.org/10.1016/j.celrep.2018.10.043
115. Zhang C, Mejia L, et al. The X-linked intellectual disability protein PHF6 associates with the PAF1 complex and regulates neuronal migration in the mammalian brain. . Neuron. 2013;78:986–93.
116. Duplomb L, Duvet S, Picot D, Jego G, El Chehadeh-Djebbar S, Marle N, et al. Cohen syndrome is associated with major glycosylation defects. Hum Mol Genet. 2014;23:2391–9. https://doi.org/10.1093/hmg/ddt630
117. Rodrigues JM, Fernandes HD, Caruthers C, Braddock SR, Knutsen AP. Cohen Syndrome: Review of the Literature. Cureus. 2018; https://doi.org/10.7759/cureus.3330
118. Roxo-Junior P, Mina I. Cohen Syndrome. Encycl Med Immunol. 2018. p. 1–2. https://doi.org/10.1007/978-1-4614-9209-2_157-1
119. Chandler KE. Diagnostic criteria, clinical characteristics, and natural history of Cohen syndrome. J Med Genet. 2003;40:233–41. https://doi.org/10.1136/jmg.40.4.233
120. Delcourt M, Riant F, Mancini J, Milh M, Navarro V, Roze E, et al. Severe phenotypic spectrum of biallelic mutations in PRRT2 gene. J Neurol Neurosurg

Psychiatry. England; 2015;86:782–5.
121. Wang H, Falk M, Wensel C, Traboulsi E. Cohen Syndrome. GeneReview. 2006.
122. Momtazmanesh S, Rayzan E, Shahkarami S, Rohlfs M, Klein C, Rezaei N. A novel VPS13B mutation in Cohen syndrome: a case report and review of literature. BMC Med Genet. 2020;21:140. https://doi.org/10.1186/s12881-020-01075-1
123. Rodrigues JM, Fernandes HD, Caruthers C, Braddock SR, Knutsen AP. Cohen Syndrome: Review of the Literature. Cureus. Cureus, Inc.; 2018;
124. Kääriäinen H, Muilu J, Perola M, Kristiansson K. Genetics in an isolated population like Finland: a different basis for genomic medicine? J Community Genet. 2017;8:319–26. https://doi.org/10.1007/s12687-017-0318-4
125. Hearn T. ALMS1 and Alström syndrome: a recessive form of metabolic, neurosensory and cardiac deficits. J Mol Med. 2019;97:1–17. https://doi.org/10.1007/s00109-018-1714-x
126. Marshall JD, Muller J, Collin GB, Milan G, Kingsmore SF, Dinwiddie D, et al. Alström Syndrome: Mutation Spectrum of *ALMS1*. Hum Mutat. 2015;36:660–8. https://doi.org/10.1002/humu.22796
127. Tsang SH, Aycinena ARP, Sharma T. Ciliopathy: Alström Syndrome. 2018. p. 179–80. https://doi.org/10.1007/978-3-319-95046-4_35
128. Roy Choudhury A, Munonye I, Sanu KP, Islam N, Gadaga C. A review of Alström syndrome: a rare monogenic ciliopathy. Intractable Rare Dis Res. 2021;10:2021.01113. https://doi.org/10.5582/irdr.2021.01113
129. Paisey R, Steeds R, Barrett T, Williams D, Geberhiwot T, Gunay-Aygun M. Alström Syndrome. GeneReviews. 2003.
130. Tahani N, Maffei P, Dollfus H, Paisey R, Valverde D, Milan G, et al. Consensus clinical management guidelines for Alström syndrome. Orphanet J Rare Dis. 2020;15:253. https://doi.org/10.1186/s13023-020-01468-8
131. Haye D, Collet C, Sembely-Taveau C, Haddad G, Denis C, Soulé N, et al. Prenatal findings in carpenter syndrome and a novel mutation in *RAB23*. Am J Med Genet Part A. 2014;164:2926–30. https://doi.org/10.1002/ajmg.a.36726
132. Jenkins D, Baynam G, De Catte L, Elcioglu N, Gabbett MT, Hudgins L, et al. Carpenter syndrome: extended *RAB23* mutation spectrum and analysis of nonsense-mediated mRNA decay. Hum Mutat. 2011;32. https://doi.org/10.1002/humu.21457
133. Khairat R, Elhossini R, Sobreira N, Wohler E, Otaify G, Mohamed AM, et al. Expansion of the phenotypic and mutational spectrum of Carpenter syndrome. Eur J

Med Genet. 2022;65:104377. https://doi.org/10.1016/j.ejmg.2021.104377
134. Richieri-Costa A, Pirolo L, Cohen MM. Carpenter syndrome with normal intelligence: Brazilian girl born to consanguineous parents. Am J Med Genet. 1993;47:281–3. https://doi.org/10.1002/ajmg.1320470228
135. Ingham PW, Nakano Y, Seger C. Mechanisms and functions of Hedgehog signalling across the metazoa. Nat Rev Genet. 2011;12:393–406. https://doi.org/10.1038/nrg2984
136. Jenkins D, Seelow D, Jehee FS, Perlyn CA, Alonso LG, Bueno DF, et al. RAB23 Mutations in Carpenter Syndrome Imply an Unexpected Role for Hedgehog Signaling in Cranial-Suture Development and Obesity. Am J Hum Genet. 2007;80:1162–70. https://doi.org/10.1086/518047
137. Hor CH, Goh EL. Small GTPases in hedgehog signalling: emerging insights into the disease mechanisms of Rab23-mediated and Arl13b-mediated ciliopathies. Curr Opin Genet Dev. 2019;56:61–8. https://doi.org/10.1016/j.gde.2019.07.009
138. Engle SE, Bansal R, Antonellis PJ, Berbari NF. Cilia signaling and obesity. Semin Cell Dev Biol. 2021;110:43–50. https://doi.org/10.1016/j.semcdb.2020.05.006
139. Twigg S, et al. Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. Am J Hum . 2012;91:897–905.
140. Engelhard C, Sarsfield S, Merte J, Wang Q, Li P, Beppu H, et al. MEGF8 is a modifier of BMP signaling in trigeminal sensory neurons. Elife. 2013;2. https://doi.org/10.7554/eLife.01160
141. Yu S, Zhou C, Cao S, He J, Cai B, Wu K, et al. BMP4 resets mouse epiblast stem cells to naive pluripotency through ZBTB7A/B-mediated chromatin remodelling. Nat Cell Biol. 2020;22:651–62. https://doi.org/10.1038/s41556-020-0516-x
142. Xu J, Chen M, Yan Y, Zhao Q, Shao M, Huang Z. The effects of altered BMP4 signaling in first branchial-arch-derived murine embryonic orofacial tissues. Int J Oral Sci. 2021;13:40. https://doi.org/10.1038/s41368-021-00142-4
143. Jacoby M, Cox JJ, Gayral S, Hampshire DJ, Ayub M, Blockmans M, et al. INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. Nat Genet. 2009;41:1027–31. https://doi.org/10.1038/ng.427
144. Hampshire DJ, Ayub M, Springell K, Roberts E, Jafri H, Rashid Y, et al. MORM

syndrome (mental retardation, truncal obesity, retinal dystrophy and micropenis), a new autosomal recessive disorder, links to 9q34. Eur J Hum Genet. 2006;14:543–8. https://doi.org/10.1038/sj.ejhg.5201577
145. Drole Torkar A, Avbelj Stefanija M, Bertok S, Trebušak Podkrajšek K, Debeljak M, Stirn Kranjc B, et al. Novel Insights Into Monogenic Obesity Syndrome Due to INPP5E Gene Variant: A Case Report of a Female Patient. Front Endocrinol (Lausanne). 2021;12. https://doi.org/10.3389/fendo.2021.581134
146. Zhang R, Tang J, Li T, Zhou J, Pan W. INPP5E and Coordination of Signaling Networks in Cilia. Front Mol Biosci. 2022;9. https://doi.org/10.3389/fmolb.2022.885592
147. Coursimault J, Guerrot A-M, Morrow MM, Schramm C, Zamora FM, Shanmugham A, et al. MYT1L-associated neurodevelopmental disorder: description of 40 new cases and literature review of clinical and molecular aspects. Hum Genet. 2022;141:65–80. https://doi.org/10.1007/s00439-021-02383-z
148. Carvalho LML, D'Angelo CS, Mustacchi Z, da Silva IT, Krepischi AC V., Koiffmann CP, et al. A novel MYT1L mutation in a boy with syndromic obesity: Case report and literature review. Obes Res Clin Pract. 2021; https://doi.org/10.1016/j.orcp.2021.01.001
149. Blanchet P, Bebin M, Bruet S, Cooper GM, Thompson ML, Duban-Bedu B, et al. MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus. PLOS Genet. 2017;13:e1006957. https://doi.org/10.1371/journal.pgen.1006957
150. De Rocker N, Vergult S, Koolen D, Jacobs E, Hoischen A, Zeesman S, et al. Refinement of the critical 2p25.3 deletion region: The role of MYT1L in intellectual disability and obesity. Genet Med. Nature Publishing Group; 2015;17:460–6.
151. Windheuser IC, Becker J, Cremer K, Hundertmark H, Yates LM, Mangold E, et al. Nine newly identified individuals refine the phenotype associated with *MYT1L* mutations. Am J Med Genet Part A. 2020;182:1021–31. https://doi.org/10.1002/ajmg.a.61515
152. Windheuser IC, Becker J, Cremer K, Hundertmark H, Yates LM, Mangold E, et al. Nine newly identified individuals refine the phenotype associated with MYT1L mutations. Am J Med Genet Part A. Wiley-Liss Inc.; 2020;182:1021–31.
153. Wöhr M, Fong WM, Janas JA, Mall M, Thome C, Vangipuram M, et al. Myt1l haploinsufficiency leads to obesity and multifaceted behavioral alterations in mice.

Mol Autism. 2022;13:19. https://doi.org/10.1186/s13229-022-00497-3
154. Miskinyte G, Devaraju K, Grønning Hansen M, Monni E, Tornero D, Woods NB, et al. Direct conversion of human fibroblasts to functional excitatory cortical neurons integrating into human neural networks. Stem Cell Res Ther. BioMed Central Ltd.; 2017;8.
155. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature. 2010;463:1035–41. https://doi.org/10.1038/nature08797
156. Liang XG, Tan C, Wang CK, Tao RR, Huang YJ, Ma KF, et al. Myt1l induced direct reprogramming of pericytes into cholinergic neurons. CNS Neurosci Ther. Blackwell Publishing Ltd; 2018;24:801–9.
157. Chen J, Yen A, Florian CP, Dougherty JD. MYT1L in the making: emerging insights on functions of a neurodevelopmental disorder gene. Transl Psychiatry. 2022;12:292. https://doi.org/10.1038/s41398-022-02058-x
158. Pappas J, Rabin R. SETD2 Neurodevelopmental Disorders. GeneReviews. Seattle; 2022.
159. Luscan A, Laurendeau I, Malan V, Francannet C, Odent S, Giuliano F, et al. Mutations in SETD2 cause a novel overgrowth condition. J Med Genet. 2014;51:512–7. https://doi.org/10.1136/jmedgenet-2014-102402
160. Lumish HS, Wynn J, Devinsky O, Chung WK. Brief Report: SETD2 Mutation in a Child with Autism, Intellectual Disabilities and Epilepsy. J Autism Dev Disord. 2015;45:3764–70. https://doi.org/10.1007/s10803-015-2484-8
161. van Rij MC, Hollink IHIM, Terhal PA, Kant SG, Ruivenkamp C, van Haeringen A, et al. Two novel cases expanding the phenotype of SETD2 -related overgrowth syndrome. Am J Med Genet Part A. 2018;176:1212–5. https://doi.org/10.1002/ajmg.a.38666
162. White SM, Thompson EM, Kidd A, Savarirayan R, Turner A, Amor D, et al. Growth, behavior, and clinical findings in 27 patients with Kabuki (Niikawa-Kuroki) syndrome. Am J Med Genet. 2004;127A:118–27. https://doi.org/10.1002/ajmg.a.20674
163. Milani D, Manzoni FMP, Pezzani L, Ajmone P, Gervasini C, Menni F, et al. Rubinstein-Taybi syndrome: clinical features, genetic basis, diagnosis, and management. Ital J Pediatr. 2015;41:4.
164. D'Angelo CS, Kohl I, Varela MC, de Castro CIE, Kim CA, Bertola DR, et al.

Obesity with associated developmental delay and/or learning disability in patients exhibiting additional features: Report of novel pathogenic copy number variants. Am J Med Genet Part A. 2013;161:479–86. https://doi.org/10.1002/ajmg.a.35761
165. Vuillaume ML, Naudion S, Banneau G, Diene G, Cartault A, Cailley D, et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. Am J Med Genet Part A. Wiley-Liss Inc.; 2014;164:1965–75.
166. Carvalho LML, D'Angelo CS, Villela D, da Costa SS, de Lima Jorge AA, da Silva IT, et al. Genetic investigation of syndromic forms of obesity. Int J Obes. 2022; https://doi.org/10.1038/s41366-022-01149-5
167. Emerson E, Robertson J, Baines S, Hatton C. Obesity in British children with and without intellectual disability: cohort study. BMC Public Health. 2016;16:644. https://doi.org/10.1186/s12889-016-3309-1
168. Hsieh K, Rimmer JH, Heller T. Obesity and associated factors in adults with intellectual disability. J Intellect Disabil Res. 2014;58:851–63. https://doi.org/10.1111/jir.12100
169. Melville CA, Cooper S-A, Morrison J, Allan L, Smiley E, Williamson A. The Prevalence and Determinants of Obesity in Adults with Intellectual Disabilities. J Appl Res Intellect Disabil. 2008;21:425–37. https://doi.org/10.1111/j.1468-3148.2007.00412.x
170. Melville CA, Oppewal A, Schäfer Elinder L, Freiberger E, Guerra-Balic M, Hilgenkamp TIM, et al. Definitions, measurement and prevalence of sedentary behaviour in adults with intellectual disabilities — A systematic review. Prev Med (Baltim). 2017;97:62–71. https://doi.org/10.1016/j.ypmed.2016.12.052
171. de Winter CF, Bastiaanse LP, Hilgenkamp TIM, Evenhuis HM, Echteld MA. Overweight and obesity in older people with intellectual disability. Res Dev Disabil. 2012;33:398–405. https://doi.org/10.1016/j.ridd.2011.09.022
172. Simon GE, Von Korff M, Saunders K, Miglioretti DL, Crane PK, van Belle G, et al. Association Between Obesity and Psychiatric Disorders in the US Adult Population. Arch Gen Psychiatry. 2006;63:824. https://doi.org/10.1001/archpsyc.63.7.824
173. Chao AM, Wadden TA, Berkowitz RI. Obesity in Adolescents with Psychiatric Disorders. Curr Psychiatry Rep. 2019;21:3. https://doi.org/10.1007/s11920-019-0990-7

174. Avila C, Holloway AC, Hahn MK, Morrison KM, Restivo M, Anglin R, et al. An Overview of Links Between Obesity and Mental Health. Curr Obes Rep. 2015;4:303–10. https://doi.org/10.1007/s13679-015-0164-9

175. Martins LB, Monteze NM, Calarge C, Ferreira AVM, Teixeira AL. Pathways linking obesity to neuropsychiatric disorders. Nutrition. 2019;66:16–21. https://doi.org/10.1016/j.nut.2019.03.017

176. Beccuti G, Pannain S. Sleep and obesity. Curr Opin Clin Nutr Metab Care. 2011;14:402–12. https://doi.org/10.1097/MCO.0b013e3283479109

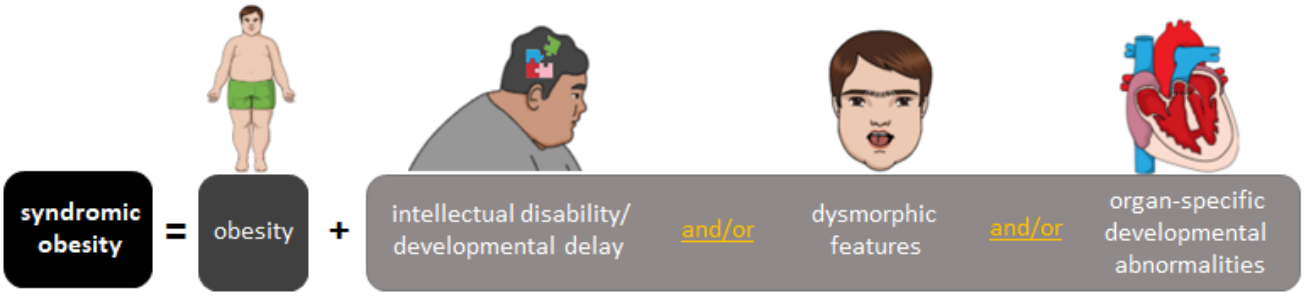
177. Mackay DJG, Temple IK. Human imprinting disorders: Principles, practice, problems and progress. Eur J Med Genet. 2017;60:618–26. https://doi.org/10.1016/j.ejmg.2017.08.014

178. Lewis EM, Kroll KL. Development and disease in a dish: the epigenetics of neurodevelopmental disorders. Epigenomics. 2018;10:219–31. https://doi.org/10.2217/epi-2017-0113

179. Blesson A, Cohen JS. Genetic Counseling in Neurodevelopmental Disorders. Cold Spring Harb Perspect Med. 2020;10:a036533. https://doi.org/10.1101/cshperspect.a036533

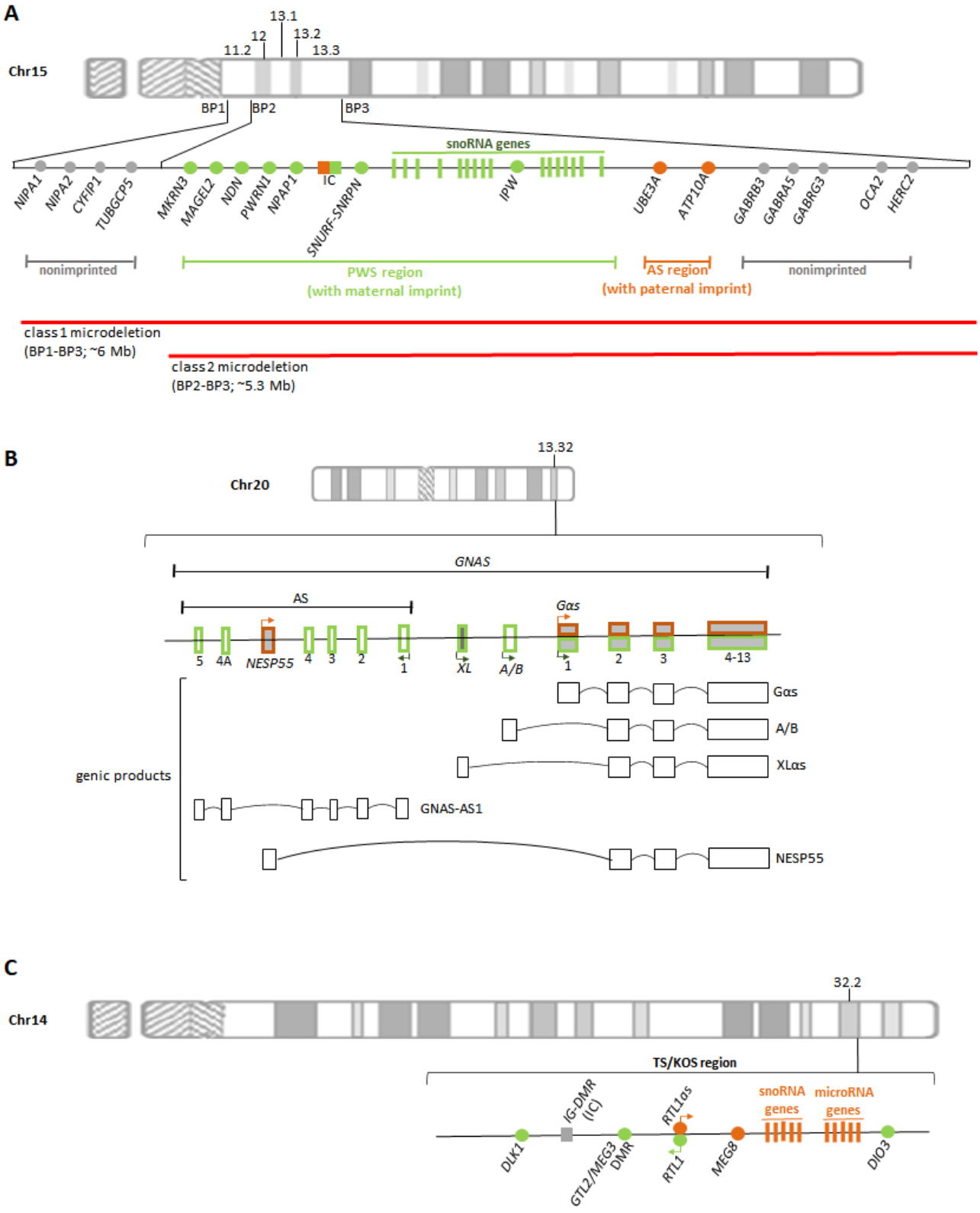
FIGURES

Figure 1 – Syndromic obesity definition



Syndromic obesity definition. Dysmorphic features are exemplified by synophrys and hypertelorism. Organ-specific developmental abnormalities are represented here by a ventricular septal defect

Figure 2 – Syndromic obesity mechanisms involving uniparental imprinting



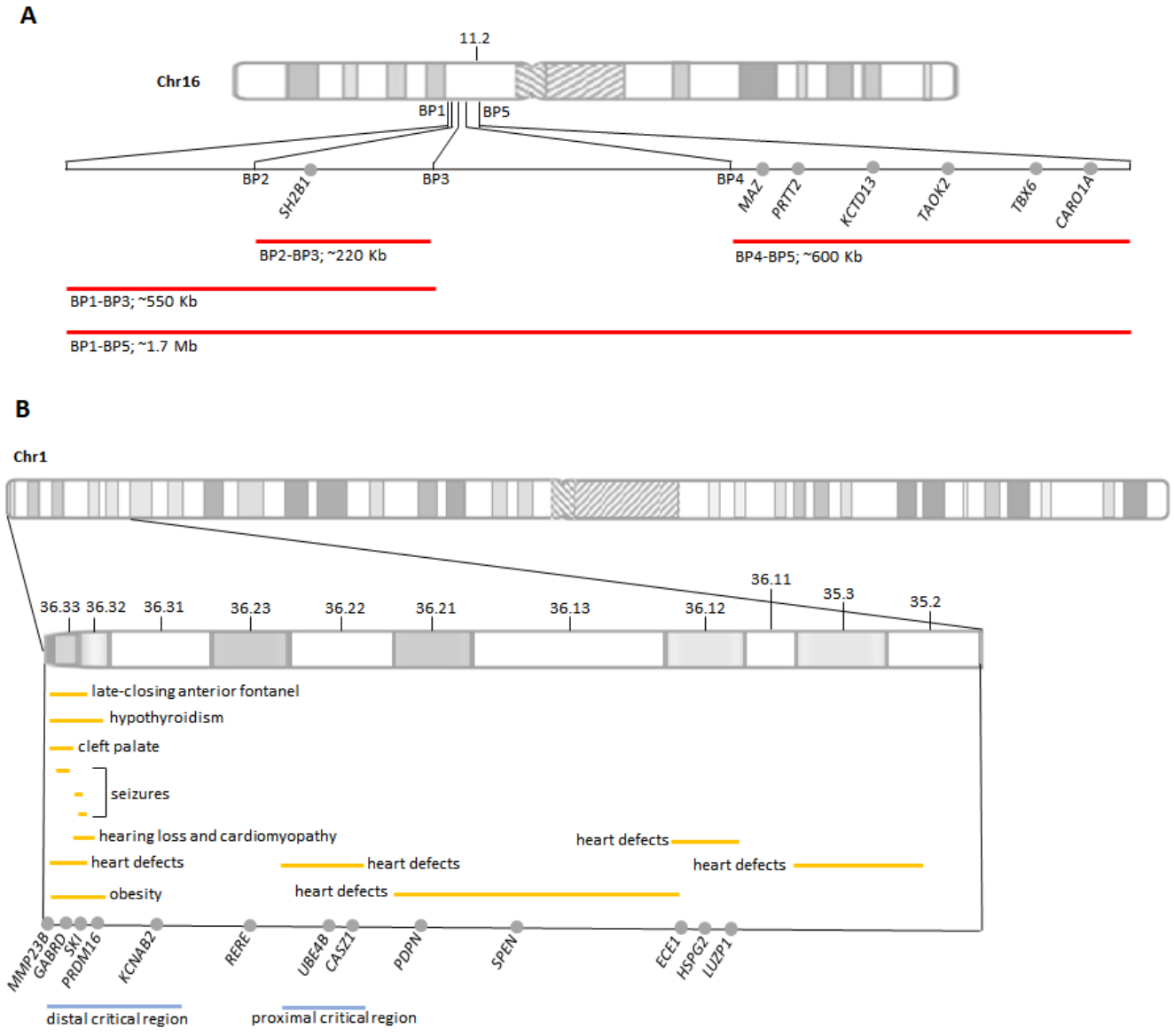
A, B, and C - The genes in this region that are expressed exclusively from the paternal chromosome (maternal imprinting) are in green, while those expressed from the maternal chromosome (paternal imprinting) are in orange. Not drawn to scale.

A - PWS – Prader-Willi syndrome; AS - Angelman syndrome; BP1 - breakpoint 1; BP2 - breakpoint 2; BP3 - breakpoint 3; IC - imprinting centers for PWS (green) and AS (orange).

B – *GNAS* locus. Exons expressed by the maternal chromosome are surrounded by the color orange, while those expressed by the paternal chromosome are surrounded by the color green; biallelic expression exons are circled in orange at the top and green at the base. Arrows indicate transcriptional start sites and transcriptional direction. Black line boxes and connecting lines depict exons and introns, respectively, that code for named proteins. Several other transcripts arise from differentially methylated promoters, including the maternally expressed NESP55 and the paternally expressed XLαs and A/B.

C - TS – Temple syndrome; KOS - Kagami-Ogata syndrome; IC - imprinting center; IG-DMR - intergenic differentially methylated region. Arrows indicate the transcriptional direction for *RTL1as* and *RTL1*. In most tissues.

Figure 3 – Schematic representation of 16p11.2 and 1p36 microdeletions



A - 16p11.2 syndromes region. Five imbalance segments (red bars) have been defined for the 16p11.2 chromosomal region. Only genes that could be related to the phenotypes are represented. Not drawn to scale.

B - 1p36 syndrome region. Critical regions (blue bars) associated with different phenotypes have been narrowed (yellow bars). Only genes that have been associated with phenotypes are represented. Not drawn to scale

Table 1: Syndromic obesity disorders described in this review

Syndrome (OMIM phenotype number)		Gene(s) involved	Inheritance	Neonatal hypotonia	Polyphagia	Obesity	Other clinical findings
Prader-Willi syndrome (#176270)		*NDN*, *SNRPN*	AD (imprinting disorder)	+	+	central; starts 1-6 years old	ID; motor DD; short stature; hypopigmentation; facial dysmorphisms; small hands; hypogonadism; behavioral problems
Bardet-Biedl syndrome		various (see Table 2)	AR/DR	-	+	72–92% of patients; central; significant weight gain in the first year of life	retinal degeneration with progressive visual impairment; blindness occurs, on average, at 15.5 years; polydactyly; ID; kidney disorders; hypogonadism (in men)
Pseudohypoparathyroidism	1a (#103580)	*GNAS*	AD (imprinting disorder)	-	+ (mild)	starts in the first 2 years of life	parathyroid hormone resistance; multiple hormone resistance; reduced erythrocyte Gsα activity; round facies; short stature; thickened calvaria; ectopic soft tissue or dermal ossification(s); brachydactyly; cognitive impairment (but ID is rare)
	1c (#612462)			-	+ (mild)	starts in the first 2 years of life	parathyroid hormone resistance; multiple hormone resistance; normal erythrocyte Gsα activity; round facies; short stature; thickened calvaria; ectopic soft tissue or dermal ossification(s); brachydactyly; cognitive impairment (ID is rare)
	Pseudopseudohypopara-thyroidism (#612463)			-	-	obesity is less common and, if present, it is not severe	without resistance to parathyroid hormone or other hormones; reduced erythrocyte Gsα activity; round facies; short stature; thickened calvaria; ectopic soft tissue or dermal ossification(s); brachydactyly; cognitive impairment (but ID is rare)
Temple syndrome (#616222)		*DLK1, RTL1, DIO3*	AD (imprinting disorder)	+	+/-	birth weight is frequently below average; starts 4-6 years old; ~49% of patients; central	small feet and hands; low weight at birth; precocious puberty; motor DD; short stature; hyperextensible joints; speech delay; facial features: broad tall forehead, nose often short with a wide, fleshy nasal tip and a short philtrum

16p11.2 deletion syndrome (593 Kb; BP4 - BP5; #611913)	*KCTD13* and possibly *MAZ, TAOK2, CORO1A, PRRT2, TBX6,* and *CTD13*	AD	-	+	evolves throughout childhood; in adulthood, ~75% of patients have obesity (approximately 45% of them with grade III)	speech impairment; motor coordination difficulties; facial dysmorphism; autism; ID
16p11.2 deletion syndrome (220 Kb; BP2 – BP3; #613444)	*SH2B1*	AD	-	+	severe early-onset obesity	cognitive deficits; DD; autism
1p36 deletion syndrome (distal; #607872)	*UBE4B, CASZ1*	AD	+	+	after infancy	DD; short stature; craniofacial dysmorphisms
Kleefstra syndrome type 1 (#610253)	*EHMT1*	AD	+	-	overweight at birth; central early onset obesity	ID (moderate to severe); severe delayed or absent speech; motor DD; sleep disorder; microcephaly; brachycephaly; facial dysmorphisms (hypertelorism, synophrys, midface hypoplasia, everted lower lip, prognathism, and macroglossia); behavioral problems; congenital heart disease; seizures
Börjeson‐Forssman‐Lehmann syndrome (#301900)	*PHF6*	XLR	+	not characterized	central obesity	ID (mild to moderate); speech delay; motor DD; facial dysmorphisms (narrow palpebral fissures and large ears); gynecomastia; characteristic digit anomalies; hyperflexible fingers; dysmorphic hands and feet; hypogonadism
Cohen syndrome (#216550)	*VPS13B*	AR	+	not characterized	increased waist circumference becomes evident at 5-12 years, but does not always lead to an obesity BMI	ID; speech development impaired; motor DD; narrow hands and feet; microcephaly; facial dysmorphisms (hick hair and eyebrows, long eyelashes, hypertelorism, arched or wave-shaped eyelids); leukopenia (mainly neutropenia); ophthalmologic problems; high-pitched crying and voice; joint hypermobility; friendly personality
Alström syndrome (#203800)	*ALMS1*	AR	-	+	central; excessive weight gain after the first year of life; BMI of most young children	early onset retinal degeneration; hearing loss observed in the first decade (progressive and severe); type 2 diabetes mellitus; dyslipidemia; arterial

					is >95th centile; in some individuals body weight tends to normalize, decreasing into the high-normal to normal range after adolescence	hypertension; low testosterone in males and hyperandrogenism in females; dilated cardiomyopathy; respiratory problems
Carpenter syndrome type 1 (#201000)	*RAB23*	AR	-	-	high birth weight; late obesity	multiple suture craniosynostosis, facial dysmorphisms; broad or bifid thumbs, absent or small middle phalanges, and polysyndactyly of hands and feet; congenital heart disease; umbilical hernia; cryptorchidism; hypoplastic testes; arched femur and tibia; nervous system malformations
Carpenter syndrome type 2 (#614976)	*MEGF8*	AR	-	-	tendency of high birth weight and childhood obesity	multisuture craniosynostosis; facial dysmorphisms; polysyndactyly of hands and feet; umbilical hernia; cryptorchidism; heart disease; heart changes; thoracic skeletal abnormalities; hypoplastic and/or supernumerary nipples
Mental retardation, truncal obesity, retinal dystrophy, and micropenis (MORM - #610156)	*INPP5E*	AR	-	-	central obesity apparent in childhood	ID; retinal dystrophy (reduced visual acuity by 3 years old; usually not progressive); micropenis; delayed language acquisition
MYT1L-related syndrome (#616521)	*MYT1L*	AD	-	+	~58% of patients have overweight/obesity (usually beginning in early childhood)	ID; behavioral disorders; language impairment; facial dysmorphisms
Luscan-Lumish syndrome (#616831)	*SETD2*	AD	+	+	obesity with generalized overgrowth	macrocephaly; ID; DD; seizures; speech delay; low sociability; behavioral problems; postnatal overgrowth; advanced carpal ossification; cryptorchidism; nasal polyps; large tongue with creases; high pain threshold; constipation

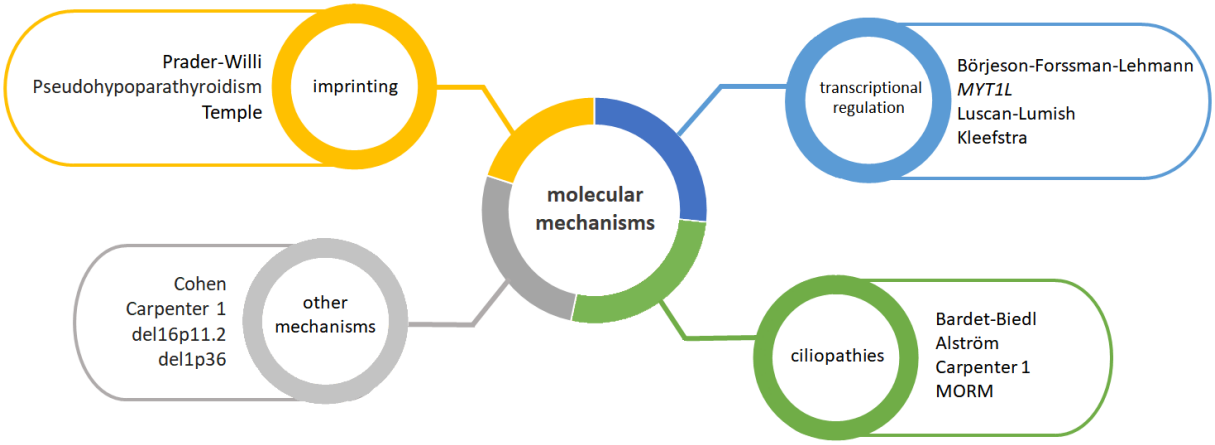
AD - autosomal dominant; DR - digenic recessive; AR - autosomal recessive; ID – intellectual disability; DD - developmental delay

Table 2: OMIM Bardet-Biedl syndrome data.

	Locus	**Genes involved**	**Inheritance**
Bardet-Biedl syndrome 1 (#209900)	1p35.2	*CCDC28B* (also *MGC1203*)	AR/DR
	3q11.2	*ARL6* (also *BBS3*)	
	11q13.2	*BBS1*	
Bardet-Biedl syndrome 2 (#615981)	16q13	*BBS2*	AR
Bardet-Biedl syndrome 3 (#600151)	3q11.2	*ARL6* (also *BBS3*)	AR
Bardet-Biedl syndrome 4 (#615982)	15q24.1	*BBS4*	AR
Bardet-Biedl syndrome 5 (#615983)	2q31.1	*BBS5*	AR
Bardet-Biedl syndrome 6 (#605231)	20p12.2	*MKKS* (also *BBS6*)	AR
Bardet-Biedl syndrome 7 (#615984)	4q27	*BBS7*	AR
Bardet-Biedl syndrome 8 (#615985)	14q31.3	*TTC8* (also *BBS8*)	AR
Bardet-Biedl syndrome 9 (#615986)	7p14.3	*PTHB1* (also *BBS9*)	AR
Bardet-Biedl syndrome 10 (#615987)	12q21.2	*BBS10*	AR
Bardet-Biedl syndrome 11 (#615988)	9q33.1	*TRIM32* (also *BBS11*)	AR
Bardet-Biedl syndrome 12 (#615989)	4q27	*BBS12*	AR
Bardet-Biedl syndrome 13 (#615990)	17q22	*MKS1* (also *BBS13*)	AR
Bardet-Biedl syndrome 14 (#615991)	8q22.1	*TMEM67* (also *MKS3*)	AR
	12q21.32	*CEP290* (also *BBS14*)	
Bardet-Biedl syndrome 15 (#615992)	2p15	*WDPCP* (also *BBS15)*	AR
Bardet-Biedl syndrome 16 (#615993)	1q43-q44	SDCCAG8 (also *BBS16*)	AR
Bardet-Biedl syndrome 17 (#615994)	3p21.31	*LZTFL1* (also *BBS17*)	AR
Bardet-Biedl syndrome 18 (#615995)	10q25.2	*BBIP1* (also *BBS18*)	AR
Bardet-Biedl syndrome 19 (#615996)	22q12.3	*IFT27* (also *BBS19*)	AR
Bardet-Biedl syndrome 20 (#619471)	2p23.3	*IFT172*	AR
Bardet-Biedl syndrome 21 (#617406)	8q22.1	*C8ORF37*	AR
Bardet-Biedl syndrome 22 (#617119)	9p21.2	*IFT74*	AR

AR - autosomal recessive; DR - digenic recessive

Supplementary figure: Molecular mechanisms of syndromic forms of obesity.



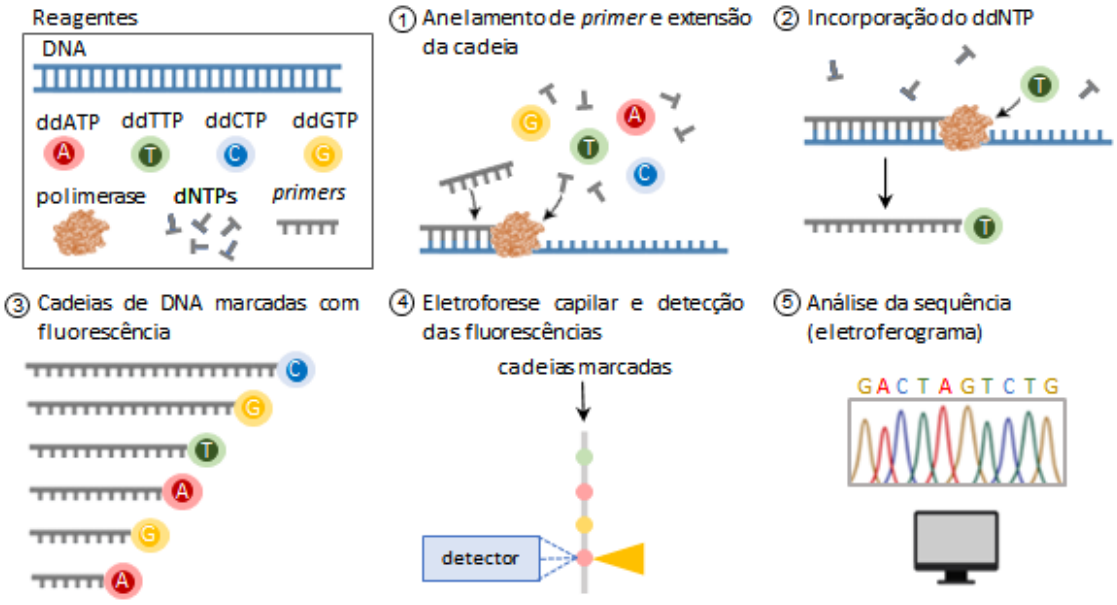
There are three main groups of mechanisms: imprinting disorders, related to genes direcly involved in regulation of transcriptional activity, and ciliopathies. For each mechanism, examples of syndromes are presented. However, not all syndromic forms of obesity seem to fit these three general mechanisms, which is represented by the gray band in the graph.

Capítulo 1.2

Tecnologias de sequenciamento e microarranjos para diagnóstico de doenças congênitas

As técnicas de sequenciamento de DNA têm grande importância para detectar variantes genéticas (KUHLENBÄUMER; HULLMANN; APPENZELLER, 2011). A primeira foi a técnica desenvolvida por Frederick Sanger na década de 1970 (SANGER; COULSON, 1975), o que lhe rendeu em 1980 o seu segundo prêmio Nobel em química (SHAMPO; KYLE, 2002). O fundamento do sequenciamento Sanger está no término da síntese de cadeias de DNA a partir da inserção de dideoxinucleotídeos (ddNTPs). A detecção de sequências terminadas era feita a partir da radiografia de géis de poliacrilamida, mas foram desenvolvidas melhorias metodológicas como a substituição de isótopos radioativos por marcadores fluorescentes específicos para cada nucleotídeo, além de eletroforese capilar (HEATHER; CHAIN, 2016). À medida que as cadeias migram, um detector registra a fluorescência (MUZZEY; EVANS; LIEBER, 2015) (**Figura 1**).

Figura 1: Sequenciamento Sanger utilizando marcações fluorescentes e eletroforese capilar.



Empresas como a Roche e a Illumina introduziram no mercado tecnologias que fazem parte do que chamamos de Sequenciamento de Nova Geração (NGS - *Next-generation Sequencing*). O NGS se baseia em sequenciamentos massivos paralelos que produzem grandes conjuntos de dados (BEHJATI; TARPEY, 2013). Em contraste, a técnica de Sanger é focada em um fragmento específico de DNA por reação e com limitação de comprimento (PETERSEN et al., 2017); como consequência, foi necessária mais de uma década (1990-2003) para que se elaborasse o rascunho do genoma humano, que envolveu mais de 20 institutos e custou cerca de 3 bilhões de dólares (BEHJATI; TARPEY, 2013; PETERSEN et al., 2017).

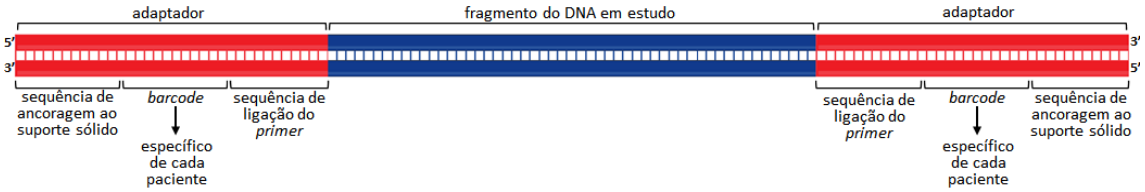
O NGS permite que o sequenciamento de genomas humanos inteiros seja feito em menos de um dia (BEHJATI; TARPEY, 2013). Além disso, o avanço tecnológico resultou em uma redução de custos (JAMUAR; TAN, 2015). Para identificação de genes candidatos em condições patológicas, a técnica de Sanger costumava ser empregada apenas após laboriosas análises de ligação baseadas em microssatélites. Em geral, esses experimentos envolviam grandes genealogias com vários indivíduos afetados para aumentar a probabilidade de se identificar regiões candidatas pequenas o suficiente para serem sequenciadas (PETERSEN et al., 2017). Sendo mais eficientes, rápidas e baratas, as tecnologias baseadas em NGS têm sido ferramentas importantes na identificação de genes candidatos e foram gradualmente introduzidas na rotina clínica (JAMUAR; TAN, 2015).

Existem várias plataformas NGS usando diferentes tecnologias de sequenciamento (BEHJATI; TARPEY, 2013), que incluem: sequenciamento por síntese, por ligação, por semicondutores, entre outros. A abordagem predominante atualmente é o sequenciamento por síntese, empregada pela Illumina (MUZZEY; EVANS; LIEBER, 2015), que segue o seguinte conceito geral: um terminador reversível marcado com fluorescência é identificado à medida que cada nucleotídeo é adicionado. Em seguida, uma porção do terminador é clivada, permitindo a incorporação da próxima base, dando continuidade à síntese (BENTLEY; ET AL., 2008). Tecnicamente, esse tipo de sequenciamento se divide nas seguintes etapas: extração do DNA, preparo de biblioteca, sequenciamento, processamento bioinformático para alinhamento genômico e chamada de variantes e, por fim, análise dos dados (CLARK; PAZDERNIK; MCGEHEE, 2019; MUZZEY; EVANS; LIEBER, 2015).

No sequenciamento Illumina, após a extração do DNA deve-se fazer o preparo da biblioteca, que envolve a quebra em fragmentos menores e adição de adaptadores às

extremidades (**Figura 2**). Em seguida, a biblioteca é adicionada a um suporte sólido (YOHE; THYAGARAJAN, 2017), no qual há sequências complementares para ancoragem dos adaptadores. Os adaptadores também contêm a sequência para ligação do *primer* de sequenciamento (CLARK; PAZDERNIK; MCGEHEE, 2019) e, ao se trabalhar com amostras de diversos pacientes ao mesmo tempo, é possível adicionar também sequências *barcode* específicas para cada indivíduo (YOHE; THYAGARAJAN, 2017).

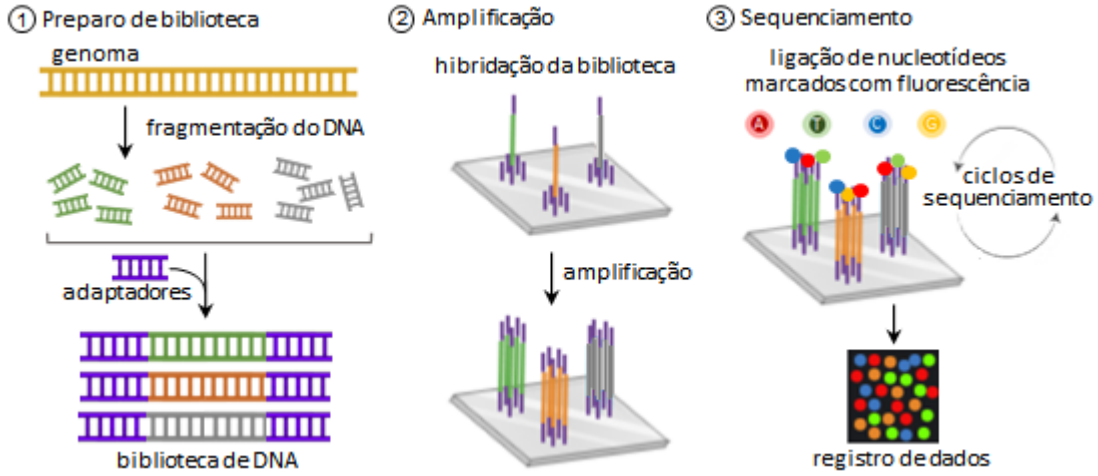
Figura 2: Representação esquemática dos adaptadores do sistema Illumina de sequenciamento por síntese.



Após o preparo da biblioteca e imobilização dos fragmentos no suporte sólido, ocorre uma amplificação clonal e, sem seguida, o sequenciamento (YOHE; THYAGARAJAN, 2017). Nessa etapa, a polimerase adiciona os nucleotídeos de maneira complementar. Porém, como os nucleotídeos são modificados para não serem extensíveis, ocorre o acréscimo de apenas um por rodada. A excitação por laser leva a uma emissão de sinal fluorescente que é registrado. Uma emissão é registrada para cada fragmento, evento que acontece simultaneamente para todos eles. Em seguida, as bases modificadas são convertidas em bases regulares, tornando-se extensíveis e não fluorescentes. A polimerase então acrescenta uma base a cada fragmento e uma nova captura de todos os sinais ocorre (MUZZEY; EVANS; LIEBER, 2015; YOHE; THYAGARAJAN, 2017). As etapas do sequenciamento Illumina são representadas esquematicamente na **Figura 3**.

No entanto, a análise do genoma completo (*whole genome sequencing* - WGS) é complexa, demorada e laboriosa. Por esse motivo, muitas vezes se opta por sequenciar seletivamente apenas éxons (TEER; MULLIKIN, 2010). Se tem mais conhecimento sobre DNA de genes codificadores de proteínas e a análise de variantes raras nessas regiões facilita sua relação com respectivos fenótipos (GLOSS; DINGER, 2018). No genoma humano, apenas cerca de 1% do material genético corresponde a éxons (BICK; DIMMOCK, 2011; NG; ET AL., 2009; RICHTER; SEXTON, 2009), de forma que a quantidade de dados gerada por WES (*whole exome sequencing*) é muito menor que a de WGS.

Figura 3: Sequenciamento por síntese (Illumina).



O WES depende de uma etapa prévia, a captura do exoma (SHIGEMIZU et al., 2015), que se baseia na identificação de sequências das regiões codificadoras consenso de humanos (CCDS) e exons de genes RefSeq (PARLA et al., 2011). Para a captura, podem ser usados kits disponíveis comercialmente e existem diferentes estratégias (SHIGEMIZU et al., 2015), porém os métodos mais utilizados são os protocolos de captura por hibridação do tipo "captura em solução" e "captura em *array*" (**Figura 4**) (SULONEN et al., 2011). Nos protocolos de "captura em solução", sondas de oligonucleotídeos marcadas são misturadas à biblioteca para capturar as sequências-alvo (exons). As sequências capturadas são recuperadas utilizando *beads* magnéticas (MARDIS; MCCOMBIE, 2017). Já no protocolo de "captura em *array*", a biblioteca é hibridada ao microarranjo de sondas complementares às sequências-alvo (exons) (SULONEN et al., 2011).

Após o sequenciamento, os dados brutos de NGS passam por um processamento bioinformático (**Figura 5**) para obter o arquivo de chamada de variantes (VCF - *Variant Call Format*) (YOHE; THYAGARAJAN, 2017).

O uso de *barcodes* para identificar as amostras de cada paciente permite que amostras de vários indivíduos sejam sequenciadas simultaneamente, diminuindo assim o custo. No entanto, esse processo requer uma etapa de desmultiplexação. O arquivo desmultiplexado com leituras brutas é o FASTQ (YOHE; THYAGARAJAN, 2017), um arquivo de texto que contém sequências e seus índices de qualidade (COCK et al., 2010). A cada sequenciamento *paired end* são gerados dois arquivos FASTQ, sendo um para cada sentido do sequenciamento (MIELCZAREK; SZYDA, 2016).

Figura 4: Preparo de biblioteca e captura de exoma.

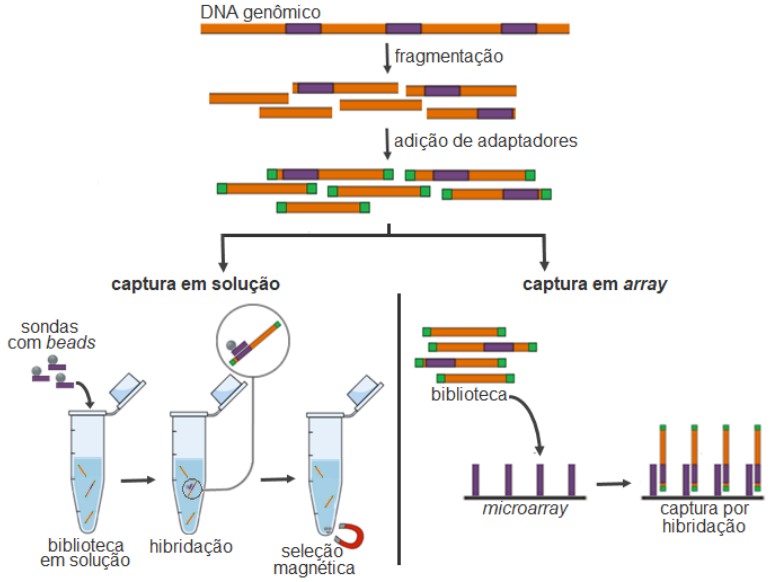
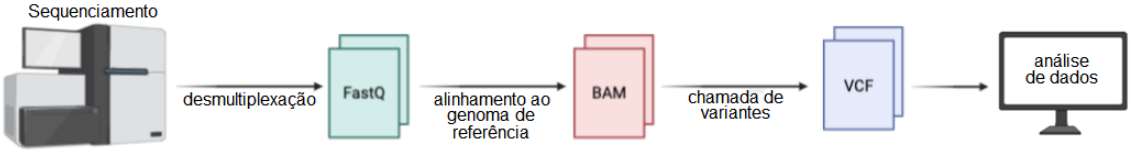


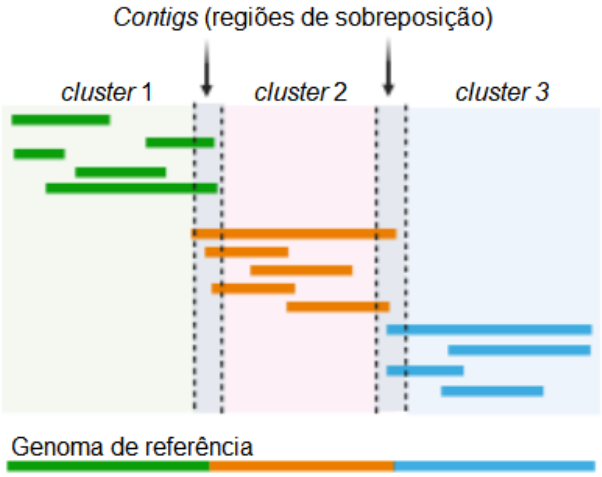
Figura 5: *Pipeline* bioinformático para gerar o arquivo VCF, a ser analisado.



A etapa seguinte é o alinhamento das leituras (*reads*) ao genoma de referência (**Figura 6**), que geralmente tem formato FASTA (MIELCZAREK; SZYDA, 2016; MUZZEY; EVANS; LIEBER, 2015). O alinhamento gera o arquivo BAM (*Binary Alignment Map*) (Yohe e Thyagarajan, 2017). O número de *reads* alinhadas em uma determinada posição é chamado de "profundidade" ou "cobertura vertical" (MUZZEY; EVANS; LIEBER, 2015).

Em seguida, algoritmos que detectam as variantes são utilizados, gerando um arquivo VCF (MIELCZAREK; SZYDA, 2016). Na prática, variantes genéticas são desvios da sequência de referência (MUZZEY; EVANS; LIEBER, 2015) e a chamada de uma variante ocorre se várias *reads* mostrarem a mesma diferença. O arquivo VCF define as variantes de uma amostra e suas frações alélicas (YOHE; THYAGARAJAN, 2017).

Figura 6: Alinhamento das *reads* ao genoma de referência.



O VCF contém todas as variantes, incluindo as comuns e, assim sendo, elas precisam ser filtradas e interpretadas (YOHE; THYAGARAJAN, 2017). Com relação à qualidade do que se observa no VCF, são critérios de filtragem úteis: cobertura vertical (*read depth*) (MUZZEY; EVANS; LIEBER, 2015) e qualidade genotípica baseada no *Phred score* (LIAO; SATTEN; HU, 2017). O nome *Phred (Phil's Read Editor*) remete ao um de seus criadores (Dr. Phil Green) (MOODY, 2004). Trata-se de um algoritmo amplamente utilizado que atribui um índice de qualidade a cada base. A pontuação é o log negativo (base 10) da probabilidade de erro Prosdocimi (2016) e WHO (2016). Dessa forma, um score de 20 (Q20) significa um erro a cada 100 bases (taxa de erro = 1%), enquanto Q30 significa um erro a cada 1000 bases (0.1%) (WHO, 2016). O *Phred score* pode ser usado para decidir quais chamadas são confiáveis (LIAO; SATTEN; HU, 2017). A relação entre o *score* de qualidade e a acurácia de chamada é apresentada no **Quadro 1**.

Quadro 1: Escores *Phred* de qualidade e sua relação logarítmica com as probabilidades de erro.

Escore	Probabilidade de chamada incorreta de base	Acurácia na chamada de base
Q10	1 em 10	90%
Q20	1 em 100	99%
Q30	1 em 1.000	99.9%
Q40	1 em 10.000	99.99%
Q50	1 em 100.000	99.999%

Para se filtrar variantes com maior probabilidade causal, convém excluir do rol de candidatas as que são frequentes de acordo com bancos de dados populacionais (RICHARDS et al., 2015). Variantes genéticas que levam a fenótipos deletérios tendem a ser eliminadas por seleção natural negativa (LOHMUELLER, 2014). Dessa forma, variantes raras têm maior probabilidade de ser patogênicas (RICHARDS et al., 2015) e, por esse motivo, a filtragem por frequência com base em grandes bancos de dados populacionais é fundamental. São exemplos de bancos de podem ser utilizados: 1000 genomes (DEVUYST, 2015 - internationalgenome.org/), gnomAD (KARCZEWSKI et al., 2020, et al., 2020 - gnomad.broadinstitute.org/) e ABraOM (NASLAVSKY et al., 2017 - abraom.ib.usp.br/).

Diversos algoritmos de predição de patogenicidade foram desenvolvidos considerando, por exemplo, questões estruturais da proteína (envolvimento de sítio catalítico ou de ligação a íons metálicos, formação de pontes dissulfeto, etc.) e conservação. Como a seleção natural tende a eliminar variantes deletérias das populações, presume-se que sequências altamente conservadas podem ter funções biológicas importantes. Dessa forma, variantes não sinônimas identificadas em códons conservados têm maior probabilidade de prejudicar a função das proteínas e causar doenças. Apesar de essas previsões *in silico* não terem a mesma robustez que ensaios funcionais, elas são ferramentas largamente utilizadas (CHAN et al., 2007).

O SIFT (*Sorting Intolerant From Tolerant*; blocks.fhcrc. org/sift/SIFT.html), por exemplo, é um programa *online* de análise de SNVs não sinônimas (CHAN et al., 2007) baseado em homologia de sequências e similaridades físico-químicas com aminoácidos alternativos (JACOBY et al., 2009) A acurácia geral é relatada como 65 a 92%. (BALASUBRAMANIAN et al., 2005; BAO; CUI, 2005; OLLILA; ET AL., 2006; RAEVAARA; ET AL., 2005; XI; JONES; MOHRENWEISER, 2004).

Outro exemplo de preditor é o PolyPhen (*Polymorphism Phenotyping*; genetics.bwh.- harvard.edu/pph), um programa *online* de análise de SNVs não sinônimas. Ele verifica se a posição do aminoácido está envolvida na ligação a íons metálicos, formação de ligações dissulfeto, se faz parte de um sítio catalítico ou núcleo hidrofóbico, além de outras avaliações relevantes de estrutura. Se a estrutura de uma proteína de consulta é desconhecida (o que é um limitante), o PolyPhen usa proteínas homólogas com estrutura conhecida (CHAN et al., 2007). Sua acurácia é estimada em 80% (MIN et al., 2016; RAMENSKY, 2002).

São outros exemplos de algoritmos de predição de patogenicidade: FATHMM (ROGERS et al., 2018), PROVEAN (CHOI et al., 2012), CADD (KIRCHER et al., 2014), MutationTaster (SCHWARZ et al., 2014), VEST (CARTER et al., 2013), DANN (QUANG; CHEN; XIE, 2015), REVEL (IOANNIDIS et al., 2016), etc.

Outra ferramenta importante na interpretação do impacto das variantes é a predição de tolerância a diferentes classes de variantes (sinônima, missense e LOF - *loss-of-function*) para cada gene no gnomAD (gnomad.broadinstitute.org). O gnomAD é um grande projeto que inclui variantes identificadas em 125,748 exomas e 15,708 genomas humanos (versão v2) e sua predição de tolerância considera o número esperado de variantes para cada gene e o número encontrado. As contagens esperadas são baseadas em um modelo mutacional que leva em consideração o contexto, a cobertura e a metilação da sequência. Valores de pLI (*probability of being loss-of-function intolerant*) ≥ 0.9 indicam que o gene é extremamente intolerante a variantes truncatórias. O preditor também apresenta o Z–score para variantes sinônimas e missenses, sendo que valores positivos indicam intolerância à variação e valores negativos foram dados a genes que tinham mais variantes que o esperado (KARCZEWSKI et al., 2020).

Em uma análise de variantes, para priorização de genes é possível utilizar painéis já descritos na literatura. Também o Varelect (varelect.genecards.org), uma ferramenta *online*, pode ser usado para priorizar genes relacionados a diferentes fenótipos (STELZER et al., 2016). Também é fundamental em uma análise de variantes consultar o OMIM (omim.org/), um catálogo *online* com descrições dos genes relacionados a doenças e fenótipos humanos, sendo baseado na literatura biomédica publicada (AMBERGER et al., 2015). Outro banco de dados fundamental é o Clinvar (ncbi.nlm.nih.gov/clinvar/), uma plataforma colaborativa com curadoria na qual são depositadas informações sobre variantes, seus significados clínicos e evidências que suportam cada interpretação (HARRISON et al., 2016).

Além da avaliação de padrões de herança autossômico recessivo, autossômico dominante e recessivo ligado ao X, cabe avaliar cuidadosamente os genes que sofrem *imprinting* (TONG et al., 2018). Nesse sentido, é útil o banco de dados geneimprint (geneimprint.com/site/genes-by-species).

O sistema mais utilizado de classificação quanto à patogenicidade das variantes mendelianas germinativas são as diretrizes da ACMG (*American College of Medical Genetics and Genomics*) (RICHARDS et al., 2015), que atribuem um valor com base em vários critérios (frequência em bancos de dados populacionais, presença em indivíduos

afetados se já relatada, segregação, estudos funcionais, informações obtidas em algoritmos preditores de efeito, etc.). Há regras para combinar todos os critérios e classificar cada variante como patogênica, provavelmente patogênica, significado incerto (VUS - *Variant of Uncertain Significance*), provavelmente benigna ou benigna. Alguns programas *online* como o Varsome (varsome.com - KOPANOS et al., 2019) e o Franklin (franklin.genoox.com) facilitam a classificação de acordo com diretrizes da ACMG. No entanto, é necessária uma cuidadosa curagem manual com dados específicos do caso clínico em estudo e seguindo as recomendações de Richards et al. (2015). Nessa curagem manual é importante, por exemplo, incluir quando disponível informação de que a variante é *de novo* ou herdada, o que interfere significativamente na classificação. Assim sendo, facilita a interpretação (embora aumente o custo) o sequenciamento de trios (paciente e seus pais), em vez de sequenciar apenas o probando. Em casos familiares, o sequenciamento de mais indivíduos também é útil para identificar a variante que causa a condição em análise.

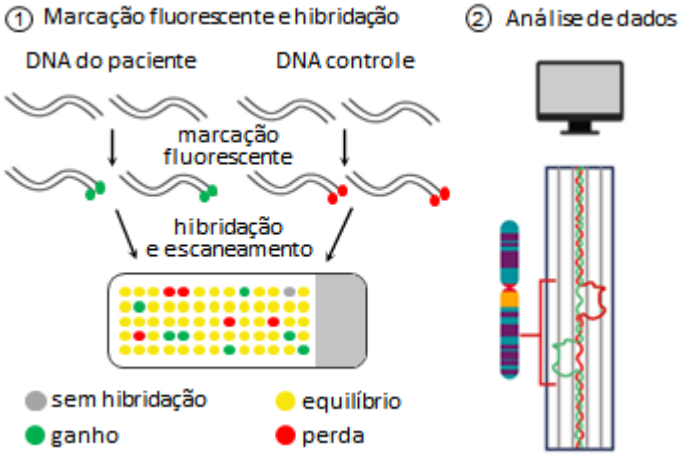
Entre as limitações das análises de exomas para detectar variantes clinicamente relevantes, está sensibilidade analítica, pois também há variantes causais em regiões não codificadoras, há áreas que são mais difíceis de sequenciar, além de limitações no conhecimento para interpretação de certas variantes (YOHE; THYAGARAJAN, 2017).

As CNVs (*Copy-Number Variations*), são variantes genéticas nas quais um segmento de DNA maior ou igual a 1 Kb está presente em número de cópias que difere do genoma de referência, podendo ser microscópicas ou – mais comumente - submicroscópicas. Assim como outros tipos de variantes genéticas, as CNVs contribuem de maneira importante para a diversidade humana, mas também podem estar relacionadas a doenças (FEUK; CARSON; SCHERER, 2006) através de vários mecanismos moleculares como interrupção ou fusão de genes, efeitos de dosagem e de posição (ZHANG et al., 2009).

Métodos de análise cromossômica por microarranjos (CMA – *Chromosome microarray analyses*), tais como a hibridação genômica comparativa em microaaranjos (CGH-*array*) e microarranjos de polimorfismo de nucleotídeo únicos (SNP-*array*), são ferramentas para a identificação de microdeleções e microduplicações relacionadas a patologias de cunho genético (BEHJATI; TARPEY, 2013; KREPISCHI; et al., 2010; MILLER et al., 2010; VOSECKÁ; et al., 2014), tendo sido utilizadas inclusive na investigação de casos de obesidade sindrômica (D'ANGELO et al., 2013, 2018).

O princípio técnico do CGH-*array* (**Figura 7**) é a competição do DNA do indivíduo analisado com um controle por sequências complementares (sondas) dispostas em uma matriz sólida. O DNA do paciente é marcado com um fluorocromo (cianina 5 ou 3), enquanto o DNA controle é marcado com o outro. A fluorescência que se sobressai no escaneamento após a hibridação indica uma maior quantidade de hibridação, o que significa que tal trecho do genoma deve estar em mais cópias na amostra com aquela marcação (AHN et al., 2015; FEUK; CARSON; SCHERER, 2006; WEISS et al., 1999).

Figura 7: CGH-*array*.



A tecnologia de SNP-*array* foi inicialmente desenvolvida para genotipagem em larga escala em estudos de associação, mas aprimorada para detecção de CNVs (MAO; YOUNG; LU, 2007). Tais matrizes permitem, além da detecção de CNVs, identificar trechos em homozigose (*regions of homozygosity* - ROH); esses blocos em homozigose podem ser indicativos de dissomia uniparental (*uniparental disomy* - UPD) ou consanguinidade (DE LEEUW et al., 2011; SONG; SHAO, 2015). O advento dos microarranjos para detectar SNPs permitiu um incremento diagnóstico. É possível ainda combinar o CGH-*array* com o SNP-*array* em um único ensaio, otimizando o teste (WISZNIEWSKA et al., 2014).

O princípio técnico do SNP-*array* é interrogar a presença de alelos em pontos em que variantes do tipo SNP são comuns. Antes de ser hibridado, o DNA genômico é processado por enzimas de restrição, há a ligação de adaptadores e PCR para produzir uma coleção de produtos com uma faixa de tamanho específico. Os produtos de PCR são então fragmentados e marcados com um fluorocromo para posterior hibridação às sondas em matriz sólida. Dependendo do seu genótipo, os fragmentos se ligam especificamente

à sonda de correspondência. É feito o escaneamento para detecção de intensidade de fluorescência, o que permite inferir número de cópias. Além dos dados do número de cópias, o SNP-*array* também produz um conjunto de dados genotípicos gerado pelas chamadas AA, AB ou BB, que representam as três possibilidades genotípicas normais para cada SNP. Isso permite a detecção ROH e UPD (NOWAK; HOFMANN; KOEFFLER, 2009).

Na análise de CNVs, os seguintes parâmetros têm grande relevância:

(1) *Log R ratio*: tal valor é obtido considerando-se a razão intensidade observada na amostra teste / intensidade da referência (WINCHESTER; YAU; RAGOUSSIS, 2009). Quando o número de cópias da amostra em estudo é similar ao da referência, o log R *ratio* fica em torno de zero. Valores inferiores indicam perda de material genético, enquanto valores superiores indicam ganho;

(2) *B-allele frequency* (BAF): quantidade relativa de um alelo B quando comparado ao total de alelos (A e B). Um valor de BAF de 0,5 indica um genótipo heterozigoto (AB), enquanto 0 e 1 indicam genótipos homozigotos (AA e BB, respectivamente) (KEARNEY; KEARNEY; CONLIN, 2011; STAAF et al., 2008).

Outro avanço importante é a possibilidade de se detectar CNVs a partir de dados de NGS. Dessa forma, é possível obter dados de diversos tipos de variantes (SNVs, indels e CNVs) a partir de uma só técnica (GROSS et al., 2019; TROST et al., 2018), o que é custo-efetivo (VILLELA et al., 2017). Porém, as tecnologias baseadas em *array* ainda são frequentemente referidas na literatura como "padrão-ouro" para detecção de CNVs (BARTHA; GYŐRFFY, 2019; KILARU et al., 2020; YANG et al., 2017; YAO et al., 2019).

Diversas metodologias computacionais podem ser aplicadas à detecção e CNVs a partir de dados de NGS, são exemplos:

(1) *Read depth* (RD - cobertura vertical): Assume que a profundidade da cobertura vertical em dada região está relacionada ao seu número de cópias. Trata-se do método de escolha para detecção de CNVs, porém dependem de alta cobertura para melhor desempenho (ZHAO et al., 2013). Muitos métodos estão disponíveis, não apenas utilizando RD, mas também em combinação com PEM: RDXplorer/EWT, mrFast, CNVnator, ReadDepth e Control-FREEC (DO NASCIMENTO; GUIMARAES, 2017).

(2) *Paired-end mapping* (PEM): As *reads* de ambas as orientações de sequenciamento são alinhadas ao genoma de referência e a maioria delas corresponde a pares concordantes. Os pares discordantes sugerem variantes estruturais. Um mapeamento discordante ocorre se a distância entre duas extremidades de um par de *reads* for significativamente diferente do tamanho médio. Se a distância mapeada de duas *reads* for maior / menor que a distância média predeterminada, isso indica um evento de exclusão / inserção (NGUYEN et al., 2016; ZHAO et al., 2013). Existem muitas ferramentas computacionais baseadas nessa abordagem (ex. PEMer, VariationHunter e BreakDancer) (DO NASCIMENTO; GUIMARAES, 2017).

(3) *Split read* (SR): A estratégia é utilizar a leitura incompleta de *reads* mapeadas de cada par para identificar pequenas CNVs. Uma das *reads* do par se alinha ao genoma de referência, enquanto a outra pode não alinhar na mesma região ou alinhar apenas parcialmente. Tais leituras mal mapeadas indicam potenciais pontos de quebra. Os métodos SR dividem as leituras mal pareadas em vários fragmentos, sendo que o primeiro e o último fragmentos são alinhados ao genoma de referência em uma etapa de remapeamento, a fim de indicar posições genômicas dos eventos mutacionais (ZHAO et al., 2013). Algumas ferramentas computacionais desenvolvidas com base em SR: Pindel, SLOPE, AGE e SRiC (DO NASCIMENTO; GUIMARAES, 2017).

(4) *de novo assembly of a genome* (AS): Os fragmentos de DNA (*contigs*) são reconstruídos a partir de *reads* curtas considerando sobreposições. Ao comparar os *contigs* reunidos ao genoma de referência, trechos com número de cópias discordantes podem ser detectados (ZHAO et al., 2013). Exemplos de ferramentas que utilizam essa abordagem: Velvet, Cortex Assembler, Magnolya e TIGRA (DO NASCIMENTO; GUIMARAES, 2017).

As estratégias computacionais para detecção de CNVs podem ser utilizadas de maneira combinada para alcançar melhor sensibilidade e especificidade ao se superar as limitações inerentes a cada abordagem (NOWAKOWSKA, 2017; ZHAO et al., 2013).

CNVs podem ser identificadas por WES, porém sua uniformidade (cobertura horizontal) é inferior à do WGS pois os exons não são uniformemente distribuídos no genoma, tornando difícil detectar alterações que envolvam longas regiões intergênicas. Além disso, há o viés introduzido pela captura do exoma, devido a variações de eficiência (WANG et al., 2015). Neste estudo, nós utilizamos a plataforma OneSeq da *Agilent*

Technologies (Santa Clara, CA), que promete contornar essas limitações, propiciando uma detecção robusta de CNVs, além das outras classes de variantes já corriqueiras na análise de exoma. Em suma, a captura OneSeq engloba três tipos de sondas: (1) sondas *backbone* para detecção de CNVs em genoma completo; (2) sondas para regiões com baixa frequência de SNVs, adequadas à detecção de ausência de heterozigose; (3) sondas para regiões exônicas, adequadas à deteção de SNVs e indels. Nosso grupo publicou um trabalho (VILLELA et al., 2017) no qual essa plataforma obteve resolução e sensibilidade semelhantes às fornecidas de ensaios baseados em *array* para detecção de CNVs.

Uma das dificuldades enfrentadas com o refinamento técnico da citogenômica é distinguir eventos patogênicos de polimorfismos. Nesse sentido, a manutenção de registros detalhados de CNVs em populações normais e afetadas têm potencial de levar a classificações mais adequadas (FERGUSON-SMITH, 2015). Há importantes bancos de dados que auxiliam na tarefa de classificação de variantes congênitas quanto à patogenicidade. Alguns desses bancos são descritos abaixo:

(1) *Database of Genomic Variant* (DGV - dgv.tcag.ca/dgv/app/home): reporta variantes estruturais identificadas em estudos com indivíduos controles, evidenciando quando se trata de uma variante comum. O DGV concentra-se em alterações genômicas que envolvem segmentos de DNA maiores que 1 Kb. Porém, alterações de 50pb ou mais também estão incluídas.

(2) *Database of Chromosomal Imbalances and Phenotype in Humans using Ensembl Resources* (DECIPHER - decipher.sanger.ac.uk): faz uma compilação de variantes identificadas em indivíduos com fenótipo de doença, auxiliando na interpretação de sua patogenicidade. O banco contém dados de mais de 30.000 pacientes.

(3) *Map of Developmental Delay:* compilação de CNVs em duas coortes, uma de indivíduos com atraso no desenvolvimento e outra de controles. O estudo de Cooper et al. (2011) analisou amostras de 15.767 crianças com vários distúrbios de desenvolvimento e as compararam com 8.329 controles adultos para produzir um mapa detalhado. O estudo de Coe et al. (2014) expandiu ainda mais o mapa de morbidade analisando 13.318 novas amostras de casos, além de 11.255 novos controles.

Cabe dizer que tais bancos podem ser consultados simultaneamente utilizando, por exemplo, o *Genome Browser* da UCSC (genome.ucsc.edu). Adicionalmente,

recomenda-se consultar na interpretação também o OMIM (omim.org) e a literatura científica. Também é relevante considerar:

(1) Tamanho da CNV (variantes maiores têm maior potencial deletério);

(2) Conteúdo gênico presente na CNV em análise (se contém uma sequência rica em genes ou não possui genes, se é composta principalmente de elementos repetitivos ou pseudogenes, investigar se os genes no intervalo são relatados como sensíveis à dosagem ou relacionados a condições clínicas);

(3) Comparação com bancos de dados internos (rastreio da própria experiência) e externos (bancos de dados publicamente disponíveis).

Nas recomendações da ACMG e *Clinical Genome Resource* (ClinGen) (RIGGS et al., 2020), para a interpretação do significado clínico das CNVs são atribuídas três categorias principais, são elas:

(1) CNVs benignas: relatadas em várias publicações revisadas por pares ou anotadas em bancos de dados como benignas, sendo documentadas em >1% da população. É importante considerar cuidadosamente a dosagem, dado que duplicações de algumas regiões podem ser benignas, enquanto deleções do mesmo intervalo podem ter relevância clínica;

(2) CNVs provavelmente benignas: se não há diferença estatisticamente significativa entre as observações em casos e controles; variantes observadas com frequência na população geral (embora inferior a 1%);

(3) VUS: se é grande o suficiente para relatar, mas não possui genes; quando contém um pequeno número de genes, mas não se sabe se são sensíveis à dosagem; se descrita em várias publicações e/ou bancos de dados de maneira contraditória; se está dentro de um gene mas com efeito pouco claro no quadro de leitura;

(4) CNVs provavelmente patogênicas: se há fortes evidências para sugerir que são causadoras de doenças, mas ainda não o suficiente para afirmar definitivamente a patogenicidade. As evidências podem incluir: deleções envolvendo a extremidade 5' (mais a sequência de codificação adicional) de genes haploinsuficientes; (2) deleções envolvendo múltiplos éxons em um gene haploinsuficiente; e deleções ou duplicações envolvendo genes com múltiplos relatos de fenótipos consistentes e altamente específicos;

(5) CNVs patogênicas: se a natureza patogênica da CNV não pode ser questionada. Exemplos de evidências: quando a relação com fenótipos clínicos é consistente em várias publicações revisadas por pares, com penetrância e

expressividade bem documentadas, mesmo que reduzidas e/ou variáveis; CNVs que se sobrepõem completamente a uma região sensível à dosagem bem estabelecida; e CNVs multigênicas nas quais pelo menos um gene é conhecido por ser sensível à dosagem, mesmo que o efeito dos outros genes seja de significância incerta.

A classificação de CNVs é limitada ao conhecimento científico existente. Uma CNV classificada como VUS pode ser reclassificada futuramente à luz de novas informações (NOWAKOWSKA, 2017). O mesmo se aplica à interpretação de SNVs e indels (RICHARDS et al., 2015).

Por vezes, a variante causal pode estar em um gene até então não relacionado na literatura à condição clínica em estudo. Após a identificação de genes candidatos, pode ser um desafio identificar o papel desses genes e determinar como sua disfunção afeta a fisiopatologia (SETH; STEMPLE; BARROSO, 2013). Um diagnóstico molecular conclusivo é fundamental para que se prestar o melhor atendimento ao paciente, prever o prognóstico e para o aconselhamento genético da família (RODENBURG, 2018). Para a validação funcional, são muito importantes, por exemplo, os estudos funcionais baseados em modelos animais (CHAO et al., 2019; PEREZ et al., 2019; SALAZAR-SILVA et al., 2021) e celulares (HAACK et al., 2015; LIM et al., 2022), além da identificação de outros casos com fenótipos sobrepostos, o que pode ser facilitado através de plataformas como o GeneMatcher (genematcher.org/), que promove contatos entre pessoas com interesse em um mesmo gene (SOBREIRA et al., 2015).

Referências

AHN, J. W.; COLDWELL, M.; BINT, S.; MACKIE OGILVIE, C. Array Comparative Genomic Hybridization (Array CGH) for Detection of Genomic Copy Number Variants. **Journal of Visualized Experiments**, n. 96, 2015. ISSN: 1940-087X. DOI: 10.3791/51718

AMBERGER, J. S.; BOCCHINI, C. A.; SCHIETTECATTE, F.; SCOTT, A. F.; HAMOSH, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders. **Nucleic Acids Research**, v. 43, n. D1, p. D789–D798, 2015. ISSN: 1362-4962. DOI: 10.1093/nar/gku1205.

BALASUBRAMANIAN, S.; XIA, Y.; FREINKMAN, E.; GERSTEIN, M. Sequence variation in G-protein-coupled receptors: analysis of single nucleotide polymorphisms. **Nucleic Acids Res** , v. 33, p. 1710–1721, 2005. DOI: 10.1093/nar/gki311

BAO, L.; CUI, Y. Prediction of the phenotypic effects of non-synonymous single

nucleotide polymorphisms using structural and evolutionary information. Bioinformatics, p. 2185–2190, 2005. DOI: 10.1093/bioinformatics/bti365

BARTHA; GYŐRFFY. Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. **Cancers**, v. 11, n. 11, p. 1725, 2019. ISSN: 2072-6694. DOI: 10.3390/cancers11111725

BEHJATI, S.; TARPEY, P. S. What is next generation sequencing? **Archives of disease in childhood - Education & practice edition**, v. 98, n. 6, p. 236–238, 2013. ISSN: 1743-0585. DOI: 10.1136/archdischild-2013-304340

BENTLEY, DR; et al. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, v. 456, p. 53–59, 2008. DOI: 10.1038/nature07517

BICK, D.; DIMMOCK, D. Whole exome and whole genome sequencing. **Current Opinion in Pediatrics**, v. 23, n. 6, p. 594–600, 2011. ISSN: 1040-8703. DOI: 10.1097/MOP.0b013e32834b20ec

CARTER, H.; DOUVILLE, C.; STENSON, P. D.; COOPER, D. N.; KARCHIN, R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. **BMC Genomics**, v. 14, n. Suppl 3, p. S3, 2013. ISSN: 1471-2164. DOI: 10.1186/1471-2164-14-S3-S3

CHAN, P. A. et al. Interpreting missense variants: comparing computational methods in human disease genes *CDKN2A* , *MLH1* , *MSH2* , *MECP2* , and tyrosinase (*TYR*). **Human Mutation**, v. 28, n. 7, p. 683–693, 2007. ISSN: 10597794. DOI: 10.1002/humu.20492

CHAO, T. et al. Precise and Rapid Validation of Candidate Gene by Allele Specific Knockout With CRISPR/Cas9 in Wild Mice. **Frontiers in Genetics**, v. 10, 2019. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00124

CHOI, Y.; SIMS, G. E.; MURPHY, S.; MILLER, J. R.; CHAN, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. **PLoS ONE**, v. 7, n. 10, p. e46688, 2012. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0046688

CLARK, D. P.; PAZDERNIK, N. J.; MCGEHEE, M. R. **Molecular Biology**. 3. ed., Academic Cell, 2019.

COCK, P. J. A.; FIELDS, C. J.; GOTO, N.; HEUER, M. L.; RICE, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. **Nucleic Acids Research**, v. 38, n. 6, p. 1767–1771, 2010. ISSN: 0305-1048. DOI: 10.1093/nar/gkp1137

D'ANGELO, C. S. et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. **Molecular Cytogenetics**, v. 11, n. 1, 2018. ISSN: 17558166. DOI: 10.1186/s13039-018-0363-7

D'ANGELO, C. S.; KOHL, I.; VARELA, M. C.; DE CASTRO, C. I. E.; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; PEREZ, A. B. A.; KOIFFMANN, C. P. Obesity with associated developmental delay and/or learning disability in patients exhibiting additional features: Report of novel pathogenic copy number variants. **American Journal of Medical Genetics Part A**, v. 161, n. 3, p. 479–486, 2013. ISSN: 15524825. DOI: 10.1002/ajmg.a.35761

DE LEEUW, N.; HEHIR-KWA, J. Y.; SIMONS, A.; GEURTS VAN KESSEL, A.; SMEETS, D. F.; FAAS, B. H. W.; PFUNDT, R. SNP Array Analysis in Constitutional and Cancer Genome Diagnostics – Copy Number Variants, Genotyping and Quality

Control. **Cytogenetic and Genome Research**, v. 135, n. 3–4, p. 212–221, 2011. ISSN: 1424-859X. DOI: 10.1159/000331273

DEVUYST, O. The 1000 Genomes Project: Welcome to a New World. **Peritoneal Dialysis International: Journal of the International Society for Peritoneal Dialysis**, v. 35, n. 7, p. 676–677, 2015. ISSN: 0896-8608. DOI: 10.3747/pdi.2015.00261

DO NASCIMENTO, F.; GUIMARAES, K. S. Copy Number Variations Detection: Unravelling the Problem in Tangible Aspects. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 14, n. 6, p. 1237–1250, 2017. ISSN: 1545-5963. DOI: 10.1109/TCBB.2016.2576441

FERGUSON-SMITH, M. A. History and evolution of cytogenetics. **Molecular Cytogenetics**, v. 8, n. 1, p. 19, 2015. ISSN: 1755-8166. DOI: 10.1186/s13039-015-0125-8

FEUK, L.; CARSON, A. R.; SCHERER, S. W. Structural variation in the human genome. **Nature Reviews Genetics**, v. 7, n. 2, p. 85–97, 2006. ISSN: 1471-0056. DOI: 10.1038/nrg1767

GLOSS, B. S.; DINGER, M. E. Realizing the significance of noncoding functionality in clinical genomics. **Experimental & Molecular Medicine**, v. 50, n. 8, p. 1–8, 2018. ISSN: 1226-3613. DOI: 10.1038/s12276-018-0087-0

GROSS, A. M. et al. Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. **Genetics in Medicine**, v. 21, n. 5, p. 1121–1130, 2019. ISSN: 1098-3600. DOI: 10.1038/s41436-018-0295-y

HAACK, T. B. et al. Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement. **Annals of Clinical and Translational Neurology**, v. 2, n. 5, p. 492–509, 2015. ISSN: 23289503. DOI: 10.1002/acn3.189

HARRISON, S. M. et al. Using ClinVar as a Resource to Support Variant Interpretation. **Current Protocols in Human Genetics**, v. 89, n. 1, 2016. ISSN: 1934-8266. DOI: 10.1002/0471142905.hg0816s89

HEATHER, J. M.; CHAIN, B. The sequence of sequencers: The history of sequencing DNA. **Genomics**, v. 107, n. 1, p. 1–8, 2016. ISSN: 08887543. DOI: 10.1016/j.ygeno.2015.11.003

IOANNIDIS, N. M. et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. **The American Journal of Human Genetics**, v. 99, n. 4, p. 877–885, 2016. ISSN: 00029297. DOI: 10.1016/j.ajhg.2016.08.016

JACOBY, M. et al. INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. **Nature Genetics**, v. 41, n. 9, p. 1027–1031, 2009. ISSN: 1061-4036. DOI: 10.1038/ng.427.

JAMUAR, S. S.; TAN, Ene-Choo. Clinical application of next-generation sequencing for Mendelian diseases. **Human Genomics**, v. 9, n. 1, p. 10, 2015. ISSN: 1479-7364. DOI: 10.1186/s40246-015-0031-5

KARCZEWSKI, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. **Nature**, v. 581, n. 7809, p. 434–443, 2020. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2308-7

KEARNEY, H. M.; KEARNEY, J. B.; CONLIN, L.K. Diagnostic Implications of Excessive Homozygosity Detected by SNP-Based Microarrays: Consanguinity,

Uniparental Disomy, and Recessive Single-Gene Mutations. **Clinics in Laboratory Medicine**, v. 31, n. 4, p. 595–613, 2011. ISSN: 02722712. DOI: 10.1016/j.cll.2011.08.003.

KILARU, V. et al. Critical evaluation of copy number variant calling methods using DNA methylation. **Genetic Epidemiology**, v. 44, n. 2, p. 148–158, 2020. ISSN: 0741-0395. DOI: 10.1002/gepi.22269

KIRCHER, M.; WITTEN, D. M.; JAIN, P.; O'ROAK, B. J.; COOPER, G. M.; SHENDURE, J. A general framework for estimating the relative pathogenicity of human genetic variants. **Nature Genetics**, v. 46, n. 3, p. 310–315, 2014. ISSN: 1061-4036. DOI: 10.1038/ng.2892

KOPANOS, C.; TSIOLKAS, V.; KOURIS, A.; CHAPPLE, C. E.; ALBARCA AGUILERA, M.; MEYER, R.; MASSOURAS, A. VarSome: the human genomic variant search engine. **Bioinformatics**, v. 35, n. 11, p. 1978–1980, 2019. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty897

KREPISCHI, A. C.; et al. Two distinct regions in 2q24.2-q24.3 associated with idiopathic epilepsy. **Epilepsia**, v. 51, n. 12, 2010. DOI: 10.1111/j.1528-1167.2010.02742.x

KUHLENBÄUMER, G.; HULLMANN, J.; APPENZELLER, S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. **Human Mutation**, v. 32, n. 2, p. 144–151, 2011. ISSN: 10597794. DOI: 10.1002/humu.21400

LIAO, P.; SATTEN, G. A.; HU, Y. PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. **Genetic Epidemiology**, v. 41, n. 5, p. 375–387, 2017. ISSN: 07410395. DOI: 10.1002/gepi.22048

LIM, E. B.; OH, H.; KIM, K. C.; KIM, M.; KIM, Y. J.; KIM, B. J.; NHO, C. W.; CHO, Y. S. Identification and functional validation of HLA-C as a potential gene involved in colorectal cancer in the Korean population. **BMC Genomics**, v. 23, n. 1, p. 261, 2022. ISSN: 1471-2164. DOI: 10.1186/s12864-022-08509-5

LOHMUELLER, K. E. The distribution of deleterious genetic variation in human populations. **Current Opinion in Genetics & Development**, v. 29, p. 139–146, 2014. ISSN: 0959437X. DOI: 10.1016/j.gde.2014.09.005

MAO, X.; YOUNG, B. D.; LU, Y. L. The Application of Single Nucleotide Polymorphism Microarrays in Cancer Research. **Current Genomics**, v. 8, n. 4, p. 219–228, 2007. ISSN: 13892029. DOI: 10.2174/138920207781386924

MARDIS, E.; MCCOMBIE, W. R. Solution-Phase Exome Capture. **Cold Spring Harbor Protocols**, v. 2017, n. 8, 2017. ISSN: 1940-3402. DOI: 10.1101/pdb.prot094680

MIELCZAREK, M.; SZYDA, J. Review of alignment and SNP calling algorithms for next-generation sequencing data. **Journal of Applied Genetics**, v. 57, n. 1, p. 71–79, 2016. ISSN: 1234-1983. DOI: 10.1007/s13353-015-0292-7

MILLER, D. T. et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. **The American Journal of Human Genetics**, v. 86, n. 5, p. 749–764, 2010. ISSN: 00029297. DOI: 10.1016/j.ajhg.2010.04.006

MIN, L.; NIE, M.; ZHANG, A.; WEN, J.; NOEL, S. D.; LEE, V.; CARROLL, R. S.; KAISER, U. B. Computational Analysis of Missense Variants of G Protein-Coupled

Receptors Involved in the Neuroendocrine Regulation of Reproduction. **Neuroendocrinology**, v. 103, n. 3–4, p. 230–239, 2016. ISSN: 0028-3835. DOI: 10.1159/000435884

MOODY, G. **Digital Code of Life: How Bioinformatics is Revolutionizing Science, Medicine, and Business.** . Wiley, 2004.

MUZZEY, D.; EVANS, E. A.; LIEBER, C. Understanding the Basics of NGS: From Mechanism to Variant Calling. **Current Genetic Medicine Reports**, v. 3, n. 4, p. 158–165, 2015. ISSN: 2167-4876. DOI: 10.1007/s40142-015-0076-8

NASLAVSKY, M. S. et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. **Human Mutation**, v. 38, n. 7, p. 751–763, 2017. ISSN: 10597794. DOI: 10.1002/humu.23220

NG, S. B,; et al. Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes. **Nature**, v. 461, p. 272–276, 2009

NGUYEN, H. T.; BOOCOCK, J.; MERRIMAN, T. R.; BLACK, M. A. SRBreak: A Read-Depth and Split-Read Framework to Identify Breakpoints of Different Events Inside Simple Copy-Number Variable Regions. **Frontiers in Genetics**, v. 7, 2016. ISSN: 1664-8021. DOI: 10.3389/fgene.2016.00160

NOWAK, D.; HOFMANN, W.; KOEFFLER, H. P. Genome-Wide Mapping Of Copy Number Variations using SNP Arrays. **Transfusion Medicine and Hemotherapy**, v. 36, n. 4, p. 246–251, 2009. ISSN: 1660-3818. DOI: 10.1159/000225372

NOWAKOWSKA, B. Clinical interpretation of copy number variants in the human genome. **Journal of Applied Genetics**, v. 58, n. 4, p. 449–457, 2017. ISSN: 1234-1983. DOI: 10.1007/s13353-017-0407-4

OLLILA, S. et al. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. **Gastroenterology**, p. 1408–1417, 2006

PARLA, J. S.; IOSSIFOV, I.; GRABILL, I.; SPECTOR, M. S.; KRAMER, M.; MCCOMBIE, W. R. A comparative analysis of exome capture. **Genome Biology**, v. 12, n. 9, p. R97, 2011. ISSN: 1465-6906. DOI: 10.1186/gb-2011-12-9-r97

PEREZ, Y.; BAR-YAACOV, R. KADIR, R.; WORMSER, O.; SHELEF, I.; BIRK, O. S.; FLUSSER, H.; BIRNBAUM, R. Y. Mutations in the microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. **Brain**, v. 142, n. 3, p. 574–585, 2019. ISSN: 0006-8950. DOI: 10.1093/brain/awz004

PETERSEN, B.; FREDRICH, B.; HOEPPNER, M. P.; ELLINGHAUS, D.; FRANKE, A. Opportunities and challenges of whole-genome and -exome sequencing. **BMC Genetics**, v. 18, n. 1, p. 14, 2017. ISSN: 1471-2156. DOI: 10.1186/s12863-017-0479-5

PROSDOCIMI, F. **Racionalizando a utilização do algoritmo PHRED para a análise de seqüências de DNA.** Belo Horizonte: Pós-graduação em Bioiformática. UFMG, 2016.

QUANG, D.; CHEN, Y.; XIE, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. **Bioinformatics**, v. 31, n. 5, p. 761–763, 2015. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu703

RAEVAARA, T. E. et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. **Gastroenterology** , v. 129, p. 537–

549, 2005. DOI: 10.1016/j.gastro.2005.06.005

RAMENSKY, V. Human non-synonymous SNPs: server and survey. **Nucleic Acids Research**, v. 30, n. 17, p. 3894–3900, 2002. ISSN: 13624962. DOI: 10.1093/nar/gkf493

RICHARDS, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **Genetics in Medicine**, v. 17, n. 5, p. 405–423, 2015. ISSN: 1098-3600. DOI: 10.1038/gim.2015.30

RICHTER, B. G.; SEXTON, D. P. Managing and Analyzing Next-Generation Sequence Data. **PLoS Comput Biol.**, v. 5, n. 6, 2009. DOI: 10.1371/journal.pcbi.1000369

RIGGS, E. R. et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). **Genetics in Medicine**, v. 22, n. 2, p. 245–257, 2020. ISSN: 10983600. DOI: 10.1038/s41436-019-0686-8.

RODENBURG, R. J. The functional genomics laboratory: functional validation of genetic variants. **Journal of Inherited Metabolic Disease**, v. 41, n. 3, p. 297–307, 2018. ISSN: 0141-8955. DOI: 10.1007/s10545-018-0146-7

ROGERS, M. F.; SHIHAB, H. A.; MORT, M.; COOPER, D. N.; GAUNT, T. R.; CAMPBELL, C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. **Bioinformatics**, v. 34, n. 3, p. 511–513, 2018. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx536

SALAZAR-SILVA, R. et al. NCOA3 identified as a new candidate to explain autosomal dominant progressive hearing loss. **Human Molecular Genetics**, v. 29, n. 22, p. 3691–3705, 2021. ISSN: 0964-6906. DOI: 10.1093/hmg/ddaa240

SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **J Mol Biol**, v. 95, n. 3, p. 441–448, 1975. DOI: 10.1016/0022-2836(75)90213-2

SCHWARZ, J. M.; COOPER, D. N.; SCHUELKE, M.; SEELOW, D. MutationTaster2: mutation prediction for the deep-sequencing age. **Nature Methods**, v. 11, n. 4, p. 361–362, 2014. ISSN: 1548-7091. DOI: 10.1038/nmeth.2890

SETH, A.; STEMPLE, D. L.; BARROSO, I. The emerging use of zebrafish to model metabolic disease. **Disease Models & Mechanisms**, v. 6, n. 5, p. 1080–1088, 2013. ISSN: 1754-8403. DOI: 10.1242/dmm.011346

SHAMPO, M. A.; KYLE, R. A. Frederick Sanger—Winner of 2 Nobel Prizes. **Mayo Clinic Proceedings**, v. 77, n. 3, p. 212, 2002. ISSN: 00256196. DOI: 10.4065/77.3.212

SHIGEMIZU, D.; MOMOZAWA, Y.; ABE, T.; MORIZONO, T.; BOROEVICH, K. A.; TAKATA, S.; ASHIKAWA, K.; KUBO, M.; TSUNODA, T. Performance comparison of four commercial human whole-exome capture platforms. **Scientific Reports**, v. 5, n. 1, p. 12742, 2015. ISSN: 2045-2322. DOI: 10.1038/srep12742

SOBREIRA, N.; SCHIETTECATTE, F.; VALLE, D.; HAMOSH, A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. **Human Mutation**, v. 36, n. 10, p. 928–930, 2015. ISSN: 10597794. DOI: 10.1002/humu.22844

SONG, J.; SHAO,. SNP Array in Hematopoietic Neoplasms: A Review. **Microarrays**, v. 5, n. 1, p. 1, 2015. ISSN: 2076-3905. DOI: 10.3390/microarrays5010001

STAAF, J.; VALLON-CHRISTERSSON, J.; LINDGREN, D.; JULIUSSON, G.; ROSENQUIST, R.; HÖGLUND, M.; BORG, Å.; RINGNÉR, M. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. **BMC Bioinformatics**, v. 9, n. 1, p. 409, 2008. ISSN: 1471-2105. DOI: 10.1186/1471-2105-9-409

STELZER, G. et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. **BMC Genomics**, v. 17, n. S2, p. 444, 2016. ISSN: 1471-2164. DOI: 10.1186/s12864-016-2722-2

SULONEN, A. et al. Comparison of solution-based exome capture methods for next generation sequencing. **Genome Biology**, v. 12, n. 9, p. R94, 2011. ISSN: 1465-6906. DOI: 10.1186/gb-2011-12-9-r94

TEER, J. K.; MULLIKIN, J. C. Exome sequencing: the sweet spot before whole genomes. **Human Molecular Genetics**, v. 19, n. R2, p. R145–R151, 2010. ISSN: 0964-6906. DOI: 10.1093/hmg/ddq333

TONG, W. et al. Whole-exome Sequencing Helps the Diagnosis and Treatment in Children with Neurodevelopmental Delay Accompanied Unexplained Dyspnea. **Scientific Reports**, v. 8, n. 1, p. 5214, 2018. ISSN: 2045-2322. DOI: 10.1038/s41598-018-23503-2

TROST, B. et al. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. **The American Journal of Human Genetics**, v. 102, n. 1, p. 142–155, 2018. ISSN: 00029297. DOI: 10.1016/j.ajhg.2017.12.007

VILLELA, D.; COSTA, S. S.; VIANNA-MORGANTE, A. M.; KREPISCHI, A. C. V.; ROSENBERG, C. Efficient detection of chromosome imbalances and single nucleotide variants using targeted sequencing in the clinical setting. **European Journal of Medical Genetics**, v. 60, n. 12, p. 667–674, 2017. ISSN: 17697212. DOI: 10.1016/j.ejmg.2017.08.020

VOSECKÁ, T.; et al. Detection of chromosome changes by CGH, array-CGH and SNP arraytechniques in tumours. **Cesk Patol**, v. 50, n. 1, p. 25–29, 2014.

WANG, X.; LI, X.; CHENG, Y.; SUN, X.; SUN, X.; SELF, S.; KOOPERBERG, C.; DAI, J. Y. Copy number alterations detected by whole-exome and whole-genome sequencing of esophageal adenocarcinoma. **Human Genomics**, v. 9, n. 1, p. 22, 2015. ISSN: 1479-7364. DOI: 10.1186/s40246-015-0044-0

WEISS, M. M.; HERMSEN, M. A.; MEIJER, G. A.; VAN GRIEKEN, N. C.; BAAK, J. P.; KUIPERS, E. J.; VAN DIEST, P. J. Comparative genomic hybridisation. **Molecular Pathology**, v. 52, n. 5, p. 243–251, 1999. ISSN: 1366-8714. DOI: 10.1136/mp.52.5.243

WHO. The use of next-generation sequeLeharancing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide. 2016. Disponível em: https://apps.who.int/iris/handle/10665/274443 Acesso em 19 de novembro de 2022

WINCHESTER, L.; YAU, C.; RAGOUSSIS, J. Comparing CNV detection methods for SNP arrays. **Briefings in Functional Genomics and Proteomics**, v. 8, n. 5, p. 353–366, 2009. ISSN: 1473-9550. DOI: 10.1093/bfgp/elp017

WISZNIEWSKA, J. et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. **European Journal of Human Genetics**, v. 22, n.

1, p. 79–87, 2014. ISSN: 1018-4813. DOI: 10.1038/ejhg.2013.77

XI, T.; JONES, I. M.; MOHRENWEISER, H. W. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. **Genomics**, v. 83, p. 970–979, 2004. DOI: 10.1016/j.ygeno.2003.12.016

YANG, A.; KIM, J.; CHO, S. Y.; JIN, D. Prevalence and risk factors for type 2 diabetes mellitus with Prader–Willi syndrome: a single center experience. **Orphanet Journal of Rare Diseases**, v. 12, n. 1, p. 146, 2017. ISSN: 1750-1172. DOI: 10.1186/s13023-017-0702-5.

YAO, R.; YU, T.; QING, Y.; WANG, J.; SHEN, Y. Evaluation of copy number variant detection from panel-based next-generation sequencing data. **Molecular Genetics & Genomic Medicine**, v. 7, n. 1, p. e00513, 2019. ISSN: 23249269. DOI: 10.1002/mgg3.513

YOHE, S.; THYAGARAJAN, B. Review of clinical next-generation sequencing. Archives of Pathology and Laboratory Medicine. **Archives of Pathology & Laboratory Medicine**, v. 141, n.11, p. 1544-1557, 2017. ISSN: 15432165. DOI: 10.5858/arpa.2016-0501-RA.

ZHANG, J.; WILSON, G. F.; SOERENS, A. G.; KOONCE, C. H.; YU, J.; PALECEK, S. P.; THOMSON, J. A.; KAMP, Timothy J. Functional Cardiomyocytes Derived From Human Induced Pluripotent Stem Cells. **Circulation Research**, v. 104, n. 4, 2009. ISSN: 0009-7330. DOI: 10.1161/CIRCRESAHA.108.192237

ZHAO, M.; WANG, Q.; WANG, Q.; JIA, P.; ZHAO, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. **BMC Bioinformatics**, v. 14, n. S11, p. S1, 2013. ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-S11-S1

Capítulo 1.3

Reprogramação celular para estudo de doenças genéticas

Células-tronco são definidas como “células com potencial prolongado ou ilimitado de autorrenovação, bem como com capacidade de originar pelo menos um tipo de célula diferenciada”. Elas estão presentes em todas as fases do desenvolvimento, mas suas potencialidades variam, sendo maiores durante a fase embrionária e menores no organismo adulto (GARCIA; FERNÁNDEZ, 2012).

Classicamente, as células-tronco são categorizadas de acordo com sua plasticidade como totipotentes, pluripotentes e multipotentes. Células totipotentes são capazes de originar qualquer tipo celular, inclusive as dos anexos extraembrionários. Células pluripotentes são as capazes de se diferenciar em qualquer célula dos folhetos embrionário (endoderma, mesoderma ou ectoderma), isto é, já não podem originar células dos anexos. Já as células multipotentes têm plasticidade mais limitada que as anteriores, podendo desenvolver apenas células de uma linhagem particular, com a qual já estão comprometidas (BOZDAĞ; YÜKSEL; DEMIRER, 2018; GARCIA; FERNÁNDEZ, 2012), ou seja, células multipotentes derivadas do mesoderma não podem gerar células dos outros folhetos embrionários. Há também as chamadas células unipotentes, que são as capazes de se diferenciar em um único tipo celular (BOZDAĞ; YÜKSEL; DEMIRER, 2018), como por exemplo as células-satélite dos músculos, que dão origem a mioblastos para que possa ocorrer miogênese e regeneração muscular (ARCHACKA K; KOWALSKI K; BRZÓSKA E., 2013).

A ampla capacidade de diferenciação das células-tronco embrionárias foi encarada pela comunidade científica como uma esperança para terapias celulares envolvendo doenças humanas como Parkinson, Alzheimer, diabetes e cardiomiopatias (ROBERTSON, 2001). O primeiro isolamento laboratorial bem-sucedido de células-tronco embrionárias humanas ocorreu em 1988 e, desde então, diversas discussões sobre entraves éticos e legais vinculados ao uso de embriões limitaram esse campo de pesquisa (WARREN, 2016). No contexto brasileiro, a Lei de Biossegurança (Lei nº 11.105/2005 – BRASIL, 2005), representou um avanço, pois o uso do embrião humano foi regulamentado – após muita discussão - para fins de pesquisa e terapia, desde que os embriões sejam inviáveis para fins reprodutivos ou congelados há três anos ou mais, mas

somente com consentimento dos genitores e aprovação por comitê de ética institucional (PEREIRA, 2008).

Em junho de 2006, Kazutoshi Takahashi e Shinya Yamanaka publicaram na revista *Cell* (TAKAHASHI; YAMANAKA, 2006) a descrição de um experimento no qual investigaram fatores de transcrição que poderiam induzir pluripotência em células somáticas. Ao induzir a expressão de quatro fatores específicos (*Oct3/4*, *Sox2*, *c-Myc* e *Klf4*), eles foram capazes de conduzir fibroblastos de camundongos (tanto embrionários quanto adultos) a expressar um fenótipo compatível com o de células-tronco embrionárias (ESC - *embryonic stem cell*), que eles denominaram iPS (*induced pluripotent stem cells*), atualmente também chamadas de iPSC. A essa técnica que faz com que células diferenciadas passem a apresentar fenótipo de pluripotênica, dá-se o nome de reprogramação celular.

O transplante subcutâneo das iPSC em camundongos nude (que são imunodeprimidos) resultou em tumores contendo uma variedade de tecidos dos três folhetos embrionários (teratomas). E a injeção das iPSC em blastocistos contribuiu para o desenvolvimento embrionário do camundongo (TAKAHASHI; YAMANAKA, 2006). Em 2007, Takahashi e Yamanaka publicaram um artigo sobre a reprogramação de fibroblastos humanos adultos em iPSC, utilizando os mesmos quatro fatores (TAKAHASHI et al., 2007), que ficaram conhecidos como “fatores Yamanaka” (HOKAYEM; CUKIER; DYKXHOORN, 2016). Em suma, as iPSC são células obtidas a partir da expressão forçada de fatores relacionados à pluripotência. Essa expressão permite que uma célula somática tenha seu estado diferenciado revertido a um estado semelhante ao embrionário.

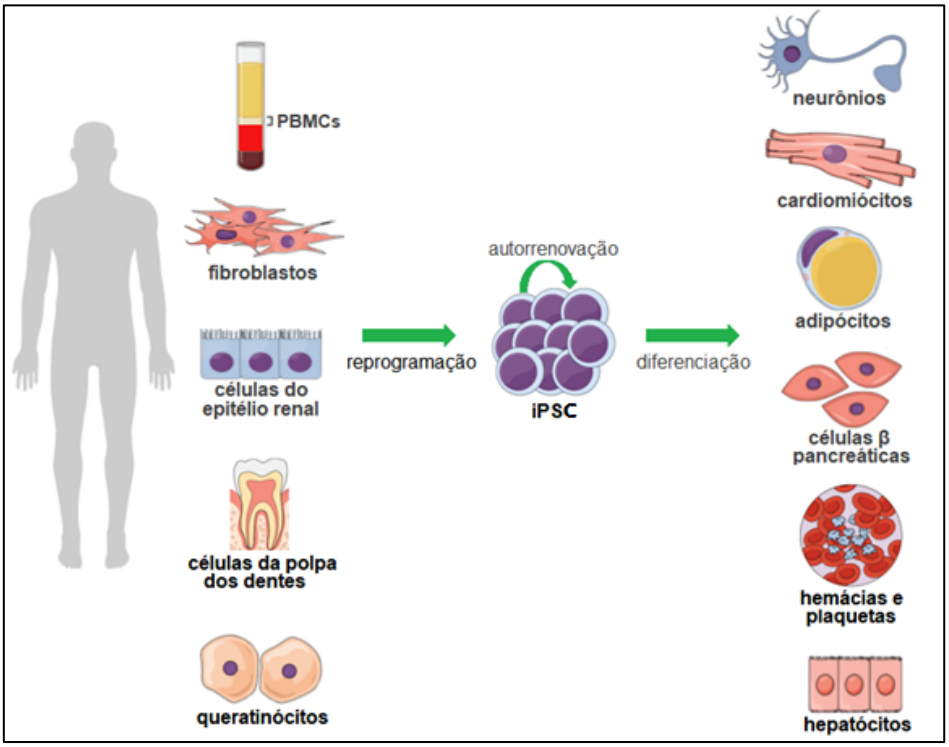
A partir da técnica proposta por Takahashi e Yamanaka, pode-se utilizar células de indivíduos já nascidos e reprogramá-las para posteriormente diferenciá-las em um tipo celular de interesse (**Figura 1**), contornando os dilemas éticos do uso de embriões. Além disso, as iPSC podem ser uma ferramenta importante para modelar e investigar doenças humanas (SCUDELLARI, 2016).

Por suas contribuições pioneiras no campo da reprogramação celular, Yamanaka foi laureado com o prêmio Nobel de Fisiologia / Medicina em 2012 (BAYART; COHEN-HAGUENAUER, 2013).

Já foram descritas na literatura reprogramações em iPSC a partir de diversos tipos celulares além dos fibroblastos, tais quais: células mononucleares de sangue periférico (PBMC - *peripheral blood mononuclear cells*) (CAI et al., 2020), células do epitélio renal

presentes na urina (WANG et al., 2016), células de polpa de dente (TAMAOKI et al., 2010), queratinócitos (AASEN et al., 2008), etc. Em condições específicas de cultura celular, é possível diferenciar as iPSC em diversos tipos celulares como neurônios (SHI; KIRWAN; LIVESEY, 2012; TANABE et al., 2018), cardiomiócitos (MUMMERY et al., 2012), adipócitos (GUÉNANTIN et al., 2017), células β pancreáticas (YABE et al., 2019), hepatócitos (TAKAYAMA; MIZUGUCHI, 2017), hemácias e plaquetas (FOCOSI; AMABILE, 2017) (**Figura 1**). Os principais fatores de crescimento e diferenciação para esses e em outros tipos celulares são apresentados no **Anexo 1** desta tese, em um esquema derivado da publicação de Williams e colaboradores (2012).

Figura 1: Reprogramação celular e diferenciação em tipos celulares de interesse.



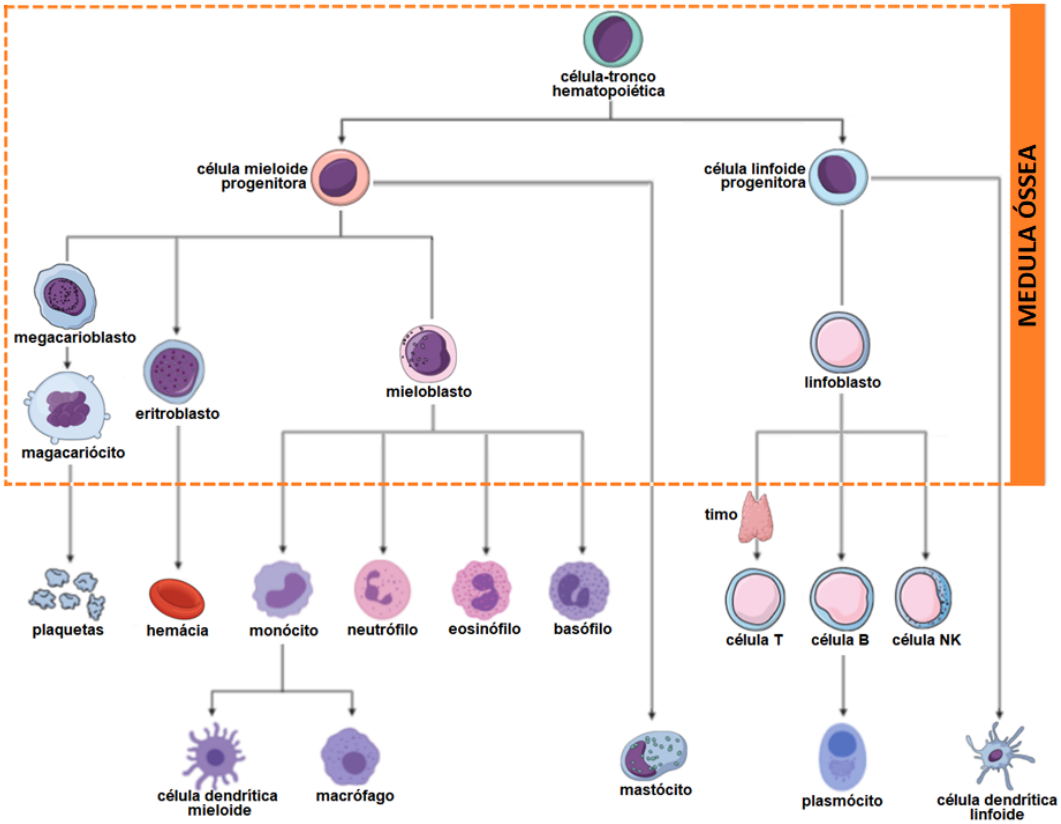
Com relação à entrega dos fatores de Yamanaka às células para reprogramação, a literatura descreve algumas estratégias: vetores virais, DNA episomal, transposons, mRNA ou entrega direta de proteínas. São preferíveis as opções com menor risco de integração ao DNA da célula hospedeira (BAYART; COHEN-HAGUENAUER, 2013).

Além dos fatores de Yamanaka (*OCT3/4, SOX2, KLF4* e *C-MYC*), outras combinações foram propostas para recombinação, como os fatores de Thomson *(SOX2, OCT3, LIN28* e *NANOG*) (XUE et al., 2012). Neste estudo, utilizamos a combinação proposta por Okita et al. (2011): *OCT3-4, SOX2, KLF4, L-MYC* e *LIN28*. Comparando-se esta combinação à de Yamanaka, *L-MYC* substitui *C-MYC* e acrescenta-se *LIN28*. C-

MYC e L-MYC são pertencentes à mesma família de proteínas, ambas formam heterodímeros com a proteína MAX e se ligam ao DNA alvo, sendo L-MYX mais potente na reprogramação (OKITA et al., 2011). *LIN28* aumenta a eficiência da reprogramação (HANNA et al., 2009) e também facilita a expressão de *OCT3/4* no nível pós-transcricional por ligação direta ao seu mRNA (QIU et al., 2010).

Por ser o sangue periférico um tecido de fácil acesso, com coleta simples e pouco invasiva, o uso de células advindas dele para gerar iPSC é tecnicamente conveniente (HOKAYEM; CUKIER; DYKXHOORN, 2016; MACK et al., 2011). Cabe dizer que o sangue periférico contém sobretudo os seguintes tipos celulares: hemácias (ou eritrócitos) e leucócitos, que por sua vez são divididos em fagócitos (neutrófilos, eosinófilos, basófilos e monócitos) e linfócitos (células B, T e *natural killer* - NK). Já as plaquetas são fragmentos de citoplasma derivados e megacariócitos da medula óssea. Esses tipos celulares de sangue são bastante diferenciados e mitoticamente inativos. As células sanguíneas são derivadas de células-tronco progenitoras (HOFFBRAND; MOSS, 2018), chamadas de PBSC (*peripheral blood stem cells*), que residem principalmente na medula óssea (**Figura 2**), sendo bastante raras em sangue periférico em condições fisiológicas (0,01-0,5% das células nucleadas) (BARNETT; ET AL., 1999; GRUBOVIC et al., 2017).

Figura 2: Representação esquemática simplificada da hematopoiese.



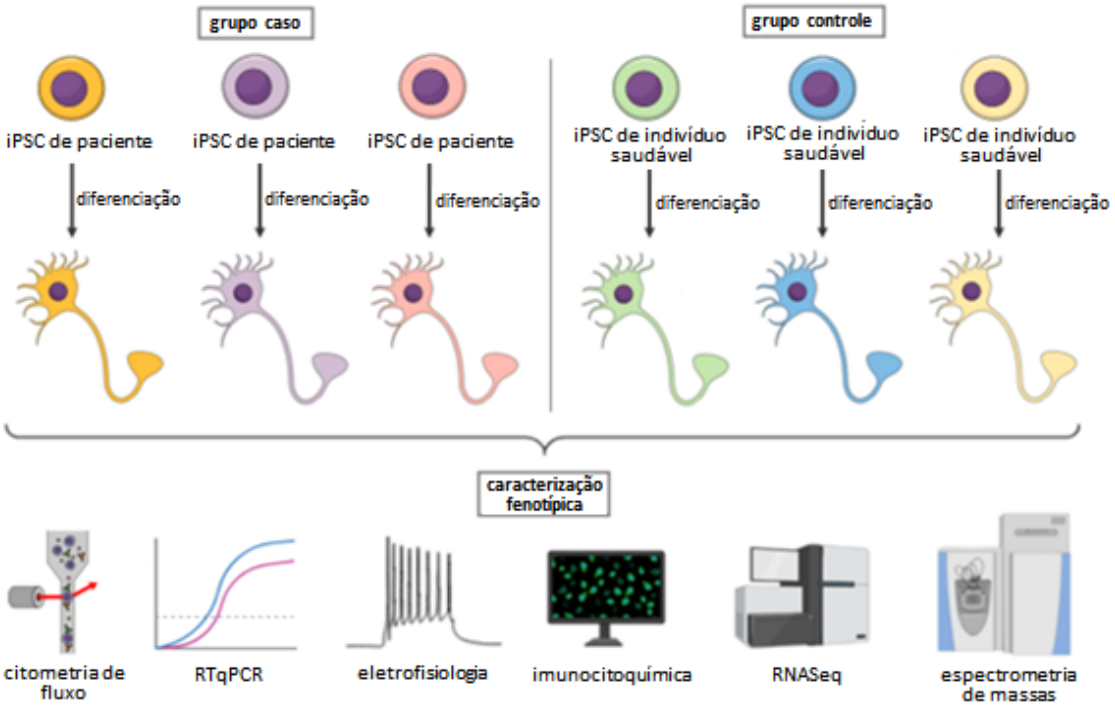
Os protocolos de reprogramação de células advindas do sangue normalmente começam com a separação das PBMC (CAI et al., 2020; VAN DEN AKKER et al., 2010; YANG; ET AL., 2008), que são uma fração sanguínea que contém células com núcleos redondos: células T, B, NK, monócitos, células dendríticas e PBSC (KLEIVELAND, 2015). Apesar de os linfócitos compreendem grande parte das PBMC, sua reprogramação tem como potencial limitação o fato deles estarem mais sujeitos a rearranjos. Além disso, as células somáticas que expressam marcadores de linhagens precoces - como CD34 - parecem mais suscetíveis à reprogramação. No entanto, como mencionado anteriormente, a quantidade de células progenitoras é muito baixa em sangue periférico, limitando o material de origem para reprogramação, sendo necessário expandi-las em condições adequadas de cultura (MACK et al., 2011). Neste projeto, utilizamos a expansão dos eritroblastos, que são células precursoras de hemácias capazes de se multiplicar *ex vivo* na presença de eritropoietina, fator de células estaminais (SCF – *stem cell factor*) e dexametasona (ENGLAND et al., 2011).

Como o sequenciamento de nova geração tem levado à identificação de variantes em diversos genes relacionados a doenças raras de fisiopatologia pouco compreendida (ADAMSON; SHERIDAN; GRIERSON, 2018), as células iPSC podem ser utilizadas para tentar responder a algumas perguntas biológicas nesse sentido. Células iPSC derivadas de pacientes com variantes genéticas permitem investigar quais vias biológicas estão alteradas. Seria impossível obter alguns tipos de tecido, como o cerebral, de pacientes vivos, mas as células iPSC permitem contornar essa dificuldade e mantém as particularidades genômicas do indivíduo em estudo, incluindo mutações causais (PERKEL, 2015). iPSC geradas a partir de células de pacientes com doenças genéticas e diferenciadas em tipos celulares de interesse são úteis na investigação de mecanismos fisiopatológicos e também testes de drogas, o que leva à expectativa de desenvolvimento de possibilidades terapêuticas (PEREIRA, 2017).

Para estudos funcionais de doenças do neurodesenvolvimento, estão estabelecidos protocolos de diferenciação de iPSC em diferentes classes de neurônios (FREEL et al., 2020; SHI et al., 2012; YANG et al., 2017) e também organoides cerebrais (BENITO-KWIECINSKI & LANCASTER, 2020). A diferenciação neuronal *in vitro* é mediada por fatores normalmente expressos no cérebro em desenvolvimento. Para estudo de fenótipos relacionados a funções cognitivas, costuma-se utilizar protocolos que mimetizam o desenvolvimento do córtex cerebral (ARDHANAREESWARAN et al., 2017). No Centro de Estudos do Genoma Humano e Células-tronco está bem estabelecido um protocolo de

diferenciação das iPSC em neurônios corticais por inibição de SMAD seguida por neurogênese *in vitro*, conforme Shi e colaboradores (2012) e adaptações introduzidas por Branco (2020). Para caracterização fenotípica, é possível utilizar diversas técnicas, a depender das perguntas biológicas do estudo, conforme representado na **Figura 3**.

Figura 3: iPSC podem ser diferenciadas em tipos celulares de interesse para serem usados como modelos de doenças e são amplas as possibilidades metodológicas para sua caracterização.



Referências

AASEN, T. et al. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. **Nature Biotechnology**, v. 26, n. 11, p. 1276–1284, 2008. ISSN: 1087-0156. DOI: 10.1038/nbt.1503

ADAMSON, K. I.; SHERIDAN, E.; GRIERSON, A. J. Use of zebrafish models to investigate rare human disease. **Journal of Medical Genetics**, v. 55, n. 10, p. 641–649, 2018. ISSN: 0022-2593. DOI: 10.1136/jmedgenet-2018-105358

ARCHACKA K.; KOWALSKI K.; BRZÓSKA E. Are satellite cells stem cells? **Postepy Biochem.** v. 59, n. 2, p. 205–2018, 2013

ARDHANAREESWARAN, K.; MARIANI, J.; COPPOLA, G.; ABYZOV, A.; VACCARINO, F. M. Human induced pluripotent stem cells for modelling neurodevelopmental disorders. **Nature Reviews Neurology**, v. 13, n. 5, p. 265–278, 2017. ISSN: 1759-4758. DOI: 10.1038/nrneurol.2017.45

BARNETT, D.; et al. Guideline for the flow cytometric enumeration of CD34+ haematopoietic stem cells. **Clin. Lab. Haem.** v. 21, p. 301–308, 1999. DOI: 10.1046/j.1365-2257.1999.00253.x

BAYART, E.; COHEN-HAGUENAUER, O. Technological Overview of iPS Induction from Human Adult Somatic Cells. **Current Gene Therapy**, v. 13, n. 2, p. 73–92, 2013. ISSN: 15665232. DOI: 10.2174/1566523211313020002

BOZDAĞ, S. C.; YÜKSEL, M. K.; DEMIRER, T. Adult Stem Cells and Medicine. **Advances in Experimental Medicine and Biology.** v. 1079, p. 17–36. DOI: 10.1007/5584_2018_184

BRANCO, Elisa Varella. **Caracterização genótipo-fenótipo em novas síndromes associadas ao Transtorno do Espectro Autista**. 2020. Universidade de São Paulo, São Paulo, 2020. DOI: 10.11606/D.41.2020.tde-20012020-101726. Disponível em: http://www.teses.usp.br/teses/disponiveis/41/41131/tde-20012020-101726/.

BRASIL. Lei nº 11.105, de 24 de março de 2005. Regulamenta os incisos II, IV e V do parágrafo 1º do art. 225 da Constituição Federal e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/l11105.htm Acesso em 19 de novembro de 2022

CAI, Y.; LIN, J.; MA, K.; HE, M.; LI, B. Generation of a human induced pluripotent stem cell line (SHUPLi001-A) from PBMCs of a healthy male donor. **Stem Cell Research**, v. 47, p. 101871, 2020. ISSN: 18735061. DOI: 10.1016/j.scr.2020.101871

ENGLAND, S. J.; MCGRATH, K. E.; FRAME, J. M.; PALIS, J. Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. **Blood**, v. 117, n. 9, p. 2708–2717, 2011. ISSN: 0006-4971. DOI: 10.1182/blood-2010-07-299743

FOCOSI, D.; AMABILE, G. Induced Pluripotent Stem Cell-Derived Red Blood Cells and Platelet Concentrates: From Bench to Bedside. **Cells**, v. 7, n. 1, p. 2, 2017. ISSN: 2073-4409. DOI: 10.3390/cells7010002

GARCIA, S. M. L.; FERNÁNDEZ, C. G. **Embriologia.** 3. ed., Porto Alegre: Artmed, 2012.

GRUBOVIC, R.; GEORGIEVSKI, B.; CEVRESKA, L.; GENADIEVA-STAVRIC, S.; GRUBOVIC, M. R. Factors Associated with Successful Mobilization and Collection of Peripheral Blood Hematopoietic Stem Cells in Autologous and Allogeneic Donors. **South East European Journal of Immunology**, v. 2017, p. 1, 2017. ISSN: 1857-9388. DOI: 10.3889/seejim.2017.20012

GUÉNANTIN, A. et al. Functional Human Beige Adipocytes From Induced Pluripotent Stem Cells. **Diabetes**, v. 66, n. 6, p. 1470–1478, 2017. ISSN: 0012-1797. DOI: 10.2337/db16-1107

HANNA, J.; SAHA, K.; PANDO, B.; VAN ZON, J.; LENGNER, C. J.; CREYGHTON, M. P.; VAN OUDENAARDEN, A.; JAENISCH, R. Direct cell reprogramming is a stochastic process amenable to acceleration. **Nature**, v. 462, n. 7273, p. 595–601, 2009. ISSN: 0028-0836. DOI: 10.1038/nature08592

HOFFBRAND, A. V.; MOSS, P. A. H. **Fundamentos em Hematologia de Hoffbrand.** . 7. ed., Artmed, 2018.

HOKAYEM, J. E.; CUKIER, H. N.; DYKXHOORN, D. M. Blood Derived Induced Pluripotent Stem Cells (iPSCs): Benefits, Challenges and the Road Ahead. **Journal of**

Alzheimer's Disease & Parkinsonism, v. 6, n. 5, 2016. ISSN: 21610460. DOI: 10.4172/2161-0460.1000275

KLEIVELAND, C. R. Peripheral Blood Mononuclear Cells. *In*: **The Impact of Food Bioactives on Health**. p. 161–167. DOI: 10.1007/978-3-319-16104-4_15

MACK, A. A.; KROBOTH, S.; RAJESH, D.; WANG, W. B. Generation of Induced Pluripotent Stem Cells from CD34+ Cells across Blood Drawn from Multiple Donors with Non-Integrating Episomal Vectors. **PLoS ONE**, v. 6, n. 11, p. e27956, 2011. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0027956

MUMMERY, C. L.; ZHANG, J.; NG, E. S.; ELLIOTT, D. A.; ELEFANTY, A. G.; KAMP, T. J. Differentiation of Human Embryonic Stem Cells and Induced Pluripotent Stem Cells to Cardiomyocytes. **Circulation Research**, v. 111, n. 3, p. 344–358, 2012. ISSN: 0009-7330. DOI: 10.1161/CIRCRESAHA.110.227512

OKITA, K. et al. A more efficient method to generate integration-free human iPS cells. **Nature Methods**, v. 8, n. 5, p. 409–412, 2011. ISSN: 1548-7091. DOI: 10.1038/nmeth.1591

PEREIRA, L. V.. A importância do uso das células tronco para a saúde pública. **Ciência & Saúde Coletiva**, v. 13, n. 1, p. 07–14, 2008. ISSN: 1413-8123. DOI: 10.1590/S1413-81232008000100002

PEREIRA, L. V. **Células-tronco: promessas e realidades**. 1. ed., São Paulo: Moderna, 2017.

PERKEL, J. M. Disease modeling with patient-specific iPS cells. **Science**, v. 347, n. 6227, p. 1271-1273 2015. DOI: 10.1126/science.347.6227.1271

QIU, C.; MA, Y.; WANG, J.; PENG, S.; HUANG, Y. Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. **Nucleic Acids Research**, v. 38, n. 4, p. 1240–1248, 2010. ISSN: 0305-1048. DOI: 10.1093/nar/gkp1071

ROBERTSON, J.A. Human embryonic stem cell research: ethical and legal issues. **Nature Reviews Genetics**, v. 2, n. 1, p. 74–78, 2001. ISSN: 1471-0056. DOI: 10.1038/35047594.

SCUDELLARI, M. How iPS cells changed the world. **Nature**, v. 534, n. 7607, p. 310–312, 2016. ISSN: 0028-0836. DOI: 10.1038/534310a

SHI, Y.; KIRWAN, P.; LIVESEY, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. **Nature Protocols**, v. 7, n. 10, p. 1836–1846, 2012. ISSN: 1754-2189. DOI: 10.1038/nprot.2012.116

TAKAHASHI, K.; TANABE, K.; OHNUKI, M.; NARITA, M.; ICHISAKA, T.; TOMODA, K.; YAMANAKA, S. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. **Cell**, v. 131, n. 5, p. 861–872, 2007. ISSN: 00928674. DOI: 10.1016/j.cell.2007.11.019

TAKAHASHI, K.; YAMANAKA, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. **Cell**, v. 126, n. 4, p. 663–676, 2006. ISSN: 00928674. DOI: 10.1016/j.cell.2006.07.024

TAKAYAMA, K.; MIZUGUCHI, H. Generation of human pluripotent stem cell-derived hepatocyte-like cells for drug toxicity screening. **Drug Metabolism and Pharmacokinetics**, v. 32, n. 1, p. 12–20, 2017. ISSN: 13474367. DOI:

10.1016/j.dmpk.2016.10.408

TAMAOKI, N. et al. Dental Pulp Cells for Induced Pluripotent Stem Cell Banking. **Journal of Dental Research**, v. 89, n. 8, p. 773–778, 2010. ISSN: 0022-0345. DOI: 10.1177/0022034510366846

TANABE, K; ANG, C. E.; CHANDA, S.; OLMOS, V. H.; HAAG, D.; LEVINSON, D. F.; SÜDHOF, T. C.; WERNIG. Transdifferentiation of human adult peripheral blood T cells into neurons. **Proceedings of the National Academy of Sciences**, v. 115, n. 25, p. 6470–6475, 2018. ISSN: 0027-8424. DOI: 10.1073/pnas.1720273115

VAN DEN AKKER, E.; SATCHWELL, T. J.; PELLEGRIN, S.; DANIELS, G.; TOYE, A. M. The majority of the in vitro erythroid expansion potential resides in CD34- cells, outweighing the contribution of CD34+ cells and significantly increasing the erythroblast yield from peripheral blood samples. **Haematologica**, v. 95, n. 9, p. 1594–1598, 2010. ISSN: 0390-6078. DOI: 10.3324/haematol.2009.019828

WANG, C.; HEI, F.; JU, Z.; YU, J.; YANG, S.; CHEN, M. Differentiation of Urine-Derived Human Induced Pluripotent Stem Cells to Alveolar Type II Epithelial Cells. **Cellular Reprogramming**, v. 18, n. 1, p. 30–36, 2016. ISSN: 2152-4971. DOI: 10.1089/cell.2015.0015

WARREN, H. Embryonic Stem Cell Research. **Plastic Surgical Nursing**, v. 36, n. 4, p. 157–161, 2016. ISSN: 0741-5206. DOI: 10.1097/PSN.0000000000000156

WILLIAMS, L. A.; DAVIS-DUSENBERY, B. N.; EGGAN, K. C. SnapShot: Directed Differentiation of Pluripotent Stem Cells. **Cell**, v. 149, n. 5, p. 1174- 1174.e1, 2012. ISSN: 00928674. DOI: 10.1016/j.cell.2012.05.015

XUE, B.; OLDFIELD, C. J.; VAN, Y.; DUNKER, A. K.; UVERSKY, Vladimir N. Protein intrinsic disorder and induced pluripotent stem cells. **Mol. BioSyst.**, v. 8, n. 1, p. 134–150, 2012. ISSN: 1742-206X. DOI: 10.1039/C1MB05163F

YABE, S. G.; FUKUDA, S.; NISHIDA, J.; TAKEDA, F.; NASHIRO, K.; OKOCHI, H. Induction of functional islet-like cells from human iPS cells by suspension culture. **Regenerative Therapy**, v. 10, p. 69–76, 2019. ISSN: 23523204. DOI: 10.1016/j.reth.2018.11.003

YANG, W.; et al. iPSC Reprogramming from Human Peripheral Blood Using Sendai Virus Mediated Gene Transfer. *In:* **StemBook.** Cambridge (MA), Harvard Stem Cell Institute, 2008. DOI: 10.3824/stembook.1.73.1

Capítulo 1.4

Zebrafish como modelo para estudo de doenças genéticas

As técnicas de sequenciamento de nova geração têm permitido a identificação de genes candidatos para diversas doenças genéticas raras (ADAMSON; SHERIDAN; GRIERSON, 2018). Modelos animais são amplamente utilizados experimentalmente para elucidar questões biomédicas (ANDERSEN; WINTER, 2019); é possível alterar seus genes para induzir doenças genéticas humanas (SIMMONS, 2008), permitindo a validação de genes candidatos e também a investigação de mecanismos fisiopatológicos. Nos últimos anos, a edição genética mediada por CRISPR-Cas9 tem se mostrado uma técnica útil para alterar genes específicos em diversos tipos celulares e modelos animais (PEREIRA, 2016).

O *zebrafish* (*Danio rerio*) é bastante utilizado como modelo animal para o estudo de doenças humanas (SANTORIELLO; ZON, 2012; SPENCE et al., 2007), o que inclui doenças metabólicas - como a obesidade (ZANG; MADDISON; CHEN, 2018) - e do neurodesenvolvimento (DE ABREU et al., 2020; LIU et al., 2018; MENEGHETTI et al., 2019; PEREZ et al., 2019). Tem considerável tolerância a grandes variações ambientais (temperatura, pH, etc.) e é facilmente reproduzido, o que também o torna atrativo para pesquisas científicas (DAMMSKI et al., 2011). Além disso, trata-se de um modelo significantemente mais barato que mamíferos (VELDMAN; LIN, 2008).

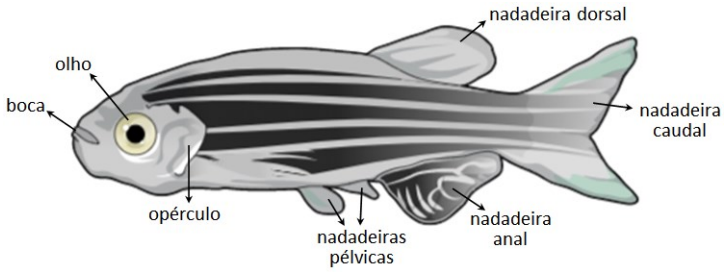
O *zebrafish* é um peixe de água doce, da ordem *Telostei*, família *Cyprinidae* (LOPES, 2018). Tem origem asiática, de rios da Índia, Bangladesh e Nepal (DAMMSKI et al., 2011). Ele é conhecido como paulistinha no Brasil (LOPES, 2018). Em ambiente natural, geralmente é encontrado em águas rasas, paradas ou de baixa movimentação. É uma espécie gregária, isto é, que se agrupa em populações (LAWRENCE, 2007). Costumam ser encontrados em pequenos cardumes de cinco a 20 indivíduos, mas também podem ser grupos maiores (DAMMSKI et al., 2011), com até centenas de indivíduos (SURIYAMPOLA et al., 2016).

O genoma do *zebrafish*, que está completamente sequenciado, apresenta uma considerável semelhança com o genoma humano Além disso, vários sistemas e órgãos desse modelo são bastante similares aos dos humanos (SETH; STEMPLE; BARROSO, 2013). Cerca de 70% dos genes associados a doenças humanas possuem ortólogos

funcionais óbvios no genoma de *zebrafish* (HOWE et al., 2013; SANTORIELLO; ZON, 2012). Outras vantagens são o rápido desenvolvimento e grande prole, o que reduz o tempo e o custo em relação a outros modelos também muito utilizados, como é o caso do camundongo (SETH; STEMPLE; BARROSO, 2013).

O corpo do *zebrafish* adulto é alongado, coberto por escamas, com cabeça curta (10-15% do corpo), narina protuberante e boca inclinada e voltada para cima. A mandíbula superior é protrusiva, o que torna a abertura da boca grande e ajuda na sucção de alimentos. As listras são uma característica marcante na espécie. Adultos têm, de cada lado, cinco listras com cor entre preto e azul, alternadas com listras mais claras. Também merecem destaque os opérculos, que são placas ósseas que protegem as brânquias, responsáveis pelas trocas gasosas, balanço osmótico e excreção de compostos nitrogenados (DAMMSKI et al., 2011). Alguns detalhes de sua anatomia são apresentados na **Figura 1**.

Figura 1: Anatomia externa do *zebrafish* adulto.



Machos são geralmente mais delgados e ligeiramente mais escuros que as fêmeas, sendo possível diferenciá-los visualmente (**Figura 2**). Internamente, os machos têm um par de testículos e as fêmeas têm ovários contendo ovócitos (DAMMSKI et al., 2011). Uma particularidade do *zebrafish* é que ele não tem cromossomos sexuais e não foi identificado um gene que possa direcionar a determinação sexual de maneira semelhante ao que ocorre em mamíferos e aves. O mecanismo de determinação sexual em *zebrafish* não é totalmente compreendido. Acredita-se que há vários genes envolvidos, mas sabe-se também que a temperatura tem influência no desenvolvimento gonadal. Temperaturas elevadas durante a embriogênese e estágio larval causam masculinização. Há também casos de fêmeas com reversão sexual induzida por temperatura elevada, que são as que possuem testículos e têm perfis de expressão gênica semelhantes aos machos normais (HOSSEINI et al., 2019).

Figura 2: Diferenciação visual de fêmeas e machos de *zebrafish*.

Superior: fêmea; inferior: macho. Animais adultos. Fonte: Braunbeck e Lammer, 2005.

Uma das principais características do *zebrafish* é sua grande prolificidade (DAMMSKI et al., 2011). Eles podem colocar de 200 a 300 ovos semanalmente (SANTORIELLO; ZON, 2012), a fertilização é externa e a liberação dos gametas depende da liberação de feromônios tanto pelos machos quanto pelas fêmeas. Pelo impacto positivo dessas interações no acasalamento, é importante manter machos e fêmeas compartilhando a mesma água por pelo menos 12 horas antes do acasalamento. Além disso, a ovulação é influenciada por fotoperíodo: ocorre normalmente um pouco antes do amanhecer e a desova começa nas primeiras horas de luz do dia (DAMMSKI et al., 2011).

O *zebrafish* é tipicamente diurno, sendo mais ativo nas primeiras horas da manhã. Eles dormem durante a noite, embora não exclusivamente. Muitos dos seus processos fisiológicos são governados por um padrão circadiano, sendo altamente dependentes do estabelecimento de um fotoperíodo regular. Interrupções no ciclo de iluminação podem acarretar problemas, especialmente no ciclo reprodutivo (DAMMSKI et al., 2011).

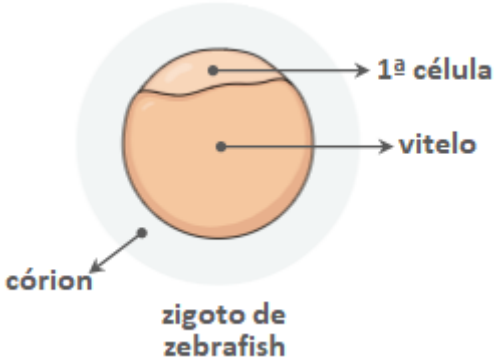
O *zebrafish* atinge maturidade sexual em três a seis meses pós-fertilização, a depender das condições ambientais (densidade populacional, temperatura e disponibilidade de alimento). Em condições apropriadas, o *zebrafish* é capaz de se reproduzir continuamente durante a maturidade sexual. As fêmeas são capazes de desovar diariamente, no entanto é recomendável um intervalo de descanso de uma semana entre posturas para produção de mais ovos viáveis (DAMMSKI et al., 2011).

Os embriões de *zebrafish* são transparentes e se desenvolvem fora do corpo, o que facilita a visualização de seu desenvolvimento (SANTORIELLO; ZON, 2012; SPENCE et al., 2007). Além disso, são visíveis a olho nu (0,7 milímetro de diâmetro no momento da fertilização) e isso, combinado com o desenvolvimento fora do corpo, facilita sua manipulação (SPENCE et al., 2007). Em contraste, o acesso a gametas, zigotos e

embriões de mamíferos é dificultado pelo desenvolvimento intrauterino, o que é um limitante técnico para edição gênica.

Uma estrutura importante do ovo de *zebrafish* (**Figura 3**) é o córion, uma membrana acelular transparente que reveste e protege o ovo até a eclosão (VALLIM, 2019), que geralmente ocorre em 2,5-3 dias pós-fertilização, quando se encerra a fase embrionária e começa a larval (DAMMSKI et al., 2011). A maior parte do ovo de *zebrafish* é ocupada pelo vitelo (GARCIA; FERNÁNDEZ, 2012), do qual o embrião se nutre e também as larvas durante algum tempo após a eclosão. As larvas são dependentes das reservas do saco vitelínico, que ainda é utilizado por um ou dois dias após o início da alimentação exógena e é totalmente absorvido até aproximadamente o sétimo dia pós-eclosão (DAMMSKI et al., 2011).

Figura 3: Morfologia do ovo de *zebrafish*.

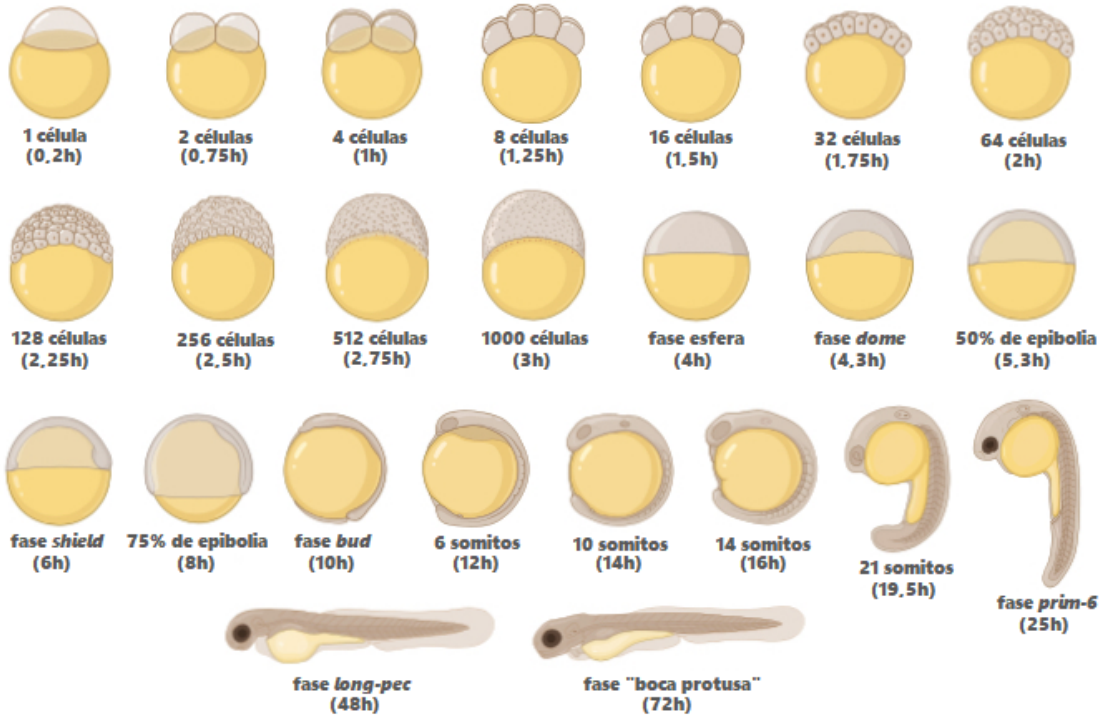


O *zebrafish* passa pelos estágios embrionário, larval (a partir da eclosão), juvenil e adulto (DAMMSKI et al., 2011). Seu desenvolvimento é bastante rápido. Precursores de todos os principais órgãos se desenvolvem dentro de 36 horas e as larvas apresentam comportamento de busca por alimentos cerca de cinco dias pós-fertilização (SPENCE et al., 2007). Com 90 dias, o *zebrafish* é considerado adulto (DAMMSKI et al., 2011). Sua expectativa de vida é de, em média, 3,5 anos, podendo viver até 5,5 anos (GERHARD et al., 2002). Aspectos importantes do desenvolvimento do *zebrafish* são apresentados na **Figura 4** e **Quadro 1**.

O *zebrafish* é onívoro. Sua dieta natural consiste principalmente de zooplâncton e insetos (SPENCE et al., 2007). No contexto experimental, costuma-se aplicar dietas industrializadas e, algumas vezes, também se utiliza o crustáceo artêmia (*Artemia nauplii*) (DAMMSKI et al., 2011; SPENCE et al., 2007). Os cistos de artêmia são metabolicamente inativos quando mantidos em ambientes secos, mas eclodem em 18 a

24 horas se mantidos em água salgada, podendo ser oferecidos como alimento ao *zebrafish* (DAMMSKI et al., 2011).

Figura 4: Representação esquemática do desenvolvimento inicial do *zebrafish*.



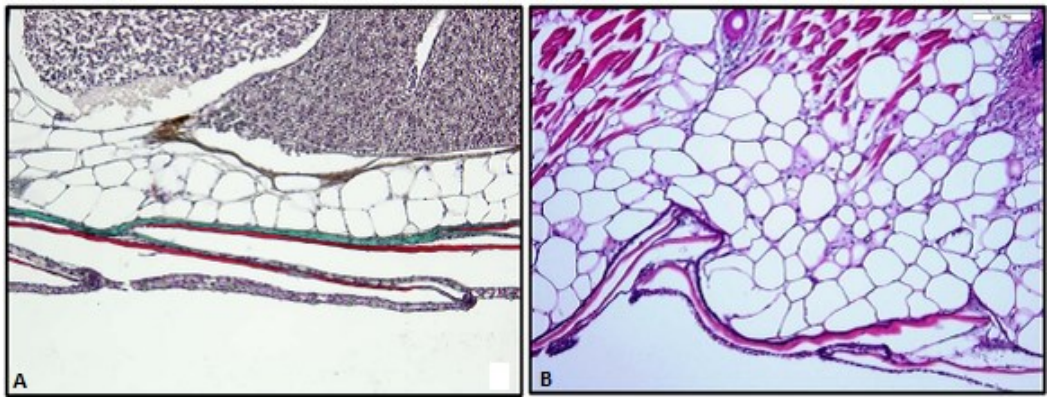
Quadro 1: Principais estágios do desenvolvimento do zebrafish pós-eclosão e respectivo tamanho

Estágio (dias)	Comprimento (mm)
larva jovem (3)	3,5
larva (14)	6
juvenil (30)	10
adulto (90)	20*

*Os adultos seguem crescendo e podem alcançar 40 a 50mm. Fonte: DAMMSKI et al., 2011.

No *zebrafish*, de maneira semelhante ao que ocorre em mamíferos, os excessos de nutrientes são armazenados sob a forma de grandes gotículas lipídicas uniloculares em adipócitos brancos (**Figura 5**). Isso contrasta com os modelos *Drosophila* e *Caenorhabditis elegans*, em que a gordura é armazenada em células não especializadas (dentro do corpo gordo ou intestino, respectivamente) que realizam várias outras funções além do armazenamento de lipídios (SETH; STEMPLE; BARROSO, 2013).

Figura 5: Tecido adiposo de *zebrafish* normal (A) e obeso (B).



Fonte: *Novel Zebrafish Model for Human Obesity* (*Epidermal Stem Cell Lab - Chung Yuan Christian University*), disponível em http://cdhsiao.weebly.com/research.html

A regulação da homeostase de energia do *zebrafish* envolve uma interação complexa entre órgãos múltiplos e sinais endócrinos. Ele é considerado um bom modelo para estudar doenças metabólicas, pois possui todos os órgãos-chave necessários para o controle metabólico em seres humanos, o que inclui os circuitos de apetite presentes no hipotálamo (o receptor de leptina e as proteínas do sistema de melanocortina são expressos), pâncreas e tecidos sensíveis à insulina (fígado, músculo e tecido adiposo branco). Os depósitos de tecido adiposo marrom, no entanto, não foram identificados nesse modelo (SETH; STEMPLE; BARROSO, 2013).

Historicamente, camundongos foram muito utilizados como organismos modelo para estudos sobre aprendizado (VISSERS; GILISSEN; VELTMAN, 2016). Mais recentemente, o *zebrafish* tem sido também utilizado como modelo para estudos de genes relacionados à deficiência intelectual (GUDMUNDSSON et al., 2019; O'RAWE et al., 2015), principalmente em função de seu baixo custo de criação e mais rápido desenvolvimento (VISSERS; GILISSEN; VELTMAN, 2016).

Para dar suporte a estudos envolvendo *zebrafish*, comumente utiliza-se as informações contidas no ZFIN (*Zebrafish Information Network* - zfin.org), no qual pesquisadores de todo o mundo depositam seus achados sobre esse modelo. Trata-se de uma ferramenta com curadoria criada em 1994 com ampla variedade de dados. Contém, por exemplo, dados genéticos, informações sobre linhas transgênicas, modelos de doenças humanas, expressão gênica, fenótipos, ortologia, regras de nomenclatura e instruções de manejo (BRADFORD; ET AL., 2017; HOWE et al., 2013).

Há protocolos bem estabelecidos para a edição de genes de *zebrafish* utilizando CRISPR-Cas9 (HRUSCHA; SCHMID, 2015; OTA et al., 2014; VEJNAR et al., 2016;

XIE et al., 2016) e essa ferramenta tem sido utilizada para gerar modelos de estudo de diversas doenças humanas (LIU et al., 2018; MENEGHETTI et al., 2019; PEREZ et al., 2019). Neste estudo, utilizamos o protocolo proposto por Vejnar e colaboradores (2016) para desenvolver um modelo para estudo da síndrome de Xia-Gibbs. Também há protocolos bem estabelecidos de criopresenvação de gametas de *zebrafish*, o que é fundamental para *backup* de recursos genéticos valiosos, como linhagens editadas geneticamente (BERGHMANS et al., 2004; DIOGO et al., 2018; MORRIS et al., 2003).

Referências

ADAMSON, K. I.; SHERIDAN, E.; GRIERSON, A. J. Use of zebrafish models to investigate rare human disease. **Journal of Medical Genetics**, v. 55, n. 10, p. 641–649, 2018. ISSN: 0022-2593. DOI: 10.1136/jmedgenet-2018-105358

ANDERSEN, M. L.; WINTER, L. M. F. Animal models in biological and biomedical research - experimental and ethical concerns. **Anais da Academia Brasileira de Ciências**, v. 91, n. suppl 1, 2019. ISSN: 1678-2690. DOI: 10.1590/0001-3765201720170238

BERGHMANS, S.; MORRIS, J. P.; KANKI, J. P.; LOOK, A. Thoma. Zebrafish Sperm Cryopreservation. **Methods in Cell Biology.** v. 77, p. 645–659, 2004. DOI: 10.1016/S0091-679X(04)77034-X

BRADFORD, Y. M.; et al. Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. **ILAR journal**, v. 58, p. 4–16, 2017. DOI: 10.1093/ilar/ilw040

BRAUNBECK, T; LAMMER, E. Detailed review paper "fish embryo toxicity assays. 2005

DAMMSKI, A. P.; MÜLLER, B. R.; GAYA, C.; REGONATO, D. **Zebrafish, manual de criação em biotério.** Curitiba: UFPR, 2011.

DE ABREU, M. S.; GENARIO, R.; GIACOMINI, A. C. V. V.; DEMIN, K. A.; LAKSTYGAL, A. M.; AMSTISLAVSKAYA, T. G.; FONTANA, B. D.; PARKER, M. O.; KALUEFF, A. V. Zebrafish as a Model of Neurodevelopmental Disorders. **Neuroscience**, 2020. ISSN: 03064522. DOI: 10.1016/j.neuroscience.2019.08.034

DIOGO, P.; MARTINS, G.; QUINZICO, I.; NOGUEIRA, R.; GAVAIA, Paulo J.; CABRITA, E. Electric ultrafreezer (− 150 °C) as an alternative for zebrafish sperm cryopreservation and storage. **Fish Physiology and Biochemistry**, v. 44, n. 6, p. 1443–1455, 2018. ISSN: 0920-1742. DOI: 10.1007/s10695-018-0500-6

GARCIA, S. M. L.; FERNÁNDEZ, C. G. **Embriologia.** 3. ed., Porto Alegre: Artmed, 2012.

GERHARD, G. S.; KAUFFMAN, E. J.; WANG, X.; STEWART, R.; MOORE, J. L.; KASALES, C. J.; DEMIDENKO, E.; CHENG, K. C. Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). **Experimental Gerontology**, v. 37, n. 8–9, p. 1055–1068, 2002. ISSN: 05315565. DOI: 10.1016/S0531-5565(02)00088-

8

GUDMUNDSSON, S. et al. TAF1, associated with intellectual disability in humans, is essential for embryogenesis and regulates neurodevelopmental processes in zebrafish. **Scientific Reports**, v. 9, n. 1, p. 10730, 2019. ISSN: 2045-2322. DOI: 10.1038/s41598-019-46632-8

HOSSEINI, S. et al. Genetic mechanism underlying sexual plasticity and its association with colour patterning in zebrafish (Danio rerio). **BMC Genomics**, v. 20, n. 1, p. 341, 2019. ISSN: 1471-2164. DOI: 10.1186/s12864-019-5722-1

HOWE, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. **Nature**, v. 496, n. 7446, p. 498–503, 2013. ISSN: 0028-0836. DOI: 10.1038/nature12111

HRUSCHA, A.; SCHMID, B. Generation of Zebrafish Zebrafish Models by CRISPRCRISPR /Cas9Cas9 Genome Editing Genome Editing. *In:* **Neuronal Cell Death. Methods in Molecular Biology**. p. 341–350. DOI: 10.1007/978-1-4939-2152-2_24

LAWRENCE, C. The husbandry of zebrafish (Danio rerio): A review. **Aquaculture**, v. 269, n. 1–4, p. 1–20, 2007.

LIU, C. X.; LI, C. Y.; HU, C. C.; WANG, Y.; LIN, J.; JIANG, Y. H.; LI, Q.; XU, Xiu. CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. **Molecular Autism**, v. 9, n. 1, p. 23, 2018. ISSN: 2040-2392. DOI: 10.1186/s13229-018-0204-x.

LOPES, F. D. **Avaliação da embriotoxidade de herbicida à base de glifosato, princípio ativo e surfactante sobre zebrafish (Danio Rerio)**. 2018. Dissertação (Mestrado em Ciências Aplicadas a Saúde) - Universidade Federal de Goiás, Jataí, 2018.

MENEGHETTI, G. et al. The epg5 knockout zebrafish line: a model to study Vici syndrome. **Autophagy**, v. 15, n. 8, p. 1438–1454, 2019. ISSN: 1554-8627. DOI: 10.1080/15548627.2019.1586247

MORRIS, J. P.; BERGHMANS, S.; ZAHRIEH, D.; NEUBERG, D. S.; KANKI, J. P.; LOOK, A. Thomas. Zebrafish sperm cryopreservation with N,N-dimethylacetamide. **BioTechniques**, v. 35, n. 5, p. 956–968, 2003. ISSN: 0736-6205. DOI: 10.2144/03355st03

O'RAWE, J. A. et al. TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. **The American Journal of Human Genetics**, v. 97, n. 6, p. 922–932, 2015. ISSN: 00029297. DOI: 10.1016/j.ajhg.2015.11.005

OTA, S.; HISANO, Y.; IKAWA, Y.; KAWAHARA, A. Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. **Genes to Cells**, v. 19, n. 7, p. 555–564, 2014. ISSN: 13569597. DOI: 10.1111/gtc.12154

PEREIRA, T. C. **Introdução à técnica de CRISPR**. 1. ed., Ribeirão Preto: SBG, 2016.

PEREZ, Y.; BAR-YAACOV, R.; KADIR, R.; WORMSER, O.; SHELEF, I.; BIRK, O. S.; FLUSSER, H.; BIRNBAUM, R. Y. Mutations in the microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. **Brain**, v. 142, n. 3, p. 574–585, 2019. ISSN: 0006-8950. DOI: 10.1093/brain/awz004

SANTORIELLO, C.; ZON, L. I. Hooked! Modeling human disease in zebrafish. **Journal of Clinical Investigation**, v. 122, n. 7, p. 2337–2343, 2012. ISSN: 0021-9738. DOI: 10.1172/JCI60434

SETH, A.; STEMPLE, D. L.; BARROSO, I. The emerging use of zebrafish to model metabolic disease. **Disease Models & Mechanisms**, v. 6, n. 5, p. 1080–1088, 2013. ISSN: 1754-8403. DOI: 10.1242/dmm.011346

SIMMONS, D. The use of animal models in studying genetic disease: transgenesis and induced mutation. **Nature Education** , v. 1, n. 1, p. 70, 2008. Disponível em: https://www.nature.com/scitable/topicpage/the-use-of-animal-models-in-studying-855/ Acesso em 20 de novembro de 2022

SPENCE, R.; GERLACH, G.; LAWRENCE, C.; SMITH, Carl. The behaviour and ecology of the zebrafish, Danio rerio. **Biological Reviews**, v. 83, n. 1, p. 13–34, 2007. ISSN: 14647931. DOI: 10.1111/j.1469-185X.2007.00030.x

SURIYAMPOLA, P. S.; SHELTON, D. S.; SHUKLA, R.; ROY, T.; BHAT, A.; MARTINS, E. P. Zebrafish Social Behavior in the Wild. **Zebrafish**, v. 13, n. 1, p. 1–8, 2016. ISSN: 1545-8547. DOI: 10.1089/zeb.2015.1159

VALLIM, J. H. **Avaliação dos efeitos toxicológicos do dimetoato e dimetoato nanoencapsulado em danio rerio (zebrafish).** 2019. Tese (Doutorado em Biologia Funcional e Molecular) - Instituto de Biologia, Universidade de Campinas (UNICAMP), Campinas, 2019.

VEJNAR, C. E.; MORENO-MATEOS, M. A.; CIFUENTES, D.; BAZZINI, A. A.; GIRALDEZ, A. J. Optimized CRISPR–Cas9 System for Genome Editing in Zebrafish. **Cold Spring Harbor Protocols**, v. 2016, n. 10, 2016. ISSN: 1940-3402. DOI: 10.1101/pdb.prot086850

VELDMAN, M. B.; LIN, S. Zebrafish as a Developmental Model Organism for Pediatric Research. **Pediatric Research**, v. 64, n. 5, p. 470–476, 2008. ISSN: 0031-3998. DOI: 10.1203/PDR.0b013e318186e609

VISSERS, L. E. L. M.; GILISSEN, C.; VELTMAN, J. A. Genetic studies in intellectual disability and related disorders. **Nature Reviews Genetics**, v. 17, n. 1, p. 9–18, 2016. ISSN: 1471-0056. DOI: 10.1038/nrg3999

XIE, S. L.; BIAN, W. P.; WANG, C.; JUNAID, M.; ZOU, J. X.; PEI, D. S. A novel technique based on in vitro oocyte injection to improve CRISPR/Cas9 gene editing in zebrafish. **Scientific Reports**, v. 6, n. 1, p. 34555, 2016. ISSN: 2045-2322. DOI: 10.1038/srep34555

ZANG, L.; MADDISON, L. A.; CHEN, W. Zebrafish as a Model for Obesity and Diabetes. **Frontiers in Cell and Developmental Biology**, v. 6, 2018. ISSN: 2296-634X. DOI: 10.3389/fcell.2018.00091

Capítulo 2
Investigação genética de casos de obesidade sindrômica: artigo original e relatos de casos

Neste segundo capítulo, apresentamos quatro artigos que versam sobre a investigação genética de casos de obesidade sindrômica. O primeiro (**2.1**) descreve o estudo de uma casuística de 20 indivíduos com obesidade sindrômica, publicado no *International Journal of Obesity*.

O segundo artigo (**2.2**) é um relato de caso envolvendo o gene *MYT1L,* derivado da casuística de **2.1**, acrescido de uma revisão de casos da literatura. Esse artigo foi publicado na revista *Obesity Research & Clinical Practice*.

O terceiro artigo (**2.3**), publicado no *Journal of Intellectual Disability Research*, relata dois casos de deficiência intelectual sindrômica envolvendo o gene *MED13L*, sendo um familial e um isolado. O caso isolado deriva da casuística de **2.1**.

O artigo apresentado em **2.4** relata um caso de obesidade sindrômica envolvendo o gene *SCAF4*. Embora não faça parte de **2.1** por ter sido analisado posteriormente, o estudo do paciente é pertinente para o tema da tese. Esse artigo foi publicado na revista *American Journal of Medical Genetics: Parte A*. Trata-se do segundo trabalho que relaciona *SCAF4* a uma doença mendeliana do neurodesenvolvimento.

Capítulo 2.1

Genetic investigation of syndromic forms of obesity

CARVALHO, L. M. L; D'ANGELO, C. S; VILLELA, D.; COSTA, S. S.; JORGE, A. A. L.; SILVA, I. T.; SCLIAR, M. O.; CHAVES, L. D.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. Genetic investigation of syndromic forms of obesity. **International Journal of Obesity**, 2022. ISSN: 0307-0565. DOI: 10.1038/s41366-022-01149-5

ARTICLE

Genetics and Epigenetics

Genetic investigation of syndromic forms of obesity

Laura Machado Lara Carvalho[1], Carla Sustek D'Angelo[1], Darine Villela[1], Silvia Souza da Costa[1], Alexander Augusto de Lima Jorge[2], Israel Tojal da Silva[3], Marília de Oliveira Scliar[1], Luiza Dias Chaves[1], Ana Cristina Victorino Krepischi[1], Celia Priszkulnik Koiffmann[1] and Carla Rosenberg[1]✉

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Background: Syndromic obesity (SO) refers to obesity with additional phenotypes, including intellectual disability (ID)/developmental delay (DD), dysmorphic features, or organ-specific abnormalities. SO is rare, has high phenotypic variability, and frequently follows a monogenic pattern of inheritance. However, the genetic etiology of most cases of SO has not been elucidated. Subjects and methods: In this study, we investigated 20 SO patients by whole-exome sequencing (WES) trios to identify causal genetic variants. Results: 4/20 patients had negative results for array comparative genomic hybridization (aCGH) analyses. In the remaining 15 patients, in addition to SNVs and indels, CNVs were also evaluated. Pathogenic/likely pathogenic (P/LP) SNVs/indels were detected in 6/20 patients (involving *MED13L*, *AHDC1*, *EHMT1*, *MYT1L*, *GRIA3*, and *SETD1A*), while two patients carried an inherited VUS. In addition, P/LP CNVs were observed in 3/15 patients (involving *SATG2*, *KIAA0442*, and *MEIS2*). Conclusions: All nine detected P/LP variants involved genes already known to lead to syndromic ID/DD; however, for only two genes (*EHMT1* and *MYT1L*) is the link with obesity well established. This is the first study applying a comprehensive genomic investigation of an SO cohort, showing a high diagnostic yield (~47%). Additionally, our findings suggested that several known ID/DD genes may also predispose individuals to SO.

International Journal of Obesity; https://doi.org/10.1038/s41366-022-01149-5

INTRODUCTION

Syndromic obesity (SO) refers to obesity occurring with additional phenotypes such as dysmorphic features or organ-specific developmental abnormalities [1, 2], but the majority of patients have intellectual disability (ID)/developmental delay (DD) [3, 4]. SO is rare [4], presents high phenotypic variability [2], and frequently follows a monogenic pattern of inheritance [2, 4].

Kaur and colleagues (2017) [4] reported that only few cases of SO described in the literature (19/79 analyzed) had their genetic etiology elucidated. The gap in knowledge regarding genetic etiologies of SO impairs genetic counseling and hinders the development of therapeutic approaches.

In the present study, we investigated the genetic etiology of 20 cases of SO based on the analysis of copy-number variations (CNVs), single-nucleotide variants (SNVs), and small insertions and deletions (indels) using data from whole-exome sequencing (WES) of parent-proband trios.

SUBJECTS AND METHODS

Patients

The cohort comprised 20 patients with a clinical diagnosis of SO (16 males and 4 females; 1.35 to 22.64 years, $\bar{x} = 9.9 \pm 5.3$ SD). Supplementary material contains clinical features data (ST1) and obesity scores (SF1).

Cognitive impairment was established in most cases based on developmental milestones, school reports, and the ability to perform daily tasks, self-care, and communication. To determine the presence of polyphagia, we considered the parents' answers regarding behavior towards food.

All patients had negative results for:

karyotype (G-banding);
methylation test of SNURF-SNRPN exon 1(Prader-Willi syndrome region);
MLPA SALSA P064, P036, and P070 (MRC-Holland);
Fragile X syndrome (AmplideX PCR/CE FMR1–Asuragen).

NGS sequencing

We extracted DNA from peripheral blood samples of the trios (60 individuals). Of the 20 patients investigated, five already had a previous array comparative genomic hybridization (aCGH) analysis with negative results. For the other 15 patients, we analyzed CNVs, in addition to SNVs and indels, from sequencing data. For the 15 patients without previous CNV analyses, we used Oneseq+SureSelect Human All Exon V6 (Agilent Technologies) for capture, which contains probes both for detection of CNVs and for exonic regions (SNVs and indels). In the other five patients (patients 7, 8, 10, 14, and 19—ST1), exome libraries were built using SureSelect Human All Exon V5 (Agilent Technologies) (Fig. 1). The samples were sequenced using HiSeq (Illumina) paired-end sequencing.

[1]Human Genome and Stem Cell Research Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo (USP), São Paulo, SP, Brazil. [2]Genetic Endocrinology Unit, Cellular and Molecular Endocrinology Laboratory (LIM/25), Faculty of Medicine, University of São Paulo (USP), São Paulo, SP, Brazil. [3]International Center for Research, A. C. Camargo Cancer Center, São Paulo, SP, Brazil. ✉email: carlarosenberg@ib.usp.br

Received: 18 February 2022 Revised: 6 May 2022 Accepted: 9 May 2022
Published online: 21 May 2022

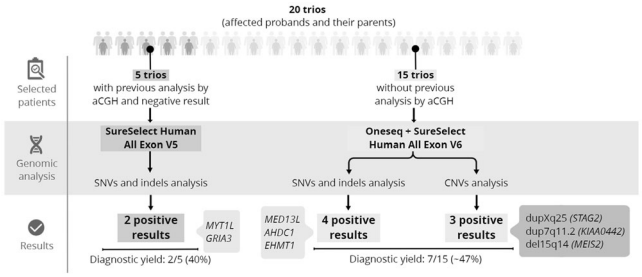


Fig. 1 Genetic investigation workflow of SO etiology in our cohort and positive results distribution. We selected 20 patients who met the inclusion criteria. NGS sequencing of these patients and their parents (trios) was performed. Two types of capture were used. Whereas 5 of the patients already had CNV analysis by aCGH with a negative result, the sequencing of the trios was performed using SureSelect Human All Exon V5 (Agilent Technologies—California, USA) capture, and for them, only SNVs and indels were analyzed, which led us to 2 positive results (2/5 = 40%). For the others (15 trios), sequencing Oneseq + V6 SureSelect Human All Exon (Agilent Technologies—California, USA) was performed to analyze CNVs, SNVs and indels. A total of 4/15 were positive for SNVs/indels, and 3/15 were positive for CNVs, totaling 7/15 (47%). The genes involved in the positive findings are those shown in italic calligraphy.

The raw data were aligned to the GRCh37/hg19 reference genome using the Burrows–Wheeler Aligner (BWA) and processed with GATK.

CNV analysis of NGS data

BAM files of 15 trios were analyzed using the Nexus Copy Number™ software (v.9-BioDiscovery) with a significance threshold of 1.0E-8. Log2 ratio and B allele frequency (BAF) values were plotted along chromosomal coordinates. As a reference, we used sex-matched BAM files.

The CNVs were classified according to the European Guidelines for constitutional cytogenomics analyses [5].

SNV and indel analysis

For analysis of SNVs and indels, the NGS data of each trio were analyzed using VarSeq software (Golden Helix). We used the following strategies:

Read depths ≥10, genotype qualities ≥17;
Low frequency on 1KG Phase3, gnomAD, and ABraOM: ≤0.01 for dominant and ≤0.03 for the recessive inheritance hypothesis;
Missense and loss-of-function variants according to Ensembl Genome Browser;
Gene panels: Panelapp—Severe early-onset obesity 2.3, Maltese et al., 2018 [6], SYSID;
VarElect program;
ClinVar, OMIM, and Geneimprint database;
UPD14 verification by genotypic comparison of mother-proband.

To classify the variants of the SNVs and indels types according to their pathogenicity, we used the American College of Medical Genetics and Genomics (ACMG) guidelines [7] using the Varsome program. Family segregation analysis involving other relatives was performed by Sanger sequencing.

RESULTS

In nine patients, pathogenic/likely pathogenic (P/LP) variants that explained the phenotype were detected (Fig. 1 and Table 1). P/LP CNVs were detected in 3/15 patients, and P/LP SNVs or indels were detected in 6/20 patients. Two patients showed missense variants classified as VUS (variant of unknown significance), all inherited. No candidate variants were identified for Patients 12–20.

DISCUSSION

The combination of CNV, SNV, and indel analysis of NGS data led to a high diagnostic yield (7/15; ~47%) in a SO cohort with previous negative results for Prader-Willi, karyotyping, and Fragile-X syndrome.

All nine P/LP variants identified involved genes known to lead to syndromic ID/DD. In only two patients (6 and 7), the causal relation between the genes involved and SO is well established (*EHMT1* and *MYT1L*). Patient 6 has a loss-of-function variant in the *EHMT1* gene, which haploinsufficiency leads to Kleefstra syndrome 1, characterized by ID/DD, childhood hypotonia, and distinct facial features; BMI > 25 is present in 30–40% of patients [8].

Patient 7 has a *MYT1L* loss-of-function variant. In a recent review of patients with *MYT1L* variants [9], we showed that 71% of individuals with heterozygous *MYT1L* SNVs or indels have obesity/overweight.

The remaining seven P/LP variants involve genes not previously associated with obesity. One aspect is that obesity is most often multifactorial and highly prevalent in the population. In Brazil, 35% of adults are overweight and 32% have obesity, while 9% of children aged 0 to 5 years have obesity/overweight (https://sisaps.saude.gov.br/sisvan/relatoriopublico/index - 2020 data). Therefore, high BMI values are likely to be present in syndromic patients, regardless of their underlying genetic defect. Accordingly, studies investigating the genetic etiology of SO also identified variants mapped to ID genes in which obesity is not a central feature [1, 4, 10].

On the other hand, there is a higher prevalence of obesity among people with cognitive impairment [11, 12]. Additionally, most patients in our cohort had behavioral problems (see ST1) and there is a relationship between obesity and psychiatric disorders [13]. Some psychiatric drugs may induce weight gain [14] and some patients in our cohort used such medications, including risperidone (patients 1 and 4), quetiapine (patient 3), periciazine (patient 15), imipramine (patient 8) and chlorpromazine (patient 8).

In a recent study in a cohort of SO conducted by us, P/LP CNVs were identified by aCGH in 67/279 (24%) patients [1]. Vuillaume and colleagues [10] obtained a similar value (23%). In the present study, we identified three patients carrying likely causative CNVs (3/15; 20%), encompassing genes not previously associated with obesity.

We detected a familial case of two maternal half brothers carrying a X-chromosome microduplication encompassing *STAG2*. Although the mother must be an obligate carrier, the variant was not detected in her peripheral blood DNA. The proband (patient 1) has obesity, but his brother BMI is in the normal range.

Patient 2 had a 4.4 Mb duplication including *KIAA0442*, whose loss-of-function are related to syndromic ID. Although the impact of duplications has not been described, several characteristics of *KIAA0442* haploinsufficiency [15] were present in patient 2, such as kyphosis and genital malformation.

Patient 3 has a microdeletion involving *MEIS2*, and the phenotype is compatible with haploinsufficiency of this gene [16].

Currently, there is no study in the SO literature reporting a genome-wide evaluation of pathogenic SNV/indel variants based on WES. In 2018, Kleinendorst and colleagues [17] investigated patients with suspected genetic obesity (syndromic and nonsyndromic) using a targeted gene panel of 55 genes, with P/LP variants in 9.3% of the patients. Most of the identified variants, however, were linked to nonsyndromic monogenic forms of obesity.

We identified six P/LP SNVs or indels, only two of which involving genes previously associated with SO. Patient 5 had *AHDC1* loss-of-function. Although obesity is not among the described features of this syndrome, some patients have obsession with food and satiety dysfunctions (personal communication: Dr. Richard Gibbs, who described this syndrome in 2014 [18] and Xia-Gibbs Society).

The clinical case of patient 4, which involves a *MED13L* frameshift variant, is described in detail in a recent publication

Table 1. Clinically significant variants—pathogenic, likely pathogenic and VUS - identified with NGS data analysis.

	Case	Sex	Variant type	Variant description	Classification	Additional data
Pathogenic/likely pathogenic						
CNVs	1	M	Microduplication	seq[GRCh37]dup(X)(q25)mat NC_000023.10:g.123056490_123433630dup size: 377 kb OMIM morbid gene: *STAG2* inheritance: maternal gonadal mosaicism assumed	Pathogenic	Included critical region of Xq25 duplication syndrome (OMIM # 300979) Affected maternal half-brother duplication carrier (aCGH analysis)
	2	M	Duplication	seq[GRCh37]dup(7)(q11.21-q11.22)pat NC_000007.13:g.66112731_70542831dup size: 4.4 Mb OMIM morbid genes: *KIAA0442* and *SBDS* inheritance: paternal	Likely pathogenic	*KIAA0442:* intellectual developmental disorder, autosomal dominant 26 (OMIM # 615834)
	3	M	Microdeletion	seq[CRCh37]del(15q14)dn NC_000015.9:g.36714646_38311846del size: 1.5 Mb OMIM genes: *MEIS2* inheritance: *de novo*	Pathogenic	cleft palate, cardiac defects, and mental retardation (OMIM # 600987)
indel	4	F	Frameshift (deletion)	GRCh37(hg19):Chr9:116444137-G- HGVS: *MED13L*(NM_015335.5):c.2318delC (p.Ser773LeufsTer28) zigozity: heterozygous inheritance: *de novo*	Pathogenic (PVS1, PS2, and PM2)	mental retardation and distinctive facial features with or without cardiac defects (OMIM # 616789)
SNVs	5	M	Nonsense	GRCh37(hg19):Chr1:27877633-G-A HGVS: *AHDC1*(NM_001029882.3):c.994C>T(p.Gln332Ter) zigozity: heterozygous inheritance: *de novo*	Pathogenic (PVS1, PS2, PM2, PP3, and PP5)	Xia-Gibbs syndrome (OMIM # 615829)
	6	M	Nonsense	GRCh37(hg19):Chr9:140695428-C-T HGVS: *EHMT1*(NM_024757.5):c.2704C>T(p.Arg902Ter) zigozity: heterozygous inheritance: *de novo*	Pathogenic (PVS1, PS2, PM2, PP3, and PP5)	Kleefstra syndrome type 1 (OMIM # 610253)
	7	M	Nonsense	GRCh37(hg19):Chr2:1843005-G-T HGVS: *MYT1L*(NM_015025.4):c.2990C>A(p.Ser997Ter) zigozity: heterozygous inheritance: *de novo*	Pathogenic (PVS1, PS2, PM2, and PP3)	mental retardation, autosomal dominant 39 (OMIM # 616521)
	8	M	Missense	GRCh37(hg19):ChrX:122561952-GGG-TGT HGVS: *GRIA3*(NM_000828.4): c.2038_2040delGGGinsTGT (p.Gly680Cys) zigozity: hemizygous inheritance: *de novo*	Pathogenic (PM2, PS2, and PP3)	Intellectual developmental disorder, X-linked, syndromic, Wu type (OMIM # 300699)
	9	M	Missense	GRCh37(hg19): Chr16:30978900-G-A HGVS: SETD1A(NM_014712.3):c.2761G>A (p.Asp921Asn) zigozity: heterozygous inheritance: paternal	Likely pathogenic (PM2, PP1, PP2, and PP4)	Neurodevelopmental disorder with speech impairment and dysmorphic facies (OMIM # 619056) patient's father has severe obesity and ID

Table 1. continued

Case	Sex	Variant type	Variant description	Classification	Additional data
VUS (variants of uncertain clinical significance)					
indel 10	M	Missense (in frame variant)	GRCh37(hg19):ChrX:79945495-AAT- HGVS: *BRWD3*(NM_153252.5): c.3697_3699delATT (p.Ile1233del) zigozity: hemizygous inheritance: maternal	VUS (PM2, PM4, and PP3)	mental retardation, X-linked 93 (OMIM # 300659) obs.: presence of the variant by sanger confirmed in maternal grandmother and maternal aunt, but absent in brother, father and maternal uncle; Llikely affected mother (obesity and speech impairment)
SNV 11	M	Missense	GRCh37(hg19):Chr7:155596307-C-T HGVS: *SHH*(NM_000193.4):c.676G>A(p.Ala226Thr) zigozity: heterozygous inheritance: maternal	VUS (PM1, PP2, PP3, and PP5)	holoprosencephaly 3 (OMIM # 142945) obs.: presence of the variant confirmed by sanger in one sister and absent in the other. *SHH* pathogenic variants have incomplete penetrance and variable expressivity

CNVs Copy-Number Variations, *SNVs* Single-nucleotide Variants, *M* male, *F* female, *aCGH* array comparative genomic hybridization, *kb* kilobases, *Mb* megabases, *ID* intellectual disability

by our group [19]. Patient 8 has a *GRIA3* pathogenic variant in hemizygosis, that explained his phenotype [20].

Patient 9 and his affected father carry a *SETD1A*(NM_014712.3): c.2761G>A(p.Asp921Asn), which could partly explain the phenotype. Deleterious — including missense—heterozygous variants in *SETD1A* cause a neurodevelopmental disorder characterized by speech impairment, dysmorphic facial features and behavioral problems [21]. In an analysis of 15 individuals with *SETD1A* variants, Kummeling and colleagues (2020) [21] found only three with obesity, which is lower than the population frequency. Additionally, 6/15 individuals analyzed by them had visual impairment, and patient 9 had mild color deficiency.

Patient 11 has a variant previously described in a publication [22] as pathogenic in the *SHH* gene, presenting incomplete penetrance and variable expressiveness. Patient´s 11 phenotype was not suggestive of variants in this gene.

Patient 10 has a VUS in *BRWD3, a* gene related to a recessive X-linked form of ID. We performed segregation analysis for brother, maternal grandmother, aunt, and uncle. The variant was identified in all women and was absent in the tested men, so it could not be excluded as VUS.

In conclusion, methods of genome-wide investigation (DNA microarray and whole exome/genome sequencing) should be prioritized over the target sequencing of well-characterized SO gene panels. The simultaneous analysis of CNVs, SNVs, and indels using NGS data is a cost-effective option for SO genetic investigation. Our study is the first to apply a comprehensive genomic investigation of an SO cohort, showing a high diagnostic yield and additionally suggesting that several genes known to cause ID may also predispose to SO.

INTERNET RESOURCES

1KG Phase3: https://phase3browser.1000genomes.org/index.html
ABESO BMI calculator: https://abeso.org.br/obesidade-e-sindrome-metabolica/calculadora-imc/
ABraOM: http://abraom.ib.usp.br/
ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
Ensembl: https://www.ensembl.org
geneimprint: https://www.geneimprint.com/
gnomAD: https://gnomad.broadinstitute.org/
OMIM: https://www.omim.org/
Panel App - Severe early-onset obesity 2.3: https://panelapp.genomicsengland.co.uk/panels/130/
SimulConsult head circumference calculator: https://simulconsult.com/resources/measurement.html
SISVAN (Food and Nutrition Surveillance System): https://sisaps.saude.gov.br/sisvan/relatoriopublico/index
SYSID: https://sysid.cmbi.umcn.nl
VarElect: https://varelect.genecards.org/
Varsome: https://varsome.com/

DATA AVAILABILITY

Data on the genetic variants relevant to this study are openly available in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). Accession codes are given below: SCV002060423; SCV002060422; SCV002060421; SCV002060420; SCV002060419; SCV002060418; SCV001478246; SCV001432217.

REFERENCES

1. D'Angelo CS, Varela MC, De Castro CIE, Otto PA, Perez ABA, Lourenço CM, et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. Mol Cytogenet. 2018;11. https://doi.org/10.1186/s13039-018-0363-7
2. Geets E, Meuwissen MEC, Van Hul W. Clinical, molecular genetics and therapeutic aspects of syndromic obesity. Clin Genet. 2019;95:23–40. https://doi.org/10.1111/cge.13367

3. Delrue M-A, Michaud J. Fat chance: genetic syndromes with obesity. Clin Genet. 2004;66:83–93. https://doi.org/10.1111/j.0009-9163.2004.00300.x
4. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. Obes Rev. 2017;18:603–34. https://doi.org/10.1111/obr.12531
5. Silva M, de Leeuw N, Mann K, Schuring-Blom H, Morgan S, Giardino D, et al. European guidelines for constitutional cytogenomic analysis. Eur J Hum Genet. 2019;27:1–16. https://doi.org/10.1038/s41431-018-0244-x
6. Maltese PE, Iarossi G, Ziccardi L, Colombo L, Buzzonetti L, Crinò A, et al. A Next Generation Sequencing custom gene panel as first line diagnostic tool for atypical cases of syndromic obesity: Application in a case of Alström syndrome. Eur J Med Genet. 2018;61:79–83. https://doi.org/10.1016/j.ejmg.2017.10.016
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A jointconsensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24. https://doi.org/10.1038/gim.2015.30
8. Willemsen MHH, Vulto-van Silfhout ATT, Nillesen WMM, Wissink-Lindhout WMM, van Bokhoven H, Philip N, et al. Update on Kleefstra syndrome. Mol Syndromol. 2011;2:202–12. https://doi.org/10.1159/000335648
9. Carvalho LML, D'Angelo CS, Mustacchi Z, da Silva IT, Krepischi AC V., Koiffmann CP, et al. A novel MYT1L mutation in a boy with syndromic obesity: Case report and literature review. Obes Res Clin Pract. 2021; https://doi.org/10.1016/j.orcp.2021.01.001
10. Vuillaume ML, Naudion S, Banneau G, Diene G, Cartault A, Cailley D, et al. New candidate loci identified by array-CGH in a cohort of 100children presenting with syndromic obesity. Am J Med Genet Part A. 2014;164:1965–75. https://doi.org/10.1002/ajmg.a.36587
11. Emerson E, Robertson J, Baines S, Hatton C. Obesity in British children with and without intellectual disability: cohort study. BMC Public Health. 2016;16:644 https://doi.org/10.1186/s12889-016-3309-1
12. Hsieh K, Rimmer JH, Heller T. Obesity and associated factors in adults with intellectual disability. J Intellect Disabil Res. 2014;58:851–63. https://doi.org/10.1111/jir.12100
13. Martins LB, Monteze NM, Calarge C, Ferreira AVM, Teixeira AL. Pathways linking obesity to neuropsychiatric disorders. Nutrition. 2019;66:16–21. https://doi.org/10.1016/j.nut.2019.03.017
14. Kleinendorst L, Abawi O, van der Voorn B, Jongejan MHTM, Brandsma AE, Visser JA, et al. Identifying underlying medical causes of pediatric obesity: Results of a systematic diagnostic approach in a pediatric obesity center. Buchner DA, editor. PLoS One. 2020;15:e0232990 https://doi.org/10.1371/journal.pone.0232990
15. Beunders G, van de Kamp J, Vasudevan P, Morton J, Smets K, Kleefstra T, et al. A detailed clinical analysis of 13 patients with AUTS2 syndrome further delineates the phenotypic spectrum and underscores the behavioural phenotype. J Med Genet. 2016;53:523–32. https://doi.org/10.1136/jmedgenet-2015-103601
16. Louw JJ, Corveleyn A, Jia Y, Hens G, Gewillig M, Devriendt K. MEIS2 involvement in cardiac development, cleft palate, and intellectual disability. Am J Med Genet Part A. 2015;167:1142–6. https://doi.org/10.1002/ajmg.a.36989
17. Kleinendorst L, Massink M, Cooiman MI, Savas M, van der Baan-Slootweg OH, Roelants RJ, et al. Genetic obesity: next-generationsequencing results of 1230 patients with obesity. J Med Genet. 2018;55:578–86. https://doi.org/10.1136/jmedgenet-2018-105315
18. Xia F, Bainbridge MN, Tan TY, Wangler MF, Scheuerle AE, Zackai EH, et al. De Novo truncating mutations in AHDC1 in individuals with syndromic expressive language delay, hypotonia, and sleep apnea. Am J Hum Genet. 2014;94:784–9. https://doi.org/10.1016/j.ajhg.2014.04.006
19. Carvalho LML, Costa SS, Campagnari F, Kaufman A, Bertola DR, Silva IT, et al. Two novel pathogenic variants in MED13L: one familial andone isolated case. J Intellect Disabil Res. 2021;65:1049–1057. https://doi.org/10.1111/jir.12891
20. Trivisano M, Santarone ME, Micalizzi A, Ferretti A, Dentici ML, Novelli A, et al. GRIA3 missense mutation is cause of an x-linkeddevelopmental and epileptic encephalopathy. Seizure. 2020;82:1–6. https://doi.org/10.1016/j.seizure.2020.08.032
21. Kummeling J, Stremmelaar DE, Raun N, Reijnders MRF, Willemsen MH, Ruiterkamp-Versteeg M, et al. Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome. Mol Psychiatry. 2020; https://doi.org/10.1038/s41380-020-0725-5
22. Roessler E, Belloni E, Gaudenz K, Vargas F, Scherer S, Tsui L, et al. Mutations in the C-Terminal Domain of Sonic Hedgehog CauseHoloprosencephaly. Hum Mol Genet. 1997;6:1847–53. https://doi.org/10.1093/hmg/6.11.1847

ACKNOWLEDGEMENTS

We are grateful to the families of patients involved in this study. We also thank the Xia-Gibbs Society (parent organization) and Dr. Richard Gibbs (Baylor College of Medicine - USA) for relevant communication. Additionally, it is important to thank Claudia Castro, Luceleni da Silva, Maraisa de Castro Sebastião, Mariana Barros, Dr. Naila Lourenço and Dr. Mustafa Zogbi for technical support. This study was supported by the São Paulo Research Foundation—FAPESP (2018/08486-3, 2012/50981-5, and 2013/08028-1), National Council for Scientific and Technological Development—CNPq (305806/2019-0) and Coordination for the Improvement of Higher Education Personnel—CAPES (1805008).

AUTHOR CONTRIBUTIONS

CR and CPK conceived and designed the study. CPK, LMLC, and CSD performed the research subject selection. LMLC performed the NGS data analysis and interpretation of genetic variants under the guidance of CR, AK, and DV. SSC and LDC provided relevant technical support. AALJ made important contributions to the discussion of the results. ITS and MOS contributed to bioinformatics processing (alignment to the reference genome and calling of variants). LMLC was responsible for the initial writing of the manuscript and produced the figures and tables under the guidance of CR and AK. All authors have read, edited, and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

This research is in accordance with ethical standards established in the Declaration of Helsinki (1964), its subsequent revisions, and Resolution 466/2012 of the Brazilian National Health Council. The project was approved by the Research Ethics Committee of the Biosciences Institute of the University of São Paulo (CEP protocol — IBUSP 021/2004). Blood samples of the patients and their parents were collected after informed consent was signed.

ADDITIONAL INFORMATION

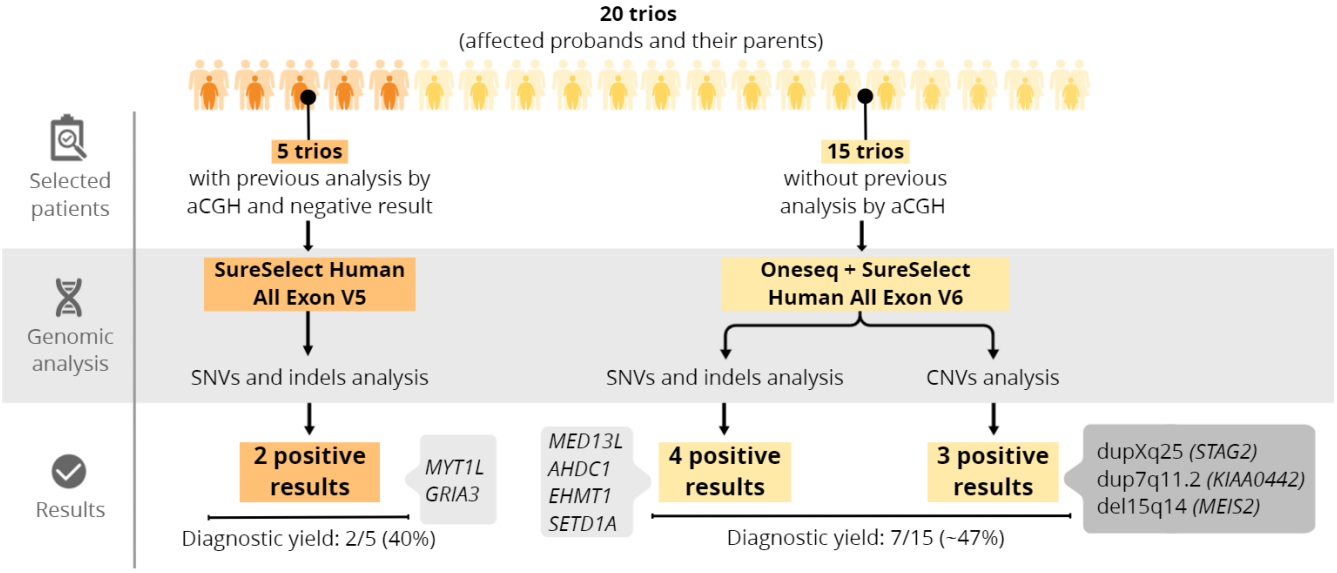
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41366-022-01149-5.

Correspondence and requests for materials should be addressed to Carla Rosenberg.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figure 1: Genetic investigation workflow of SO etiology in our cohort and positive results distribution.



We selected 20 patients who met the inclusion criteria. NGS sequencing of these patients and their parents (trios) was performed. Two types of capture were used. Whereas 5 of the patients already had CNV analysis by aCGH with a negative result, the sequencing of the trios was performed using SureSelect Human All Exon V5 (Agilent Technologies—California, USA) capture, and for them, only SNVs and indels were analyzed, which led us to 2 positive results (2/5 = 40%). For the others (15 trios), sequencing Oneseq + V6 SureSelect Human All Exon (Agilent Technologies—California, USA) was performed to analyze CNVs, SNVs and indels. A total of 4/15 were positive for SNVs/indels, and 3/15 were positive for CNVs, totaling 7/15 (47%). The genes involved in the positive findings are those shown in italic calligraphy.

Nota da autora: reapresentamos aqui a Figura 1 do Capítulo 2.1 em cores e maior tamanho. Também foi acrescentado o gene *SETD1A* ao esquema, que por um equívoco faltava na versão que foi publicada. No texto do artigo, no entanto, esse gene estava apresentado corretamente.

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

case	age	sex	anthropometric data	cognition and communication	eating behavior	social behavior	other clinical features	family history
1	12.4y	M	BMI z core 3.63 (obesity); OFC 56cm at birth: 2.28kg, 49cm; OFC 34cm	ID, speech impairment	polyphagia; picks up hidden food; he vomits after eating	anxious; nervous; gnaws nails; bite himself; cries a lot without apparent motivation; autism diagnosis	motor-coordination difficulties; narrow forehead; low implantation of hair on the forehead; big ears; retrognathism; short neck; bilateral clinodactyly (4th finger); short thumb; hypospadias	non-consanguineous parents; his father has obesity; his maternal half-brother has speech delay, ID, motor DD, hypospadias, and autism diagnosis; there are other cases of ID in the mother's family
2	16.5y	M	BMI z score 3.9 (obesity) at birth: 2.66kg, 45.5cm; OFC 34cm obs.: started to gain a lot of weight at 18 months of age	speech impairment	polyphagia; binge eating since 18 months of age; frequently vomiting	skin lesions suggestive of self-harm and skin picking; stubbornness; attacks of aggressiveness	few fetal movements; neonatal hypotonia; suction deficit with need for tube feeding; his mother had gestational diabetes and the insulin was not properly administered; was born with hypoglycemia and has diabetes; perspires a lot; arthrosis in both knees; vision problems; motor DD; already had a seizure without fever; kyphosis; hypogenitalism; testicles were not found by ultrasound; large arms; hepatomegaly and grade II liver steatosis; renal glomerular function preserved on the left and depressed on the right; bone age incompatible (advanced); macrostomy; macrognathism; carp mouth; frontal bossing	non-consanguineous parents; no similar cases in the family; the patient has a 7-year-old brother with partial paralysis (neonatal anoxia); three abortions between the siblings; has a healthy paternal half-brother (12-year-old)
3	13y	M	BMI z score 2.61;	moderate ID; speech impairment	at 2 years old he had reflux and lost weight very fast; he does not currently overeat	is organized, thorough, had a habit of washing his hands repeatedly; was aggressive for no apparent reason	bilateral cleft palate and absent uvula; karyotype 46,XY,inv(9) - normal variant; turricephaly; neonatal hypotonia with difficulty	non-consanguineous parents; an abortion after 4,5 months of

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

			OFC 59.8cm (+3.77 SD – macrocephaly) at birth: 3.8kg, 51cm			before he started taking psychiatric drugs; does not make friends at school	sucking; do not sleep without medication; decreased fetal movements; use of forceps; did not cry at birth; cyanosis at birth; motor DD (walked with more than 4 years of age); has urinary incontinence; recurrent infections (pneumonia, sinusitis); heart murmurs; big ears; auricular appendages (had surgery to correct); myopia; nasal voice; everted upper lip; inverted and spread nipples; bilateral clinodactyly (5th finger); *café au lait* spots; has difficulty walking	pregnancy; a healthy 18-year-old sister
4	22.6y	F	BMI 40.16 (obesity - grade IIII) at birth: 3.26kg, 48cm; OFC 34cm	mild ID	polyphagia after 17 years old; vomits	low social skills; after 17 years old, she presented obsessive-compulsive aggressive behaviour, anxiety, trichotillomania, skin picking, and self-harm; there were also reports of hallucinations	congenital clubfoot; motor DD; motor coordination problems; macrostomy; macroglossia; short neck; hypotonia since birth; excessive salivation; sleeps a lot	non-consanguineous parents; she has a healthy 19-year-old sister; there were no abortions
5	4y	M	BMI z score 5.70 (obesity) at birth: 3.35kg obs.: at 1 year-6 months his OFC was 54cm (+4.67 SD - macrocephaly)	global DD; do not speak	childhood feeding problems (used tube feeding) and little weight gain; presently presents polyphagia; episodes of choking mainly with liquids	autism diagnosis; repetitive attitudes; agitation; self-mutilation; is afraid of people; he has attacks of aggression but at other times he is affectionate	reduced fetal movements; neonatal hypotonia; suction difficulty; hypogenitalism; motor DD; body thermal instability in childhood; snores; small mouth; low nasal bridge; protruding ears; almond-shaped eyes; club-foot; exacerbated patellar reflex; *corpus callosum* dysgenesis; presence of mega cisterna magna; accentuation of the grooves between the cortical gyres, volume reduction of the periventricular and subcortical white matter with irregular ex-	non-consanguineous parents; has a brother and two half-siblings, all healthy

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

							vacum dilation of the lateral ventricles and the 3rd ventricle	
6	15.8y	M	BMI z score 3.32 (obesity) at birth: 3.9kg, 49cm obs.: obesity concentrated mainly on thighs and abdomen; rapid and excessive weight gain between 3 and 6 years of age	ID; speech impairment	polyphagia; gets angry when food is denied	aggressiveness; pokes nails of hands and feet; skin picking; stereotyped movements; autism diagnosis	decreased fetal movements; transverse palmar creases; he did not suck and had to be fed with a syringe; hypotonia up to 8 months; motor DD; has tachycardia even at rest; turricephaly; narrow forehead; posteorized and malformed ears; synophrys; narrow eyelid slits; microstomy; thick lips; carp-shaped mouth; constantly keeps his mouth open; strabismus of the right eye; wide neck; macroglossia; decreased pain sensitivity; hearing loss; volumetric reduction of the brain; telecanthus; high myopia; absence seizures; sleeps little and sometimes does not sleep	non-consanguineous parents; two apparently normal paternal half-sisters; report of a case of deaf individual in the family
7	10.3y	M	BMI z score: 3.49 (obesity) OFC: 58.5 cm (+3.92 SD – macrocephaly) at birth: 3.32kg, 47cm. obs.: abdominal fat	ID	polyphagia	anxious; aggressive behavior when food is denied	did not cry at birth; gynecomastia; divergent strabismus; frontal bossing; telecanthus; hypogenitalism; cryptorchidism; ↑LDL; had seizures at age three with fever; asthma	non-consanguineous parents; four healthy siblings; maternal grandfather had heart and mental problems
8	6.9y	M	BMI z score 4.62; OFC 57.5cm (+4.03 SD – macrocephaly)	global DD; speech impairment	polyphagia; looks for food at night; vomits frequently	skin picking and self-mutilation; he has tried to mutilate his penis a few times; stubborn; he has fits of violence over food; rapid	reduction of fetal movements in relation to his sister; didn't suck; motor DD; big ears; synophrys; enophthalmia (both eyes), straight eyebrows; triangular mouth, short	non-consanguineous parents; no history of abortions; the patient has a healthy 13-year-old sister

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

			at birth: 3.46kg, 47cm; OFC 39cm (+1.92 SD) obs.: fast weight gain before one year of age; at 8 months of age already had obesity			mood swings; hyperactivity; usually puts objects in the ear and mouth; often wheezing	neck; hypogenitalism; transitional palmar folds; healing difficulty; hepatomegaly; moderate hepatic steatosis; ↑gamma-GT; ↑total cholesterol, ↑HDL; ↑LDL ↑triglycerides; ↑insulin; incompatible bone age (+4.01 SD); apnea; only sleeps with medication	
9	15.5y	M	BMI z score 2.51 (obesity); OFC 55.5cm at birth: 3.08kg, 46,8cm; OFC 33cm	mild ID; speech impairment	-	has friends at school and is sociable; anxious; psychomotor agitation; repetitive movements and speeches (repeats what people say); suspected autistic spectrum; until the age of 7 he walked with his arms raised; he is not aggressive	motor DD; mild color blindness (Ishihara test); hypogenitalism; unilateral cryptorchidism; gynecomastia; normal suction at birth; presented hypotonia; brachycephaly; microstomy; macrognathism; thick lips; short neck; bilateral clinodactyly of the 5th fingers; hypothyroidism; *café au lait* spot in the cervical region; white hair early	non-consanguineous parents; report of spontaneous abortion; the father has ID and obesity (he weighed 220kg, today weighs 110kg); the father's aunt is deaf and has ID; the patient has an apparently healthy 25-year-old sister despite her short stature (1.45m)
10	6.9y	M	BMI z score 4.82; OFC 59cm (+5.1 SD – macrocephaly) at birth: 4.55kg, 56cm. obs.: gained weight from the first month	talk a lot (all the time); cannot read; finds it difficult to pronounce some words	polyphagia; looks for food at night	agitated; aggressiveness; hit people; scratches himself	does not control urination; hypogenitalism; brachycephaly; protruding ears (7 cm); strabismus of the left eye; sunken eyes; micrognathism; short neck; arms measuring 16cm; sydney palmar crease; the mother had urinary infections in the 2nd and 3rd trimester and used antibiotics	non-consanguineous parents; the patient's mother has obesity and finds it difficult to pronounce some words; the mother's brother is infertile; the patient has a healthy 3-year-old brother; there are no similar cases in the family

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

11	6.8y	M	BMI z-score 4.48; OFC 53.5cm at birth: 2.18kg, 44cm; OFC 30cm (-2.58 SD - microcephaly) obs: protruding abdomen and thick thighs	global DD; speech impairment	polyphagia	anxiety; aggressiveness; agitation; stubbornness; lies; beat the sisters; obsessive-compulsive behavior; skin picking	began to suck two days after birth; hypotonia; cryptorchidism (right); motor DD; hypogenitalism; narrow forehead; low implantation of hair on the forehead; small ears; transitional palmar crease; decreased pain sensitivity; slept a lot, got to sleep 26 hours straight	non-consanguineous parents; patient has two sisters and one of them presented learning disability; parents don't know if there are similar cases in the family
12	4.7y	M	BMI z score 6.27 (obesity) at birth: normal weight obs.: started to gain weight at the age of two; abdominal and thigh fat	speech impairment	polyphagia; cries and screams if food is denied; looks for food at night	does not accept routine changes; anxious; does not let it be examined; skin-picking; obsessive-compulsive behavior; already showed aggressiveness	suction difficulty; cyanosis at birth; uses diapers outside the home; already had reflux; hypopigmentation in relation to parents; short neck, brachycephaly, telecanthus, decreased pain sensitivity; motor DD (firm head at 1 year-2 months old).	parents are first cousins; autistic brother; paternal grandfather has epilepsy since birth; mother's cousin has microcephaly.
13	13.3y	F	BMI z score 2.49 at birth: 2.5kg, 53cm obs.: started to gain weight at the age of seven	speech impairment	polyphagia; gets angry when food is denied	gnaws nails; skin picking; attacks family members	was slow to start crying at birth; started to suck after leaving the maternity hospital; cyanosis and jaundice at birth; presented hypotonia; needed to be woken up to suckle; decreased pain sensitivity; hyperextensible fingers; at the age of 6 she had developed breasts; menarche at 7 years; frontal bossing; small ears; straight eyebrows; thick lips; small hands; single palmar fold in both hands	4th cousin parents; there are no similar cases in the family; the patient has two sisters, (18 and 14 years old), both healthy

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

14	9.2y	M	BMI z score 4.51; OFC 57.5cm (+3.61 SD - macrocephaly) at birth: 2.75kg, 49cm. obs.: started to gain weight from the age of two	learning disability	polyphagia from the age of five; he vomits; as a baby, breastfeeding had to be supplemented with a bottle	anxious; rides a bicycle all day; the use of sibutramine to treat obesity made him more anxious and aggressive and had crying spells	the mother had a urinary tract infection during pregnancy and use of antibiotics; hypertension and fever spikes during pregnancy; reduction of fetal movements in relation to siblings; he cried at birth; reflux; sunken eyes; narrow eyelid cleft; small nose; microstomy; macrognathism; short neck; short fingers	non-consanguineous parents; the patient has two healthy paternal half-brothers and three healthy maternal half-siblings (two women and one man); paternal half-sister of the patient's mother has epilepsy
15	5.6y	F	BMI z score 5.18 at birth: 3.37kg, 48.5cm; OFC 34.5cm	speech impairment	-	didn't socialize when she went to school; is hyperactive; does not like to stay indoors; imitates people; she didn't follow the rules at school and just wanted to stay in the play area; she likes water a lot, goes swimming every day	46,XY,inv(9) - normal variant; has difficulty controlling sphincters; big ears; straight eyebrows; synophrys; small nose; macrognathism; tent pitch; *café au lait spots*; sleep problems, she changed the day for the night for 6 months	non-consanguineous parents; there is no report of abortions; has no siblings
16	1.4y	M	BMI z score 8.38; OFC 50.5cm (+2.4 SD – macrocephaly). at birth: 3.42kg, 51cm	-	-	-	when he was born, his temperature dropped and he had tremors; congenital atrial septal defect, but closed spontaneously; hypogenitalism; high and wide forehead; high implantation of hair on the forehead; big ears; small saddle nose with reversed nostrils; microstomy; macrognathism; hallux facing upwards; trident-looking hand; nystagmus; to fix the vision, he turns his head to the side; visual acuity within the lower normal limit; retinal mapping found that both eyes have pale optic nerves, faint outlines, blemishes with	non-consanguineous parents; report of spontaneous abortion; the patient has one brother and one sister (both healthy)

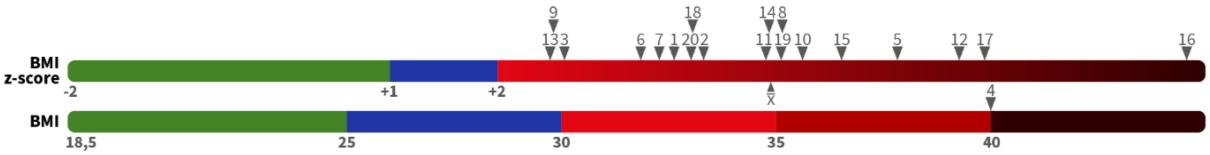
ST1 – Clinical features of the group of 20 patients with syndromic obesity.

							preserved brightness, vessels of normal conformation; US revealed a choroid plexus cyst on the right	
17	3.2y	F	BMI z score 6.51 at birth: 3.20kg, 51cm; OFC 33.5cm obs.: at 9 months she was 18kg; has typical PWS fat distribution	-	polyphagia since 2 years of age; gets angry when food is denied	lies saying that her parents don't feed her; rapid mood swings; difficulty with routine changes; skin picking	didn't suck when she was born; has hypertension and diabetes; hepatic steatosis; heart murmur; apnea; constipation; she has had pneumonia several times and a severe kidney infection; frontal bossing; sunken eyes; saddle nose; microstomy; macrognathism; triangular mouth, short neck; clinodactyly; hypopigmentation (skin and hair are light in relation to the family); viscous saliva	non-consanguineous parents; the patient has no siblings and there is no history of abortions
18	10.4y	M	BMI z score 3.80; OFC 53cm at birth: 2.7kg, 47cm. obs.: typical PWS fat distribution	global DD; speech impairment; doesn't read, doesn't write and doesn't do math	polyphagia	he has no bouts of violence; is stubborn; skin picking	motor DD (sat unsupported over a year old and walked almost two years old); had seizures with fever at 2 years of age; frontal bossing; low implantation of hair on the forehead; big ears; reversed nostrils; macrostomy; macrognathism; thin upper lip; tent-shaped mouth; short neck; hypogenitalism; unilateral cryptoquidism; retracted penis; clinodactyly (4th and 5th fingers); apnea; breath through mouth only	non-consanguineous parents; the patient has no siblings; a niece’s daughter of the patient's father was born blind and has ID
19	10.1y	M	BMI z score 4.63; OFC 54cm at birth: 2.47kg, 47cm	-	polyphagia; looks for food at night; vomits	behaves calmly; bites the lower lip compulsively	hydronephrosis discovered during pregnancy; one-year-old nephrectomy; malformed left ear; reversed nostrils; thick lips; short neck; gynecomastia; hypomelanosis of Ito; transverse palmar fold; strabismus;	non-consanguineous parents; there are no similar cases in the family; the patient has a healthy five-year-old sister

ST1 – Clinical features of the group of 20 patients with syndromic obesity.

			obs.: started to gain weight at the age of 5; has a protruding abdomen				hyperopia; ↓FSH, ↓LH, ↓testosterone		
20	9.5y	M	BMI z score 3.79 (obesity); OFC 57cm at birth: almost 5kg	-	polyphagia; food anxiety	quiet; cries when stressed; not aggressive	turricephaly; diabetes; hypothyroidism; asthma; normal suction at birth; did not cry at birth; wear a diaper at night; posteorized ears; narrow eyes; thick lips; short neck; transitional palmar crease; vitiligo	non-consanguineous parents; has a 19-year-old sister and a 15-year-old brother, both healthy	

SF1 - Obesity score measurements



Black arrows indicate the obesity level of each patient (1 to 20). The blue color represents overweight scores, while the green color represents a normal range. The red to black color range indicates increasing levels of obesity. (A) BMI z scores for children. Obesity is diagnosed when the BMI z score > 2 SD. x̄ = mean. (B) BMI of the only adult patient. Obesity is diagnosed when BMI > 30, with values ≥ 40 corresponding to grade III obesity. The mean BMI z score was +4.45 (±1.50 SD),

Capítulo 2.2

A novel MYT1L mutation in a boy with syndromic obesity: Case report and literature review

CARVALHO, L. M. L; D'ANGELO, C. S; MUSTACCHI, Z.; SILVA, I. T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. A novel *MYT1L* mutation in a boy with syndromic obesity: Case report and literature review. **Obesity Research & Clinical Practice**, 2021. ISSN: 1871-403X. DOI: 10.1016/j.orcp.2021.01.001

Obesity Research & Clinical Practice 15 (2021) 124–132

Contents lists available at ScienceDirect

Obesity Research & Clinical Practice

journal homepage: www.elsevier.com/locate/orcp

Original Article

A novel *MYT1L* mutation in a boy with syndromic obesity: Case report and literature review

Laura M.L. Carvalho[a], Carla S. D'Angelo[a], Zan Mustacchi[b,c], Israel T. da Silva[d], Ana Cristina V. Krepischi[a], Celia P. Koiffmann[a], Carla Rosenberg[a,*]

[a] *Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo (USP), Sao Paulo, SP, Brazil*
[b] *Genetic Outpatient Clinic of the Darcy Vargas Children's Hospital (HIDV), São Paulo, SP, Brazil*
[c] *São Paulo Center for Clinical Studies and Research – CEPEC-SP, São Paulo, SP, Brazil*
[d] *International Center for Research, A. C. Camargo Cancer Center, Sao Paulo, Brazil*

ARTICLE INFO

Article history:
Received 24 October 2020
Received in revised form
26 December 2020
Accepted 1 January 2021

Keywords:
Syndromic obesity
MYT1L
Whole exome sequencing
Polyphagia
Intellectual disability
Neurodevelopment

ABSTRACT

Background: Pathogenic variants involving the *MYT1L* gene lead to an autosomal dominant form of syndromic obesity, characterized by polyphagia, intellectual disability/developmental delay, and behavioral problems, and that a characteristic facial phenotype does not seem to be recognizable.
Methods: Trio whole exome sequencing was performed in a 10-year-old Brazilian male presenting polyphagia, severe early-onset obesity, intellectual disability, speech delay, macrocephaly, frontal bossing, telecanthus, strabismus, and hypogenitalism. Additionally, we performed a literature review of patients carrying non-copy number *MYT1L* variants.
Results: A *de novo* genetic variant not previously reported in *MYT1L* (NM_015025.4:c.2990C > A) was identified in the proband and classified as pathogenic. From a literature search, 22 further patients carrying non-copy number *MYT1L* variants were identified, evidencing that although the associated phenotype is quite variable, intellectual disability/developmental and speech delays are always present. Further, most patients have obesity or overweight due to polyphagia. Macrocephaly, strabismus, behavioral problems, and hand/feet malformations are also recurrent features.
Conclusions: We described the first Brazilian case of *MYT1L* related syndrome and highlighted clinical characteristics based on the literature. Other syndromic forms of obesity such as Prader-Willi, Bardet-Biedl, Börjeson-Forssman-Lehmann, MORM, Cohen, Alstrom, and Kleefstra type 1 syndromes should be considered in the differential diagnosis. Further, although obesity is frequent, it is not an obligatory feature of all carriers of *MYT1L* mutations.
© 2021 Published by Elsevier Ltd on behalf of Asia Oceania Association for the Study of Obesity.

Abbreviations: ABESO, Brazilian Association for the Study of Obesity and Metabolic Syndrome; ACMG, American College of Medical Genetics and Genomics; BAM, Binary Alignment Map; BMI, Body mass index; CGH, Comparative genomic hybridization; CNV, Copy Number Variation; DD, Developmental Delay; EEG, Electroencephalogram; ID, Intellectual Disability; LDL, Low Density Lipoproteins; MLPA, Multiplex Ligation-dependent Probe Amplification; MORM, Mental retardation, truncal obesity, retinal dystrophy and micropenis syndrome; MRI, Magnetic Resonance Imaging; MYT1L, Myelin Transcription Factor 1-Like; myt1la, *MYT1L* zebrafish orthologue A; myt1lb, *MYT1L* zebrafish orthologue; NR, Not Reported; OFC, Occipitofrontal Circumference; pLI, probability of being loss-of-function intolerant; SD, Standard Deviation; SNV, Single Nucleotide Variation; VCF, Variant Call Format; WES, Whole Exome Sequencing; y, year.

* Corresponding author at: University of Sao Paulo, Institute of Biosciences, Department of Genetics and Evolutionary Biology, Human Genetics Research Laboratory – LIGH, Matão Street, lane 14, number 277 – room 350, São Paulo, SP 05508-090, Brazil.
E-mail addresses: lauralara@usp.br (L.M.L. Carvalho), carlasustek@gmail.com (C.S. D'Angelo), drzan@drzan.com.br (Z. Mustacchi), itojal@accamargo.org.br (I.T. da Silva), ana.krepischi@ib.usp.br (A.C.V. Krepischi), cpkoiffm@ib.usp.br (C.P. Koiffmann), carlarosenberg@ib.usp.br (C. Rosenberg).

https://doi.org/10.1016/j.orcp.2021.01.001
1871-403X/© 2021 Published by Elsevier Ltd on behalf of Asia Oceania Association for the Study of Obesity.

Introduction

Syndromic obesity is an umbrella term used to describe cases of obesity occurring with additional phenotypes [1,2], including intellectual disability, dysmorphic features, or organ-specific developmental abnormalities [2]. These conditions are rare [3,4] and highly variable [5,6]. In a significant number of patients, syndromic obesity follows a Mendelian pattern of inheritance and may occur by the altered expression of genes with pleiotropic effects. Although these genes are usually autosomal or X-linked [4] in most cases, their etiology cannot be defined [4,7].

Prader-Willi and Bardet-Biedl are the best known genetic syndromes in which obesity is a central feature, but there are many other syndromic forms of obesity described in the literature, such as type 1, Kleefstra, Börjeson-Forssman-Lehmann, MORM (mental retardation, truncal obesity, retinal dystrophy and micropenis), Alstrom, and Cohen syndromes, amongst others [4].

Copy number variants (CNV), indels and single-nucleotide variants (SNVs) involving *MYT1L* gene (Myelin Transcription Factor 1-Like – OMIM*613084) can cause an autosomal dominant syndromic form of obesity, characterized by polyphagia, intellectual disability/developmental delay, and behavioral problems. However, a characteristic facial phenotype does not seem to be recognizable [8–10] (OMIM#616521).

Using parents-proband trio and whole exome sequencing (WES), a novel genetic variant (NM_015025.4:c.2990C > A) disrupting the *MYT1L* gene was detected in a patient presenting severe early-onset syndromic obesity; this nonsense variant was not detected in his parents.

Additionally, we reviewed 22 further cases of indel/SNV *MYT1L* variants to improve the clinical spectrum description of the syndrome.

Patient and methods

Case report

Clinical features

The proband was part of a cohort originally referred for testing for Prader-Willi syndrome in which all patients presented obesity and intellectual disability. The Patient (Fig. 1A and B) is the child of a non-consanguineous healthy couple and has four healthy half-siblings (Fig. 1C). There are no similar cases in the family. Birth occurred at 39 weeks of pregnancy and he was born with 3.315 kg weight (35 percentile) and 47 cm height (12 percentile). He didn't cry at birth and, although he was able to suck, he had to stay in hospital for the first 10 days because of respiratory distress and jaundice. Early developmental milestones were not documented in medical records, but were eventually obtained by familial reporting: he was described as having normal developmental milestones, supporting his head at 3 months, sitting without support at 6 months, and walking at 1-year-1-month. He had seizures at the age of three during a fever.

On physical examination at 10 years and 4 months of age, his weight was 71.5 kg and height 146 cm [BMI (Body Mass Index) *z*-score: 3.49 – obesity]. He exhibited telecanthus (internal intercanthal distance 3.8 cm; external intercanthal distance 11 cm), divergent strabismus, macrostomy, macrognathism, frontal bossing, macrocephaly [58.5 cm of occipitofrontal circumference – OFC (>99 percentile; +3.92 SD)], gynecomastia (possibly due to obesity), and hypogenitalism without cryptorchidism.

He is not able to read or grasp any mathematical principles. Although intellectual development is clearly delayed, he has not been formally evaluated. He has speech delay, is very anxious, wants to eat all the time (polyphagia), and can be aggressive when food is denied.

Doppler echocardiography, electroencephalogram (EEG), brain Magnetic Resonance Imaging (MRI) and Computed Tomography results were all normal. LDL values were high (131 mg/dL, values less than 110 mg/dL are normal in healthy children and adolescents [11]).

Genetic investigation

The patient was seen at the syndromic obesity outpatient service of our University, where samples of peripheral blood from the patient and his parents were collected for genetic investigation.

Because the patient has been referred for Prader-Willi investigation, initially the following genetic tests were performed, all with normal results: MLPA (salsa P064 for the main microdeletion syndromes, P036 and P070 to examine subtelomeric regions – MRC-Holland, Amsterdam, Netherlands), and methylation test of the Prader-Willi region (exon 1 of SNURF-SNRPN). After obtaining negative results in these tests, we performed an array-CGH (60K, Agilent Technologies, Santa Clara, California, USA), allowing identification of CNVs throughout the genome. However, we were unable to identify CNVs that could explain the patient's clinical status.

Whole-exome sequencing (WES) from the proband and his unaffected parents (SureSelectXT Human All Exon V5 – Agilent Technologies, Santa Clara, California, USA) was performed for the detection of SNVs and indels. The raw paired-end reads were aligned to the GRCh37/hg19 reference genome using the Burrows-Wheeler Aligner – BWA [12], and pre-processed according to GATK toolkit [13].

The trio-WES was analyzed using the VarSeq software (Golden Helix, Bozeman, MT, USA). VCF data was used for annotation and filtering of genetic variants. Visual verification of the findings was made using data from the BAM files. The following annotation and filtering strategies were used: read depths ≥10, genotype qualities ≥17 (based on the Phred score), population frequency (1KG Phase3, gnomAD, NHLBI and ABraOM), missense and loss-of-function variants were selected. We examined the genes from syndromic obesity (Severe early-onset obesity 2.3, and Maltese et al., 2018 [14]), intellectual disability (SYSID) and developmental delay (DD – gene2phenotype) panels. We tested all modes of inheritance: starting by a hypothesis of dominant (investigation of variants *de novo*), followed by recessive autosomal or X-linked (homozygosity, compound heterozygosis, and hemizygosity), and imprinting (using geneimprint database). The approach used does not allow a secure evaluation of mitochondrial variants.

Compilation of literature data to help establishing the phenotypic spectrum

It is known that CNVs at 2p25.3 can lead to a clinical condition characterized by intellectual disability with early-onset of obesity [1,16,17]. The most frequent type of variation described for *MYT1L* are microdeletions [8,9]. In 2015, De Rocker et al. (2015) [8] studied 20 patients with microdeletions or microduplications in 2p25.3. The authors concluded that since *MYT1L* was the only gene that was commonly disrupted or lost in these cases, this indicated that this gene was the probable cause of the observed phenotypes, with haploinsufficiency as the most likely mechanism.

Additionally, nonsense, frameshift, missense and splice-site variants were described in the literature [8–10,18–21]. We reclassified these variants according to the ACMG criteria [15].

In Table 1 we present the BMI *z*-scores for individuals under 18 years of age. For children, the diagnosis of obesity occurs when the BMI *z*-score >+2 SD (standard deviation) for age and sex and is considered overweight when the BMI *z*-score is between +1 and +2 [22]. For adults, overweight is defined as a BMI between 25.0 and 29.99, and obesity as a BMI ≥30 [23]. For BMI and BMI *z*-score calculations, we used the ABESO

Table 1
Characteristics of 22 patients with SNV/indel *MYT1L* variants and pathogenicity classification according to ACMG criteria.

	Reference	Age	Sex	Genetic variant	Classification (ACMG criteria)	ID/DD	Speech Delay	Anthropometric features	Polyphagia	Behavior	Dysmorphic facial features	Other features
Nonsense variants	Patient of this publication	10.3y	M	NM_015025.4:c.2990C > A p.Ser997* - exon 21 of 25 - nonsense - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 3.49 (obesity) OFC: 58.5 cm (+3.92 SD – macrocephaly)	+	Aggressive behavior when food is denied, anxious	Divergent strabismus, frontal bossing, telecanthus.	Hypogenitalism, high LDL values.
	de Rocker et al., (2015)[8] – Patient 15	13.5y	M	NM_015025.4:c.1917T > G rs869320676 p.Tyr639* - exon 14 of 25 - nonsense - inheritance not determined	Pathogenic (PVS1, PM2)	+	+	BMI *z*-score: 1.52 (overweight) OFC: 54 cm (−0.23 SD)	NR	Autism	Brachycephaly, upward slant, narrow eyes	NR
	Blanchet et al. (2017)[10] - Patient 7	5y	F	NM_015025.4:c.2671C > T p.Arg891* - exon 18 of 25 - nonsense - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 3.05 (obesity) OFC: 51.5 cm (+0.84 SD)	+	NR	Strabismus	Hyperopia
	Windheuser et al. (2020)[9] – Individual 1	12.4y	M	NM_015025.2:c.1531G > T p.Gly511* - exon 11 of 25 - nonsense - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	Small for gestational age. Primary microcephaly. BMI *z*-score −0.52 (normal weight) and OFC 52.5 (−0.94 SD)	Absent(*)	Aggressive outbursts.	NR	Mild hypothyroidism at 2 years.
	Windheuser et al. (2020)[9] – Individual 4	6.3y	F	NM_015025.2:c.223C > T rs1258062046 p.Arg75* - exon 9 of 25 - nonsense - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 1.32 (overweight) OFC: 50 cm (−0.77 SD)	NR	Autism, aggressiveness and self-injurious behavior, sleep disturbance	Long Face, thick and horizontal eyebrows, small vermillion, pointed chin, simple ears	Bilateral inguinal hernia, joint hypermobility and instability, gait abnormality, muscular hypotonia, limb hypertonia, increased tendon reflexes
	Windheuser et al. (2020)[9] – Individual 6a	27y	M	NM_015025.4:c.3118A > T p.Lys1040* - exon 22 of 25 - nonsense - inherited from affected father.	Pathogenic (PVS1, PM2)	+	+	BMI: 24.34 (normal weight) OFC information not available	NR	ADHD, aggressive and self-injurious behavior.	Broad nasal bridge, wide nasal tip, short philtrum, large ears, hyperplastic ear lobes, wide mouth, irregularly spaced teeth	Hearing impairment, systolic heart murmurs, mild gastroesophageal reflux, relatively small hands, micro-orchidism, myopia.

Table 1 (*Continued*)

	Reference	Age	Sex	Genetic variant	Classification (ACMG criteria)	ID/DD	Speech Delay	Anthropometric features	Polyphagia	Behavior	Dysmorphic facial features	Other features
	Windheuser et al. (2020)[9] – Individual 6b (father of 6a)	56y	M	NM_015025.4:c.3118A > T p.Lys1040* - exon 22 of 25 - nonsense - inheritance not determined	Pathogenic (PVS1, PM2)	+	NR	BMI: 26.26 (overweight) OFC information not available	NR	Aggressive outbursts. Hyperactivity.	Mild strabismus, bulbous and rounded nasal tip, short and broad philtrum, wide mouth, widely spaced teeth, thick lower lip.	Epilepsy until puberty.
Frameshift variants	Loid et al. (2018)[18]	13.8y	M	NM_015025.2: c.2215_2224delACGCGCTGCC rs1253072668 p.Thr739Alafs*7 - exon 15 of 25 - frameshift deletion - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 2.61 (obesity) Normal OFC at birth	+	Aggressive behavior, ADHD	Strabismus	Problems with balance
	Blanchet et al. (2017)[10] - Patient 1	10y	F	NM_015025.4:c.1298delT p.Leu433Argfs*16 - exon 10 of 25 - frameshift deletion - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 3.29 (obesity) OFC: 55.6 cm (+2.56 SD – macrocephaly)	+	Autism	Bilateral strabismus	Echinovarus, camptodactyly gastroesophageal reflux, never walked independently
	Blanchet et al. (2017)[10] - Patient 6	15y	F	NM_015025.4: c.2220_2229 delCTGCCGCGAG p.Glu743Argfs*3 - exon 15 of 25 - frameshift deletion - *de novo*	Pathogenic (PVS1, PM2, PM6)	+	+	BMI *z*-score: 0.64 (normal weight) OFC: 52.5 cm (−1.52 SD)	Absent	NR	Not dysmorphic	
Splice variants sdddsvariants	de Ligt et al. (2012)[19] – Trio 92/de Rocker et al., (2015)[8] – Patient 14	12.7y	F	NM_015025.4: c.2636+1G > A rs869320675 - intron 17 of 24 - splice_donor_variant - *de novo*	Pathogenic (PVS1, PM2, PM6, PP3)	+	+	Central obesity	NR	Stereotypic movements, autism	Broad base and bridge of the nose, large mouth	Neonatal feeding difficulties
	Blanchet et al. (2017)[10] - Patient 4	28y	F	NM_015025.4: c.1703+1G > A - intron 12 of 24 position - splice donor variant - *de novo*	Pathogenic (PVS1, PM2, PM6, PP3)	+	+	Obesity OFC: 55 cm (+0.62 SD)	+	Autism, aggressive and self-injurious behavior, and ADHD	NR	Hydrocephalus, ventricular shunt, spina bifida, spastic paraparesis
	Windheuser et al. (2020)[9] – Individual 2(**)	8y	F	NM_015025.4: c.2769-2A > G - intron 19 of 24 - splice_acceptor_variant - *de novo*	Pathogenic (PVS1, PM2, PM6, PP3)	+	+	BMI *z*-score: 4.1 (obesity) OFC: 54 cm (+1.21 SD)	+	Aggressive outbursts sleep disturbances	Narrow palpebral fissures	Increased noise and pain sensitivity, rough voice, short neck, tapering fingers, sandal gap, short distal phalanges of the toes with small nails
Missense variants	Blanchet et al. (2017)[10] - Patient 2	9y	M	NM_015025.4: c.1700G > A(***) rs878853045 p.Arg567Gln - exon 12 of 25 - missense - *de novo*	Pathogenic (PM1, PM2, PM6, PP2, PP3, PP5)	+	NR	BMI *z*-score: 1.54 (overweight) OFC: 54.5 cm (+1.45 SD)	+	NR	NR	Posterior plagiocephaly, 5th finger clinodactyly, ataxic gait, cerebral atrophy

Table 1 (*Continued*)

Reference	Age	Sex	Genetic variant	Classification (ACMG criteria)	ID/DD	Speech Delay	Anthropometric features	Polyphagia	Behavior	Dysmorphic facial features	Other features
Windheuser et al. (2020)[9] – Individual 5	1.6y	M	NM_015025.4: c.1700G > C(***) p.Arg567Pro - exon 12 of 25 - missense - *de novo*	Likely Pathogenic (PM1, PM2, PM5, PM6, PP2, PP3)	+	+	BMI *z*-score: 1.85 (overweight) OFC: 50 cm (+1.55 SD)	NR	Friendly, poor eye fixation	Strabismus, high and broad forehead, flat midface, deep-set eyes, broad and flat nasal bridge, small mouth with downturned corners	Puffy hands and feet, tapering fingers, rather short fifth fingers
Blanchet et al. (2017)[10] - Patient 3	10y	M	NM_015025.4: c.1672C > T rs1558371790 p.His558Tyr - exon 12 of 25 - missense - *de novo*	Pathogenic (PM1, PM2, PM6, PP2, PP3)	+	+	BMI *z*-score: 0.86 (normal weight) OFC: 54 cm (+0.84 SD)	+	ADHD	NR	Dyspraxia, astigmatism
Blanchet et al. (2017)[10] - Patient 5	17y	F	NM_015025.4: c.1553-T > C rs1057519560 p.Leu518Pro - exon 11 of 25 - missense - *de novo*	Likely Pathogenic (PM2, PP2, PM6, PP3)	+	+	BMI *z*-score: 2.26 (obesity) OFC information not available	+	Autism	NR	Complex partial seizures from 11 years old
Blanchet et al. (2017)[10] - Patient 8	3y	F	NM_015025.4: c.1564C > A p.His522Asn - exon 11 of 25 - missense - *de novo*	Likely Pathogenic (PM2, PM6, PP2, PP3)	+	+	BMI *z*-score: 0.67 (normal weight) OFC: 51 cm (+1.59 SD)	Absent	Autism	NR	Absence seizures, thinning of the corpus callosum
Blanchet et al. (2017)[10] - Patient 9	13y	M	NM_015025.4: c.1579G > A p.Gly527Arg - exon 11 of 25 - missense - *de novo*	Pathogenic (PM2, PM6, PP2, PP3, PP5):	+	+	BMI *z*-score: 0.38 (normal weight) OFC: 53 cm (−0.80 SD)	NR	ADHD, sleep disturbances	NR	
al Tuwaijri and Alfadhel (2019)[20]	10y	F	NM_001303052.2: c.1585G > A rs1275489527 p.Gly529Arg - exon 11 of 25 - missense - *de novo*	Pathogenic (PM2, PM6, PP2, PP3, PP5)	+	+	BMI *z*-score: 3.26 (obesity) OFC: 57 cm (+3.61 SD – macrocephaly)	+	Autism, self-injury behavior, ADHD, sleep disturbances	NR	Inverted nipples, bilateral amblyopia with astigmatism.
Windheuser et al. (2020)[9] – Individual 3	3.4y	F	NM_015025.4: c.1711G > A rs1330054460 p.Gly571Arg - exon 13 of 25 - missense - *de novo*	Likely pathogenic (PM1, PM2, PM6, PP2, PP3)	+	+	BMI *z*-score: 3.10 (obesity) OFC: 52.6 cm (+2.39 SD – macrocephaly)	NR	Agitated	Low-set ears, large palpebral fissures, long eyelashes	Febrile and nonfebrile seizures
Loid et al. (2020)[21] – Patient 5	12y	F	NM_015025.4: c.2414G > A rs200841614 p.Arg805Gln - exon 16 of 25 - missense - inheritance not determined	Benign (PP2, BS1, BS2, BP4)	NR	NR	BMI *z*-score: 3.8 (obesity)	NR	Aggressive behavior	NR	

ID – Intellectual Disability; DD – Developmental Delay; NR – Not Reported; ADHD – Attention Deficit Hyperactivity Disorder; SD – Standard Deviation; OFC – Occipitofrontal Circumference; LDL – Low Density Lipoproteins; y – year; * – stop codon; fs – frameshift.
ACMG (American College of Medical Genetics and Genomics) guidelines can be checked in Richards et al. [15].
(*) Engels (author) personal communication.
(**) Patient has also a variant in *SETD1B*.
(***) Variants in the same position.

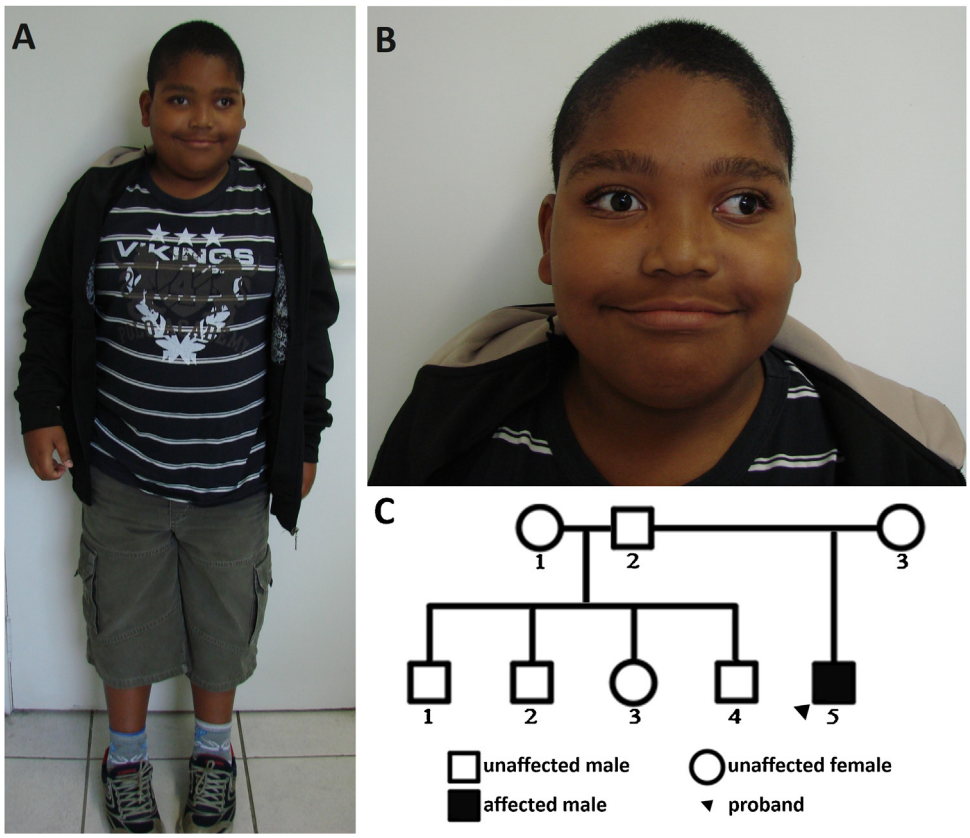


Fig. 1. Patient's features and pedigree.
(A) Full body image of the patient; (B) Facial characteristics of the patient (strabismus, telecanthus and frontal bossing); (C) Pedigree of the family showing the proband as an isolated case.

(Brazilian Association for the Study of Obesity and Metabolic Syndrome) online program (abeso.org.br/obesidade-e-sindrome-metabolica/calculadora-imc/).

We used the SimulConsult tool (https://simulconsult.com/resources/measurement.html) to classify each patient based on their OFC measurements presented in Table 1; macrocephaly was considered established if the OFC was greater than two SDs above the mean, and microcephaly if the OFC was less than two SDs below the mean [24].

Results

Genetic diagnostic profile of our patient

We identified a novel nonsense variant (NM_015025.4: c.2990C > A – chr2:1843005-G-T – exon 21 of 25 – p.Ser997Ter) in the *MYT1L* gene. This variant was not present in his parents nor in any of the populational databases consulted. Following the ACMG (American College of Medical Genetics and Genomics) guidelines [15], the variant was classified as pathogenic (PVS1, PM2, and PM6). We were unable to identify other variants that could justify the patient's phenotype.

Fig. 2A shows that the one nucleotide substitution leads to an early stop codon, causing loss of protein function. Images of the BAM files showing the region of the *MYT1L* variant in the proband and his parents are shown in Fig. 2 (B, C and D).

Phenotypic variation of patients with MYT1L *SNVs/indels variants*

The compilation of data is shown in Table 1. The information presented is mostly limited to published data, complemented with data provided by some authors. A phenotypic spectrum associated with pathogenic or likely pathogenic SNVs/indels in *MYT1L* (21/22) and the frequency of clinical characteristics shown in Table 1 are compiled in Table 2.

Table 2
Frequency of common features in patients with SNV/indels pathogenic/likely pathogenic *MYT1L* variants. Data from Table 1.

Clinical feature	Number of patients	Percentage
ID or DD	21/21	100%
Speech delay	19/19	100%
Polyphagia	10/13	76.9%
Obesity	10/21	47.6%
Strabismus	6/13	46.2%
Overweight	5/21	23.8%
Macrocephaly	4/17	23.5%

ID – Intellectual Disability; DD – Developmental Delay.
The nominator has the count of cases in which the characteristic is present over the total number of cases that we could consider (denominator).

Although the phenotypes associated with mutations in the *MYT1L* gene are generally heterogeneous, ID/DD and speech delay are always present. Even though hypotonia was described in ten patients, it was possibly not evaluated or reported in the remaining cases.

Obesity or overweight are present in 71.4% of the affected individuals, and in all of them polyphagia, when evaluated, was observed, suggesting that polyphagia is an over-arching mechanism leading to the excessive weight gain in these patients.

Although strabismus is a relatively common feature (46.2%), a generic facial phenotype does not seem to be recognizable, as previously pointed out by Windheuser et al. (2020) [9]. Macrocephaly was also prevalent (23.5%). Other features present in more than

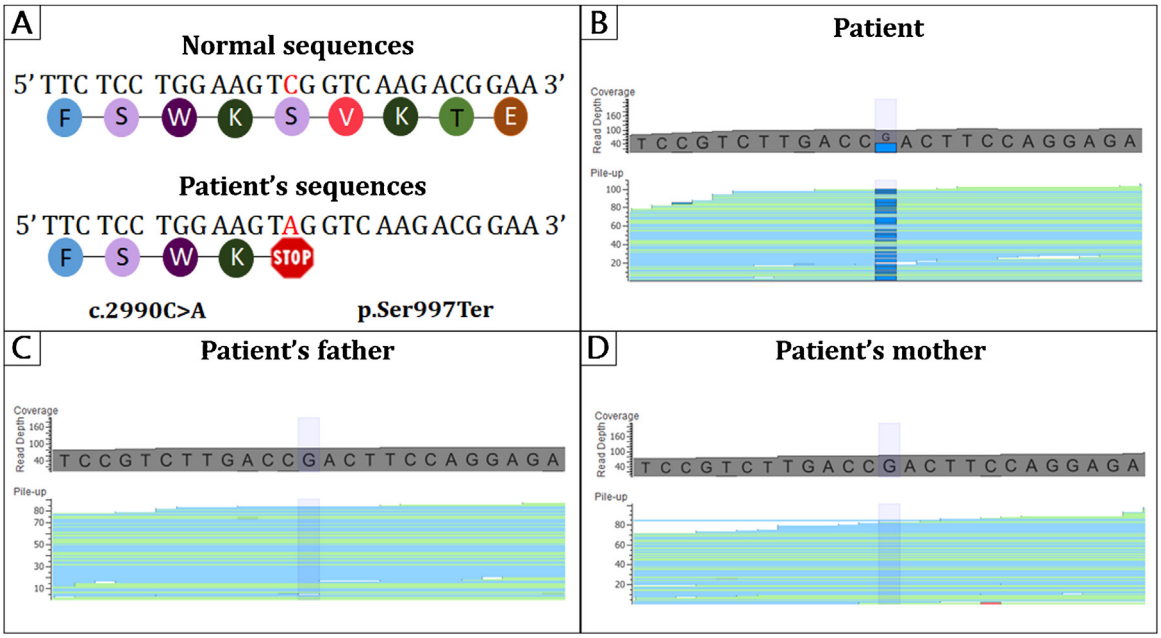


Fig. 2. Schematic representation of the *MYT1L* variant and BAM images of this variant region in the parents-proband trio.
(A) Schematic representation of a normal and mutated (NM_015025.4:c.2990C > A) segments of the *MYT1L* sequence and its consequence. The change of a C nucleotide (cytosine) by A (adenine) leads to an early stop codon. (B-D) BAM images of the *MYT1L* variant region showing its presence in the Patient and absence in his parents.

one patient include seizures and EEG abnormalities and anomalies of hands and feet.

As shown in Table 1, two patients had significant morphological changes in the central nervous system (Patients 4 and 8 of Blanchet et al., 2017 [10]). However, it is not possible to establish a reliable frequency for these features since no information was provided on which particular cases these tests were performed. However, a reasonable conclusion is that investigating brain anomalies could also be informative in the characterization of these patients.

The patient reported here presented hypogonadism. Although there is no information in the literature about infertility in these patients, it is notable that individual 6b of the Windheuser et al. (2020) [9] study fathered individual 6a, and both had pathogenic variants in *MYT1L*.

Autism and aggressiveness are relatively frequent, and both were present in eight patients. Sleep disturbances were described in four patients.

Except for the Patient reported here, delayed developmental milestones have been reported in all patients from whom data was available (18/19 – 94.7%). Importantly, there were no medical records certifying the initial normal developmental of our patient, and this information relied solely on the 10-year-old verbal recollections of his mother.

Discussion

Molecular diagnosis of patients with MYT1L *SNVs/indels*

The general phenotypic spectrum of patients with *MYT1L* mutations is broad. *MYT1L* mutations should be considered in the differential diagnosis of cases of ID and early-manifestation obesity together with other etiologies, including Prader-Willi (OMIM#176270), Bardet-Biedl [25], Börjeson-Forssman-Lehmann (OMIM#301900), MORM (OMIM#610156), Cohen (OMIM#216550), Alstrom (OMIM#203800), and Kleefstra type 1 syndrome (OMIM#610253).

Among the 21 cases shown in Table 1, only 7 benefitted from a molecular diagnosis before the age of 10, and 3 of them only in adulthood. The average age of molecular diagnosis should decrease in the coming years due to increasing access to targeted NGS panels, WES and WGS.

Having parental sequence data in addition to that of the proband, should ideally be performed to establish the inheritance pattern of each variant. It also provides information needed for genetic counseling of the family.

However, in Brazil, sequencing is relatively more expensive than in developed countries, primarily because of importation costs and heavy taxation on sequencing products. Therefore, whole exome sequencing is not routinely performed in public health facilities, and even private diagnostic labs often sequence only the proband DNA, sometimes complemented with Sanger sequencing of parental DNA to determine whether the mutation in the proband is *de novo*.

MYT1L *function and impact of variants on phenotype*

Table 1 shows seven nonsense, three frameshift, three splice and nine missense *MYT1L* variants. Using the ACMG criteria, we classified all but one of the variants as pathogenic ($n = 17$) or likely pathogenic ($n = 4$). Indeed, the human *MYT1L* gene appears to be greatly intolerant to loss of function (pLI score = 1) and to missense variants (z-score = 4.77), according to the prediction tools of gnomAD (Genome Aggregation Database).

MYT1L is a pro-neuronal transcription factor and, in combination with other transcription factors, can reprogram different types of cells [26–31]. In humans, *MYT1L* is expressed only in different parts of the brain [8,10] and appears to be notably high in fetal brain [8] and restricted to neurons [32].

In adults, it was not possible to demonstrate significant expression of *MYT1L* in hypothalamic structures relevant to appetite and satiety. However, *MYT1L* is expressed in several hypothalamic structures at 15–16 weeks post-conception, with significant reduction at week 21 [10]. This suggests that *MYT1L* mainly plays a role in fetal hypothalamic development, with no postnatal hypothalamic function activity.

Similar expression of *MYT1L* zebrafish orthologs (*myt1la/b*) was observed during neurogenesis. A zebrafish knockdown for *myt1la* and *myt1lb* had less oxytocin expression in the preoptic neuroendocrine area, which is the functional equivalent of the human hypothalamic paraventricular nucleus [10]. In humans, damage to the paraventricular nucleus leads to hypothalamic obesity, characterized by marked polyphagia [33]. Further, variants in the oxytocin gene and its receptor are associated with an increased risk of autism [34], which is a frequent condition associated with *MYT1L* variants [8,10]. Therefore, it is plausible that the intellectual disability and variable obesity observed in patients with *MYT1L* mutations is caused by variable dysregulating of gene expression and development of the neuroendocrine hypothalamus [10].

Conclusions

Considering the quite unspecific phenotype and the various forms of syndromic obesity that must be considered for a differential diagnosis, non-focused exploratory methods of genetic analysis, such as genome or exome sequencing and chromosome microarray analysis are advisable.

We described the first Brazilian case of a *MYT1L* mutation, a nonsense *de novo* variant not previously reported (NM_015025.4:c.2990C > A).

The literature review provides a phenotypic spectrum of patients with *MYT1L* SNVs/indels variants, in which ID/DD and speech delay are always present, and the majority of patients exhibit obesity or overweight associated with polyphagia. *MYT1L* haploinsufficiency impacts the neuroendocrine hypothalamic development, possibly resulting in polyphagia. Other recurrent clinical features are: macrocephaly, strabismus, behavioral problems, and hand and foot malformations.

Ethical standards

The research project was approved by the ethics committee (protocol number CEP 021/2004). The informed consent was obtained from the parents and al decisions were taken in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki, its subsequent revisions, and the Resolution 466/2012 of the Conselho Nacional de Saúde (National Health Council – Brazil) [35]. The signed informed consent explicitly allow publication of photographs of the patient.

Funding

Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (CEPID number 2013/08028-1; LMLC research fellowship number 2018/08486-3), Conselho Nacional de Desenvolvimento Científico e Tecnológico- CNPq (AK research fellowship 305806/2019-0), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (LMLC research fellowship number 1805008).

Internet resources

1KG Phase3: https://phase3browser.1000genomes.org/index.html
ABESO BMI calculator: https://abeso.org.br/obesidade-e-sindrome-metabolica/calculadora-imc/
ABraOM: http://abraom.ib.usp.br/
ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
Ensembl: https://www.ensembl.org
gene2phenotype: https://www.ebi.ac.uk/gene2phenotype
geneimprint: https://www.geneimprint.com/
gnomAD: https://gnomad.broadinstitute.org/
MYT1L on OMIM: https://omim.org/entry/613084?search=myt1l&highlight=myt1l
NHLBI: evs.gs.washington.edu/EVS
OMIM: https://www.omim.org/
Severe early-onset obesity 2.3: https://panelapp.genomicsengland.co.uk/panels/130/
SimulConsult head circumference calculator: https://simulconsult.com/resources/measurement.html
SYSID: https://sysid.cmbi.umcn.nl/a

Author contributions

Laura Carvalho: conceptualization, investigation; formal analysis; data curation; writing – original draft
Carla D'Angelo: conceptualization; resources; writing – review & editing
Zan Mustacchi: resources; writing – review & editing
Israel da Silva: software
Ana Krepischi: conceptualization; supervision; writing – review & editing
Celia Koiffmann: conceptualization; resources; supervision; writing – review & editing
Carla Rosenberg: conceptualization; project administration; funding acquisition; writing – review & editing

Conflicts of interest statement

All authors declare that there are no competing financial interests.

Acknowledgment

The authors are grateful to the patient and his family for their generous collaboration. We thank Prof. Peter Pearson for kindly reviewing the manuscript for improving the English.

References

[1] D'Angelo CS, Varela MC, de Castro CIE, Otto PA, Perez ABA, Lourenço CM, et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. Mol Cytogenet 2018;11, http://dx.doi.org/10.1186/s13039-018-0363-7.
[2] Huvenne H, Dubern B, Clément K, Poitou C. Rare genetic forms of obesity: clinical approach and current treatments in 2016. Obes Facts 2016;9:158–73, http://dx.doi.org/10.1159/000445061.
[3] Campbell LVC. Genetics of obesity focus. Aust Fam Physician 2017;46:456–9.
[4] Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. Obes Rev 2017;18:603–34, http://dx.doi.org/10.1111/obr.12531.
[5] Geets E, Meuwissen MEC, van Hul W. Clinical, molecular genetics and therapeutic aspects of syndromic obesity. Clin Genet 2019;95:23–40, http://dx.doi.org/10.1111/cge.13367.
[6] Wiegand S, Krude H. Monogene und syndromale Krankheitsbilder bei morbider Adipositas: Selten aber wichtig. Internist 2015;56:111–20, http://dx.doi.org/10.1007/s00108-014-3532-8.
[7] Vuillaume ML, Naudion S, Banneau G, Diene G, Cartault A, Cailley D, et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. Am J Med Genet Part A 2014;164:1965–75, http://dx.doi.org/10.1002/ajmg.a.36587.
[8] de Rocker N, Vergult S, Koolen D, Jacobs E, Hoischen A, Zeesman S, et al. Refinement of the critical 2p25.3 deletion region: the role of MYT1L in intellectual disability and obesity. Genet Med 2015;17:460–6, http://dx.doi.org/10.1038/gim.2014.124.
[9] Windheuser IC, Becker J, Cremer K, Hundertmark H, Yates LM, Mangold E, et al. Nine newly identified individuals refine the phenotype associated with MYT1L mutations. Am J Med Genet Part A 2020;182:1021–31, http://dx.doi.org/10.1002/ajmg.a.61515.
[10] Blanchet P, Bebin M, Bruet S, Cooper GM, Thompson ML, Duban-Bedu B, et al. MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus. PLoS Genet 2017;13, http://dx.doi.org/10.1371/journal.pgen.1006957.
[11] Faludi A, Izar M, Saraiva J, Chacra A, Bianco H, Afiune Neto A, et al. Atualização da diretriz brasileira de dislipidemias e prevenção da aterosclerose – 2017. Arq Bras Cardiol 2017;109, http://dx.doi.org/10.5935/abc.20170121.

[12] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60, http://dx.doi.org/10.1093/bioinformatics/btp324.
[13] Depristo MA, Banks E, Poplin R, Garimella Kv, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–501, http://dx.doi.org/10.1038/ng.806.
[14] Maltese PE, Iarossi G, Ziccardi L, Colombo L, Buzzonetti L, Crinò A, et al. A next generation sequencing custom gene panel as first line diagnostic tool for atypical cases of syndromic obesity: application in a case of Alström syndrome. Eur J Med Genet 2018;61:79–83, http://dx.doi.org/10.1016/j.ejmg.2017.10.016.
[15] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24, http://dx.doi.org/10.1038/gim.2015.30.
[16] Bonaglia MC, Giorda R, Zanini S. A new patient with a terminal de novo 2p25.3 deletion of 1.9 Mb associated with early-onset of obesity, intellectual disabilities and hyperkinetic disorder. Mol Cytogenet 2014;7, http://dx.doi.org/10.1186/1755-8166-7-53.
[17] Doco-Fenzy M, Leroy C, Schneider A, Petit F, Delrue MA, Andrieux J, et al. Early-onset obesity and paternal 2pter deletion encompassing the ACP1, TMEM18, and MYT1L genes. Eur J Hum Genet 2014;22:471–9, http://dx.doi.org/10.1038/ejhg.2013.189.
[18] Loid P, Mäkitie R, Costantini A, Viljakainen H, Pekkinen M, Mäkitie O. A novel MYT1L mutation in a patient with severe early-onset obesity and intellectual disability. Am J Med Genet Part A 2018;176:1972–5, http://dx.doi.org/10.1002/ajmg.a.40370.
[19] de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. New Engl J Med 2012;367:1921–9, http://dx.doi.org/10.1056/NEJMoa1206524.
[20] al Tuwaijri A, Alfadhel M. MYT1L mutation in a patient causes intellectual disability and early onset of obesity: a case report and review of the literature. J Pediatr Endocrinol Metab 2019;32:409–13, http://dx.doi.org/10.1515/jpem-2018-0505.
[21] Loid P, Mustila T, Mäkitie RE, Viljakainen H, Kämpe A, Tossavainen P, et al. Rare variants in genes linked to appetite control and hypothalamic development in early-onset severe obesity. Front Endocrinol 2020;11, http://dx.doi.org/10.3389/fendo.2020.00081.
[22] Mancini MC. Tratado de obesidade. 2nd ed. Rio de Janeiro – RJ: Guanabara Koogan; 2015.
[23] WHO Technical Counsultation. Obesity: preventing and managing the global epidemic. WHO Consultation 2000; 2000, 0512-3054 (Print).
[24] Fenichel GM. Clinical pediatric neurology: a signs and symptoms approach. 6th ed. Philadelphia, PA: Saunders/Elsevier; 2009.
[25] Guo DF, Rahmouni K. Molecular basis of the obesity associated with Bardet-Biedl syndrome. Trends Endocrinol Metab 2011;22:286–93, http://dx.doi.org/10.1016/j.tem.2011.02.009.
[26] Miskinyte G, Devaraju K, Grønning Hansen M, Monni E, Tornero D, Woods NB, et al. Direct conversion of human fibroblasts to functional excitatory cortical neurons integrating into human neural networks. Stem Cell Res Therapy 2017;8, http://dx.doi.org/10.1186/s13287-017-0658-3.
[27] Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature 2010;463:1035–41, http://dx.doi.org/10.1038/nature08797.
[28] Liang XG, Tan C, Wang CK, Tao RR, Huang YJ, Ma KF, et al. Myt1l induced direct reprogramming of pericytes into cholinergic neurons. CNS Neurosci Ther 2018;24:801–9, http://dx.doi.org/10.1111/cns.12821.
[29] Chuang W, Sharma A, Shukla P, Li G, Mall M, Rajarajan K, et al. Partial reprogramming of pluripotent stem cell-derived cardiomyocytes into neurons. Sci Rep 2017;7, http://dx.doi.org/10.1038/srep44840.
[30] Yang Y, Jiao J, Gao R, Yao H, Sun XF, Gao S. Direct conversion of adipocyte progenitors into functional neurons. Cell Reprogram 2013;15:484–9, http://dx.doi.org/10.1089/cell.2013.0013.
[31] Zhang SZ, Ma LX, Qian WJ, Li HF, Wang ZF, Wang HX, et al. Modeling neurological disease by rapid conversion of human urine cells into functional neurons. Stem Cells Int 2016 2016, http://dx.doi.org/10.1155/2016/2452985.
[32] Kim JG, Armstrong RC, Agoston Dv, Robinsky A, Wiese C, Nagle J, et al. Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system, vol. 50. Wiley-Liss, Inc. †; 1997.
[33] Jennifer Abuzzahab M, Roth CL, Shoemaker AH. Hypothalamic obesity: prologue and promise. Horm Res Paediatr 2019;91:128–36, http://dx.doi.org/10.1159/000496564.
[34] Cataldo I, Azhari A, Esposito G. A review of oxytocin and arginine-vasopressin receptors and their modulation of autism spectrum disorder. Front Mol Neurosci 2018;11, http://dx.doi.org/10.3389/fnmol.2018.00027.
[35] Brasil. Resolução 466 do Conselho Nacional de Saúde 2012.

Glossary

SNVs (Single-nucleotide Variants): changes in a base pair.
Indels: small insertions or deletions (usually between 1 and 50 bp).
CNVs (Copy-Number Variations): genetic variants in which a segment of DNA greater than or equal to 1 kb is present in a number of copies that differs from the reference genome. The limits for indels and CNVs vary a lot in the literature, depending on the lower sensitivity for detecting array and higher for detecting indels, but up to 50 bp is an acceptable threshold for in/dels while larger than 1 kb is the minimum for chromosome microarray analyses. There is a gray area in between which boundaries will be defined as technology changes.

Capítulo 2.3

Two novel pathogenic variants in MED13L: one familial and one isolated case

CARVALHO, L. M. L; COSTA, S. S.; CAMPAGNARI, F.; KAUFMAN, A; BERTOLA, D. R., da SILVA, I. T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. Two novel pathogenic variants in *MED13L*: one familial and one isolated case. **Journal of Intellectual Disability Research**, 2021. ISSN: 1365-2788. DOI: 10.1111/jir.12891

JIDR Journal of Intellectual Disability Research
Published on behalf of mencap and in association with IASSID

Journal of Intellectual Disability Research doi: 10.1111/jir.12891

Two novel pathogenic variants in *MED13L*: one familial and one isolated case

L. M. L. Carvalho,[1] S. S. da Costa,[1] F. Campagnari,[2] A. Kaufman,[3] D. R. Bertola,[1] I. T. da Silva,[4] A. C. V. Krepischi,[1] C. P. Koiffmann[1] & C. Rosenberg[1]

1 *Human Genome and Stem Cell Research Centre, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo (USP), São Paulo, SP, Brazil*
2 *STgenetics, Indaiatuba, SP, Brazil*
3 *Department of Psychiatry, Faculty of Medicine, University of São Paulo (USP), São Paulo, SP, Brazil*
4 *International Centre for Research, A. C. Camargo Cancer Centre, São Paulo, SP, Brazil*

Abstract

Background Genetic variants involving the *MED13L* gene can lead to an autosomal dominant syndrome characterised by intellectual disability/developmental delay and facial dysmorphism.
Methods We investigated two cases (one familial and one isolated) of intellectual disability with speech delay and dysmorphic facial features by whole-exome sequencing analyses. Further, we performed a literature review about clinical and molecular aspects of *MED13L* gene and syndrome.
Results Two *MED13L* variants have been identified [*MED13L*(NM_015335.5):c.4417C>T and *MED13L* (NM_015335.5):c.2318delC] and were classified as pathogenic according to the ACMG (American College of Medical Genetics and Genomics) guidelines. One of the variants was present in sibs.
Conclusions The two pathogenic variants identified have not been previously reported. Importantly, this is the first report of a familial case of *MED13L* nonsense mutation. Although the parents of the affected children were no longer available for analysis, their apparently normal phenotypes were surmised from familial verbal descriptions corresponding to normal mental behaviour and phenotype. In this situation, the familial component of mutation transmission might be caused by gonadal mosaicism of a *MED13L* mutation in a gonad from either the father or the mother. The case reports and the literature review presented in this manuscript can be useful for genetic counselling.

Keywords developmental delay, gonadal mosaicism, intellectual disability, MED13L, speech delay, whole-exome sequencing

Introduction

Copy number variations (CNVs), single-nucleotide variations (SNVs) and indels involving the *MED13L* gene (mediator complex subunit 13 like - OMIM*608771) can lead to an autosomal dominant clinical condition (mental retardation and distinctive

Correspondence: Dr Carla Rosenberg, Human Genome and Stem Cell Research Centre, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, Matão Street, Lane 14, Number 277 – Room 350, São Paulo, SP 05508-090, Brazil (e-mail: carlarosenberg@ib.usp.br).

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

facial features with or without cardiac defects - MRFACD, OMIM#616789) characterised by intellectual disability (ID) associated with facial dysmorphisms in which heart defects have incomplete penetrance (Cafiero *et al.* 2015; Asadollahi *et al.* 2017). Both the *MED13L* haploinsufficiency mutation and missense mutation have been reported to be causative (Yi *et al.* 2020).

The protein encoded by *MED13L* is a component of the large Mediator complex, an important regulator of transcription mediated by RNA polymerase II (Cafiero *et al.* 2015). Clinical and experimental data demonstrate that *MED13L* plays an important role in development, especially in the regulation of the neural crest cells and neurogenesis (Asadollahi *et al.* 2017).

We report two novel variants in *MED13L*. The first variant is apparently *de novo* (patient 1 – isolated case), while the second is present in two siblings (patients 2 and 3) who supposedly inherited the pathogenic variant from a parent with gonadal mosaicism.

Figure 1. Craniofacial features of patients 1 (a, b), 2 (c, d) and 3 (e, f). Facial features of the three individuals reported: (a, b) patient 1 at the age of 22 – note long columella, short philtrum, retrognathia and prominent antihelix; (c, d) patient 2 at the age of 33 showing mild bitemporal narrowing, upslanting palpebral fissures, midface retraction, low nasal bridge, long tip of the nose and micrognathia; and (e, f) patient 3 presenting similar facial features of her brother (patient 2). [Colour figure can be viewed at wileyonlinelibrary.com]

Patients and methods

Isolated case: patient 1

A 22-year-old Brazilian woman (Fig. 1a,b) was referred from the Psychiatry Department of the Academic Hospital of the University of São Paulo - USP (Hospital das Clínicas de São Paulo) to the Genetic Service of the Human Genome and Stem Cell Research Centre (USP, São Paulo, Brazil) for presenting syndromic obesity with ID.

The patient is the first child (Fig. 1a,b – patient 1) of a non-consanguineous healthy couple and has a 19-year-old healthy sister. There are no similar cases in the family.

She was born at term, after an uneventful pregnancy, with normal anthropometric measurements (birth weight of 3.260 g, length of 48 cm and occipital head circumference - OFC - of 34 cm). At birth, she presented with clubfeet. She used orthopaedic cast and had foot surgery at the age of 1.

She was hypotonic and evolved with developmental delay (DD): sat with support at 8-month and walked at 2-year-8-month old. At the age of 7, she lost night-time urine control and continued to wear a night diaper.

Facial features at 22-year-old exhibited (Fig. 1a,b) prominent, long columella, short philtrum, retrognathia, prominent antihelices and macroglossia. She frequently kept her mouth open and drooled excessively. Her hands and feet were small relative to body size.

At age 7, the patient underwent screening tests for inborn errors of metabolism with normal results. Her electroencephalogram (EEG) showed a mild

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

electrical disorganisation with signs of irritative activity in the left temporal region. However, she never had seizures and there was no history of cardiac anomalies.

Cranial magnetic resonance imaging at 21-year-old indicated a nodular image with cystic foci measuring about 7 × 10 mm in the pineal gland.

Ophthalmological evaluation disclosed astigmatism and myopia. Hearing assessment and bone densitometry test had normal results. She experienced menarche at 11 years of age.

A formal cognitive profile made at 15 years revealed mild ID (estimated intelligence quotient - IQ - of 56). She currently (26 years of age) attends a special school dedicated to people with disabilities.

She is illiterate and is very dependant on assistance with daily personal hygiene, clothing choice and household tasks, has impaired interactions and has no friends.

Between the ages of 17 and 22 years old, she presented obsessive–compulsive aggressive behaviour, anxiety, trichotillomania, skin picking, self-harm and polyphagia. There were also reports of hallucinations.

On physical examination at 22 years old, she weighed 94 kg and a height of 1.53 m (−1.58 SD); [body mass index (BMI) = 40.2 kg/m^2 – grade III obesity]. Her obesity was probably caused to the hyperphagic behaviour that the patient presented between 17 and 22 years of age, in combination with some medications that she used, which can lead to weight gain (antipsychotics risperidone and quetiapine).

Blood samples were collected from the proband and her parents for genetic evaluation. Previous genetic testing included G-banding karyotype, multiplex ligation-dependant probe amplification (MLPA - P036, P070, P245 and GOD – MRC-Holland, Amsterdam, the Netherlands) for analysis of subtelomeric regions, as well as known regions for prevalent microdeletion syndromes, and methylation test to Prader–Willi region; all analyses were normal.

Familial case: patients 2 and 3

Patient 2 (33-year-old man) and patient 3 (27-year-old woman) are siblings presenting with ID and facial dysmorphisms that were initially evaluated as part of an old project aimed at validating the use of SNP arrays in routine chromosome diagnosis. They both attended a clinic for dental care for people with special needs run by a non-profit organisation (Dental Assistance Centre for People with Deficiency – CAOE). This is part of the Faculty of Dentistry – Araçatuba Campus – São Paulo State University (UNESP), Brazil.

Both patients were born to healthy, non-consanguineous parents, who had a total of 11 children, two of them affected (pedigree on Fig. 2 – patients 2 and 3). There was no history of miscarriages.

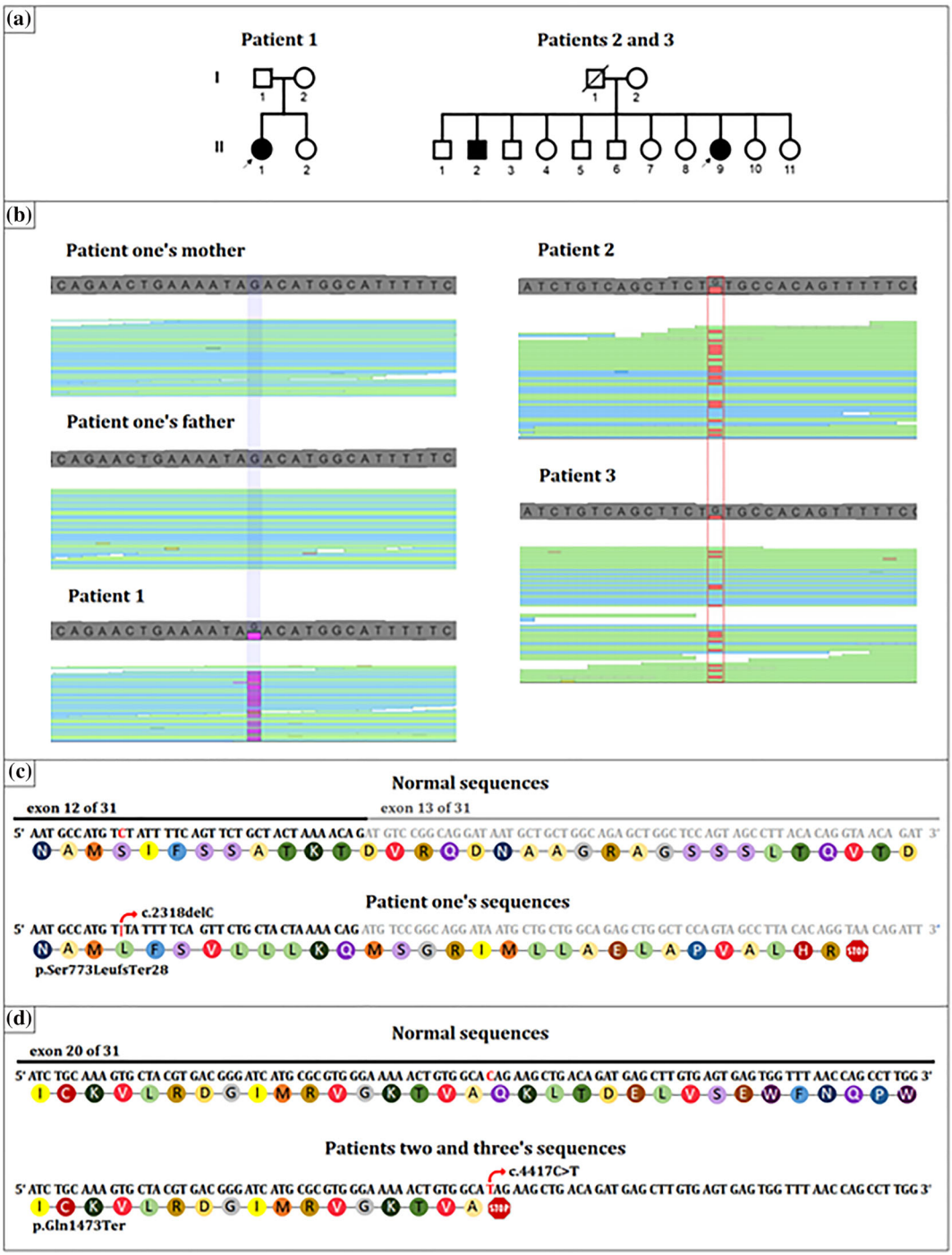
Patient 2 (Fig. 1c,d) was born at term, by vaginal delivery. Motor milestones were delayed: he sat unsupported at 1-year, crawled at 2-year and walked at 3-year. He has moderate ID and lack of speech. He is not able to read but understands commands and is independent in daily household tasks, including washing his own clothes. He has normal sphincter control. He has collaborative and sociable behaviour. Audiological evaluation was normal. There is no history of seizures and cardiac anomalies.

Physical examination disclosed mild bitemporal narrowing, upslanting palpebral fissures, midface retraction, low nasal bridge, long tip of the nose and micrognathia.

Patient 3 (Fig. 1e,f) was also born at term by vaginal delivery. She also presented with DD: sat unsupported at 1-year, crawled at 1-year-2-month and walked at 2-year. Similar to her brother, she has moderate ID, is independent in daily tasks and has normal sphincter control. However, she has more difficulty understanding commands than her brother. Her speech development is severely delayed (speaks just a few words), but her hearing is normal. There is no history of seizures and cardiac anomalies. She also had facial dysmorphisms, similar to her brother: upslanting palpebral fissures, low nasal bridge, long tip of the nose and micrognathism.

Saliva DNA from the affected siblings was obtained for genetic analyses. We were unable to obtain DNA from the parents once the father was deceased, and we were unable to re-contact the family. The fragile X syndrome (AmplideX® FMR1; Asuragen, Austin, TX, USA) and genomic imbalances (850K platform; Illumina, San Diego, CA, USA) had previously been excluded in these siblings.

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd



© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

Figure 2. (a) Patient's pedigrees, (b) BAM files highlighting the variants identified in *MED3L* and (c, d) related genetic mechanisms. (a) Patient's pedigrees. (b) On the left, we present the BAM files referring the parents–proband trio of isolated case (patient 1). As can be seen, the *MED13L*:c.2318delC variant was identified only in patient 1. On the right, we present patients 2 and 3 BAM files. As can be seen, the *MED13L*:c4417C>T variant was identified in both siblings. (c) Schematic representation of the *MED13L*(NM_015335.5):c.2318delC variant (patient 1) and its consequence in the nucleotide and amino acid sequences. The deletion of 1 bp leads to a shift in the reading frame and forms an early stop codon (p.Ser773LeufsTer28). (d) Schematic representation of the *MED13L*:c4417C>T variant (patients 2 and 3) and its consequence. As can be seen, the change of a C nucleotide (cytosine) by T (thymine) leads to an early stop codon, causing loss of protein function. [Colour figure can be viewed at wileyonlinelibrary.com]

Genetic diagnosis

For patient 1 and her parents, CNV analyses and whole-exome sequencing (WES) were simultaneously performed using the V6 + Oneseq from Agilent Technologies (Santa Clara, CA, USA), which, in addition to targeting exome regions (WES), includes a backbone for genome-wide copy number estimates (Villela *et al.* 2017).

The CNV analyses (850K cytochip Illumina) of patients 2 and 3 showed no relevant copy number alterations. WES of these individuals was performed using Nextera Rapid Capture Expanded Exome Kit (Illumina).

For all WES analyses, the raw paired-end reads were aligned to the GRCh37/hg19 reference genome using the Burrows–Wheeler Aligner - BWA (Li & Durbin 2009) and preprocessed according to GATK toolkit (DePristo *et al.* 2011). Exome data were analysed using the VarSeq software (Golden Helix, Bozeman, MT, USA).

In patient 1, we identified a novel *de novo* frameshift variant [*MED13L*(NM_015335.5):c.2318delC – p.Ser773LeufsTer28] in the *MED13L* gene. The variant is not present in control population databases, including the Brazilian population (gnomAD, 1KG and AbraOM). Following the ACMG (American College of Medical Genetics and Genomics) guidelines (Richards *et al.* 2015), the variant was classified as pathogenic (PVS1, PS2 and PM2).

In patients 2 and 3, we identified a novel nonsense variant [*MED13L*(NM_015335.5):c.4417C>T-p.Gln1473Ter] in the same gene. The variant is not present in control population databases and was classified as pathogenic (PVS1 and PM2). Segregation analyses could not be performed because the father was deceased and the family could not be contacted. There is no report of any of the parents presenting mental impairment or dysmorphic features.

Images of the BAM files from all patients showing the region of the pathogenic genetic variants identified are shown in Fig. 2b.

A deletion of a single base pair in patient 1 leads to a frameshift and an early stop codon, resulting in loss of protein function, as shown on Fig. 2c. The change of a C nucleotide (cytosine) by T (thymine) in patients 2 and 3 leads to an early stop codon, causing loss of protein function, as shown on Fig. 2d.

Literature review and discussion

The human gene *MED13L* has a chromosomal location at 12q24.21, size about 317 kb and 31 exons. *MED13L* gene encodes a 2210aa protein (NP_056150.1), which has two LxxLL (where L is leucine and x is any amino acid) motifs, which have a biological function of participating in many protein–protein interactions, associated with activation or repression of transcription (Plevin *et al.* 2005).

MED13L protein is a subunit of the multiprotein Mediator complex (Yamamoto *et al.* 2017) that interacts with RNA polymerase II, transcription factors and enhancers. The Mediator complex is considered an essential transcriptional co-activator of protein-coding genes. The presence of the Mediator complex is characteristic in eukaryotes, and although the number of subunits varies among species, the general structure and subunit organisation are relatively conserved (Jeronimo & Robert 2017).

ID/DD was related to variants in genes that encode the subunits MED12 (Risheg *et al.* 2007; Schwartz *et al.* 2007), MED13 (Snijders Blok *et al.* 2018), MED13L (Asadollahi *et al.* 2013; Cafiero *et al.* 2015; van Haelst *et al.* 2015), MED17 (Agostini *et al.* 2018), MED20 (Vodopiutz *et al.* 2015), MED23 (Hashimoto *et al.* 2011), MED25 (Basel-Vanagaite *et al.* 2015; Figueiredo *et al.* 2015) and CDK19 (Mukhopadhyay *et al.* 2010).

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

CDK8, *MED12* and *MED13* encode proteins that are part of the CDK kinase module of the Mediator complex and went through independent gene duplication events, originating their paralogs in vertebrates: *CDK19*, *MED12L* and *MED13L* (Allen & Taatjes 2015). The CDK kinase module may or may not be present in the Mediator complex, and each of the paralogs is mutually exclusive with each other; for example, MED13L is not present in MED13 isolations and vice versa (Daniels *et al.* 2013). This is probably due to the relay of information from particular temporal/spatial signals (Asadollahi *et al.* 2017).

MED13L protein is considerably conserved among mammalian species. In primates, the similarity is even more evident (>99%). Highly conserved sequences usually have crucial biological functions and new genetic variants that are compatible with life can produce functional changes causing genetic disease. However, the human *MED13L* gene appears to be poorly tolerant to non-synonymous variants – especially loss of function (pLi score = 1) according to the prediction tools of gnomAD (Genome Aggregation Database – a megaproject which currently includes 125 748 exomes and 15 708 genomes).

Human *MED13L* exon 1 contains the start codon, which is embedded in a CpG island (Muncke *et al.* 2003; Musante *et al.* 2004), which suggests an important control of gene expression, because the inhibitory effect of the intragenic DNA methylation on transcription elongation is larger when it occurs nearer the start codon (Choi *et al.* 2009). In fact, *MED13L* is expressed in several human tissues; the mRNA has been detected in foetal brain and heart, and adult tissues including skeletal muscle, brain, heart, aorta, kidney and peripheral blood leukocytes, but appears to have a relatively greater expression in the foetal and adult brain – especially cerebellum –, heart and skeletal muscle (Muncke *et al.* 2003; Musante *et al.* 2004; Asadollahi *et al.* 2017).

Utami *et al.* (2014) published a relevant functional study showing a defective migration of cranial neural crest cells following knockdown of the of a *MED13L* orthologue in zebrafish (*med13b*), which leads to inappropriate development of branchyal and pharyngeal arches. At 2 days postfertilisation, more than 95% of morphants showed a curved body axis, underdeveloped head, pericardial oedema and microphthalmia. Additionally, they performed a transcriptomic analysis of *MED13L*-knockdown neurons derived from embryonic stem cells showing a differential expression of components of the Wnt and FGF signalling pathways, both crucial for directing neural crest induction and neurogenesis. In summary, clinical and experimental data demonstrate that *MED13L* plays an important role in development, especially in the regulation of neural crest cells and neurogenesis. Accordingly, the syndrome involved is a probable neurocristopathy (Asadollahi *et al.* 2017).

MED13L mutations were related to the clinical condition OMIM#616789 (mental retardation and distinctive facial features with or without cardiac defects) and with several types of genetic variants involving *MED13L* gene: nonsense (Cafiero *et al.* 2015; Caro-Llopis *et al.* 2016), frameshift (Hamdan *et al.* 2014; Redin *et al.* 2014; Adegbola *et al.* 2015; Cafiero *et al.* 2015; Codina-Solà *et al.* 2015; Asadollahi *et al.* 2017; Yamamoto *et al.* 2017), CNV deletion including *MED13L* (Adegbola *et al.* 2015), CNV duplication including *MED13L* (Adegbola *et al.* 2015), CNV triplication including *MED13L* (Asadollahi *et al.* 2013), intragenic CNV deletion (Asadollahi *et al.* 2013; Adegbola *et al.* 2015; van Haelst *et al.* 2015; Yamamoto *et al.* 2017), intragenic CNV duplication (Adegbola *et al.* 2015), disruptive translocation (Muncke *et al.* 2003; Utami *et al.* 2014) and missenses (Gilissen *et al.* 2014; Asadollahi *et al.* 2017; Yi *et al.* 2020). In general, they are cases of heterozygous variants. However, Najmabadi *et al.* (2011) reported a missense variant in homozygosity.

The observed/expected missense variants ratio for the gene is 0.7 (z-score = 3.69), according to gnomAD. The positive z-score indicates intolerance to this class of variant. However, the clinical relevance of missense mutations in *MED13L* is poorly understood, and only a limited number of patients have been described (Jiménez-Romero *et al.* 2018; Smol *et al.* 2018). The majority of *MED13L* missense variants reported are of uncertain significance (Asadollahi *et al.* 2017). Therefore, additional functional studies are necessary to reclassify them as to their pathogenicity.

Recently, Smol *et al.* (2018) published a study proposing two mutational hotspots for missense variants: exons 15–17 and 25–31. To conclude this, they considered seven missense variants identified in

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

their study as well as the other five *MED13L*-missense variants previously reported in the literature. However, according to these same authors, precise roles of these domains and specific residues remain to be determined to better understand the molecular mechanisms underlying *MED13L*-related ID. They also suggest that patients carrying missense variants develop epilepsy earlier, compared with patients with protein-truncating variants (4/9 vs. 1/26). Further, according to Yi *et al.* (2020), patients with missense mutations in *MED13L* also have a more severe phenotype in other aspects, including hypotonia, absent speech and severely delayed motor function, compared with patients with truncating variants. The occurrence of truncating variants involving the entire *MED13L* gene would lead to the conclusion that the clinical condition is due to haploinsufficiency. However, the clinical status of our patients (1, 2 and 3) can be explained by the haploinsufficiency mechanism because the variants are loss of function. Effectively, *MED13L* appears to be a dose-sensitive gene, because duplications (Adegbola *et al.* 2015) and triplications including this gene (Asadollahi *et al.* 2013) also lead to an overlapping clinical condition.

With respect to the phenotypic profile, according to Asadollahi *et al.* (2017), 100% of patients with pathogenic variants in *MED13L* have DD/ID and facial dysmorphism (although these dysmorphisms are variable). Other common characteristics are hypotonia (60%), severe speech delay (50%), abnormal growth parameters (50%), coordination problems (38%), abnormal magnetic resonance imaging (38%), autistic features (20%), seizures or abnormal EEG (16%), persistent foramen ovale/ventricular septal defect (17%), complex congenital heart defects (13%) and aggressive behaviour (9%). The following frequencies are described for specific facial dysmorphisms: depressed/flat nasal bridge (86%), broad/prominent forehead (76%), bulbous nose (71%), upslanting palpebral fissures (62%), abnormal chin (micrognathia/prognathia/retrognathia) (55%), low set ears (52%), bitemporal narrowing (50%), horizontal eyebrows (29%) and macroglossia (29%). The following frequencies are described for specific abnormal growth parameters: microcephaly (15%), low birth weight (15%), postnatal underweight (15%), macrocephaly (5%) and postnatal overweight (5%).

We would like to point out that clubfoot, which is a characteristic presented by patient 1, does not seem to be common in the phenotype of the *MED13L* syndrome. However, this characteristic was also reported by van Haelst *et al.* (2015) and Cafiero *et al.* (2015). It is worth noting that our patient 1 was referred to a genetic investigation service for syndromic obesity. A genetic investigation of Prader–Willi syndrome was carried out (with negative results), because of the marked polyphagia and obesity. Previously, Cafiero *et al.* (2015) also described a patient with central obesity (90th centile). Further, the literature suggests differential diagnosis for the syndrome caused by variants in *MED13L*: 22q11.2 (Asadollahi *et al.* 2013, 2017), 1p36 (Cafiero *et al.* 2015) and Kleefstra syndromes (Asadollahi *et al.* 2017).

We describe the first familial case report with a *MED13L* nonsense variant. Previously, Yamamoto *et al.* (2017) described a familial case involving two brothers with intragenic deletion involving exons 3–14 of *MED13L*, and the variant was inherited from the mother (mosaic carrier).

Considering that only two (patients 2 and 3) of the 11 siblings in the family are affected (Fig. 2a) and the parents are apparently clinically normal, the most likely explanation is gonadal mosaicism in one of the parents. It should be noted that, despite the variable expressiveness of the syndrome caused by the haploinsufficiency of *MED13L*, ID and facial dysmorphisms are present in 100% of the cases, according to the wide review published by Asadollahi *et al.* (2017). Thus, the hypothesis of incomplete penetrance to explain familial cases is unlikely.

Ethical Standards

The research projects in which the patients reported were involved were approved by the ethics committee of our institution. The informed consent was obtained, and our decisions were taken in accordance with the ethical standards as laid down in the Declaration of Helsinki and Resolution 466/2012 of the Conselho Nacional de Saúde (National Health Council – Brazil).

Acknowledgements

The authors are grateful to the patients and their families for precious cooperation in this study. We

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

thank Prof. Peter Pearson for kindly reviewing the manuscript.

Source of Funding

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (process numbers 2018/08486-3, 2012/50981-5 and 2013/08028-1), Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (305806/2019-0 - A. K.) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (process number 1805008).

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

Adegbola A., Musante L., Callewaert B., Maciel P., Hu H., Isidor B. *et al.* (2015) Redefining the MED13L syndrome. *European Journal of Human Genetics* **23**, 1308–17.

Agostini A., Marchetti D., Izzi C., Cocco I., Pinelli L., Accorsi P. *et al.* (2018) Expanding the phenotype of MED 17 mutations: description of two new cases and review of the literature. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **177**, 687–90.

Allen B. L. & Taatjes D. J. (2015) The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology* **16**, 155–66.

Asadollahi R., Oneda B., Sheth F., Azzarello-Burri S., Baldinger R., Joset P. *et al.* (2013) Dosage changes of *MED13L* further delineate its role in congenital heart defects and intellectual disability. *European Journal of Human Genetics* **21**, 1100–4.

Asadollahi R., Zweier M., Gogoll L., Schiffmann R., Sticht H., Steindl K. *et al.* (2017) Genotype-phenotype evaluation of *MED13L* defects in the light of a novel truncating and a recurrent missense mutation. *European Journal of Medical Genetics* **60**, 451–64.

Basel-Vanagaite L., Smirin-Yosef P., Essakow J. L., Tzur S., Lagovsky I., Maya I. *et al.* (2015) Homozygous *MED25* mutation implicated in eye–intellectual disability syndrome. *Human Genetics* **134**, 577–87.

Cafiero C., Marangi G., Orteschi D., Ali M., Asaro A., Ponzi E. *et al.* (2015) Novel *de novo* heterozygous loss-of-function variants in *MED13L* and further delineation of the *MED13L* haploinsufficiency syndrome. *European Journal of Human Genetics* **23**, 1499–504.

Caro-Llopis A., Rosello M., Orellana C., Oltra S., Monfort S., Mayo S. *et al.* (2016) *De novo* mutations in genes of mediator complex causing syndromic intellectual disability: mediatorpathy or transcriptomopathy? *Pediatric Research* **80**, 809–15.

Choi J., Bae J. B., Lyu J., Kim T. Y. & Kim Y. J. (2009) Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biology* **10**, 1–10, R89.

Codina-Solà M., Rodríguez-Santiago B., Homs A., Santoyo J., Rigau M., Aznar-Laín G. *et al.* (2015) Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Molecular Autism* **6**, 1–16, 21.

Daniels D. L., Ford, M., Schwinn, M. K., Benink, H., Galbraith, M. D., Amunugama, R. *et al.* (2013) Mutual exclusivity of MED12/MED12L, MED13/13L, and CDK8/19 paralogs revealed within the CDK-Mediator kinase module. *Journal of Proteomics and Bioinformatics* **01**, 1–7.

DePristo M. A., Banks E., Poplin R., Garimella K. V., Maguire J. R., Hartl, C. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–8.

Figueiredo T., Melo U. S., Pessoa A. L. S., Nobrega P. R., Kitajima J. P., Correa I. *et al.* (2015) Homozygous missense mutation in *MED25* segregates with syndromic intellectual disability in a large consanguineous family. *Journal of Medical Genetics* **52**, 123–7.

Gilissen C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H. *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–7.

Hamdan F. F., Srour M., Capo-Chichi J.M., Daoud H., Nassif C., Patry L. *et al.* (2014) *De novo* mutations in moderate or severe intellectual disability. *PLoS Genetics* **10**, e1004772. Edited by G. M. Cooper.

Hashimoto S., Boissel S., Zarhrate M., Rio M., Munnich A., Egly J. M. *et al.* (2011) *MED23* mutation links intellectual disability to dysregulation of immediate early gene expression. *Science* **333**, 1161–3.

Jeronimo C. & Robert F. (2017) The mediator complex: at the nexus of RNA polymerase II transcription. *Trends in Cell Biology* **27**, 765–83.

Jiménez-Romero S., Carrasco-Salas P. & Benítez-Burraco A. (2018) Language and cognitive impairment associated with a novel p.Cys63Arg change in the MED13L transcriptional regulator. *Molecular Syndromology* **9**, 83–91.

Li H. & Durbin R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–60.

Mukhopadhyay A., Kramer J. M., Merkx G., Lugtenberg D., Smeets D. F., Oortveld M. A. *et al.* (2010) *CDK19* is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. *Human Genetics* **128**, 281–91.

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

Muncke N., Jung C., Rudiger H., Ulmer H., Roeth R., Hubert A. *et al.* (2003) Missense mutations and gene interruption in *PROSIT240*, a novel *TRAP240*-like gene, in patients with congenital heart defect (transposition of the great arteries). *Circulation* **108**, 2843–50.

Musante L., Bartsch O., Ropers H. H. & Kalscheuer V. M. (2004) cDNA cloning and characterization of the human *THRAP2* gene which maps to chromosome 12q24, and its mouse ortholog *Thrap2*. *Gene* **332**, 119–27.

Najmabadi H., Hu H., Garshasbi M., Zemojtel T., Abedini S. S., Chen W. *et al.* (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63.

Plevin M. J., Mills M. M. & Ikura M. (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends in Biochemical Sciences* **30**, 66–9.

Redin C., Gérard B., Lauer J., Herenger Y., Muller J., Quartier A. *et al.* (2014) Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics* **51**, 724–36.

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* **17**, 405–24.

Risheg H., Graham Jr, J. M., Clark, R. D., Rogers, R. C., Opitz, J. M., Moeschler, J. B. *et al.* (2007) A recurrent mutation in *MED12* leading to R961W causes Opitz-Kaveggia syndrome. *Nature Genetics* **39**, 451–3.

Schwartz C. E., Tarpey P. S., Lubs H. A., Verloes A., May M. M., Risheg H. *et al.* (2007) The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the *MED12* gene. *Journal of Medical Genetics* **44**, 472–7.

Smol T., Petit F., Piton A., Keren B., Sanlaville D., Afenjar A. *et al.* (2018) *MED13L*-related intellectual disability: involvement of missense variants and delineation of the phenotype. *Neurogenetics* **19**, 93–103.

Snijders Blok L., Hiatt S. M., Bowling K. M., Prokop J. W., Engel K. L., Cochran J. N. *et al.* (2018) De novo mutations in MED13, a component of the Mediator complex, are associated with a novel neurodevelopmental disorder. *Human Genetics* **137**, 375–88.

Utami K. H., Winata C. L., Hillmer A. M., Aksoy I., Long H. T., Liany H. *et al.* (2014) Impaired development of neural-crest cell derived organs and intellectual disability caused by *MED13L* haploinsufficiency. *Human Mutation*, **35**, 1311–20.

van Haelst M. M., Monroe G. R., Duran K., Van Binsbergen E., Breur J. M., Giltay J. C. *et al.* (2015) Further confirmation of the *MED13L* haploinsufficiency syndrome. *European Journal of Human Genetics* **23**, 135–8.

Villela D., Costa S. S., Vianna-Morgante A. M., Krepischi A. C. V. & Rosenberg C. (2017) Efficient detection of chromosome imbalances and single nucleotide variants using targeted sequencing in the clinical setting. *European Journal of Medical Genetics*, **60**, 667–74.

Vodopiutz J., Seidl R., Prayer D., Khan M. I., Mayr J. A., Streubel B. *et al.* (2015) *MED20* mutation associated with infantile basal ganglia degeneration and brain atrophy. *European Journal of Pediatrics* **174**, 113–8.

Yamamoto T., Shimojima K., Ondo Y., Shimakawa S. & Okamoto N. (2017) *MED13L* haploinsufficiency syndrome: a de novo frameshift and recurrent intragenic deletions due to parental mosaicism. *American Journal of Medical Genetics Part A* **173**, 1264–9.

Yi Z., Zhang Y., Song Z., Pan H., Yang C., Li F. *et al.* (2020) Report of a de novo c.2605C > T (p.Pro869Ser) change in the *MED13L* gene and review of the literature for MED13L-related intellectual disability. *Italian Journal of Pediatrics* **46**, 1–10.

Internet Resources

1KG Phase3: https://phase3browser.1000genomes.org/index.html

ABraOM: http://abraom.ib.usp.br/

MED13L on OMIM: https://omim.org/entry/608771

gnomAD: https://gnomad.broadinstitute.org/

Accepted 19 September 2021

© 2021 MENCAP and International Association of the Scientific Study of Intellectual and Developmental Disabilities and John Wiley & Sons Ltd

Capítulo 2.4

SCAF4-related syndromic intellectual disability

CARVALHO, L. M. L; PINTO, C. F.; SCLIAR, M. O.; OTTO, P. A.; KREPISCHI, A. C. V.; ROSENBERG, C. Case Report: *SCAF4*-related syndromic intellectual disability. **American Journal of Medical Genetics Part A**, 2022. ISSN: 1552-4833. DOI: 10.1002/ajmg.a.63032

Received: 24 April 2022 | Revised: 13 August 2022 | Accepted: 13 October 2022

DOI: 10.1002/ajmg.a.63032

CASE REPORT

AMERICAN JOURNAL OF medical genetics PART A WILEY

SCAF4-related syndromic intellectual disability

Laura Machado Lara Carvalho[1] | Carla Franchi Pinto[2] | Marília de Oliveira Scliar[1] | Paulo A. Otto[1] | Ana Cristina Victorino Krepischi[1] | Carla Rosenberg[1]

[1]Human Genome and Stem Cell Research Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo (USP), Sao Paulo, Sao Paulo, Brazil

[2]Department of Pathological Sciences, Faculty of Medical Sciences of Santa Casa de São Paulo, Sao Paulo, Sao Paulo, Brazil

Correspondence
Carla Rosenberg, Department of Genetics and Evolutionary Biology, Human Genetics Research Laboratory – LIGH, Institute of Biosciences, University of Sao Paulo, Matão Street 277 – Room 350, Sao Paulo 05508-090, SP, Brazil.
Email: carlarosenberg@ib.usp.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 305806/2019-0; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 1805008; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Numbers: 2018/08486-3, 2013/08028-1

Abstract

The causal link between variants in the *SCAF4* gene and a syndromic form of intellectual disability (ID) was established in 2020 by Fliedner et al. Since then, no additional cases have been reported. We performed exome sequencing in a 16-year-old Brazilian male presenting with ID, epilepsy, behavioral problems, speech impairment, facial dysmorphisms, heart malformations, and obesity. A de novo pathogenic variant [*SCAF4*(NM_020706.2):c.374_375dup(p.Glu126LeufsTer20)] was identified. This is the second study reporting the involvement of *SCAF4* in syndromic ID, and the description of the patient's clinical features contributes to defining the phenotypic spectrum of this recently described Mendelian disorder.

KEYWORDS
epilepsy, exome sequencing, human genetics, intellectual disability, obesity, *SCAF4*

1 | INTRODUCTION

Recently, Fliedner et al. (2020) published a study describing heterozygous loss-of-function (LoF) variants in the *SCAF4* gene associated with a novel autosomal dominant neurodevelopmental disorder characterized by intellectual disability (ID), seizures, behavioral abnormalities, facial dysmorphisms, and skeletal anomalies. Nine LoF *SCAF4* variants classified as likely pathogenic were reported, eight of which were de novo and one inherited. An additional LoF variant present in two siblings with neurodevelopmental phenotypes was classified as a variant of unknown significance (VUS) because it mapped to the C-terminal part of the protein. A missense variant was also described by Fliedner et al. as a VUS.

Here, we report a de novo frameshift *SCAF4* variant mapped to exon 5/20 in a Brazilian patient with clinical features overlapping those described by Fliedner et al. (2020). This is the second publication supporting the involvement of *SCAF4* in a new Mendelian neurodevelopmental disease.

2 | CASE REPORT

The patient was first referred to the Institute of Biosciences of the University of Sao Paulo for genetic evaluation at the age of 4.4 years due to presentation of microcephaly and speech delay. He was born at term gestation from consanguineous parents (double first cousins) and exhibited feeding difficulties in early infancy, apparently without a significant degree of neonatal hypotonia. His motor development progressed without significant delay (seated and walking without support at the ages of 9 months and 1.5 years, respectively). His physical examination confirmed the presence of microcephaly (occipitofrontal circumference 45.5 cm, −3.43 *SD*) and slightly long protruding ears with a moderate degree of cartilage hypoplasia, epicanthus, slight convergent strabismus, bulbous nasal tip, high palate, joint hyperextensibility, pes planus, and obesity (20 kg, 1.03 m—BMI *z*-score 2.38); the patient did not exhibit brachydactyly (Figure 1a). Magnetic resonance imaging and cranial tomography were performed, both with

 © 2022 Wiley Periodicals LLC.

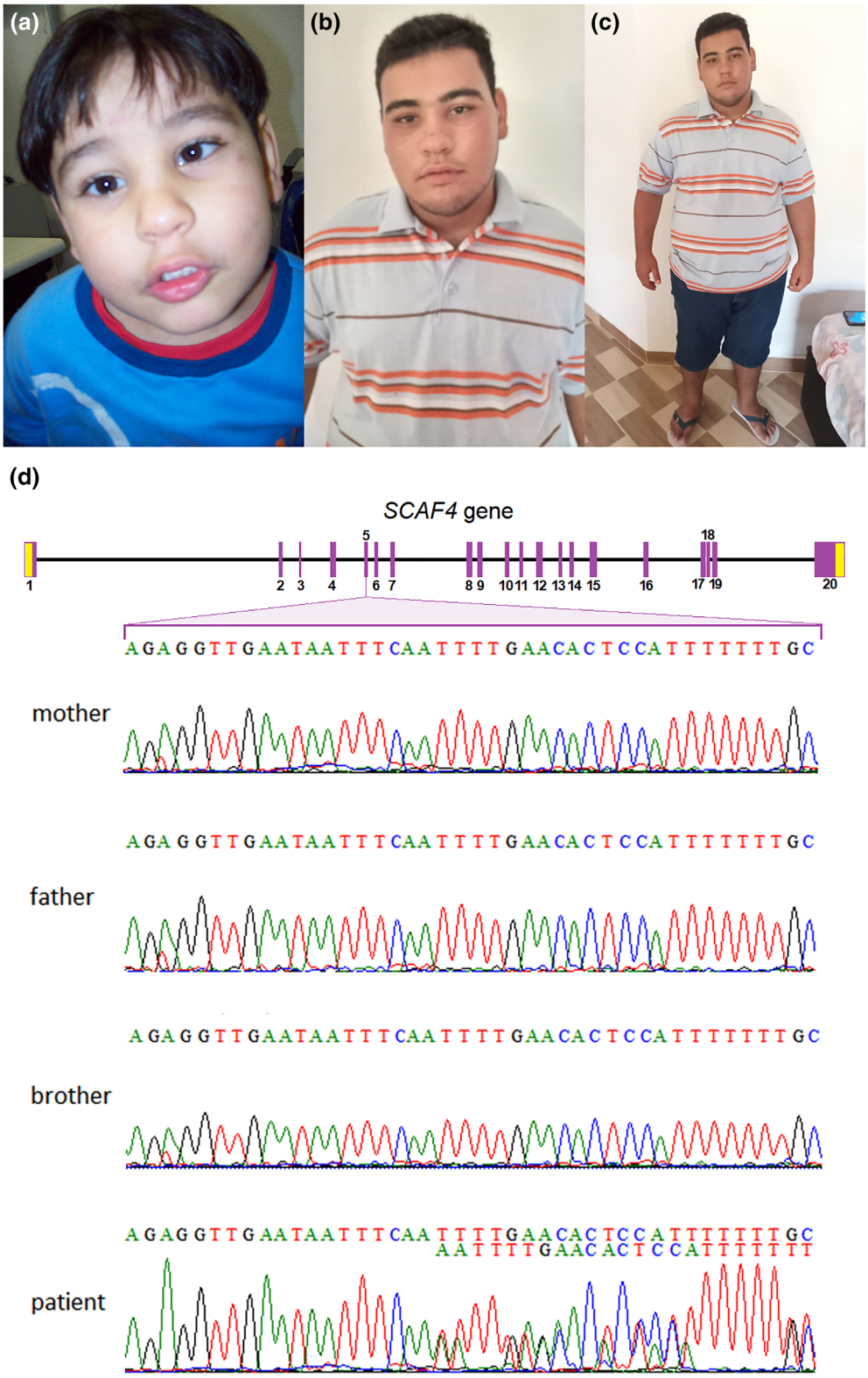


FIGURE 1 Photographs and *SCAF4* sequencing results of the patient. (a) Patient at 4.4 years old. (b, c) Patient at 16.6 years old. Note the presence of slightly long protruding ears, epicanthus, convergent strabismus, prominent nasal bridge, bulbous nasal tip, long philtrum and obesity (BMI z-score: 2.27). (d) In the top of the image, a schematic representation of *SCAF4* shows the localization of the patient's pathogenic variant at exon 5 of the gene. Underneath the *SCAF4* (NM_020706.2):c.374_375dup variant segregation analysis by Sanger sequencing, showing the duplication of two nucleotides (AA) in the patient's sample and the absence is noted in the other family members' samples.

normal results. The patient has no history of kidney problems, although renal ultrasound was not performed. Brazil has a high degree of ethnic admixture, but the patient is phenotypically Caucasian. His family has no previous history of individuals with ID or syndromic characteristics.

The family reported that at 6 years of age, he underwent cardiac surgery to correct a congenital atrial communication. Currently, he has mild mitral and tricuspid regurgitation. At the age of 8 years, he presented with epileptic seizures that are currently well-controlled with anticonvulsants.

TABLE 1 Clinical characteristics of the patients with LoF *SCAF4* variants

	Fliedner et al. (2020) patients[a]									
	1	2	3	5	6	7	8	9	This publication	Frequency
Genotype	c.321+1G>T	c.453_456delTGAA	c.1028delC	c.1423C>T	c.1614+1G>C	c.1649dupT	c.1812G>A	c.1889G>A	c.374_375dup	
DD/ID	+	+	+	+	+	+	+	+	+	9/9
Speech impairment	+	n/a	+	+	+	n/a	+	+	+	7/7
Behavioral problems	+	n/a	+	+	+	−	+	+	+	7/8
Epilepsy	+	−	−	+	−	+	+	−	+	5/9
MRI anomalies	+	n/a	n/a	−	n/a	+	+	+	−	4/6
Muscular hypotonia	+	+	−	−	n/a	n/a	Hypertonia	+	n/a	3/6
Heart malformations	n/a	−	+	+	n/a	n/a	n/a	+	+	4/5
Microcephaly	−	−	n/a	−	−	n/a	−	−	+ (in infancy)	1/7
Obesity[b] (BMI *z*-score)	− (+1.04)	− (+1.28)	− (+1.43)	− (−0.46)	− (+0.19)	+ (+2.86)	− (+1.44)	n/a	+ (+2.38; +2.27)	2/8
Renal anomalies	n/a	n/a	+	+	−	n/a	n/a	+	n/a	3/4
Urogenital anomalies	−	−	+	−	−	n/a	n/a	+	n/a	2/6
Gastrointestinal anomalies	−	−	+	−	−	n/a	n/a	−	n/a	1/5
Skeletal anomalies	−	−	+	+	−	−	+	−	n/a	3/8

Abbreviations: DD, developmental delay; ID, intellectual disability; LoF, loss-of-function; MRI, magnetic resonance imaging; n/a, not available or not applicable.

[a]We only considered Patients 1–3 and 5–9; the variants of Patients 10 and 11 were classified as VUS, and Patient 4 also had an *NSD1* pathogenic variant.

[b]Obesity was considered when the BMI *z*-score was greater than +2.

Further clinical evaluation was performed at 16.6 years, in which facial dysmorphisms (convergent strabismus, epicanthus, prominent nasal bridge, bulbous nasal tip, long philtrum, and long ears—Figure 1b) and obesity (87 kg, 1.7 m—BMI *z*-score 2.27—Figure 1c) were observed. Notably, his father and brother also presented obesity. Microcephaly was no longer present (occipitofrontal circumference 54 cm, −0.74 *SD*).

The patient exhibited ID associated with speech impairment and behavioral problems, including anxiety and aggressiveness, particularly related to feeding. The following genetic tests were performed previously, all with normal results: G-banding karyotype, chromosome microarray analyses (CMA—Agilent 180K), and PCR for Fragile-X syndrome investigation (Haddad et al., 1996).

Genomic libraries for exome sequencing were produced from the proband's blood DNA using SureSelect QXT Human All Exon V6 technology (Agilent). The raw paired-end reads (Fastq) were aligned with the reference genome GRCh38 (hg38) using the Burrows–Wheeler Aligner (BWA; Li & Durbin, 2009) and processed with GATK tools (Depristo et al., 2011). The annotation was performed using Varseq software (Golden Helix), and the filtering involved the following criteria: read depth, population frequency—1KG Phase3 (The 1000 Genomes Project Consortium, 2015), gnomAD (Cummings et al., 2020), and ABraOM (Naslavsky et al., 2022)-, clinical data—ClinVar 2018 (Landrum et al., 2018)-, OMIM (Amberger et al., 2015), functional prediction—Ensembl (Hunt et al., 2018)-, and pathogenicity prediction algorithms. Filtered variants were classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015) using Varsome with manual review (Kopanos et al., 2019).

Segregation analyses were performed by Sanger sequencing with genomic DNA samples extracted from saliva from the patient's parents and unaffected brother.

3 | RESULTS

We identified a pathogenic variant in the *SCAF4* gene [*SCAF4* (NM_020706.2):c.374_375dup(p.Glu126LeufsTer20)] of a patient with syndromic ID. Figure 1d shows the presence of the variant in the proband and the absence in his parents and unaffected brother. The variant, which leads to a frameshift and consequent premature stop codon, was classified as pathogenic (ACMG criteria: PVS1, PM2, and PM6). Other clinically relevant variants were not identified.

4 | DISCUSSION

There is a significant overrepresentation of epigenetic machinery-encoding genes in ID etiology in general (Scandaglia & Barco, 2019), with *SCAF4* falling into this category. *SCAF4* is an exclusively nuclear anti-terminator mRNA protein associated with RNA polymerase II; it inhibits the use of early alternative polyA sites to diminish the production of nonfunctional mRNAs and proteins (Gregersen et al., 2019). An RNAseq assay performed by Fliedner et al. (2020) using blood lymphocytes of *SCAF4*-deficient individuals revealed a broad deregulation of more than 9000 genes and a significant differential splicing of more than 2900 genes.

Our patient presents a partial clinical overlap with those described by Fliedner et al. (2020) carrying likely pathogenic *SCAF4* variants (Table 1). Variable levels of cognitive and speech impairment were present in all cases. Other features described in some patients reported by Fliedner et al. (2020), including epilepsy and heart malformations, were also present in this patient. Behavioral problems were frequent, and our patient exhibited anxiety and aggressiveness. Notably, we report the first observations of microcephaly (in early development) associated with *SCAF4* LoF variant.

Patients 7 and 11 from the study of Fliedner et al. (2020) also had obesity, but Patient 11 carried a *SCAF4* VUS. Obesity is a common feature and has a high prevalence among people with cognitive impairment (Carvalho et al., 2022; Emerson et al., 2016; Hsieh et al., 2014; Melville et al., 2008). There are many factors that may contribute to this higher prevalence in people with ID, such as low levels of physical activity, poor nutritional habits (Melville et al., 2017), and the use of psychiatric drugs (de Winter et al., 2012). In particular, our patient does not exercise, is very anxious regarding food and uses some medications that can lead to weight gain (currently: chlorpromazine, periciazine, and clobazam; in previous years: risperidone). In addition, he has a family history of obesity, which is a condition with high heritability (Sandholt et al., 2012).

This publication substantiates the role of *SCAF4* LoF variants as a cause of syndromic ID and contributes to broadening the associated clinical spectrum.

AUTHOR CONTRIBUTIONS

Laura Machado Lara Carvalho performed NGS data analysis, Sanger sequencing, and the interpretation of genetic variants, was responsible for the initial writing of the manuscript and produced the figures under the guidance of Carla Rosenberg and Ana Cristina Victorino Krepischi, who conceived and designed the study. Carla Franchi Pinto and Paulo A. Otto clinically evaluated the patient and contributed to the interpretation of the results. Marília de Oliveira Scliar contributed to bioinformatics processing (alignment to the reference genome and calling of variants). All authors have read, edited, and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are grateful to the patient's family for collaborating in this study. The authors also thank Dr. Peter Pearson for kindly reviewing the manuscript to improve the English and Maraisa de Castro Sebastião for technical support. This study was supported by the São Paulo Research Foundation—FAPESP (2018/08486-3 and 2013/08028-1), the National Council for Scientific and Technological Development—CNPq (305806/2019-0), and the Coordination for the Improvement of Higher Education Personnel—CAPES (1805008).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ClinVar at https://www.ncbi.nlm.nih.gov/clinvar/, reference number accession VCV001334441.1.

ORCID

Laura Machado Lara Carvalho https://orcid.org/0000-0001-5599-2236

Ana Cristina Victorino Krepischi https://orcid.org/0000-0003-2931-8605

REFERENCES

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*, *43*(D1), D789–D798. https://doi.org/10.1093/nar/gku1205

Carvalho, L. M. L., D'Angelo, C. S., Villela, D., da Costa, S. S., de Lima Jorge, A. A., da Silva, I. T., de Oliveira Scliar, M., Chaves, L. D., Krepischi, A. C. V., Koiffmann, C. P., & Rosenberg, C. (2022). Genetic investigation of syndromic forms of obesity. *International Journal of Obesity*, *46*, 1582–1586. https://doi.org/10.1038/s41366-022-01149-5

Cummings, B. B., Karczewski, K. J., Kosmicki, J. A., Seaby, E. G., Watts, N. A., Singer-Berk, M., Mudge, J. M., Karjalainen, J., Satterstrom, F. K., O'Donnell-Luria, A. H., Poterba, T., Seed, C., Solomonson, M., Alföldi, J., Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Daly, M. J., & MacArthur, D. G. (2020). Transcript expression-aware annotation improves rare variant interpretation. *Nature*, *581*(7809), 452–458. https://doi.org/10.1038/s41586-020-2329-2

de Winter, C. F., Bastiaanse, L. P., Hilgenkamp, T. I. M., Evenhuis, H. M., & Echteld, M. A. (2012). Overweight and obesity in older people with intellectual disability. *Research in Developmental Disabilities*, *33*(2), 398–405. https://doi.org/10.1016/j.ridd.2011.09.022

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–501. https://doi.org/10.1038/ng.806

Emerson, E., Robertson, J., Baines, S., & Hatton, C. (2016). Obesity in British children with and without intellectual disability: Cohort study. *BMC Public Health*, *16*(1), 644. https://doi.org/10.1186/s12889-016-3309-1

Fliedner, A., Kirchner, P., Wiesener, A., van de Beek, I., Waisfisz, Q., van Haelst, M., Scott, D. A., Lalani, S. R., Rosenfeld, J. A., Azamian, M. S., Xia, F., Dutra-Clarke, M., Martinez-Agosto, J. A., Lee, H., Noh, G. J., Lippa, N., Alkelai, A., Aggarwal, V., Agre, K. E., Agre, K. E., Gavrilova, R., Mirzaa, G. M., Straussberg, R., Cohen, R., Horist, B., Krishnamurthy, V., McWalter, K., Juusola, J., Davis-Keppen, L., Ohden, L. van Slegtenhorst, M.., Man S. S., B. Ekici, A. G., van de Laar, I., Zweier, C. (2020). Variants in SCAF4 cause a neurodevelopmental disorder and are associated with impaired mRNA processing. *The American Journal of Human Genetics*, *107*(3), 544–554. https://doi.org/10.1016/j.ajhg.2020.06.019

Gregersen, L. H., Mitter, R., Ugalde, A. P., Nojima, T., Proudfoot, N. J., Agami, R., Stewart, A., & Svejstrup, J. Q. (2019). SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell*, *177*(7), 1797–1813.e18. https://doi.org/10.1016/j.cell.2019.04.038

Haddad, L. A., Mingroni-Netto, R. C., Vianna-Morgante, A. M., & Pena, S. D. J. (1996). A PCR-based test suitable for screening for fragile X syndrome among mentally retarded males. *Human Genetics*, *97*(6), 808–812. https://doi.org/10.1007/BF02346194

Hsieh, K., Rimmer, J. H., & Heller, T. (2014). Obesity and associated factors in adults with intellectual disability. *Journal of Intellectual Disability Research*, *58*(9), 851–863. https://doi.org/10.1111/jir.12100

Hunt, S. E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I. M., Trevanion, S. J., Flicek, P., & Cunningham, F. (2018). Ensembl variation resources. *Database*, *2018*, bay119. https://doi.org/10.1093/database/bay119

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2019). VarSome: The human genomic variant search engine. *Bioinformatics*, *35*(11), 1978–1980. https://doi.org/10.1093/bioinformatics/bty897

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., … Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Melville, C. A., Cooper, S.-A., Morrison, J., Allan, L., Smiley, E., & Williamson, A. (2008). The prevalence and determinants of obesity in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *21*(5), 425–437. https://doi.org/10.1111/j.1468-3148.2007.00412.x

Melville, C. A., Oppewal, A., Schäfer Elinder, L., Freiberger, E., Guerra-Balic, M., Hilgenkamp, T. I. M., Einarsson, I., Izquierdo-Gómez, R. H., Sansano-Nadal, O., Rintala, P., Cuesta-Vargas, A., & Giné-Garriga, M. (2017). Definitions, measurement and prevalence of sedentary behaviour in adults with intellectual disabilities—A systematic review. *Preventive Medicine*, *97*, 62–71. https://doi.org/10.1016/j.ypmed.2016.12.052

Naslavsky, M. S., Scliar, M. O., Yamamoto, G. L., Wang, J. Y. T., Zverinova, S., Karp, T., Nunes, K., Ceroni, J. R. M., de Carvalho, D. L., da Silva Simões, C. E., Bozoklian, D., Nonaka, R., dos Santos Brito Silva, N., da Silva Souza, A., de Souza Andrade, H., Passos, M. R. S., Castro, C. F. B., Mendes-Junior, C. T., Mercuri, R. L. V., … Zatz, M. (2022). Whole-genome sequencing of 1,171 elderly admixed individuals from São Paulo, Brazil. *Nature Communications*, *13*(1), 1004. https://doi.org/10.1038/s41467-022-28648-3

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–423. https://doi.org/10.1038/gim.2015.30

Sandholt, C. H., Hansen, T., & Pedersen, O. (2012). Beyond the fourth wave of genome-wide obesity association studies. *Nutrition & Diabetes*, *2*(7), e37. https://doi.org/10.1038/nutd.2012.9

Scandaglia, M., & Barco, A. (2019). Contribution of spurious transcription to intellectual disability disorders. *Journal of Medical Genetics*, *56*(8), 491–498. https://doi.org/10.1136/jmedgenet-2018-105668

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

How to cite this article: Carvalho, L. M. L., Pinto, C. F., de Oliveira Scliar, M., Otto, P. A., Krepischi, A. C. V., & Rosenberg, C. (2022). *SCAF4*-related syndromic intellectual disability. *American Journal of Medical Genetics Part A*, 1–5. https://doi.org/10.1002/ajmg.a.63032

Capítulo 3

Establishment of iPSC lines and zebrafish with loss-of-function AHDC1 variants: models for Xia-Gibbs syndrome

---

Artigo submetido à revista *Gene.* Descreve o desenvolvimento dos primeiros modelos biológicos para estudo da síndrome de Xia-Gibbs, que ainda tem mecanismos fisiopatológicos pouco compreendidos.

Nosso primeiro paciente com a síndrome de Xia-Gibbs teve seu diagnóstico molecular a partir da investigação genética da casuística de obesidade sindrômica apresentada no **Capítulo 1.1.** Posteriormente, recrutamos mais dois pacientes com essa condição. Dois desses três pacientes têm obesidade. O paciente que não tem obesidade apresenta uma importante ansiedade por comida, segundo a genitora, sendo necessário controlar sua alimentação.

Não há na literatura científica referências de que hiperfagia e obesidade poderiam ser características da síndrome de Xia-Gibss, porém essas e outras observações nos levaram à hipótese de que variantes em *AHDC1* poderiam interferir em vias neurais de fome e saciedade, podendo causar quadros de obesidade sindrômica, conforme apresentado na **Introdução Geral**.

Establishment of iPSC lines and zebrafish with loss-of-function *AHDC1* variants*:* models for Xia-Gibbs syndrome

running title: **Xia-Gibbs syndrome models**

Authors: Laura Machado Lara Carvalho[1], Elisa Varella Branco[1], Raquel Delgado Sarafian[2], Gerson Shigeru Kobayashi[1], Fabiano Tófoli de Araújo[2], Lucas Santos Souza[1], Danielle de Paula Moreira[1], Gabriella Shih Ping Hsia[1], Eny Maria Goloni Bertollo[3], Cecília Barbosa Buck[4], Silvia Souza da Costa[1], Davi Mendes Fialho[1], Felipe Tadeu Galante Rocha de Vasconcelos[1], Luciano Abreu Brito[1], Luciana Elena de Souza Fraga Machado[1], Igor Cabreira Ramos[1], Lygia da Veiga Pereira[2], Celia Priszkulnik Koiffmann[1], Maria Rita dos Santos e Passos-Bueno[1], Tiago Antonio de Oliveira Mendes[5], Ana Cristina Victorino Krepischi[1], Carla Rosenberg[1]*

Affiliations:

[1] Human Genome and Stem Cell Research Center. Institute of Biosciences - University of São Paulo. São Paulo-SP, Brazil;

[2] National Embryonic Stem Cell Laboratory Institute of Biosciences - University of São Paulo. São Paulo-SP, Brazil;

[3] Faculty of Medicine of São José do Rio Preto. São José do Rio Preto-SP, Brazil;

[4] Anhembi Morumbi University. Piracibaba-SP, Brazil;

[5] Department of Biochemistry and Molecular Biology. Viçosa Federal University. Viçosa-MG, Brazil.

*Correspondence: Carla Rosenberg, Department of Genetics and Evolutionary Biology, Laboratory of Human Genetics – LGH, Institute of Biosciences, University of Sao Paulo, Matão Street 277 – Room 350, São Paulo, SP, Brazil. Zip code: 05508-090. Email: carlarosenberg@ib.usp.br

Abstract:

Xia-Gibbs syndrome (XGS) is a syndromic form of intellectual disability caused by heterozygous *AHDC1* variants, but the pathophysiological mechanisms underlying this syndrome are still unclear. In this manuscript, we describe the development of two different functional models: three induced pluripotent stem cell (iPSC) lines with different loss-of-function (LoF) *AHDC1* variants, derived by reprogramming peripheral blood mononuclear cells from XGS patients, and a zebrafish strain with a LoF variant in the ortholog gene (*ahdc1*) obtained through CRISPR/Cas9-mediated editing. The three iPSC lines showed expression of pluripotency factors (*SOX2, SSEA-4, OCT3/4*, and *NANOG*). To verify the capacity of iPSC to differentiate into the three germ layers, we obtained embryoid bodies (EBs), induced their differentiation, and confirmed the mRNA expression of ectodermal, mesodermal, and endodermal markers using the TaqMan hPSC Scorecard. The iPSC lines were also approved for the following quality tests: chromosomal microarray analysis (CMA), mycoplasma testing, and short tandem repeat (STR) DNA profiling. The zebrafish model has an insertion of four base pairs in the *ahdc1* gene, is fertile, and breeding between heterozygous and wild-type (WT) animals generated offspring in a genotypic proportion in agreement with Mendelian law. The established iPSC and zebrafish lines were deposited on the hpscreg.eu and zfin.org platforms, respectively. These biological models are the first for XGS and will be used in future studies that investigate the pathophysiology of this syndrome, unraveling its underlying molecular mechanisms.

Keywords: Xia-Gibbs syndrome, *AHDC1*, neurodevelopment, iPSC, zebrafish, CRISPR–Cas9.

INTRODUCTION

Xia-Gibbs syndrome (XGS – OMIM # 615829) is a rare autosomal dominant form of syndromic intellectual disability (ID) that was first described by Xia et al. (2014), who identified four probands with *de novo AHDC1* loss-of-function (LoF) variants. Furthermore, another study (Yang et al., 2015) identified LoF *AHDC1* variants by whole-exome sequencing (WES) in another seven patients from a cohort of 2157 ID/developmental delay individuals (0.32%). The Xia-Gibbs society, a parents' organization, reports on its website (https://xia-gibbs.org/- accessed in November/2022) that it had contact with 390 families of patients.

The XGS patients' characteristics frequently reported are ID, hypotonia, motor development delay, speech delay (severely limited/absent speech), structural brain anomalies (such as corpus callosum and myelination changes), and nonspecific facial dysmorphisms. Other features may include sleep apnea, movement disorders (ataxia, tremors, and bradykinesias), short stature (because of growth hormone deficiency), behavioral disturbance, and autism (Goyal et al., 2020). There is phenotypic variability even among patients with the same mutation. Although the occurrence of most features could not be directly related to the causal mutation, it was observed that seizures and scoliosis are more frequently associated with N-terminal AHDC1 variants (arising in the upstream half of the gene) (Khayat et al., 2021).

The gibbin protein (name recently proposed by Collier et al., 2022, to the AHDC1 protein) has two AT-hook motifs (see **Supplementary Document 1** for a schematic representation). Most eukaryotic AT-hook proteins have chromatin regulatory functions, but there is also evidence that these proteins can bind RNA (Filarsky et al., 2015). There is also evidence that gibbin can interact with nuclear proteins (Collier et al., 2022; Shalaby et al., 2011; Vandamme et al., 2011), which reinforces role as an epigenetic regulator. Nonetheless, the mechanisms by which *AHDC1* variants cause XGS phenotypes are still unknown, and the development of biological models to investigate the pathophysiological mechanisms of XGS are needed. Then, the aim of this work was to establish cellular and animal models to study this syndrome.

METHODS

Establishment and characterization of XGS iPSC lines

Induced pluripotent stem cell (iPSC) lines were obtained from reprogramming peripheral blood mononuclear cells (PBMCs) of three unrelated XGS patients (P1, P2, and P3) through nucleofection with nonintegrating plasmids, which led to the expression of pluripotency factors. Additionally, we developed an iPSC line derived from cells of a nonaffected individual (C1) to be used as a control in future functional studies. Two other iPSC lines (C2 and C3), both already characterized and published (Moreira et al., 2022; Sarafian et al., 2021), can be used as controls. All cell lines were obtained from males.

A workflow regarding the establishment of the three XGS iPSC lines and one control (C1) is presented in **Figure 1A**. The methodology is detailed below.

Subjects

Patients P1, P2, and P3 were molecularly diagnosed with XGS by identifying *AHDC1* LoF variants by WES. They have been referred for WES for presenting with syndromic ID. P1 is part of a syndromic obesity cohort previously published by us (Carvalho et al., 2022, patient 5). For P1 and P3, the next-generation sequencing (NGS) methodology used was SureSelect Human All Exon V6 (Agilent Technologies). For P2, Illumina NextSeq (Illumina) was used. Considering the GRCh38(hg38) reference genome, the causal variants identified are presented below:

P1 - Chr1:27551122-G-A/ / *AHDC1*(NM_001371928.1)c.994C>T(p.Gln332Ter)
P2 - Chr1:27550463-A- / *AHDC1*(NM_001371928.1):c.1653del(p.Lys552ArgfsTer33)
P3 – Chr1:27550634-TT- / *AHDC1*(NM_001371928.1):c.1481_1482del(p. Lys494SerfsTer22)

The positions of the patients' mutations in the gibbin protein are presented in **Supplementary Document 1**.

Sanger sequencing of DNAs from the trios (proband, mother, father) did not detect variants in their parents, showing that *AHDC1* variants were *de novo*. The primers used in this sequencing are shown in **Supplementary Table 1**. Such variants were classified as pathogenic according to the guidelines of the American College of Medical Genetics and Genomics (Richards et al., 2015).

We collected peripheral blood from P1, P2, P3, and C1 and separated their PBMCs using Leucosep tubes and a Ficoll density gradient, according to the protocol described in the **Supplementary Document 2**. PBMCs were used in cell reprogramming.

Reprogramming of PBMCs

Erythroblasts were expanded from PBMCs in StemSpan medium containing stem cell factor (SCF - 50 ng/mL), erythropoietin (EPO - 2 U/mL), dexamethasone (1 µM), insulin growth factor-1 (IGF1 - 40 ng/mL), interleukin-3 (IL3 -10 ng/mL), and gentamicin (10 µg/mL). At this stage, the medium was changed every other day, but daily mixing “up and down” with the micropipette was performed to avoid the formation of lumps that could induce differentiation. Reprogramming was performed on 1-1.2x10^6 cells.

The cells were reprogrammed using plasmids described by Okita et al. (2011) (pCXLE-hOCT3/4-shp53, pCXLE-hSK, and pCXLE-hUL – Addgene #27077, #27078, and #27080, respectively - **Figure 1B**) containing the transcription factors *OCT3/4*, *SOX2*, *KLF4*, *L-MYC*, and *LIN28*, in addition to a shRNA for *p53* suppression. Nucleofection was performed using the Amaxa Human CD34+ Cell Nucleofection Kit and the Amaxa Nucleofector Instrument (Lonza).

Reprogrammed cells were cocultivated with irradiated murine embryonic fibroblasts (MEFs, Millipore) in DMEM/F12 supplemented with 1% GlutaMAX, 1% nonessential amino acids, 50 µM 2-mercaptoethanol, 20% knockout serum replacement, and 100 µg/mL normocin. MEF is a limited cell line that simulates an embryonic microenvironment.

iPSC colonies with typical morphology (clearly defined borders, uniform structure, adherent to the plate, and clustered small cells) were then transferred to Matrigel-coated plates,

fed E8 medium containing Normocin (1:500), and cultivated at 37 °C and 5% CO_2 for five to eight passages for characterization tests. All lines were clonally derived by manual selection of single colonies.

Evaluation of pluripotency status

The morphology of the obtained cell lines was evaluated using an EVOS XL microscope (Life Technologies).

We evaluated pluripotency markers expression using qualitative (Immunocytochemistry - ICC - for OCT3/4, SSEA-4, and SOX2) and quantitative (RTqPCR for *OCT3/4* and *NANOG*) analyses. Immunostaining protocol is described in **Supplementary Document 3**. The ICC images were obtained with a confocal microscope (LSM 800, Zeiss).

RNA was extracted from PBMCs and iPSC lines using RNeasy Plus Mini Kit (Qiagen), and the synthesis of first-strand cDNA was performed using SuperScript Vilo Mastermix (Invitrogen). RTqPCR compared reprogrammed cells with PBMCs and was performed with technical triplicates using Fast SYBR Green Master Mix and QuantStudio 5 (both from Applied Biosystems). The primers used in these reactions (**Supplementary Table 1)** presented efficiency close to 100% (±10%). The $2^{-\Delta\Delta \boldsymbol{CT}}$ method was used as a quantification strategy relative to housekeeping genes (*TBP* and *HBMS*).

Evaluation of differentiation capacity

iPSC lines differentiation into the three germ lineages was evaluated by *in vitro* embryoid bodies (EBs) formation and differentiation (as described by Tofoli et al., 2016 with posterior adaptations) followed by TaqMan Scorecard.

The *in vitro* EBs formation took place in five days of suspension of square clippings from iPSC colonies that had previously reached more than 75% confluence. These iPSC colonies clippings were placed on plates covered with 1% agarose film with antibiotic-free E8 medium. During the five days of EBs formation, there was no change of medium.

On the 6th day, the EBs were transferred to plates treated with gelatin (0.1%), and from that day onward, a differentiation medium - DMEM enriched with FBS ES cell qualified (20%), GlutaMax (2 mM), Pen/Strep (100 U/mL), Antibiotic Antimycotic Solution - Sigma Aldrich (1X), nonessential amino acids (2 mM), HEPES (10 mM), sodium pyruvate (200 mM), and β-mercaptoethanol (0.1 mM) – was used and changed every other day until the 17th day.

EBs photographs were obtained on Days 6 (culture of EBs in suspension) and 17 (spontaneously differentiated cells adhered to the gelatin). At Day 17, EBs were harvested with TRIzol, and RNA extraction was performed using RNeasy Plus Mini Kit (Qiagen). The synthesis of cDNA was performed using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems). RTqPCR was performed using the TaqMan hPSC Scorecard kit and cDNAs of EBs at Day 17, which were compared with cDNA of iPSC lines. The analysis and representation of the results were performed using the hPSC Scorecard Analysis software (Applied Biosystems).

Quality control of iPSC lines

All iPSC lines were tested for plasmid integration into the genome by conventional PCR applied to the *OriP* and *EBNA-1* loci (**Figure 1B**) with the primers presented in **Supplementary Table 1**. Copy number alterations were investigated by chromosome microarray analyses (CMA - 60K - Agilent Technologies). We also verified by Sanger sequencing of iPSC lines' DNAs if the patients LoF variants were present.

A short tandem repeat (STR) DNA profiling assay was performed using AmpFLSTR Identifiler kit (Applied Biosystems), which evaluates amelogenin + 15 *loci* (*D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818,* and *FGA*) for the detection of eventual cellular cross-contamination. STR data were analyzed with GeneMapper software (Applied Biosystems) to compare each iPSC line profile with the respective PBMCs.

A mycoplasma test was performed with the MycoScope PCR Detection Kit (Genlantis).

Establishing the zebrafish XGS model

For gene editing, we used zygotes of the AB zebrafish strain. To generate a new zebrafish (*Danio rerio*) strain with an *ahdc1* LoF variant, we used CRISPR–Cas9 editing without a DNA template donor, which leads to nonhomologous end joining repair, i.e., with the production of random indels (**Figure 2**). The methodology is described below.

sgRNA synthesis and its efficiency evaluation

The sgRNA for *ahdc1* editing was designed using CRISPRscan online software (https://www.crisprscan.org/- Moreno-Mateos et al., 2015). We selected a sgRNA directed to an initial portion of the coding region (PAM at position 589-591 of 4191 on coding sequence considering the XM_005159774.4 transcript), with a good efficiency score (97) and low probabilities of occurrence of off targets according to the cutting frequency determination(CFD) values given by CRISPRscan.

For the *in vitro* sgRNA synthesis, we followed the protocol proposed by Vejnar et al. (2016) with some adaptations. The first step of this protocol is annealing and extension of oligonucleotides under PCR conditions to generate a DNA template. Then, two oligonucleotides with sequences described in **Supplementary Table 1** were used: the first oligonucleotide contains the portion that gives rise to the sgRNA scaffold, and the second oligonucleotide contains a promoter for T7 RNA polymerase (necessary for *in vitro* transcription), the portion that gives rise to the target sgRNA region and a complementary portion to the first oligonucleotide. The structure of these oligonucleotides is schematically explained in **Figure 2**.

We applied the PCR product to a 2% agarose gel and observed a band of 117 bp. The PCR product was purified using the Easy Pure PCR Purification Kit (Transgen Biotech). Then, we performed *in vitro* transcription using the MEGAshortscript T7 Transcription Kit (Invitrogen) according to the manufacturer's instructions with the following specificities:

approximately 160 ng of purified template DNA, incubation for transcription for 7 h 30 min, and incubation with TURBO DNase for 20 min.

As shown in **Figure 2A**, the beginning of the target portion (GG) does not have perfect complementarity with the *ahdc1* gene, which does not significantly interfere with the efficiency of Cas9 cleavage. On the other hand, the *in vitro* transcription kit has greater efficiency for sequences that start with GG.

After *in vitro* transcription, we performed sgRNA precipitation according to the protocol of Vejnar et al. (2016) but with resuspension in 50 µL of nuclease-free water and fluorometric quantification using the Qubit® RNA HS kit (Thermo Fisher Scientific).

Before performing injections into zebrafish embryos for *ahdc1* editing, we evaluated the quality of the sgRNA produced. The sgRNA integrity was verified by 1.4% agarose gel electrophoresis. We also performed an *in vitro* sgRNA efficiency test that evaluated the ability of Cas9 to cut the target sequence being guided by the synthesized sgRNA. For this step, we performed PCR using the sequence primers presented in **Supplementary Table 1** and DNA from WT zebrafish. The PCR product was subjected to this sgRNA efficiency test, as follows: 120 ng of PCR product, 35 ng of sgRNA, 600 ng of Cas9 protein (Sigma–Aldrich), 1 µL buffer (200 mM Hepes, 1 M NaCl, 50 nM $MgCl_2$, 1 mM EDTA - pH 6.5), 10 µg of BSA (BioLabs), nuclease-free water in volume needed to complete 10 µL, and incubation at 37° for one hour followed by 80° for 5 min. The reaction product was subjected to electrophoresis in a 13% polyacrylamide gel stained with silver nitrate. This assessment of sgRNA quality is important to avoid misuse of animals in research. Only after verifying that the PCR product (441 bp) was cut into 245 bp and 196 bp fragments (**Supplementary Figure 2A**) did we proceed to gene editing involving animals.

Editing control using sgRNA targeting the *slc45a2* gene

Before *ahdc1* editing in the animals, we performed an assay targeting the *slc45a2* gene (involved in pigmentation) to evaluate the efficiency of the system. *slc45a2* editing is a fast protocol, and a lack of pigmentation is easily visually verifiable. For the synthesis of the sgRNA targeting *slc45a2*, we used the oligonucleotides sequences indicated by Vejnar et al. (2016) (**Supplementary Table 1)** and the same protocol used for the synthesis of sgRNA targeting *ahdc1*. After 48 hours of injection, 15 of the 18 (83%) viable embryos that had the control sgRNA and Cas9 injection showed the albino phenotype (**Supplementary Figure 2B**). After this verification, we proceeded with the injection of embryos to edit *ahdc1*.

Embryo injection for *ahdc1* editing

To carry out the injections for gene editing, zygotes were obtained by breeding in a commercial mating system, in which the males are initially separated from the females by a partition. We used a ratio of four males to three females for mating. After approximately 15 hours in this system and at a time compatible with the reproductive habits of the species, the divider was pulled to allow breeding.

Visually viable eggs (not whitish) were positioned into depressions in an agarose disk (3%) for microinjection using a solution containing TrueCut Cas9 Protein v2 (400 ng/µL - Invitrogen), sgRNA synthesized in house as described above (80 ng/µL), phenol red (0.2 µL - Sigma–Aldrich), and nuclease-free water (qsp 5 µL). The microinjection was performed while the embryos still had only one cell using borosilicate glass needles connected to a microinjector (Harvard Apparatus) powered by compressed O_2 gas.

After injection, 71 eggs were placed in petri dishes containing E3 embryo medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$) and kept at 28.5 °C in an incubator with circadian light control. The unviable embryos were removed daily, and the E3 medium was exchanged on alternate days. Finally, the animals were transferred to the aquarium until the 6^{th} day.

ahdc1 editing verification

After 100 days of injections, the animals were anesthetized by immersion in tricaine solution (Sigma–Aldrich) at 150 mg/L. As soon as signs of anesthesia were observed (loss of the righting reflex and decreased movement of the gills), the caudal fin fragment was cut for genotyping. Each animal was kept isolated to control the origin of the collected samples until the genotyping results were obtained.

DNA from each fragment was extracted using the NucleoSpin Tissue kit (Macherey-Nagel). Sanger sequencing was performed with the primers described in **Supplementary Table 1**, which are the same as those used in the *in vitro* efficiency test of sgRNA directed to *ahdc1*.

Mating to obtain *ahdc1*$^{+/-}$

The animals with indel mutations upstream of the PAM were mosaics (example of sequencing results in **Supplementary Figure 2**), which can be verified by the presence of multiple overlapping peaks in the electropherogram, evidencing the presence of more than two alleles. These animals were paired with adult wild-type (WT) fish of the AB lineage to obtain animals with heterozygous LoF *ahdc1* mutations.

The selection of heterozygous animals was performed after genotyping by Sanger sequencing at least 100 days after fertilization.

Genotypic proportion assessment

Twenty-nine adult offspring of the animals chosen to give rise to the *ahdc1*$^{+/-}$ strain were genotyped by Sanger sequencing to assess the genotypic proportions (*ahdc1*$^{+/-}$ versus *ahdc1*$^{+/+}$).

RESULTS

iPSC characterization

All iPSC colonies from patients and control exhibited typical morphology: a clearly defined border and a uniform structure of tightly clustered small cells (**Figure 3A)**. The pluripotent phenotype was also evidenced by ICC and RTqPCR assays. It is possible to observe in the ICC images (**Figure 3B**) the nuclear expression of OCT3/4 and SOX2, which colocalize with DAPI labeling, while SSEA4 is expressed in the cell membrane. The *OCT3/4* and *NANOG* mRNA quantification had substantially greater values in the reprogrammed cells than in the PBMCs (**Figure 3C**). Thus, the reprogrammed cells had a status compatible with pluripotency.

On Day 6 of EBs formation (**Figure 4A)** it is possible to observe spheroid structures. On the 17th day, we observed variable cellular morphology: different sizes, with some groups of cells being closer together and others being more widely spaced; some cell types had more prominent nuclei than others. This morphological heterogeneity is a "mix of cells" containing early ectoderm-, mesoderm-, and endoderm-like cells, as demonstrated by a panel of the three germ layers markers (TaqMan hPSC scorecard assay - **Figure 4B**).

OriP and *EBNA-1* PCRs and mycoplasma tests had negative results (**Supplementary Figures 1A and 1B**). No relevant CNVs were detected by CMA, which is in agreement with the normal 46,XY karyotypes observed for all iPSC lines. In the STR analysis, all iPSC lines matched PBMC samples (data submitted to the journal for verification but not published to protect patients' identity). The iPSC lines Sanger sequencing results showed the maintenance of patients LoF variants (**Figure 4C**).

Establishment of a zebrafish *ahdc1*$^{+/-}$ model

We obtained heterozygous animals and selected one female with an indel *ahdc1* variant that leads to a frameshift and premature stop codon: a 4 base pair insertion (+TCGT). The electropherogram resulting from the Sanger sequencing of this female (which we consider F0) is shown in **Figure 2**. The position of the mutation and predicted consequence in the gibbin protein sequence is shown in **Figure 5A**.

A total of 14/29 (~48%) of the F0 offspring were heterozygous (**Figure 5B**), which is expected by Mendel's 1st law considering mating involving a WT animal with a heterozygous animal.

DISCUSSION

Here, we obtained four iPSC lines, three of which were derived from XGS patients and one from a control individual. These four iPSC lines showed a pluripotency phenotype and the capacity to differentiate into the three germ layers. These results indicate that the cells we obtained could be differentiated into neurons and/or other cellular types for future functional XGS studies.

We propose that efforts to understand the pathophysiology of XGS could be diverted from the protein per se and directed toward altered biological pathways in complex biological models with *AHDC1* LoF variants. According to the IUPred2A prediction (https://iupred2a.elte.hu - Mészáros et al., 2018), the amino acid residue scores of the human gibbin tend to be high (mean 0.62; SD 0.17 - **Supplementary Figure 3**), which is also observed in the zebrafish protein (mean 0.68; SD 0.14 - **Supplementary Figure 3**). Residues with scores above 0.5 are predicted to be structurally disordered (Erdős & Dosztányi, 2020). These data indicate that gibbin must be an intrinsically disordered protein, a rather common class of proteins in eukaryotic proteomes with a highly flexible and dynamic structure related to many biological functions involving different pathways and diseases (Skrabana et al., 2006; Punta et al., 2015; Schor et al., 2016). However, the characterization of intrinsically disordered proteins remains a challenge due to their complex and promiscuous nature (Weng & Wang, 2020). Many of the proteins that misfold and aggregate *in vivo* are intrinsically disordered (Breydo et al., 2017) and gibbin protein show high probability of fibril and/or aggregation.

Recently, Khayat et al. (2021) and Chander et al. (2022) presented arguments against the haploinsufficiency mechanism of disease for *AHDC1* mutations, namely: *de novo* missense *AHDC1* variants result in XGS-like clinical manifestations; the *AHDC1* mutated transcript escapes from nonsense-mediated mRNA decay; and there is phenotypic heterogeneity associated with the position of the variant in the gene. The hypothesis of protein aggregation for gibbin truncated forms could be explored by investigating a dominant-negative or gain-of-function mechanism for XGS, as proposed by these authors.

Using models such as those presented in this manuscript, OMICs technologies can be used to propose pathophysiological pathways involved in XGS, for example, to identify differentially expressed genes and proteins (e.g., RNAseq and proteomics by mass spectrometry, respectively), chromatin regions with alterations in accessibility or methylation (ATAC-Seq and genome bisulfite sequencing), among other possibilities. Considering the clinical spectrum of XGS, we plan to prioritize the investigation of neurodifferentiation, plasticity, synapse function, myelination, chromatin remodelers, and ID/neurodevelopmental gene-related pathways. Because our XGS patients have increased food anxiety, we also plan to specifically investigate neural pathways of hunger-satiety. In particular, for *in vitro*-derived neurons, electrophysiological recordings of action potentials, synaptic currents, and response to neurotransmitters are pivotal for validation of the neuronal fate and could also provide information about dysfunction linked to XGS mechanisms.

Zebrafish has many advantages as model for diseases. More than 70% of human genes have ortholog genes in zebrafish (Lawson, 2022), zebrafish maintenance is relatively cheap compared to mammalian models, it presents a fast development and numerous offspring (Seth et al., 2013), eggs are naked eye visible facilitating handling

- (Spence et al., 2007), and embryos are transparent and develop outside the female's body (Santoriello & Zon, 2012; Spence et al., 2007). Due to these characteristics, the zebrafish has been used as a model to study neurodevelopmental diseases (de Abreu et al., 2020; Meshalkina et al., 2018; Perez et al., 2019; Sakai et al., 2018). A key advantage for our study is that the *ahdc1* gene is present in a single copy in the zebrafish genome (data from ncbi.nlm.nih.gov/gene).

Regarding the zebrafish model with an LoF *ahdc1* mutation (+TCGT), we did not observe obvious phenotypic differences between heterozygous and WT animals, but it should be noted that XGS human patients usually have mild dysmorphisms. We verified that the heterozygous animals do not have lower viability than WT animals since the genotypic proportion follows what was expected by Mendel's first law. For a more detailed characterization of this animal model, we intend to verify the viability of homozygotes and the life expectancy of the model and to evaluate behavior, brain morphology, and gene expression in future experiments.

As commented recently by Khayat and colleagues (Khayat et al., 2021), biological models for XGS study have been challenging to implement. According to them, recombinant mice do not show an obvious phenotype and are not able to breed. In contrast, Collier et al. (2022) edited *Ahdc1* of mice, and the heterozygous and homozygous mutant mice failed to survive past birth. E18 mosaic mutants reveal developmental patterning defects affecting craniofacial structure, abdominal wall closure, and epidermal stratification. Our zebrafish *ahdc1*$^{+/-}$ model is fertile and viable. Human and mouse exhibit 94% identity, while human and zebrafish present 34%. However, as previously published by Xia et al. (2014), two regions, named by them as CR1 and CR2 have clustered conserved amino acids considering the protein of human, mouse, zebrafish and western clawed frog. These authors argued that these conserved regions should be the two functional gibbin domains. Additionally, as presented in the **Supplementary document 1**, the two AT-Hook DNA motifs are very similar between human, mouse and zebrafish protein sequences. We believe that the lower identity of complete AHDC1 protein between human and zebrafish compared to the identity between human and mouse could result in a higher tolerance to *AHDC1* mutations in zebrafish. *Ahdc1* mutations in the mouse model, which shows higher sequence identity with human protein, could be more deleterious.

In conclusion, we established a zebrafish *ahdc1*$^{+/-}$ model and three iPSC lines with different LoF *AHDC1* variants, which could be differentiated into cell types of interest for XGS studies, such as neurons. OMIC analyses of these models in a case–control approach have important potential for the identification of altered biological pathways, contributing to the understanding of the pathophysiology of XGS. iPSC and zebrafish lines with *AHDC1/ahdc1* LoF variants were deposited on the hpscreg.eu and zfin.org platforms, respectively, to facilitate the scientific community's access to their information and enable collaborationsfor future functional XGS studies.

Contact information distributor

Dr. Ana Cristina Victorino Krepischi (ana.krepischi@ib.usp.br)

Data availability

The iPSC lines established from patients and unaffected individual were registered on https://hpscreg.eu/, and the QR codes to access the lines are below:

P1 USPi001-A

P2 USPi002-A

P3 USPi003-A

C1 CEGHUi001-A

The zebrafish mutant line was submitted to the Zebrafish Information Network (ZFIN) database - https://zfin.org - with the designation *ahdc1*^{dsao1} (ZDB-ALT-221025-2).

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the project reported in this paper.

Ethical statement

For human participants

This study was carried out in accordance with the guidelines for research on human subjects established in the Declaration of Helsinki (1964), its subsequent revisions, and Resolution 466/2012 of the Brazilian National Health Council. The ethics committee of the Institute of Bioscience from the University of Sao Paulo gave ethical approval for this study (CAAE 80921117.5.0000.5464 and CAAE 43559314.0.0000.546). Informed consents were obtained.

For animal experiments

All procedures involving zebrafish were conducted in accordance with the provisions of Brazilian law 11,794/2008, Normative Resolutions 34/2017 and 44/2019 of the National Council for the Control of Animal Experimentation (CONCEA). The Ethics Committee on the Use of Animals of the Faculty of Medicine and Institute of Bioscience from the University of

Sao Paulo gave approval for the procedures used in this study (certificates 1415/2019 and 397/2022, respectively).

Acknowledgments

We are especially grateful to the families of patients involved in this study and Dr. Regina Mingroni for providing the structure of her laboratory for molecular biology assays. We also thank Claudia Castro, Luceleni da Silva, Maraisa Sebastião, Mariana Barros, Dr. Naila Lourenço, Dr. Patricia Semedo, Camila Fénero, Thays Amorim, Juliana Silva, Julia de França, Simone Ferreira, Lais Paredes, Letícia Feitosa, Antonio Ribeiro, and São Lucas Laboratory (Rio Claro – SP, Brazil) for technical support.

Funding

Financial support was provided by FAPESP (2013/08028-1; 2019/02605-3) and CNPq (305806/2019-0). We are grateful for the Doctorate Scholarships (LMLC - FAPESP 2018/08486-3 and CAPES 1805008; LSS – FAPESP 2020/04744-8; FTGRV – CAPES 88882.377407/2019-01; RDS – FAPESP 2020/03108-0). LESFM is grateful for her Young Investigator Grant (FAPESP - 2020/10168-0).

Author contributions

CR, ACVK, and LMLC conceived of and designed the study. EVB, MRSPB, LVP, TAOM, FTGRV, and LAB contributed to the experimental design. CR, ACVK, LMLC, CPK, EMGB, and CBB were involved in the patients' genetic diagnosis. LMLC, EVB, DPM, GSK, and GSPH reprogrammed the cells and, with the support of FTA and ICR, maintained them in culture. LMLC, EVB, RDS, and FTGRV performed ICC assays. LMLC, RDS, EVB, and LSS performed RTqPCR experiments. LMLC, FTA, and RDS were responsible for EBs formation and differentiation. FTGRV and LMLC carried out the *in vitro* sgRNA efficiency test. LMLC and RDS performed the STR profiling. The CMA was performed by SSC. LMLC and EVL carried out the PCR for *OriP* and *EBNA-1* targets. LMLC and DMF cared for, genotyped, and mated the animals. The LESFM made contributions to the discussion. LMLC also carried out the mycoplasma test and patients/iPSC Sanger sequencing; she was responsible for the initial writing of the manuscript. All authors have read, edited, and approved the final manuscript.

REFERENCES

Breydo, L., Redington, J. M., & Uversky, V. N. (2017). *Effects of Intrinsic and Extrinsic Factors on Aggregation of Physiologically Important Intrinsically Disordered Proteins* (pp. 145–185). https://doi.org/10.1016/bs.ircmb.2016.08.011

Carvalho, L. M. L., D'Angelo, C. S., Villela, D., da Costa, S. S., de Lima Jorge, A. A., da Silva, I. T., de Oliveira Scliar, M., Chaves, L. D., Krepischi, A. C. V., Koiffmann, C. P., & Rosenberg, C. (2022). Genetic investigation of syndromic forms of obesity. *International Journal of Obesity*. https://doi.org/10.1038/s41366-022-01149-5

Chander, V., Mahmoud, M., Hu, J., Dardas, Z., Grochowski, C. M., Dawood, M., Khayat, M. M., Li, H., Li, S., Jhangiani, S., Korchina, V., Shen, H., Weissenberger, G., Meng, Q., Gingras, M., Muzny, D. M., Doddapaneni, H., Posey, J. E., Lupski, J. R., … Gibbs, R. A. (2022). Long read sequencing and expression studies of AHDC1 deletions in Xia-Gibbs syndrome reveal a novel genetic regulatory mechanism. *Human Mutation*, *43*(12), 2033–2053. https://doi.org/10.1002/humu.24461

Collier, A., Liu, A., Torkelson, J., Pattison, J., Gaddam, S., Zhen, H., Patel, T., McCarthy, K., Ghanim, H., & Oro, A. E. (2022). Gibbin mesodermal regulation patterns epithelial development. *Nature*, *606*(7912), 188–196. https://doi.org/10.1038/s41586-022-04727-9

de Abreu, M. S., Genario, R., Giacomini, A. C. V. V., Demin, K. A., Lakstygal, A. M., Amstislavskaya, T. G., Fontana, B. D., Parker, M. O., & Kalueff, A. V. (2020). Zebrafish as a Model of Neurodevelopmental Disorders. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2019.08.034

Erdős, G., & Dosztányi, Z. (2020). Analyzing Protein Disorder with IUPred2A. *Current Protocols in Bioinformatics*, *70*(1). https://doi.org/10.1002/cpbi.99

Goyal, C., Naqvi, W. M., Sahu, A., & Aujla, A. S. (2020). Xia-Gibbs Syndrome: A Review of Literature. *Cureus*. https://doi.org/10.7759/cureus.12352

Khayat, M. M., Li, H., Chander, V., Hu, J., Hansen, A. W., Li, S., Traynelis, J., Shen, H., Weissenberger, G., Stossi, F., Johnson, H. L., Lupski, J. R., Posey, J. E., Sabo, A., Meng, Q., Murdock, D. R., Wangler, M., & Gibbs, R. A. (2021). Phenotypic and protein localization heterogeneity associated with AHDC1 pathogenic protein-truncating alleles in Xia-Gibbs syndrome. *Human Mutation*, *42*(5), 577–591. https://doi.org/10.1002/humu.24190

Lawson, N. D. (2022). Decoding the zebrafish genome. *Nature Genetics*, *54*(7), 917–919. https://doi.org/10.1038/s41588-022-01080-5

Meshalkina, D. A., N. Kizlyk, M., V. Kysil, E., Collier, A. D., Echevarria, D. J., Abreu, M. S., Barcellos, L. J. G., Song, C., Warnick, J. E., Kyzar, E. J., & Kalueff, A. V. (2018). Zebrafish models of autism spectrum disorder. *Experimental Neurology*, *299*, 207–216. https://doi.org/10.1016/j.expneurol.2017.02.004

Mészáros, B., Erdős, G., & Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, *46*(W1), W329–W337. https://doi.org/10.1093/nar/gky384

Moreira, D. de P., Suzuki, A. M., Silva, A. L. T. e, Varella-Branco, E., Meneghetti, M. C. Z., Kobayashi, G. S., Fogo, M., Ferrari, M. de F. R., Cardoso, R. R., Lourenço, N. C. V., Griesi-Oliveira, K., Zachi, E. C., Bertola, D. R., Weinmann, K. de S., Lima, M. A. de, Nader, H. B., Sertié, A. L., & Passos-Bueno, M. R. (2022). Neuroprogenitor Cells From Patients With TBCK Encephalopathy Suggest Deregulation of Early Secretory Vesicle Transport. *Frontiers in Cellular Neuroscience*, *15*. https://doi.org/10.3389/fncel.2021.803302

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods*, *12*(10), 982–988. https://doi.org/10.1038/nmeth.3543

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., & Yamanaka, S. (2011). A more efficient method to generate integration-free human iPS cells. *Nature Methods*, *8*(5), 409–412. https://doi.org/10.1038/nmeth.1591

Perez, Y., Bar-Yaacov, R., Kadir, R., Wormser, O., Shelef, I., Birk, O. S., Flusser, H., & Birnbaum, R. Y. (2019). Mutations in the microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. *Brain*, *142*(3), 574–585. https://doi.org/10.1093/brain/awz004

Punta, M., Simon, I., & Dosztányi, Z. (2015). Prediction and Analysis of Intrinsically Disordered Proteins. In *Methods Mol Biol* (Vol. 1261, pp. 35–59). https://doi.org/10.1007/978-1-4939-2230-7_3

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–423. https://doi.org/10.1038/gim.2015.30

Sakai, C., Ijaz, S., & Hoffman, E. J. (2018). Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Frontiers in Molecular Neuroscience*, *11*. https://doi.org/10.3389/fnmol.2018.00294

Santoriello, C., & Zon, L. I. (2012). Hooked! Modeling human disease in zebrafish. *Journal of Clinical Investigation*, *122*(7), 2337–2343. https://doi.org/10.1172/JCI60434

Sarafian, R., Morato-Marques, M., Borsoi, J., Tófoli-Araújo, F., Pinheiro, M., Medina-Armenteros, Y., Lotufo, P. A., Bensenor, I. M., & Pereira, L. V. (2021). Generation of 6 lines of human pluripotent stem cells from hypertensive patients and 3 lines of human pluripotent stem cell from normotensive patients. *Stem Cell Research*, *53*, 102384. https://doi.org/10.1016/j.scr.2021.102384

Schor, M., Mey, A. S. J. S., & MacPhee, C. E. (2016). Analytical methods for

structural ensembles and dynamics of intrinsically disordered proteins. *Biophysical Reviews*, *8*(4), 429–439. https://doi.org/10.1007/s12551-016-0234-6

Seth, A., Stemple, D. L., & Barroso, I. (2013). The emerging use of zebrafish to model metabolic disease. *Disease Models & Mechanisms*, *6*(5), 1080–1088. https://doi.org/10.1242/dmm.011346

Shalaby, M. A., Hampson, L., Oliver, A., & Hampson, I. (2011). Identification of PlexinD1 and AHDC1 as a putative interactors for Tip-1 protein. *Genes & Genomics*, *33*(4), 399–405. https://doi.org/10.1007/s13258-011-0033-1

Skrabana, R., Sevcik, J., & Novak, M. (2006). Intrinsically Disordered Proteins in the Neurodegenerative Processes: Formation of Tau Protein Paired Helical Filaments and Their Analysis. *Cellular and Molecular Neurobiology*, *26*(7–8), 1083–1095. https://doi.org/10.1007/s10571-006-9083-3

Spence, R., Gerlach, G., Lawrence, C., & Smith, C. (2007). The behaviour and ecology of the zebrafish, Danio rerio. *Biological Reviews*, *83*(1), 13–34. https://doi.org/10.1111/j.1469-185X.2007.00030.x

Tofoli, F. A., Dasso, M., Morato-Marques, M., Nunes, K., Pereira, L. A., da Silva, G. S., Fonseca, S. A. S., Costas, R. M., Santos, H. C., da Costa Pereira, A., Lotufo, P. A., Bensenor, I. M., Meyer, D., & Pereira, L. V. (2016). Increasing The Genetic Admixture of Available Lines of Human Pluripotent Stem Cells. *Scientific Reports*, *6*(1), 34699. https://doi.org/10.1038/srep34699

Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P., & Angrand, P.-O. (2011). Interaction Proteomics Analysis of Polycomb Proteins Defines Distinct PRC1 Complexes in Mammalian Cells. *Molecular & Cellular Proteomics*, *10*(4), M110.002642. https://doi.org/10.1074/mcp.M110.002642

Vejnar, C. E., Moreno-Mateos, M. A., Cifuentes, D., Bazzini, A. A., & Giraldez, A. J. (2016). Optimized CRISPR–Cas9 System for Genome Editing in Zebrafish. *Cold Spring Harbor Protocols*, *2016*(10), pdb.prot086850. https://doi.org/10.1101/pdb.prot086850

Weng, J., & Wang, W. (2020). Dynamic multivalent interactions of intrinsically disordered proteins. *Current Opinion in Structural Biology*, *62*, 9–13. https://doi.org/10.1016/j.sbi.2019.11.001

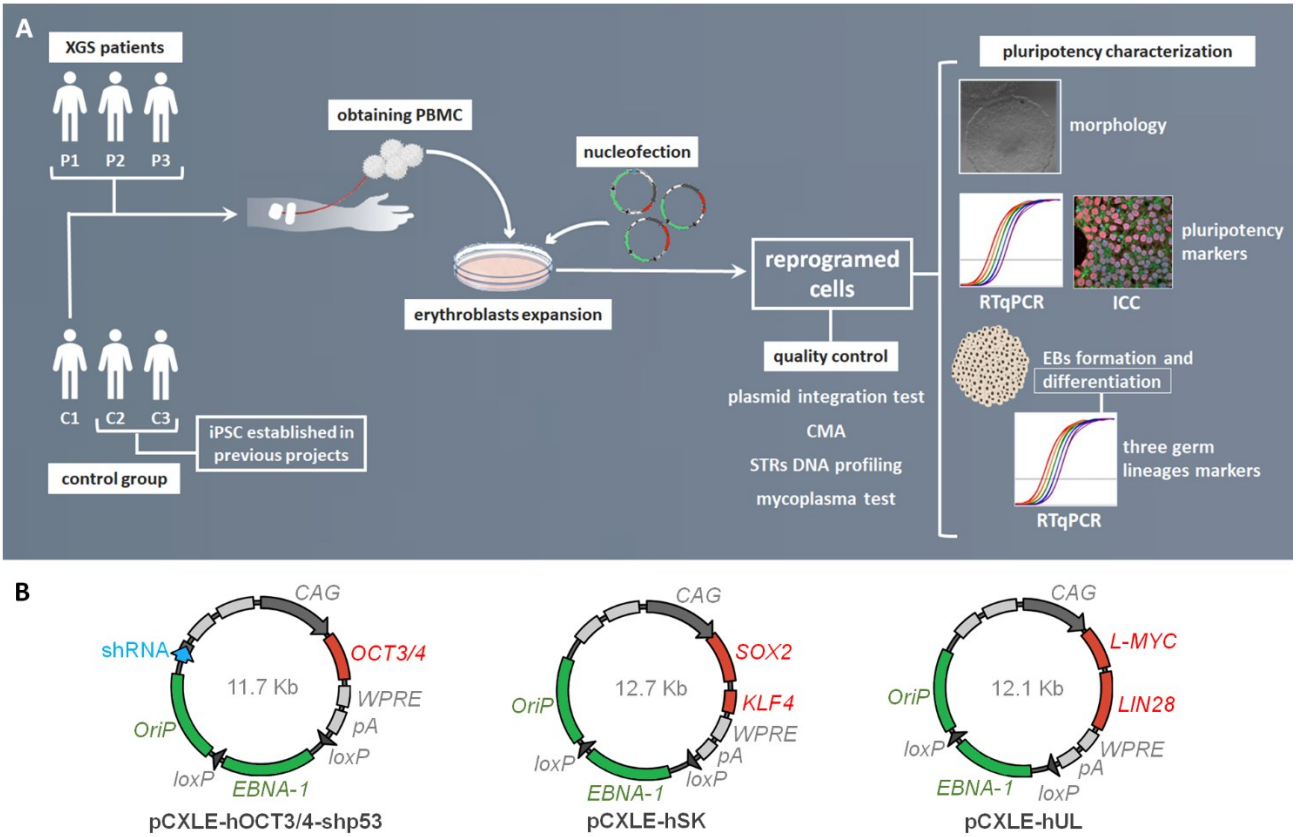
Xia, F., Bainbridge, M. N., Tan, T. Y., Wangler, M. F., Scheuerle, A. E., Zackai, E. H., Harr, M. H., Sutton, V. R., Nalam, R. L., Zhu, W., Nash, M., Ryan, M. M., Yaplito-Lee, J., Hunter, J. V., Deardorff, M. A., Penney, S. J., Beaudet, A. L., Plon, S. E., Boerwinkle, E. A., … Gibbs, R. A. (2014). De Novo Truncating Mutations in AHDC1 in Individuals with Syndromic Expressive Language Delay, Hypotonia, and Sleep Apnea. *The American Journal of Human Genetics*, *94*(5), 784–789. https://doi.org/10.1016/j.ajhg.2014.04.006

Yang, H., Douglas, G., Monaghan, K. G., Retterer, K., Cho, M. T., Escobar, L. F., Tucker, M. E., Stoler, J., Rodan, L. H., Stein, D., Marks, W., Enns, G. M., Platt,

J., Cox, R., Wheeler, P. G., Crain, C., Calhoun, A., Tryon, R., Richard, G., … Chung, W. K. (2015). De novo truncating variants in the *AHDC1* gene encoding the AT-hook DNA-binding motif-containing protein 1 are associated with intellectual disability and developmental delay. *Molecular Case Studies*, *1*(1), a000562. https://doi.org/10.1101/mcs.a000562

FIGURES

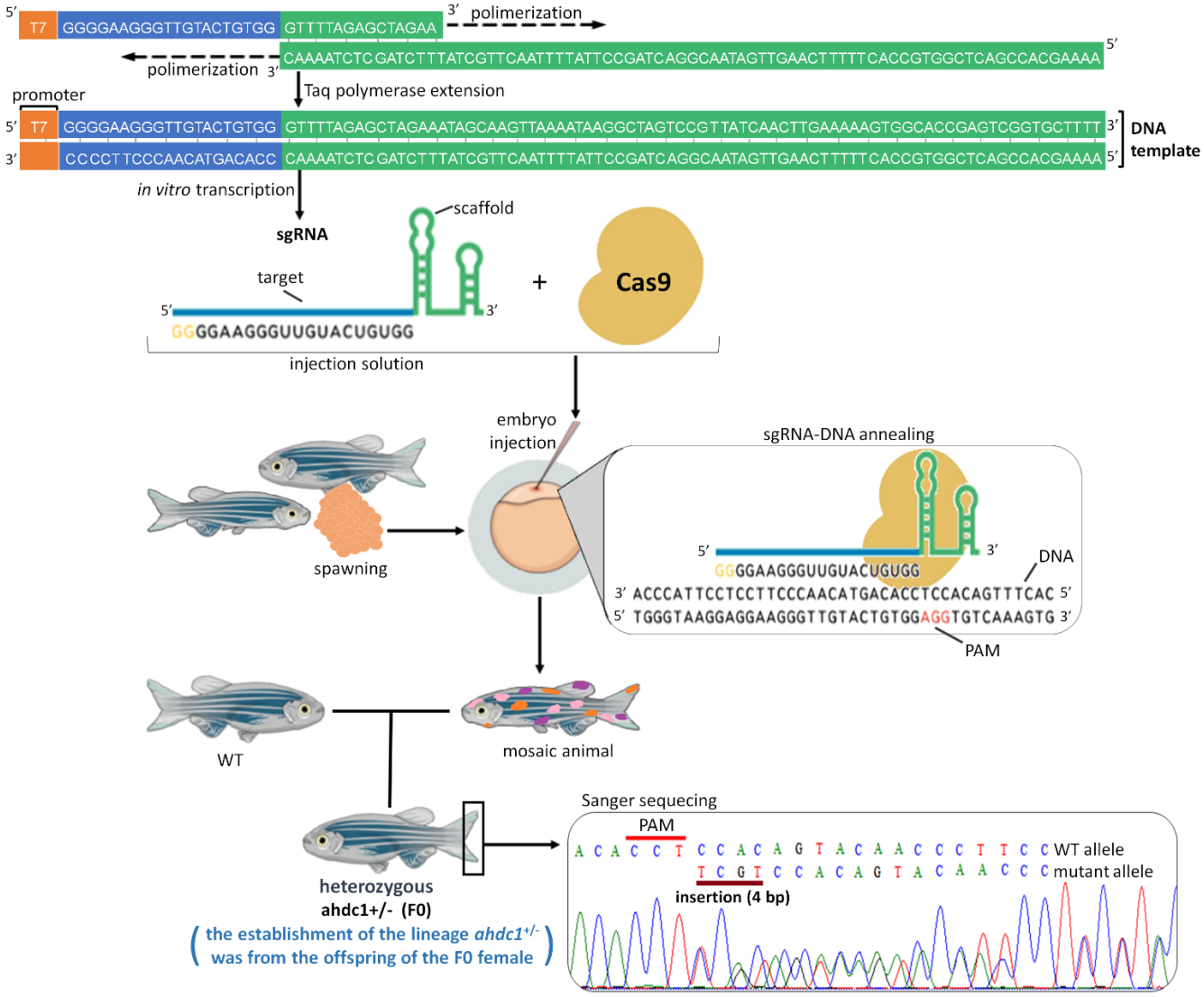
Figure 1 - Methodological design to establish the iPSC lines



A – Methodology to establish the IPSC lines. After *in vitro* erythroblast expansion, we nucleofected the cells with plasmids containing sequences to express pluripotency factors. The reprogramed cells were evaluated regarding the pluripotency phenotypes, assays for plasmid integration, copy number alteration (Chromosomal Microarray Analysis - CMA), short tandem repeat (STR) profiling, and mycoplasma contamination. PBMCs (Peripheral Blood Mononuclear Cells); ICC (Immunocytochemistry); EBs (Embryoid Bodies).

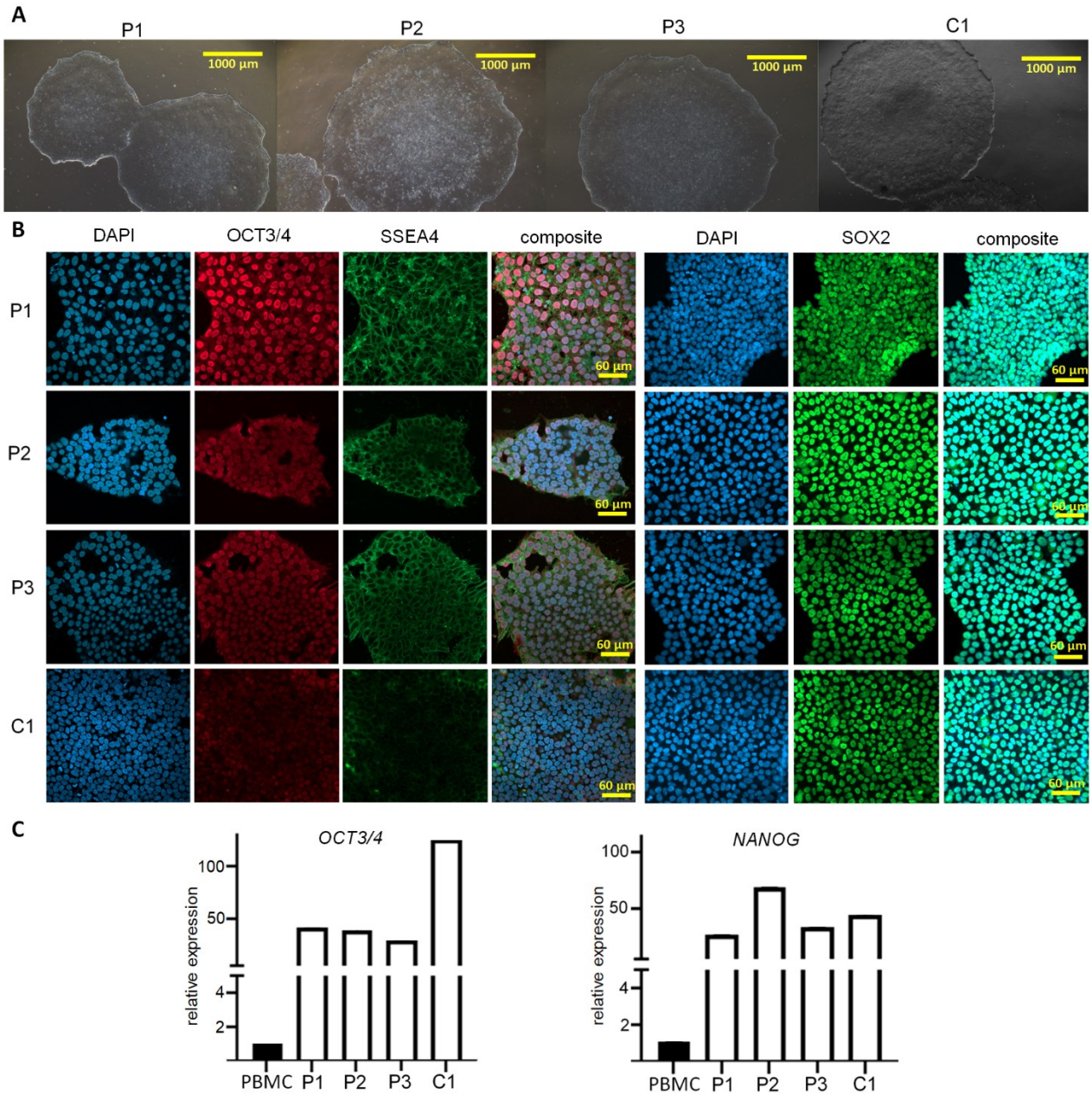
B – Structural details of plasmids used in cellular reprogramming. The sequences that lead to the expression of pluripotency factors are marked in red. The sequence for the shRNA for *p53* suppression is shown in blue. *OriP* and *EBNA-1* (in green) are the targets for the PCR plasmid integration test.

Figure 2 - Workflow of the establishment of the zebrafish *ahdc1*$^{+/-}$ model



The methodology for developing the zebrafish *ahdc1*$^{+/-}$ model involved sgRNA synthesis (annealing between two oligonucleotides, extension under PCR conditions, and *in vitro* transcription), its injection together with the Cas9 protein into the embryos, the detection of mosaic animals, and their mating with WT to obtain heterozygous animals. We chose a female (F0) with an insertion of four base pairs (TCGT) to give rise to the zebrafish line. The colors of the DNA sequences correspond to the regions of the synthesized sgRNA. Designing sgRNA with a target region starting with GG (in orange color) increases the *in vitro* transcription yield. The different colored spots on the mosaic animal represent cell populations with different genotypes. The electropherogram with overlapping peaks shows the presence of two alleles (heterozygosity).

Figure 3 - Pluripotency characteristics of the iPSC lines.

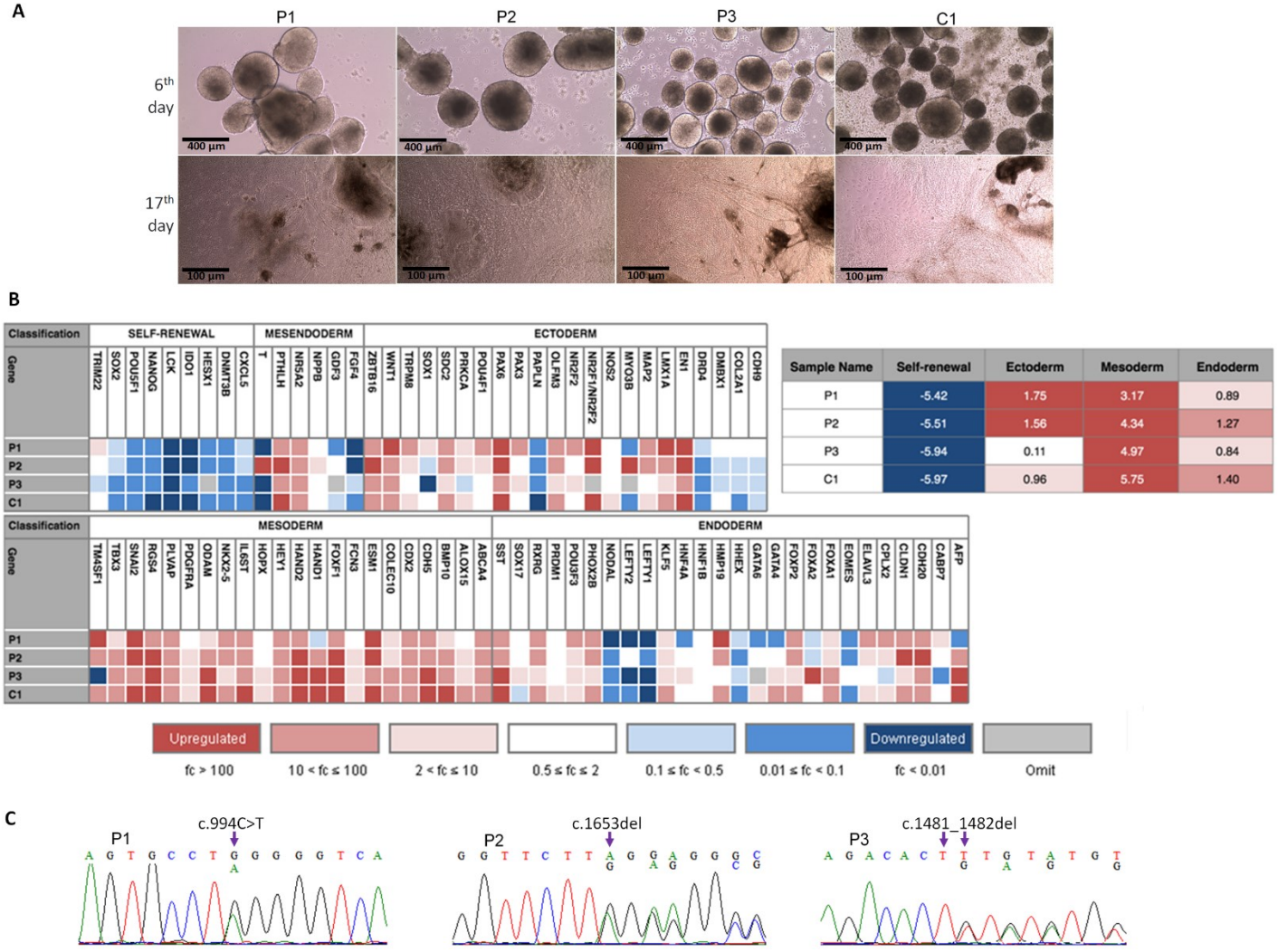


A – iPSC colonies with typical morphology: clearly defined borders, uniform structure, and clustered small cells adherent to the plat.

B –Detection by immunofluorescence of pluripotency markers. It is possible to observe the nuclear expression of OCT3/4 and SOX2, whereas SSEA4 is expressed in the cell membrane.

C – RTqPCR results demonstrating the gene expression of the pluripotency markers *OCT3/4* and *NANOG* in iPSC lines compared to PBMCs.

Figure 4 - Embryoid bodies assay and Sanger sequencing

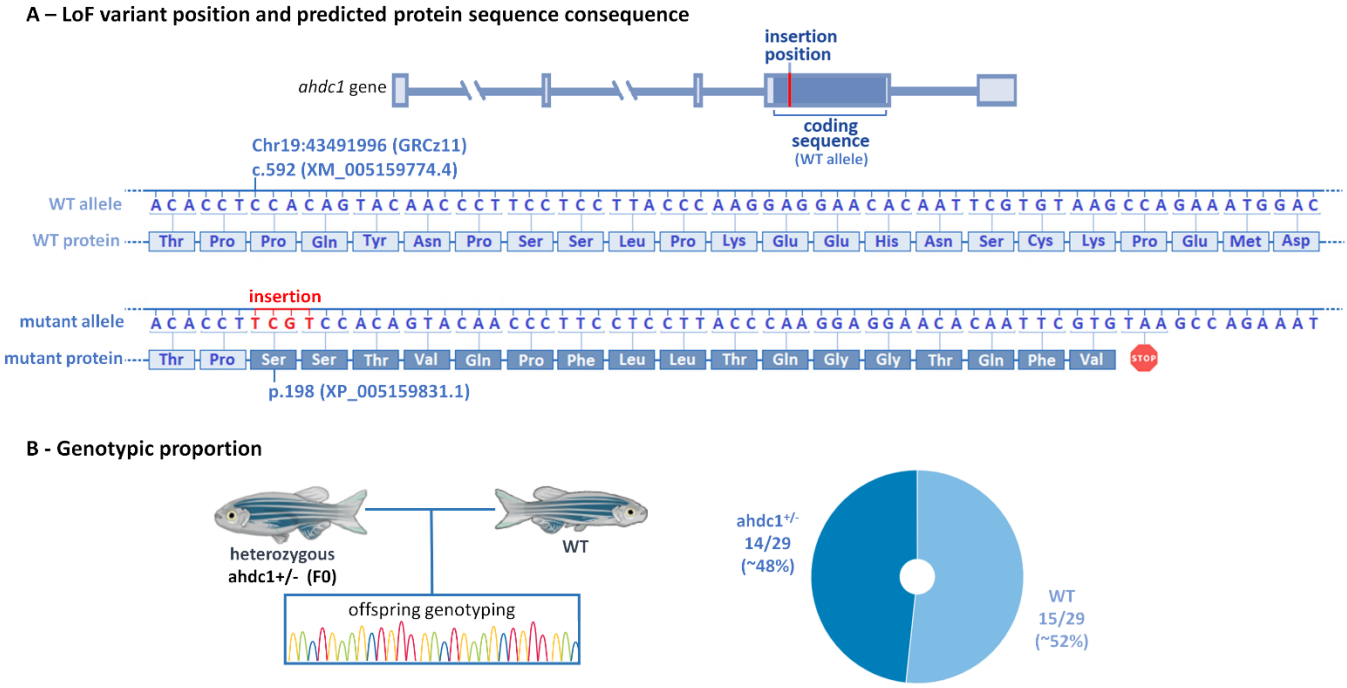


A – Spheroid structure of EBs at Day 6 and differentiation, shown by different cellular morphologies, of EBs at Day 17.

B – Differentiation capacity was confirmed by evaluation of the mRNA expression of markers of three germ layers by TaqMan Scorecard. Expression data are presented for each gene and each germline. The legend of colors indicates the correlation with the fold change in gene expression relative to the undifferentiated reference set.

C - Electropherograms generated by Sanger sequencing of iPSC lines DNAs showing the maintenance of patients LoF variants.

Figure 5 – *ahdc1* zebrafish LoF variant and its consequences

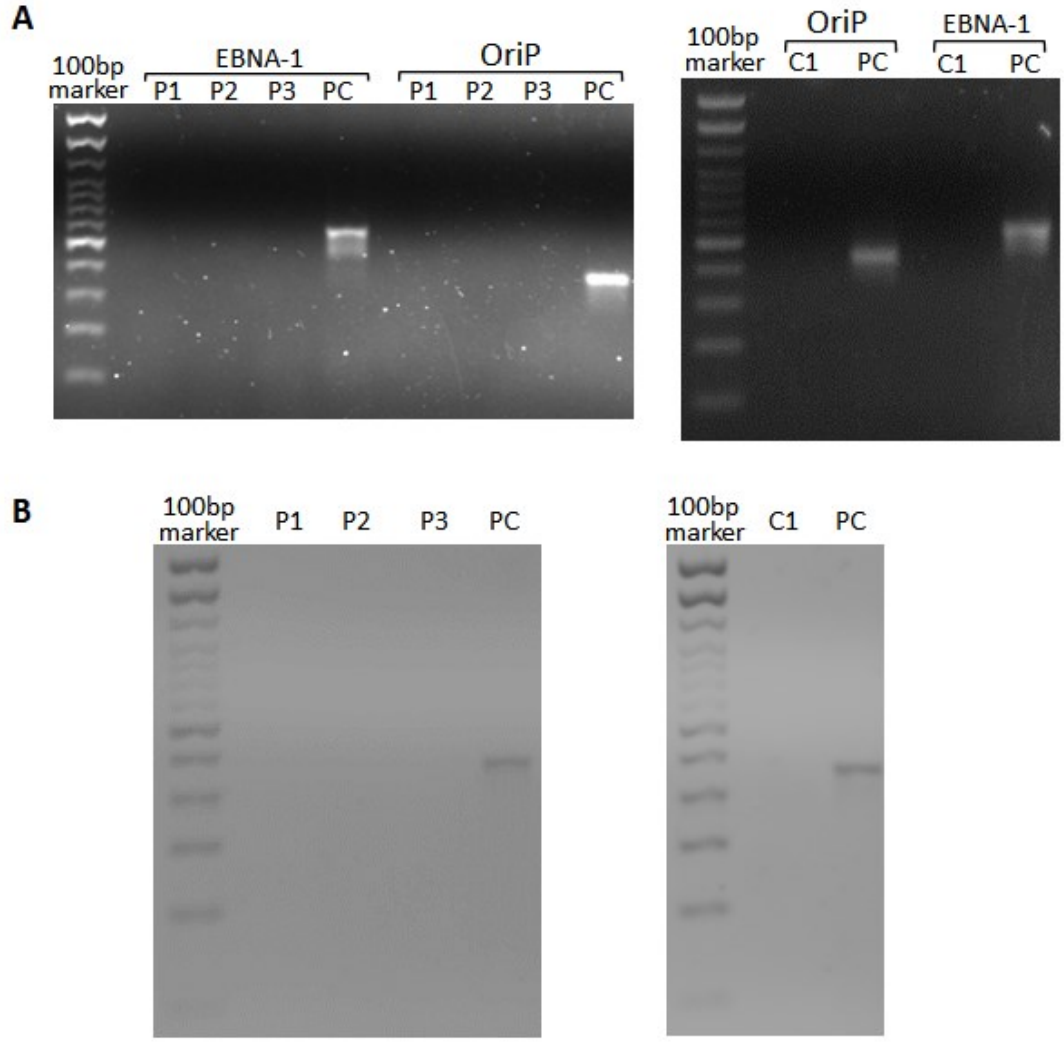


A – The position of the variant is shown in a schematic representation of the *ahdc1* gene, which has the coding region concentrated in a single exon. The consequence in the protein is also presented, with the frameshift leading to a premature stop codon. The position of the variant is shown in the nucleotide and protein sequences.

B - Mating between a heterozygous animal and WT leads to offspring with a genotypic proportion compatible with Mendel's 1st law. Approximately 48% of the offspring animals were heterozygous.

SUPPLEMENTARY FIGURES/DOCUMENTS

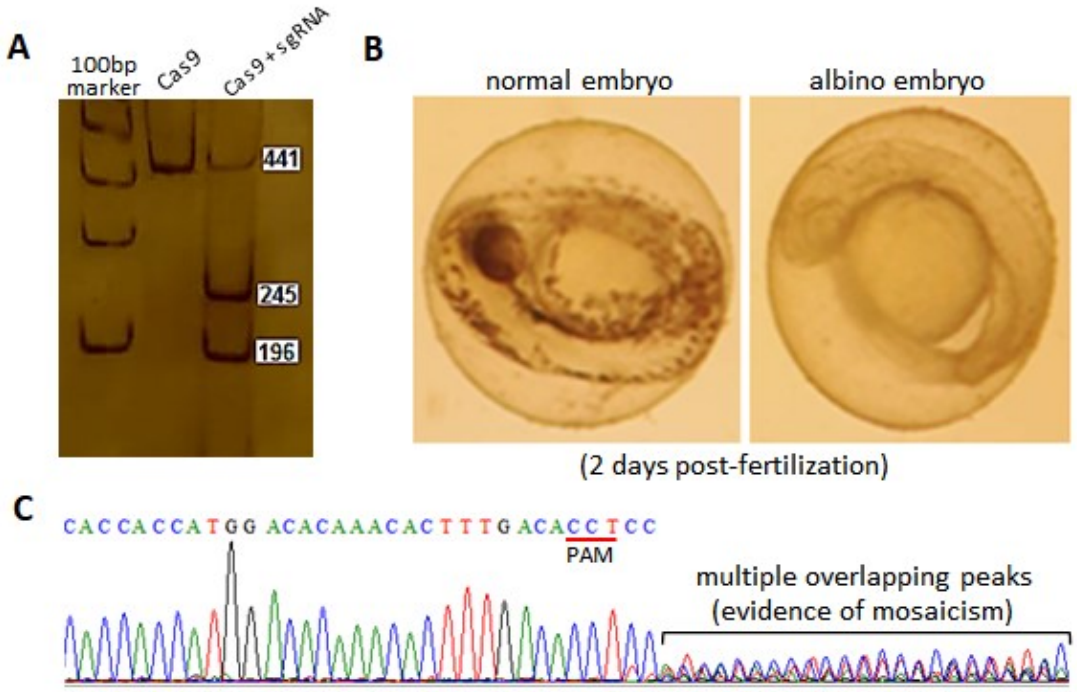
Supplementary Figure 1: iPSC quality control tests



A – The genomes of iPSC lines are free of reprogramming plasmids; PC: positive control; product size for *OriP* positive reaction: 544 bp; product size for *EBNA-1* positive reaction: 666 bp.

B – Negative results on the mycoplasma test; PC – positive control; product size for positive reaction: 500 bp.

Supplementary Figure 2: Experimental checkpoints for the zebrafish model

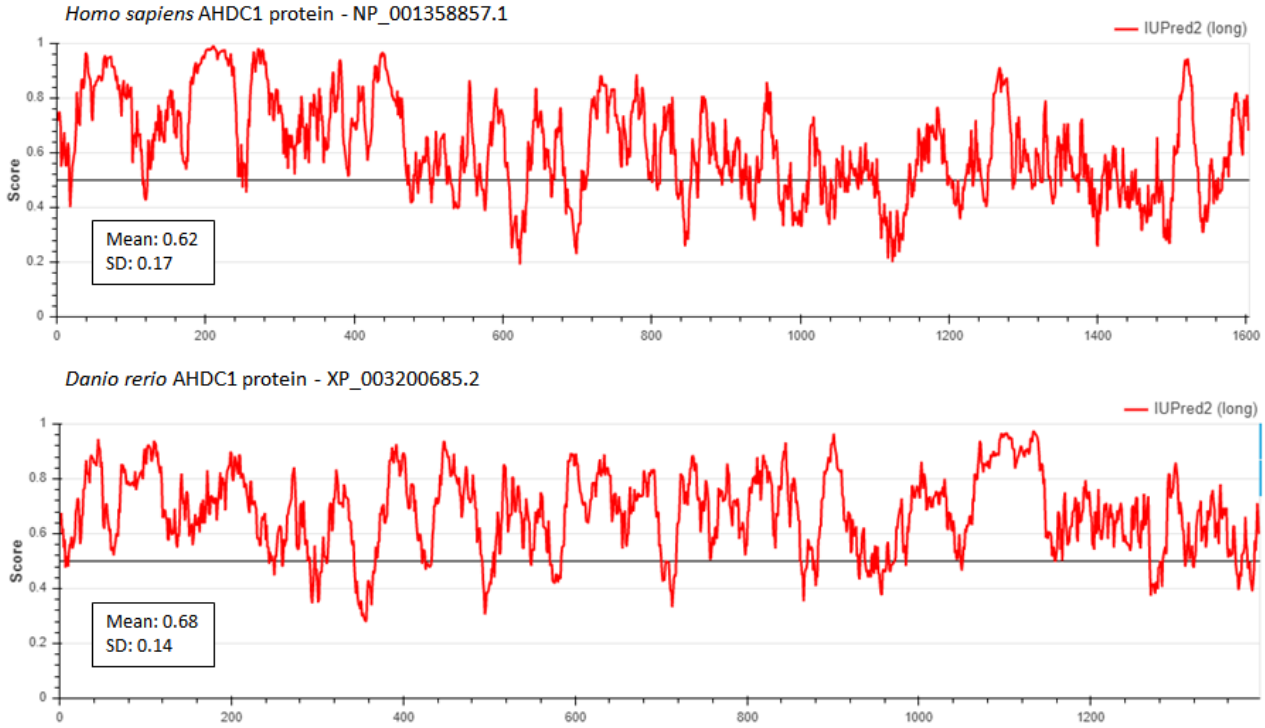


A – The sgRNA that we synthesized in house to edit *ahdc1* was able to guide Cas9 to cut specifically at the expected point of the PCR product (which was 441 bp), generating fragments of 245 and 196 bp. It has been shown that Cas9 without sgRNA is not able to cut the PCR product. In the presence of the specific sgRNA, Cas9 was able to cut the PCR product generating fragments with the expected sizes.

B – The sgRNA control (targeted to the *slc45a2* gene) showed efficiency demonstrated by the albino phenotype (absence of black dots in the embryos).

C - Example of an electropherogram that shows the condition of genetic mosaicism. The multiple peaks upstream of the PAM sequence indicate the presence of more than two alleles in the analyzed sample.

Supplementary Figure 3: Prediction of protein disorder using IUPred2A



IUPred2A is one of the most widely used tools for the prediction of disordered regions/proteins. Residues with scores above 0.5 are predicted to be disordered, and scores below 0.5 indicate order. Predictions for human and zebrafish gibbin suggest that these proteins are intrinsically disordered proteins.

Supplementary document 1: Comparison of the sequences of the gibbin protein of *Homo sapiens, Mus musculus* and *Danio rerio,* with indication of the AT-hook DNA motifs and the patients' mutations in the human sequence.

AT-hook DNA motif (A/T rich regions):
P-R-G-R-P
Positively charged Lys and Arg on either side of the core sequence

AT-hook extended: R/K - X_{12-15} – GRP - X_{12-15} – R/K
P1 - *AHDC1*(NM_001371928.1):c.994C>T(p.Gln332Ter)
P2 – *AHDC1*(NM_001371928.1):c.1653del(p.Lys552ArgfsTer33)
P3 – *AHDC1*(NM_001371928.1):c.1481_1482del(p.Lys494SerfsTer22)

```
danio       ------------------------------------------------------------
          0
human       MRVKPQGLVVTSSAVCSSPDYLREPKYYPGGPPTPRPLLPTRPPASPPDKAFSTHAFSEN
         60
musmus      MRVKPQGLVVTSSAVCSSPDYLREPKYYPGGPPTPRPLLPTRPPASPPDKAFSTHTFSEN
         60


danio       ------------------------------------------------------------
          0
human       PRPPPRRDPSTRRPPVLAKGDDPLPPRAARPVSQARCPTPVGDGSSSRRCWDNGRVNLRP
        120
musmus      PRPPPRRDPSSRRPPVLAKGDDLLPPRAARPVSQAHCPSPAPDNS-SLRHWDNGRVNLRP
        119


danio       ----------MSRLSGRLRSGGVRLTCEALPEKDCNDDEGPSLWEGDAEPEATSMGPSDA
         50
human       VVQLIDIMKDLTRLSQDLQHSGVHLDCGGLRL----------------------SRPPA
        157
musmus      VVQLIDIMKDLTRLSQDLQHSGVHLDCGGLRL----------------------SRPPA
        156
                     ::***  *: .**:* * .*                        .   *

danio       NPASALHYNFYSPPSLANGLRNHEEQLRRRARRRRPTEYVEQIHTQEQTR--IQQNTQPQ
        108
human       PPPGDLQYSFFSSPSLANSIRSPEERATPHAKSERPSHPLYEPEPEPRDSPQPGQGHSPG
        217
musmus      PPPGDLQYSFFSSPSLANSIRSPEERANPHTKSERPSHPLYEPEPEPRDSPQPGQGHGPG
        216
             * . *:*.*:* *****.:*. **:   ::: .**:. : : . : :      *.  *

danio       MHSLTQPSTHLQPHTH-------IPVDAHLWVDTNTLTHKHSQK----GMCTP---TSAS
        154
human       ---ATAAATGLPPEPEPDSTDYSELADADILSELASLTCPEAQLLEAQALEPPSPEPEPQ
        274
musmus      ---AAATATGLPPEPEPDGPDYSELADADILSELASLTCPEAQLLEAQALEPPSPQPEPQ
        273
               :  :* * *. .         .**.:  :  :**  .:*     .:  *    . .

danio       LMELDPSLIP--------EDLSLPSKTE------VHSVQVQPAPLLHTTLDTNTLTPPQY
        200
human       LLDPQPRFLDPQALEPLGEALELPPLQPLADPLGLPGLALQALDTLPDSLESQLLDPQAL
        334
musmus      LLDPQPRFLDPQALEPLGEGLELPPLQPLADPLGLPSLTLQALDTLPDSLESQLLDPQAL
        333
            *:: :* ::         * *.**          : .: :*    *  :*::: * *
```

```
danio       NPSS----L------PKEEHNSCKPEMDDSTGLEHIKGQAAAMTSVEAERKYTLRSSGRP
          250
human       DPLPKLLDVPGRRLEPQQPLGHCPLA--EPLRLDL--CSPHGPPGPEGHPKYALRRTDRP
          390
musmus      DPLPKLLDVPGRRLEPQQSLGHCQLA--EPLRLDL--CSPHGPPGPEGHPKYALRRTDRP
          389
            :*      :      *::  . *     :   *:    .  .  . *.. **:** :.**

danio       RFPCHLRKSSRLRRAPED---HMFKRELDQKDEEEEENIVWRTEDISQIVEHLPEVCAAE
          307
human       KILCRRRKAGRGRKADAGPEGRLLPLP-------------MPT----GLVAALAEPP---
          430
musmus      KILCRRRKAGRGRKADSGPEGRLLPLP-------------MPT----GLAAALAEPP---
          429
            :: *: **:.* *:*  .   :::                  *     :.  * *

danio       SSFEVLPPTDTSVDITPNHSMITSEAEGSVTQSVRGKRQGRYIGVKRIVVKVARIPVHMS
          367
human       ----PPPPPPPPALPGPGPVSVPELKPESSQTPVVSTRKGKCRGVRRMVVKMAKIPVSLG
          486
musmus      ----PLPPPPPPTLSGPGP--VPELEPESSQTPMVPTRKGKCRGVRRMVVKMAKIPVSLG
          483
                  **    .   *.   : .    *    :  .*:*:  **:*:***:*:*** :.

danio       RRQK-SYKISSLEPVSAPPRGEGVTTGEGPEGGVGSEPVSNVPREPTALLRMKNNGKSVM
          426
human       RRNKTTYKVSSLSSSLS---------------VEGKELGLRVSAEPTPLLKMKNNGRNVV
          531
musmus      RRNKTTYKVSSLSSSLS---------------VEGKELGLRVSSEPTPLLKMKNNGRNVV
          528
            **:* :**:***.   :                 *.*   .*  *** **:*****:.*:

danio       VMFPPGELPVILKRRRGRPPKQALPGQPDMHETRVGAAN---------AAEPKKIRRRRT
          477
human       VVFPPGEMPIILKRKRGRPPKNLLLGPGKPKEPAVVAAEAATVAAATMAMPEVKKRRRRK
          591
musmus      VVFPPGEMPIILKRKRGRPPKNLLLGPGKPKEPTVVAAEAATVTAATMAMPEVKKRRRRK
          588
            *:*****:*:****:******: * *  . :*  * **:         *    * ****.

danio       VKLPSPQPSYVNDTNDVKVEYADVLSKLAFLNRQPPSTGRCSPPRCWTPTEPETFNIPPE
          537
human       QKLASPQPSYAADANDSKAEYSDVLAKLAFLNRQSQCAGRCSPPRCWTPSEPESVHQAPD
          651
musmus      QKLASPQPSYAADANDSKAEYSDVLAKLAFLNRQSQCAGRCSPPRCWTPSEPESVHQAPD
          648
             ** ******. *:** *.**:***:********  .:***********:***:.:  *:

danio       NPSLSTLLHRLTGFRRRGGRAGCMGSRGGGAAGASSSFKRSFSDFFETIGKKRKVPASE-
          596
human       TQSISHFLHRVQGFRRRGGKAGGFGGRGGGHAAK--SARCSFSDFFEGIGKKKKVVAVAA
          709
musmus      TQSISQFLHRVQGFRRRGGKTGGFGGRGGGHAAK--AARCSFSDFFEGIGKKKKVVAVAA
          706
            . *:* :***: *******::* :*.**** *.   : : ******* ****:** *

danio       ----------PGTPRKRGKGAAGGINRAALADSAQGEKVRKRRPRKNGALKNGPGVQEQD
          646
human       AGVGGPGLTELGHPRKRGRGEV---------DAVTGKPKRKRRSRKNGTLFPEQVPSGPG
          760
musmus      PGLVGPGLTELGHPRKRGRGEV---------DAVTGKPKRKRRSRKNGTLFPEQVPSGPG
          757
                      * *****:* .         *:. *:  **** ****:*      .  .

danio       WQNENSSWIGK--------GDIPEKAGSYQSPCSPRGS---------FQSSDGTKGGMYH
          689
human       FGEAGAEWAGDKGGGWAPHHGHPG--GQAGRNCGFQGTEARAFASTGLESGASGRGSYYS
          818
```

```
musmus      FGEAGAEWVGDKGGGWAPHHGHPG--GQAGRNCGFQGTEARAFASTGLESGASGRGSYYA
          815
              : : .:.* *.         . *   *.    *. :*:         ::*. . :*. *

danio       SPGM-RGVGSGEESQGLFAGYFRSLLDSDDSSDLLDISMSSPAGRQESRKLTPGYEGGSP
          748
human       TGAPSGQTELSQERQNLFTGYFRSLLDSDDSSDLLDFALSAS--RPESRKASGTYAGPPT
          876
musmus      -GAPSGQTELSQERQNLFTGYFRSLLDSDDSSDLLDFALSAS--RPESRKASGTYAGPPS
          872
                .    .  .:* *.**:******************:::*:   * **** :  * *

danio       GA--AHRWSPAFPKRSPKGA----ACLGDGAQLSSPQTQGS-----STTKSSYSYS--VS
          795
human       SALPAQRGLATFPSRGAKASPVAVGSSGAGADPSFQPVLSARQTFPPGRAASYGLTPAAS
          936
musmus      SALPAQRGLATFPSRGAKASPVAVGSSGAGADPSFQPVLPSRQTFPPGRATSYGITPATS
          932
              .*  *:*   :**.*. *.:    .. * **: *   .  :         :**. :  .*

danio       QTSPTPSFPKS--PALLLSRSPSSPHPSGGFPQYPPNYSAGVPQGGSIYP--LPQQQQQQ
          851
human       DCRAAETFPKLVPPPSAMARSPTTHPPANT---YLPQYG-GYGAGQSVFAPTKPFTGQDC
          992
musmus      DCRAAETFPKLAPPPSAVARSPTTHPPANT---YPPQYG-GYGAGQSVFASAKPFSGQDC
          988
              :   : :***   *   ::***::  *:.    * *:*. *   * *::    *   *:

danio       QRPSDCSFTYGTKTPS--VPSAQCQMSYSNYQ-------GSAKRNYSSYPGPPHST----
          898
human       ANSKDCSFAYGSGNSLPASPSSAHSAGYAPPPTGGPCLPPSKASFFSSSEGAPFSGSAPT
         1052
musmus      ANSKDCSFAYGSGNSLPASPSSAHSAGYAPPPTGGPCLPPSKASFFNSSEGGPFSGSAPT
         1048
               . .****:**: .     **:  . .*:           *    :.*  * *.*

danio       --TQRGESGPTSPGGSYMSMS-----------KSSPYSLSSSPPEGCRQYT--AQWGYRQ
          943
human       PLRCDSRASTVSPGGYMVPKGTTASATSAASAASSSSSSFQPSPENCRQFAGASQWPFRQ
         1112
musmus      PLRCDSRASTVSPGGYMVPKGTTASAAS---VASSSSSSFQPSPENCRQFVGASQWPFRQ
         1105
                    ..:. .****  :  .            **  *  .  **.***:.  :** :**

danio       G--GNNWGGDAYGSHQFHGYSEYGVAGTGSESKDILDISNYTPQKAKQRPCID-TLSESS
         1000
human       GYGGLDWASEAFS-----QLYNPSFDCHVSEPNVILDISNYTPQKVKQQTAVSETFSESS
         1167
musmus      GYGGLDWASEAFS-----QLYNPNFDCHGSEPNVILDISNYTPQKVKQQTAVSETFSESS
         1160
              *  * :*..:*:.        : ..    ** : ***********.**: .:. *:****

danio       SDSSHTGGGGGGASVVGAAFRPRDVPMPEGQSSLSSLEKLMLDWNENSAGPSYNWSQNVL
         1060
human       SDSTQFNQPVGGGGF---RRANSEASSSEGQSSLSSLEKLMMDWNEASSAPGYNWNQSVL
         1224
musmus      SDSTQFSQPVGGGGF---RRANSEASSSEGQSSLSSLEKLMMDWNEASSAPGYNWNQSVL
         1217
              ***:: .   **...        :.   *************:**** *:.*.***.*.**

danio       FQGGAKPGRGRRKKSEAHNEKESCSLPPGSPASPPMQGAGPKRSSTGGRQPRGARGRGGF
         1120
human       FQSSSKPGRGRRKKVDLFEASHL-GFPTSA--S-AAASGYPSKRSTGPRQPRGGRGGGAC
         1280
musmus      FQSSSKPGRGRRKKVDLFEASHL-GFSTST--S-ATASGYPSKRSTGPRQPRGGRGSGAC
         1273
              **..:********* : .: ..  .:  .:  *    .. *.: *** *****.** *.
```

```
danio       SPCQRDRPPPKTKSQKPSAPSGSGQMGSGAVYQEALDYYSGDSSSLSPLS--------HA
         1172
human       SAKKERGG-AAAKAKFIPKPQPVNP-LFQDSPDLGLDYYSGDSSM-SPLPSQSRAFGVGE
         1337
musmus      SAKKERGG-TAAKAKFIPKPQPVNP-LFQDSPDLGLDYYSGDSSM-SPLPSQSRAFGVGE
         1330
               *  :.      :*::    *.  .        : .*********  ***

danio       PESCEYPSPYSVHTSTPSSDERFAHVYPPDSASVSPSLSIQSDALKQFPKSGPTAQTYG-
         1231
human       RDPCDFIGPYSMNPSTPSDG-TFGQGFHCDSPS----LGAPELDGKHFPPLAHPPTVFDA
         1392
musmus      RDPCDFMGPYSMNPSTPSDG-TFGQGFHCDSPS----LGAAELDGKHFPPLAHPPTVFDA
         1385
             : *:: .***:: ****..  *.: :  ** *    *.  .   *:**  .    .:.

danio       ----------HAARTFSPNLSPTPRLLPQCGSAMS-----------------PHRVPKDQ
         1264
human       GLQKAYSPTCSPTLGFKEELRPPPTKLAACEPLKHGLQGASLGHAAAAQAHLSCRDLPLG
         1452
musmus      GLQKAYSPTCSPTLGFKEELRPPPSKLTACEPLKHGLQGASLSHA--AQAHLSCRDLPLG
         1443
                       :  *. :* * *  *  *                       *

danio       FSQYDSPSYSGSPCWYGQGGSVAGSPQNYEEQRTPAVSLPSQKRDMSLMVSGMRIASHTS
         1324
human       QPHYDSPSCKGTAYWYPPGSAARSPPYEG--KVGTGLLADFLGRTE--------------
         1496
musmus      QPHYDSPSCKGTAYWYPPGSAARSPPYEG--KVGSGLLADFLGRTE--------------
         1487
              :***** .*:  **  *.:. . * :   :   .:      *

danio       YPSPLQRGPSMSTSCVSGTLDSSPQHEEMGYHGNLESYAPVGQRYAPQTARGGVLC----
         1380
human       ------------AACLSAPHLASPPATPKADKEPLEMARPPGPPRGPAAAAAGYGCPLLS
         1544
musmus      ------------AVCLSAPHLASPPATPKADKEPLEMARPPGPPRGPAAATAGYGCPLLS
         1535
                       : *:*.   :**     . :  **   * *   .* :* .*  *

danio       -------------------------------------------QLLDQPSDEGFTVTSL
         1396
human       DLTLSPVPRDSLLPLQDTAYRYPGFMPQAHPGLGGGPKSGFLGPMAEPHPEDTFTVTSL
         1603
musmus      DLTLSPVPRDSLLPLQDTAYRYPGFMPQAHPGLGGGPKSGFLGPMAEPHPEDTFTVTSL
         1594
                                                      : :   :: ******
```

Supplementary document 2 - Protocol for Peripheral Blood Mononuclear Cells (PBMC) isolation

1 – Prepare the media:

Medium 1: 2mM EDTA in PBS; filter and store at 4ºC

Medium 2: RPMI 1640 (90%) and DMSO (10%)

2 – Add 2mL of medium 1 to 5mL of blood.

3 – Prepare the Leucosep tube: put 3mL of Ficoll-Paque PLUS (GE Healthcare) in a Leucosep tube of 12mL.

4 – Centrifuge at 1000g for 10min (18ºC).

5 – Transfer the diluted blood to the Leucosep tube and centrifuge at 400g for 30min (18ºC).

6 – Transfer, very carefully, using Pasteur pipette, the PBMC ring to a 15mL Falcon tube.

7 - Complete the Falcon tube with medium 1.

8 – Centrifuge at 300g for 10min, at 18ºC.

9 – Discard the supernatant.

10 – Suspend the pellet in 15mL of medium 1.

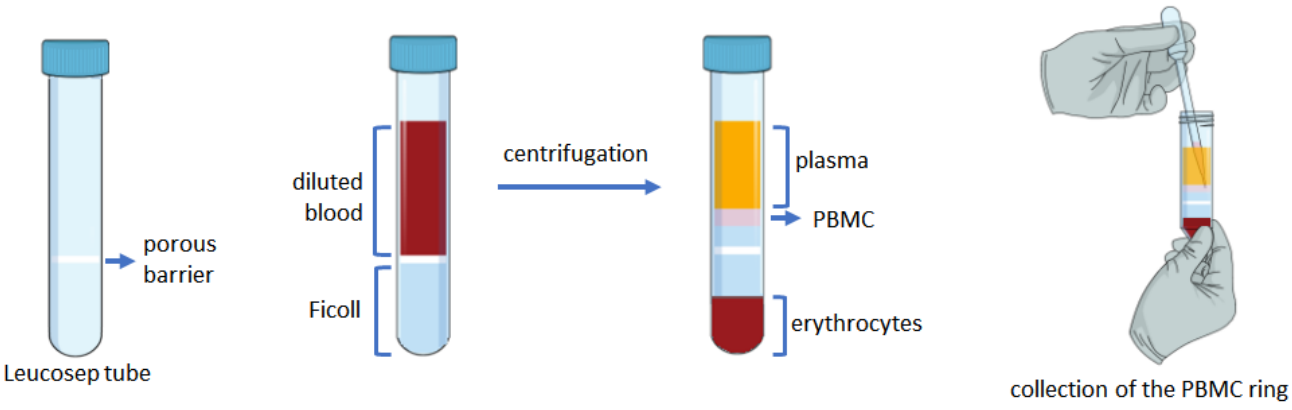
11 – Centrifuge at 200g for 10-15min, at 18ºC.

12 – Discard the supernatant.

13 – Suspend the cells in 5mL of medium 1 and count the cells.

14 – Centrifuge again at 200g for 10-15min; discard the supernatant and suspend in medium 2 (3-5x10^6 cells per cryovial).

15 - Freeze in -80ºC freezer using Mr. Frosty or Corning CoolCell according to manufacturer's instructions. Between 48 and 72 hours later, transfer the cryovials to liquid nitrogen storage.



Supplementary document 3 - Immunostaining Protocol

Markes used:
OCT4 - Also known as POU5F1 (POU class 5 homeobox 1); transcription factor containing a POU homeodomain that plays a key role in embryonic development and stem cell pluripotency. Subcellular localization: nucleus.

SSEA4 - Stage-specific embryonic antigen 4 is a globo-series carbohydrate core structure of glycoproteins. It has been commonly used as a pluripotent human stem cell marker. Subcellular localization: plasma membrane.

SOX2 - sex determining region Y-box 2 is a transcription factor that is essential for maintaining self-renewal, or pluripotency, of undifferentiated embryonic stem cells. Subcellular localization: nucleus.

Antibodies used for iPSC characterization:

	Antibody	Dilution	Company Cat #
Pluripotency Marker	rabbit anti-OCT4	1:200	Abcam #ab19857
Pluripotency Marker	mouse anti-SSEA4	1:200	Abcam #ab16287
Pluripotency Marker	rat anti-SOX2	1:100	Invitrogen #A24759
Secondary antibody	goat anti-rabbit, Alexa fluor 647	1:1000	Invitrogen #A27040
Secondary antibody	goat anti-Mouse, Alexa Fluor 488	2 drops/mL	Invitrogen #R37120
Secondary antibody	donkey anti-rat, Alexa Fluor 488	1:250	Invitrogen #A24876

Antibodies #A24759 (anti-SOX2) and #A24876 (anti-rat, Alexa Fluor 488) are part of an iPSC characterization kit (Pluripotent Stem Cell 4-Marker Immunocytochemistry Kit, Invitrogen, Catalog number: A24881). For SOX2 analysis, the reagents from this kit were used according to the manufacturer's instructions. For the other markers, we followed the following immunostaining steps:
1 – Cultivate the iPSC in Chamber slides (8-well configuration) treated with matrigel and using E8 as medium;
2 – Prepare blocking solution: 90% PBS 1X : 10% BSA.
3 – Prepare permeabilization solution: 99.8% blocking solution : 0.2% Triton X.
4 – Prepare fixation solution: 96% PBS 1X : 4% PFA.
5 – Fix cells at desired confluence for immunocytochemistry assay. For this, remove the culture medium, wash the cells with PBS 1X and incubate for 15 minutes with the fixation solution.
6 – Wash twice with PBS 1X.
7 – Incubate for 15 minutes with permeabilization solution.
8 – Wash twice with PBS 1X.
9 – Incubate 30 minutes with blocking solution.
10 – Dilute the primary antibody in blocking solution according to the dilution factor indicated in the table of this document.

11 – Incubate overnight at 4ºC.
12 – Wash three times with PBS 1X.
13 – Dilute the secondary antibody in blocking solution according to the dilution factor indicated in the table of this document.
14 – Incubate one hour at room temperature.
15 – Dilute DAPI (1:1000) in 1X PBS and incubate for two minutes at room temperature.
16 – Wash three times with PBS 1X.
17 – Remove chamber walls.
18 – Add Fluoromount-G Mounting Medium and a cover slip.

Supplementary table 1 - Oligonucleotides used in this study

Primers for iPSC model			
	Target	**Product size**	**Forward/Reverse primers (5′-3′)**
P1 Sanger sequencing	*AHDC1* (P1 mutation)	499 bp	(F) CGACACAGGATCTTTGGCCT
			(R) CTGAGCCAGACAGCACTGAT
P2 and P3 Sanger sequencing	*AHDC1* (P2 and P3 mutations)	468 bp	(F) TCTGCTGCATAGGATGGCTG
			(R) CCCAGAGTTGAAGCCGGAAT
Plasmid integration into the iPSC genome test by PCR - episomal Plasmids pCXLE-hOCT3/4-shp53, pCXLE-hSK, and pCXLE	*OriP*	544 bp	(F) TTCCACGAGGGTAGTGAACC
			(R) TCGGGGGTGTTAGAGACAAC
Plasmid integration into the iPSC genome test by PCR - episomal Plasmids pCXLE-hOCT3/4-shp53, pCXLE-hSK, and pCXLE	*EBNA-1*	666 bp	(F) ATCGTCAAAGCTGCACACAG
			(R) CCCAGGAGTCCCAGTAGTCA
House-Keeping Gene (RT-qPCR)	*TBP*	138 bp	(F) GTGACCCAGCATCACTGTTTC
			(R) GCAAACCAGAAACCCTTGCG
House-Keeping Gene (RT-qPCR)	*HMBS*	157 bp	(F) AGCTTGCTCGCATACAGACG
			(R) AGCTCCTTGGTAAACAGGCTT
Pluripotency marker (RT-qPCR)	*OCT3/4* (*POU5F1*)	103 bp	(F) TCCCATGCATTCAAACTGAGG
			(R) CCAAAAACCCTGGCACAAACT
Pluripotency marker (RT-qPCR)	*NANOG*	142 bp	(F) TGGACACTGGCTGAATCCTTC
			(R) CGTTGATTAGGCTCCAACCAT
Oligonucleotides for sgRNA synthesis (zebrafish model)			
	Oligo size	**Sequence (5′-3′)**	
Referent to scaffold portion (universal)	80 nt	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA ACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCT AAAAC	
Referent to T7 + target (*ahdc1*) + scaffold anchoring portions	52 nt	TAATACGACTCACTATAGGGGAAGGGTTGTACTGTGG GTTTTAGAGCTAGAA	
Referent to T7 + target (*slc45a2*) + scaffold anchoring portions	52 nt	TAATACGACTCACTATAGGGGAAGGTTGATTATGCAC GTTTTAGAGCTAGAA	
Primers for zebrafish model			
	Target	**Product size**	**Forward/Reverse primers (5′-3′)**
I - Zebrafish genotyping by Sanger sequencing II – PCR for sgRNA efficiency test	*ahdc1*	441 bp	(F) GGACACGAACACGCTAACAC
			(R) TCGTCCTTTTGGTCCAGTTCC

Capítulo 4
Artigos de educação em genética e divulgação científica

Apresentamos três artigos voltados para educação em genética e divulgação científica, com conteúdos redigidos em conjunto com outros colegas do programa de pós-graduação. Esses artigos versam sobre alguns dos temas que foram trabalhados pela candidata em função desta tese, são eles: deficiência intelectual, modelos animais e CRISPR-Cas9.

Os artigos apresentados nos **capítulos 4.1** ("Inteligência e deficiência intelectual: bases genéticas e fatores ambientais") e **capítulo 4.2** ("Modelos animais no estudo de doenças genéticas humanas") foram publicados na revista Genética na Escola em 2020 e 2022, respectivamente. O **capítulo 4.3** é um manuscrito que ainda pretendemos submeter a uma revista de escopo compatível e se intitula "CRISPR-Cas9: conceitos fundamentais e aplicações em saúde humana".

Cabe mencionar que a revista Genética na Escola é voltada sobretudo para alunos e professores de ensino médio, de linguagem acessível ao público geral. Já o artigo apresentado no capítulo **4.3** é voltado para um público de graduação/pós-graduação, por descrever mecanismos que exigem conhecimento prévio de biologia molecular, sobretudo de estrutura e função de ácidos nucleicos, e por isso será submetido a outra revista.

Capítulo 4.1

Inteligência e deficiência intelectual: bases genéticas e fatores ambientais

TOLEZANO, G. C.; CARVALHO, L. M. L; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. Inteligência e deficiência intelectual: bases genéticas e fatores ambientais. **Genética na Escola.** 2020. ISSN: 1980-3540. DOI: 10.55838/1980-3540.ge.2020.332

Inteligência e deficiência intelectual: bases genéticas e fatores ambientais

Giovanna Cantini Tolezano[1*], Laura Machado Lara Carvalho[1*], Ana Cristina Victorino Krepischi[2], Carla Rosenberg[2]

[1] Pós-graduanda do Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP
[2] Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP
* As autoras Giovana Tolezano e Laura Carvalho contribuíram igualmente para o artigo

Autor para correspondência - carlarosenberg@ib.usp.br

Palavras-chave: inteligência, deficiência intelectual, genética humana, neurodesenvolvimento, fatores ambientais, herança multifatorial

"Inteligência" é um termo genérico que abrange diferentes habilidades cognitivas. Pessoas com deficiência intelectual apresentam limitações identificáveis tanto no desempenho **cognitivo** quanto no **comportamento adaptativo** (por exemplo, no cuidado pessoal, uso de dinheiro e meios de transporte). Será que existe um gene da inteligência que, quando não funciona direito, causa deficiência intelectual? O ambiente também contribui para a inteligência? Este artigo explora a relação entre inteligência, deficiência intelectual e fatores genéticos e ambientais.

Cognitivo – a cognição refere-se ao processo de aquisição e construção do conhecimento.

Comportamento adaptativo – atitude que resulte em adequação a diferentes contextos.

Albert Einstein é considerado um gênio da física; Marie Curie um gênio da química; Beethoven, um gênio da música e J. K. Rowling, um gênio dentre os escritores de livros juvenis. Tais pessoas destacam-se por capacidades em áreas muito distintas, porque não existe uma única forma de inteligência. Inteligência é um termo genérico que abrange diferentes habilidades cognitivas como lógica, memorização, comunicação, aprendizado, resolução de problemas, planejamento e **abstração**.

Mas, será que é possível medir a inteligência de alguém? E, se existem áreas distintas de habilidades, como saber em qual delas uma pessoa é "mais inteligente"? Uma das maneiras mais utilizadas para avaliar as capacidades cognitivas de um indivíduo é estimar o Quociente de Inteligência, ou QI, cujo resultado é um valor numérico obtido a partir do desempenho de uma pessoa em diferentes testes. Ou seja, aplica-se um modelo matemático que permite, grosso modo, medir o quão inteligente uma pessoa é. No entanto, todos os testes têm algumas limitações: por exemplo, a maioria deles apenas avalia pessoas alfabetizadas e que conheçam as operações matemáticas básicas.

Além disso, como dito acima, existem diferentes áreas de inteligência e, dessa forma, os testes mais modernos de QI subdividem o resultado de modo a identificar os pontos mais fortes e os mais fracos de cada um. Vale dizer que os testes de QI devem ser aplicados por profissionais de saúde devidamente habilitados para esse tipo de avaliação como, por exemplo, psicólogos.

A figura 1 mostra a distribuição das faixas de valores de QI na população geral. A maior parte das pessoas tem QI entre 85 e 115, sendo mais comuns valores em torno de 100. Indivíduos com QI acima de 130 são considerados superdotados, enquanto aqueles com resultado abaixo de 70 apresentam o que chamamos de deficiência intelectual, classificada como leve (QI 50-70), moderada (QI 35-50) e grave/profunda (QI <35).

Abstração – ação de isolar mentalmente um objeto de reflexão para considerá-lo individualmente, independente do todo.

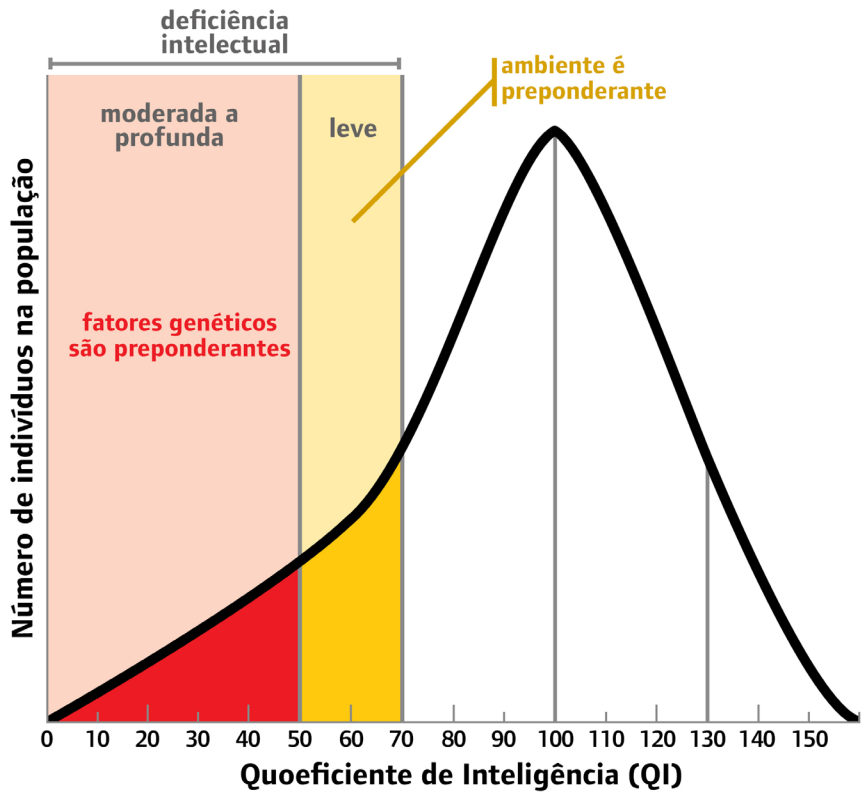


Figura 1.
Gráfico apresentando a distribuição de QI na população geral, sendo que a maior parte das pessoas tem QI próximo de 100. Fonte dos dados: BHATE, S.; WILKINSON, S. Aetiology of learning disability. *Aetiology and epidemiology*, v. 5, n. 9, p. 298-301, 2006.

A INTELIGÊNCIA É MULTIFATORIAL

Niels, Aage e Tomas Bohr (Figura 2) são avô, filho e neto, respectivamente, e assumiram grande destaque em suas carreiras como cientistas. Será que Niels passou o **alelo** da inteligência para Aage, que por sua vez o passou para Tomas? A resposta é muito mais complexa do que sim ou não. Diversas pesquisas científicas sugerem que a genética é responsável por grande parte das diferenças cognitivas entre os indivíduos. Entretanto, não há um único **gene** que seja responsável pela inteligência. Sabemos que, dos aproximadamente 20 mil genes que os seres humanos têm, cerca de mil estão associados ao **neurodesenvolvimento** e, portanto, contribuem para o desempenho cognitivo. Cada um deles apresenta uma pequena contribuição para a composição do que chamamos de inteligência.

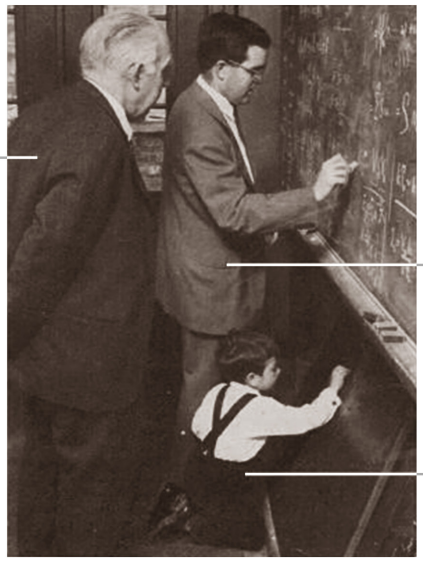
Alelo – alelos são formas alternativas de um mesmo gene.

Gene – Sequência de DNA contendo informações necessárias para a produção de proteínas ou regulação de processos biológicos.

Neurodesenvolvimento – desenvolvimento físico e psíquico do sistema nervoso.

Por outro lado, Niels, Aage e Tomas Bohr também compartilharam o mesmo ambiente. Como saber se foi a herança genética ou a similaridade no ambiente que promoveu o alto desempenho científico dos três? Em combinação com os fatores genéticos, fatores ambientais também são importantes no desenvolvimento da inteligência. Diz-se então que a inteligência tem um padrão de **mecanismo multifatorial** (Figura 3), isto é, há vários genes e também muitas variáveis ambientais que em conjunto influenciam o neurodesenvolvimento e o desempenho intelectual de cada indivíduo.

Mecanismo multifatorial – diz respeito a características cujas manifestações resultam de interações entre múltiplos fatores genéticos e ambientais.



Niels Bohr
Físico teórico dinamarquês e ganhador do Prêmio Nobel de Física de 1922

Aage Bohr
Físico nuclear e ganhador do Prêmio Nobel de Física de 1975

Tomas Bohr
É professor de Física na Universidade Técnica da Dinamarca

Figura 2.
Niels Bohr, seu filho Aage Bohr e Tomas Bohr, filho de Aage. Os três cientistas desenvolveram importantes pesquisas em seus campos de estudo dentro da Física.

Sabe-se que um ambiente propício para estudos permite um maior aprendizado e desempenho intelectual. Fatores como o relacionamento da criança com sua família, recursos educacionais, desenvolvimento de atividades extracurriculares, boa nutrição, entre outros, influenciam diretamente o desempenho cognitivo do indivíduo e, como consequência, potencializam suas chances de obter bons resultados em testes de inteligência.

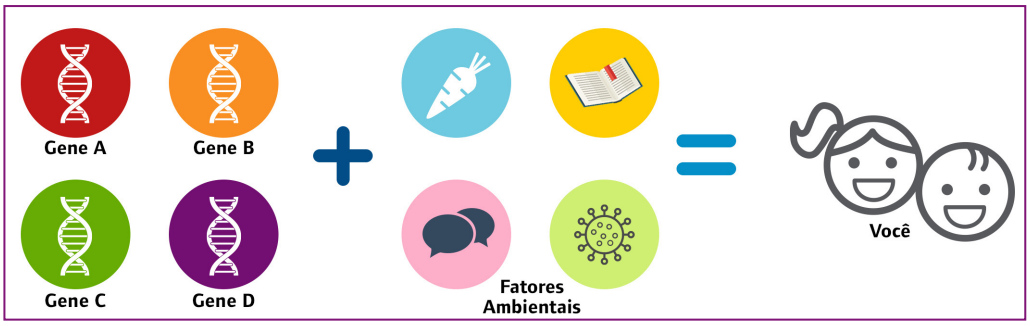


Figura 3.
Modelo de mecanismo multifatorial, no qual o **fenótipo** é dado pela atuação conjunta de fatores genéticos e ambientais. Diversos genes em associação com fatores ambientais como, por exemplo, nutrição, educação, interações sociais e infecções, têm impacto no desenvolvimento cognitivo. Então, qual é a contribuição da genética e do ambiente para a inteligência em termos quantitativos? Muitos estudos mostram que fatores genéticos são responsáveis por aproximadamente 70% da variação de QI observada entre indivíduos. No entanto, essa contribuição varia dependendo da idade e das condições socioeconômicas da população averiguada.

ENTENDENDO O QUE É A DEFICIÊNCIA INTELECTUAL

A deficiência intelectual é uma condição que afeta de 1 a 3% da população global, trazendo um alto custo financeiro, social e emocional, afetando tanto os sistemas de saúde quanto as famílias dos indivíduos afetados, já que é uma condição irreversível; os tratamentos disponíveis visam apenas amenizar os sintomas e melhorar a qualidade de vida. O termo anteriormente utilizado para essa condição era "retardo mental" mas, em 2004, a Organização Mundial da Saúde (OMS) aprovou a Declaração de Montreal sobre Deficiência Intelectual, que prioriza o uso do termo "deficiência intelectual".

Conforme mencionado no início deste artigo, considera-se que indivíduos com QI abaixo de 70 têm deficiência intelectual. A deficiência intelectual manifesta-se antes dos 18 anos e é caracterizada por limitações significativas no desempenho intelectual (raciocínio, aprendizagem e resolução de problemas) e em pelo menos duas habilidades do chamado comportamento adaptativo: comunicação, autocuidado, vida em casa e em comunidade, aptidão social, autonomia, saúde e segurança, desempenho acadêmico, lazer e trabalho. Casos de declínio cognitivo com início tardio não são considerados deficiência intelectual e se originam, sobretudo, por **processos neurodegenerativos** (como na **Doença de Alzheimer**, por exemplo).

Para crianças de até cinco anos, o termo preferencialmente usado para diagnóstico é "atraso global de desenvolvimento" ao invés de "deficiência intelectual". Esta distinção se faz porque os testes de QI só têm robustez quando aplicados após os cinco anos de idade.

No caso da deficiência intelectual leve, os fatores ambientais apresentam maior contribuição para o quadro clínico. Em países com menores níveis socioeconômicos, fatores ambientais, tais como: desnutrição, ausência de assistência pré-natal e baixa escolaridade resultam em uma prevalência maior de deficiência intelectual leve. Por outro lado, a deficiência intelectual de moderada a profunda é determinada preponderantemente por fatores genéticos. As frequências de deficiência intelectual moderada a grave são similares entre populações, pois frequentemente, nestes casos, a deficiência intelectual resulta de mutações ou alterações cromossômicas que perturbam gravemente o neurodesenvolvimento. No gráfico apresentado anteriormente (Figura 1) há um desvio à esquerda da curva mostrando um aumento de indivíduos com deficiência intelectual moderada a profunda em relação à distribuição dos valores de QI da população.

Frequentemente, indivíduos diagnosticados com deficiência intelectual também apresentam outras alterações do neurodesenvolvimento, como **transtorno do espectro autista**, **epilepsia** e distúrbios comportamentais ou psiquiátricos. Adicionalmente, a deficiência intelectual pode se apresentar isolada ou associada a **malformações congênitas**. Nesses casos em que há outros sinais físicos, diz-se que se trata de uma deficiência intelectual sindrômica.

Fenótipo – características observáveis de um indivíduo.

Processos neurodegenerativos – neurodegeneração é a perda progressiva da estrutura ou funcionamento dos neurônios.

Doença de Alzheimer – transtorno neurodegenerativo que causa a deterioração cognitiva e da memória, além de alterações comportamentais que comprometem a vida diária.

Transtorno do espectro autista – o termo engloba diferentes distúrbios do neurodesenvolvimento caracterizados por problemas de interação e comunicação social, padrões estereotipados e repetitivos de comportamento.

Epilepsia – distúrbio no qual os neurônios se comportam de maneira hiperexcitável, levando a convulsões e perda de consciência.

Malformações congênitas – qualquer defeito de nascença na constituição de um ou mais órgãos, que alterem sua estrutura física e, possivelmente, seu funcionamento.

HÁ FATORES GENÉTICOS E AMBIENTAIS QUE ISOLADAMENTE SÃO SUFICIENTEMENTE FORTES PARA CAUSAR DEFICIÊNCIA INTELECTUAL

Alguns fatores ambientais isolados (ou seja, sem estarem associados a fatores genéticos) têm impacto muito forte no neurodesenvolvimento, podendo levar então à deficiência intelectual (Tabela 1), como é o caso de algumas infecções maternas no período de gestação (por exemplo, **rubéola**), ou durante a infância (como a **meningite**); problemas durante o parto (**hipóxia neonatal**); lesões traumáticas decorrentes de acidentes; subnutrição dentre outros. Um dos fatores mais importantes como causa de deficiência intelectual é o consumo de álcool durante a gestação, que pode causar a síndrome alcoólica fetal; não há consumo de álcool considerado seguro durante a gravidez.

Rubéola – Infecção viral contagiosa conhecida pela sua erupção vermelha característica. É possível de prevenir por meio de vacinação. Durante os primeiros meses de gestação, pode causar aborto ou deficiências no feto.

Meningite – Inflamação das meninges (membranas que revestem o cérebro e medula espinhal). É geralmente causada por infecções, que podem ser virais, bacterianas ou fúngicas.

Hipóxia neonatal – Também chamada de "sofrimento fetal", consiste na diminuição ou ausência de absorção de oxigênio pelo feto.

Tabela 1.
Principais causas de deficiência intelectual.

Principais causas de deficiência intelectual		
Genéticas	**Ambientais**	
	Pré-natal	**Pós-natal**
Alterações cromossômicas (ex.: síndrome de Down) 21	Infecções (ex.: Zica)	Desnutrição
	Toxinas (ex.: álcool)	Trauma mecânico
	Irradiação (ex.: raios X)	
Mutações em genes únicos (ex.: síndrome do X Frágil) *FMR1*	Desnutrição materna	Infecção (ex.: meningite)
	Oxigenação insuficiente no parto O_2	Baixa escolaridade
	Trauma mecânico	

Alteração cromossômica – considerando que as células humanas têm 46 cromossomos (ou 23 pares), é a alteração do número ou morfologia dos cromossomos de uma célula.

Fatores genéticos são responsáveis por grande parte dos casos de deficiência intelectual moderada a profunda, sejam decorrentes de **alterações cromossômicas** (numéricas e estruturais), ou devido a **mutações** em genes específicos que têm grande impacto no neurodesenvolvimento. Estima-se que existam mais de 1000 genes que, quando mutados individualmente, estão relacionados à deficiência intelectual.

Mutação –qualquer alteração na sequência do DNA.

Trissomia – alteração genética na qual a célula tem três cópias de um cromossomo específico em vez de duas.

A **trissomia** do cromossomo 21 (Figura 4), isto é, a presença de três cópias do cromossomo 21, causa a **Síndrome de Down**, que é a causa genética mais comum de deficiência intelectual. Em cerca de 95% dos casos, a chance de haver recorrência na família é ínfima. Cabe dizer que a idade materna avançada no momento da concepção (a partir de 35 anos) é um fator de risco para a ocorrência de síndrome de Down.

Síndrome de Down – condição genética causada pela presença de um cromossomo 21 a mais. Causa uma variedade de sinais clínicos, incluindo deficiência intelectual.

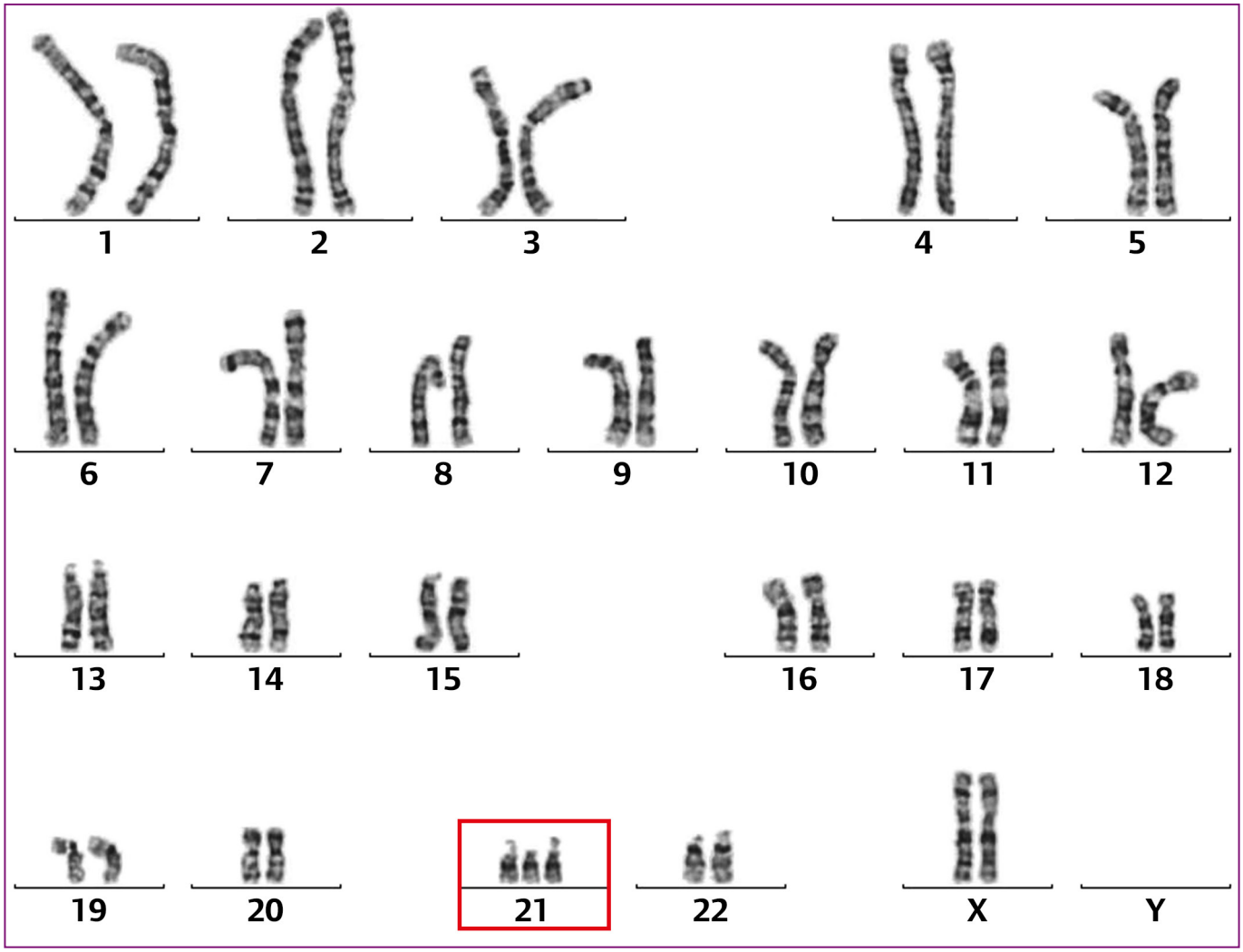


Figura 4. **Cariótipo** humano de uma menina (note a presença de dois cromossomos X e nenhum cromossomo Y) com trissomia do 21. Imagem cedida pelo Laboratório de Genética Humana do Instituto de Biociências da USP.

Cariótipo – conjunto de cromossomos de uma espécie, ordenados de acordo com seu número, forma e tamanho.

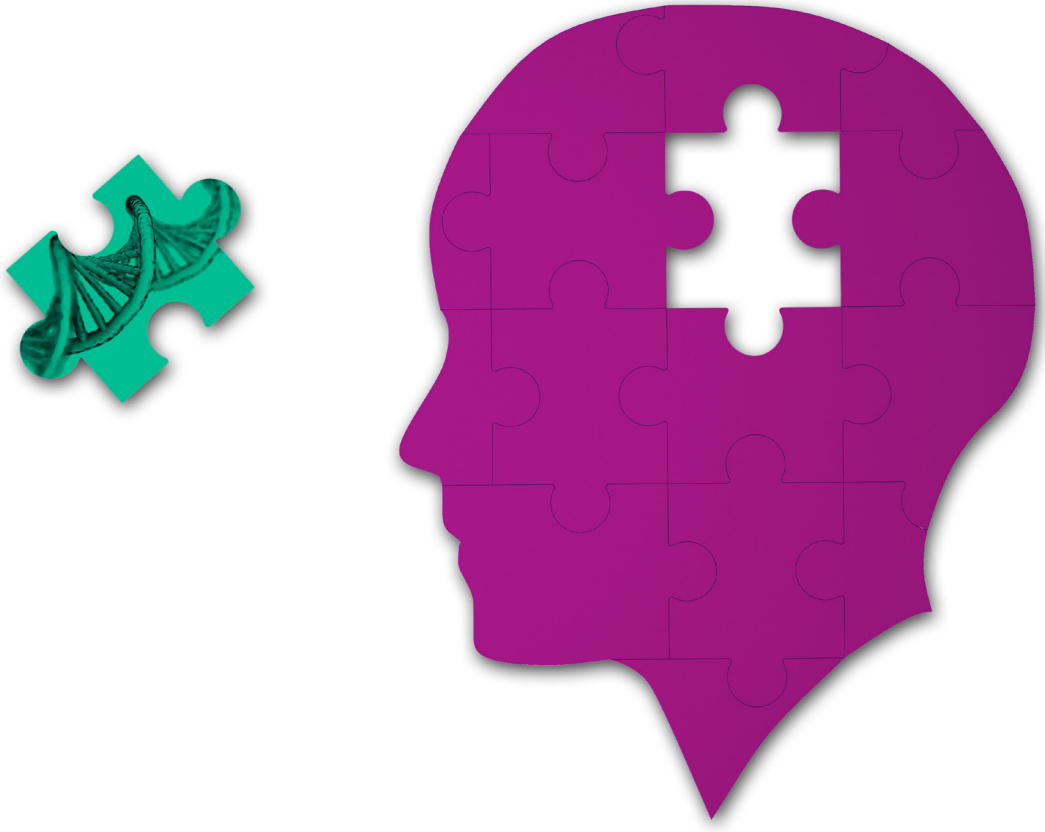
A segunda causa genética mais comum de deficiência intelectual – e primeira dentre os casos hereditários – é a Síndrome do X-Frágil. Esta síndrome é causada por mutações no gene *FMR1*, localizado no cromossomo X, e que exerce um papel importante no funcionamento dos neurônios. Por terem apenas uma cópia do cromossomo X, homens são mais frequentemente afetados. As mulheres podem não manifestar quadro clínico ou apresentar grau variável de deficiência intelectual; no entanto, independentemente do quadro clínico, mulheres portadoras de mutações no gene *FMR1* têm risco alto de terem filhos com a síndrome do X-Frágil.

QUAIS SÃO OS GENES ASSOCIADOS À DEFICIÊNCIA INTELECTUAL E QUAL O SEU PAPEL NO ORGANISMO?

Genes relacionados à capacidade cognitiva desempenham função relevante no nosso organismo e só causam problemas caso apresentem algum defeito e tenham sua função alterada ou anulada. Foi demonstrado, por exemplo, que o gene *MEF2C* atua no processo de aprendizagem e de formação de memória; mutações nesse gene estão relacionadas à deficiência intelectual.

Mencionamos aqui o exemplo do gene *FMR1*, que causa a Síndrome do X-Frágil, e demos algumas estimativas da quantidade de genes já associados à deficiência intelectual. Mas, afinal, o que eles têm em comum? Em linhas gerais, são genes que, embora desempenhem funções diferentes, apresentam papel essencial no cérebro em desenvolvimento. Assim, quando a função de algum deles está comprometida por uma mutação, o processo de formação e funcionamento do sistema nervoso é alterado, trazendo consequências como a deficiência intelectual.

As pesquisas que se dedicam a identificar quais são os fatores genéticos e também ambientais associados à deficiência intelectual são essenciais para que se possa entender melhor o funcionamento do cérebro, incluindo os processos de aprendizagem, formação de memória e diferentes tipos de inteligência. No futuro, esperamos que essas pesquisas possam ser base para o desenvolvimento de possíveis terapias para indivíduos com deficiência intelectual.



REFERÊNCIAS

BHATE, S.; WILKINSON, S. Aetiology of learning disability. *Aetiology and epidemiology*, v. 5, n. 9, p. 298-301, 2006.

ROPERS, H. H. Genetics of early onset cognitive impairment. *Annual Review of Genomics and Human Genetics*, v. 11, p. 161-187, 2010.

Universia Brasil (2014). *QI e hereditariedade: qual é a relação?* Disponível em: <https://noticias.universia.com.br/ciencia-tecnologia/noticia/2014/04/09/1094577/qi-hereditariedade-relaco.html>. Acesso em: 22 set. 2019.

PARA SABER MAIS

REGAN, R.; WILLATT, L. Mental retardation: definition, classification and etiology. *Monographs in Human Genetics*, v.18, p. 16-30, 2010.

Capítulo 4.2

Modelos animais no estudo de doenças genéticas humanas

VASCONCELOS, F. T. G. R.; CARVALHO, L. M. L; SOUZA, L. S.; BARBOSA, I. N.; DÁVILA, C. K.; ZATZ, M.; VAINZOF, M. Modelos animais no estudo de doenças genéticas humanas. **Genética na Escola.** 2022. ISSN: 1980-3540. DOI: 10.55838/1980-3540.ge.2022.420

Genética na Escola – ISSN: 1980-3540 201

Modelos animais no estudo de doenças genéticas humanas

Felipe Tadeu Galante Rocha de Vasconcelos[1*], Laura Machado Lara Carvalho[1*], Lucas Santos e Souza[1*], Igor Neves Barbosa[2*], Carolini Kaid Dávila[3], Mayana Zatz[3], Mariz Vainzof[3]

[1] Doutorando(a) do Programa de Pós-graduação em Ciências Biológicas (Biologia-Genética), Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP
[2] Mestrando do Programa de Pós-graduação em Ciências Biológicas (Biologia-Genética), Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP
[3] Pesquisadora do Departamento de Genética e Biologia Evolutiva, Centro de Estudos do Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo, SP
[*] Felipe, Laura, Lucas e Igor contribuíram igualmente para o artigo.

Autor para correspondência - mvainzof@usp.br

Palavras-chave: modelos animais, genética humana, doenças genéticas, *zebrafish*, camundongo, cão, 3R's

Modelos animais são espécies não humanas que têm sido utilizadas em inúmeras pesquisas científicas, permitindo a compreensão do funcionamento de organismos vivos complexos, além do desenvolvimento de tecnologias aplicadas à saúde humana, como medicamentos, soros e vacinas. Esses modelos são utilizados na mimetização de doenças, na expectativa de se poder posteriormente utilizar o conhecimento gerado em humanos. Nesta leitura você poderá conhecer alguns dos principais modelos utilizados em experimentação, implicações éticas e legais do seu uso, além de exemplos de aplicações que contribuíram significativamente para a sociedade, principalmente no campo da genética humana.

Por que utilizar animais para estudar doenças humanas?

Ao longo da história, os estudos com animais permitiram melhor compreensão de aspectos de fisiologia, patologia, **farmacologia** e genética. Os modelos animais dão aos pesquisadores a possibilidade de experimentar em situações controladas e mimetizar condições biológicas complexas de um ser vivo, contribuindo para encontrar soluções para questões de saúde humana. A Figura 1 apresenta uma linha do tempo com alguns exemplos de avanços importantes a partir de estudos com animais no campo das ciências biomédicas.

Os experimentos com modelos animais contribuem para o conhecimento científico e melhorias na qualidade de vida, através do desenvolvimento de novos medicamentos e vacinas, bem como novas técnicas cirúrgicas e protocolos de administração de fármacos. Praticamente 90% das pesquisas laureadas com o Prêmio Nobel de Fisiologia e Medicina utilizaram experimentos com animais.

Em genética, costuma-se usar modelos animais para identificar a função de genes e entender como estão relacionados às doenças e a outras características fenotípicas. A partir da década de 1980, com o aperfeiçoamento das tecnologias de **biologia molecular** e **edição gênica**, houve uma explosão nas possibilidades de manipulação genética de animais modelo. Assim, pode-se, por exemplo, silenciar um gene específico e verificar o efeito resultante em escala molecular e fisiológica, como também observar as características visíveis a olho nu. Dessa forma, é possível avaliar se um **gene candidato** realmente é causador de uma doença. No extremo oposto, podemos aumentar a expressão, ou mesmo adicionar cópias de um ou mais genes, inclusive genes não pertencentes à espécie, e observar as consequências. Outras possibilidades são os **estudos funcionais** que investigam os mecanismos fisiopatológicos das doenças, além dos testes de novos tratamentos.

Com o advento recente de uma nova técnica muito eficaz e específica de edição gênica (**CRISPR**), aumentou consideravelmente o potencial de se obter resultados promissores para a sociedade a partir de estudos de doenças genéticas humanas utilizando modelos animais. Sobre a técnica de CRISPR, esta revista publicou dois artigos esclarecedores em 2017 (Carli e colaboradores; Souza e Pereira), de maneira que você pode consultá-los para entender melhor os mecanismos moleculares envolvidos e as implicações sociais.

Farmacologia - ciência que estuda como diferentes substâncias, os fármacos, interagem com o organismo e que resposta desencadeiam.

Biologia molecular - ciência que busca compreender a estrutura e o funcionamento de moléculas biológicas, em especial as que constituem o material genético.

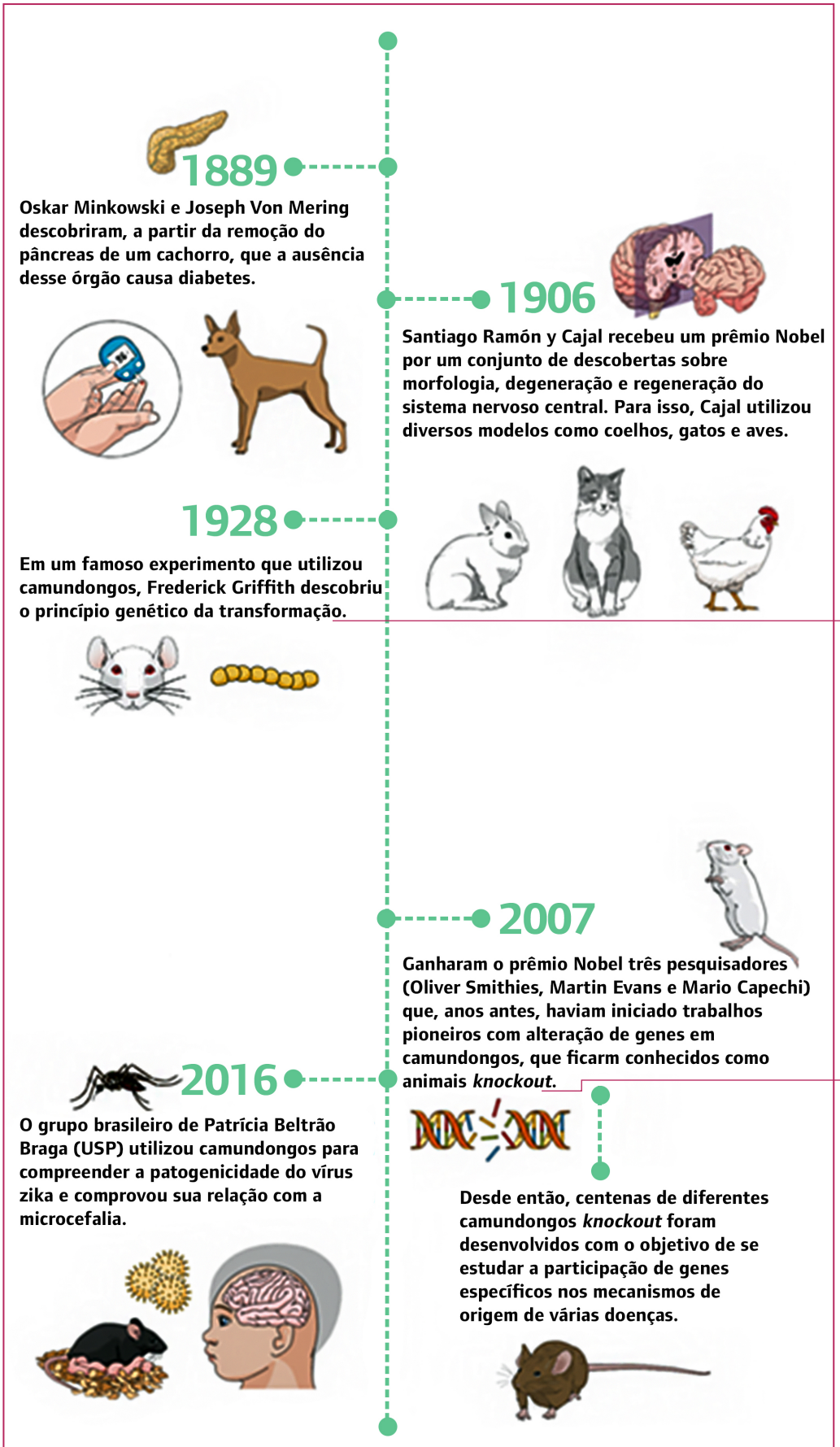
Edição gênica - alteração artificial de sequência de DNA em organismo vivo.

Gene candidato - considera-se como candidato um gene para o qual há evidências preliminares de que possa estar relacionado mecanisticamente à manifestação de determinada doença ou característica. A relação do gene candidato com uma doença é uma hipótese que precisa ser confirmada ou refutada a partir de evidências experimentais complementares.

Estudos funcionais - no âmbito da genética, são estudos que buscam entender mecanisticamente a relação entre genes e seus fenótipos.

CRISPR - técnica de edição gênica que mimetiza mecanismos bacterianos de clivagem de DNA. A clivagem é seguida de reparo do DNA pela célula utilizada no experimento, o que frequentemente resulta em mutações. A sigla significa Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas (do inglês *Clustered Regularly Interspaced Short Palindromic Repeats*), em referência ao lócus bacteriano que desencadeia tal mecanismo de combate a DNAs exógenos.

Genética na Escola – ISSN: 1980-3540



Princípio genético da transformação - também chamado de transformação bacteriana. Por esse processo, ocorre a captura de pedaços de DNA do ambiente e incorporação ao genoma da bactéria.

Animais *knockout* - animais em que ocorreu a inativação ou deleção de um gene.

Figura 1.
Exemplos de experimentos com modelos animais que foram importantes para a humanidade.

Genética na Escola – ISSN: 1980-3540

Figura 2.
Monumento ao camundongo de laboratório, em um parque em frente ao Instituto de Citologia e Genética da Academia de Ciências da Rússia, na cidade de Novosibirsk. Observa-se um camundongo tricotando uma molécula de DNA, que representa a gratidão da humanidade pelo animal que é utilizado em pesquisas genéticas para entender mecanismos biológicos e desenvolver novas drogas e tratamentos para doenças. Autora: Irina Gelbukh (2013). Licença *Creative Commons* atribuída.

Apresentaremos a seguir três modelos animais muito utilizados nas pesquisas de doenças genéticas humanas: o *zebrafish*, o camundongo e o cão.

Zebrafish ou peixe-zebra (*Danio rerio*)

O *zebrafish*, também conhecido no Brasil como paulistinha, é um peixe de água doce que tem origem asiática.

À primeira vista, o *zebrafish* não se parece muito conosco, mas as aparências enganam. Cerca de 70% dos genes humanos têm pelo menos um **ortólogo** em *zebrafish* e, vários órgãos desse modelo são bastante similares aos dos humanos. Isso quer dizer que, se nos perguntarmos para que servem determinados genes humanos, muitas vezes estudá-los no modelo *zebrafish* poderá nos levar a respostas. Além disso, trata-se de um modelo relativamente barato e de rápido desenvolvimento. Em cerca de três meses, o *zebrafish* atinge a maturidade sexual e cada casal pode produzir de 200 a 300 ovos semanalmente. Outra vantagem é que os embriões são transparentes e se desenvolvem fora do corpo, o que permite a visualização macroscópica para estudo desse desenvolvimento.

Não só o desenvolvimento, mas também a fecundação ocorre fora do corpo, fazendo com que a edição gênica seja metodologicamente facilitada. Em contraste, a fecundação e o desenvolvimento de mamíferos ocorrem dentro do corpo das fêmeas, dificultando sua manipulação. A edição gênica é realizada a partir da injeção em ovos de substâncias necessárias para modificar parte específica do gene de interesse. Assim, é possível obter animais com mutação em um gene de interesse, o que é importante, por exemplo, para desenvolver um modelo de *zebrafish* com determinada doença, no qual podem ser testados novos tratamentos que poderão em longo prazo ser aplicados em humanos acometidos pela mesma patologia.

Ortólogo - genes que estão em espécies diferentes, mas possuem funções semelhantes e origem evolutiva comum.

Genética na Escola – ISSN: 1980-3540

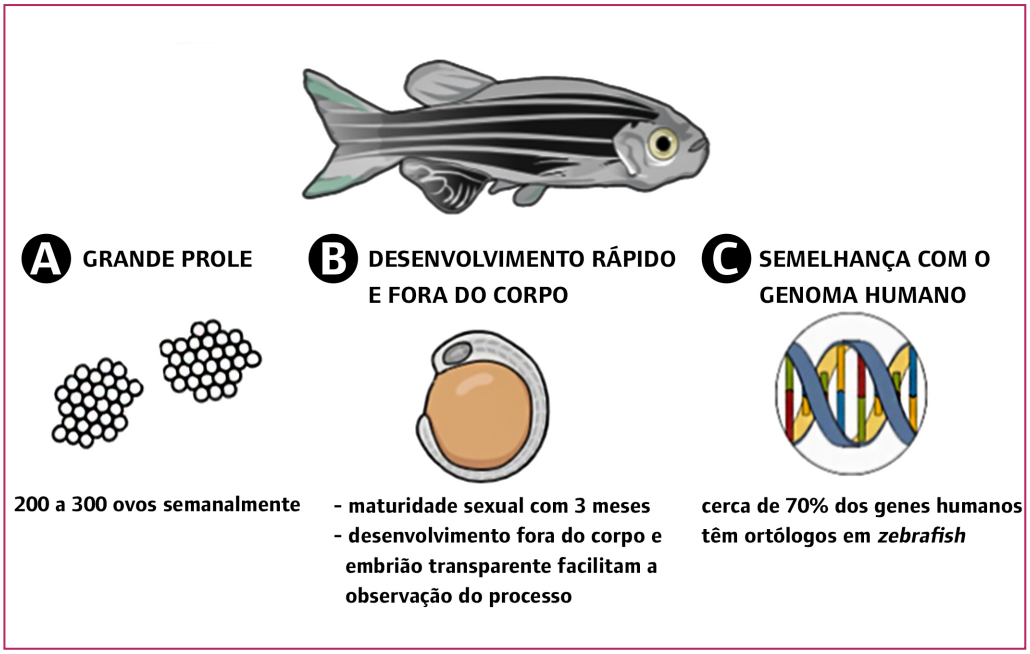


Figura 3.
Alguns benefícios da utilização do *zebrafish* como modelo animal.

Outra utilidade é validar genes candidatos a doenças. Por exemplo, se for observado que alguns recém-nascidos com microcefalia têm mutação em certo gene, essa causalidade pode ser verificada em *zebrafish*. Pode-se editar o gene candidato no embrião de *zebrafish* para verificar se o peixinho expressará tal fenótipo (cabeça pequena) (Figura 4). Por exemplo, mutações nos genes *map11* e *cdc6* induzidas pela metodologia CRISPR causaram microcefalia em *zebrafish* e o fenótipo em humanos é correspondente.

O *zebrafish* tem sido utilizado, por exemplo, em estudos sobre **distrofias musculares**, obesidade, câncer, **albinismo ocular**, microcefalia, **fendas orofaciais**, doenças hepáticas, surdez etc.

Distrofias musculares - grupo de doenças caracterizadas por fraqueza progressiva e perda de massa muscular.

Albinismo ocular - forma de albinismo que leva à falta de pigmentação da íris, prejudicando a visão.

Fendas orofaciais - malformações no palato (céu da boca) e/ou lábio.

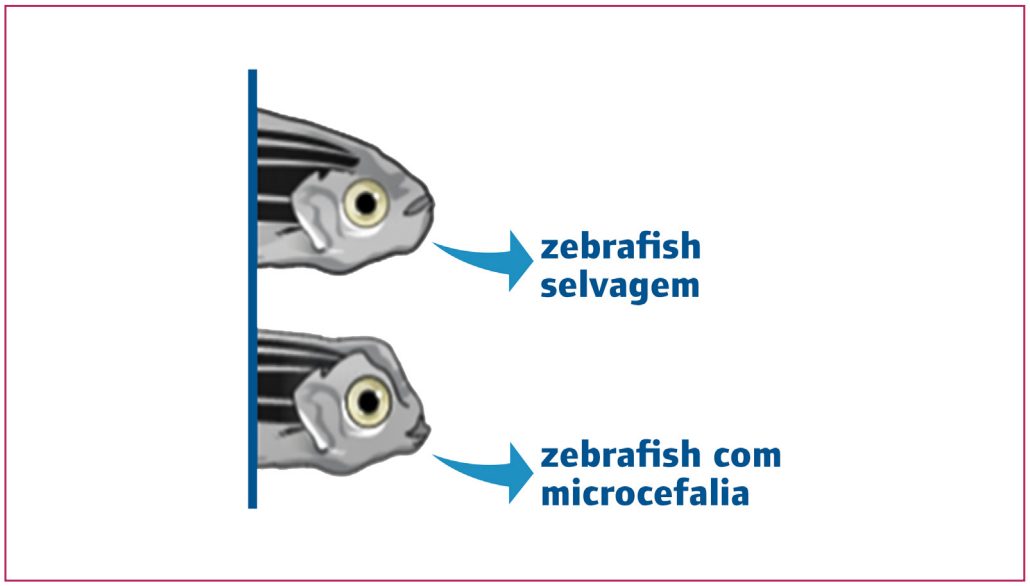


Figura 4.
Microcefalia em *zebrafish*.

Genética na Escola – ISSN: 1980-3540

Roedores

O uso de roedores como modelo em pesquisa experimental já data de mais de um século e há muitas razões para isso: disponibilidade, facilidade de manuseio e cuidado, alta prolificidade (com gestações que resultam em vários filhotes por ninhada), e relativo baixo custo. As **linhagens selvagens** de camundongo, por exemplo, têm entre cinco e dez filhotes por ninhada e cada fêmea pode ter até cinco ninhadas ao longo da vida.

Entre os roedores utilizados estão os camundongos, ratos, hamsters, porquinhos-da-índia e gerbis. Mas os mais utilizados são ratos e camundongos.

Linhagens - grupos de indivíduos de uma espécie cujos organismos apresentam a mesma composição genotípica e que geram descendentes com essa mesma composição quando entrecruzados.

Linhagens selvagens - formas típicas, como ocorrem na natureza.

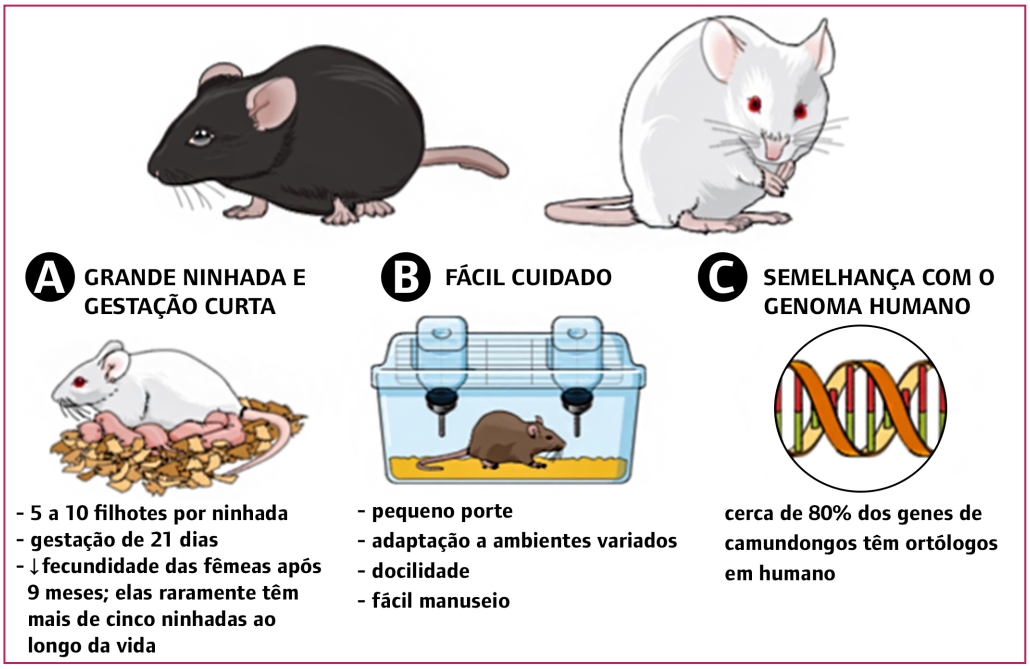


Figura 5.
Benefícios de se utilizar o camundongo (*Mus musculus*) como modelo animal.

As linhagens dos animais de laboratório podem ser divididas em dois grandes grupos:

1. heterogênicas - *outbred*, não consanguíneas
2. isogênicas - *inbred* ou consanguíneas

Animais não consanguíneos são os que apresentam na constituição genética uma alta taxa de heterozigose (~99%), o que faz com que seja mantida uma grande diversidade genética e fenotípica num grupo de animais.

Um animal isogênico é o produto de pelo menos 20 gerações consecutivas do acasalamento entre irmãos, ou pais e filhos. Utilizando esse tipo de acasalamento, conseguimos obter um índice de homozigose de 99%, o que torna tais animais os mais semelhantes entre si que se pode obter.

As linhagens isogênicas são de especial interesse para os grupos de pesquisa pois apresentam um conjunto único de características. Mas, por que isso é interessante? Imagine que se pretende comparar um grupo de animais em que foi utilizada determinada terapia com um grupo em que não se utilizou. É importante, nesse caso, que ambos os grupos sejam geneticamente parecidos, porque variações genéticas poderiam interferir nos resultados do experimento. Ao se comparar um grupo de animais que teve um gene editado com um grupo não editado, deve-se seguir o mesmo raciocínio: é importante que os animais que os originaram tenham uma constituição genética muito próxima pois, do contrário, variantes em outros genes que não o editado podem interferir fenotipicamente, atrapalhando a análise dos resultados do experimento.

Genética na Escola – ISSN: 1980-3540

Doenças neurodegenerativas - doenças nas quais ocorre perda progressiva da estrutura e do funcionamento dos neurônios.

Alzheimer - transtorno que causa a deterioração cognitiva e da memória, além de alterações comportamentais que comprometem a vida diária.

Neurodesenvolvimento - desenvolvimento físico e psíquico do sistema nervoso.

Imunodeprimido - indivíduo que tem baixa resposta imunológica devido a uma condição patológica.

Existem mais de mil linhagens diferentes de camundongos. E muitas delas são modelos para doenças humanas como câncer, **doenças neurodegenerativas** (como **doença de Huntington** e o **Alzheimer**); **neuromusculares** (como as distrofias musculares); do **neurodesenvolvimento** (como autismo); imunodeficiências, **doenças autoimunes**, obesidade, entres outras.

Uma outra razão pela qual camundongos são amplamente utilizados como modelos em pesquisa experimental é a alta similaridade com o genoma do ser humano, havendo semelhança de aproximadamente 85% nas regiões codificadoras de proteínas. Alguns genes apresentam quase 100% de equivalência. Além disso, cerca de 80% dos genes desses dois genomas são ortólogos.

Uma linhagem relevante nas pesquisas biomédicas é a *nude* (Figura 6). Ela apresenta uma mutação espontânea em homozigose no gene *Foxn1*, o que acarreta defeitos na produção das células do timo - glândula localizada no tórax que é responsável pela maturação de células do sistema imune. Em decorrência da ausência do timo, trata-se de um animal **imunodeprimido** e isso faz com que esse modelo seja amplamente utilizado para estudar diversos mecanismos, sem interferência do sistema imune. Além disso, apresenta deficiência na produção dos pelos, por isso a ausência de pelagem.

Doença de Huntington - doença genética que acomete progressivamente o sistema nervoso central, causando movimentos involuntários, alterações comportamentais e declínio cognitivo.

Neuromusculares - doenças neuromusculares são aquelas que afetam, direta ou indiretamente, a função muscular, isto é, incluem patologias intrínsecas do músculo e também aquelas do sistema nervoso que afetam o comando das contrações musculares.

Doenças autoimunes - são caracterizadas por respostas imunológicas contra estruturas do próprio organismo.

Figura 6.
Camundongo *nude*.

Os pesquisadores aproveitam que os camundogos *nude* não apresentam uma boa resposta imunológica para fazer experimentos sobre câncer. É possível inserir um pedaço de tumor humano no animal, pois o organismo do modelo *nude* não reage adequadamente para combatê-lo. Assim, pode-se observar tanto o desenvolvimento de tumores derivados de pacientes humanos quanto testar a resposta a diferentes tratamentos.

Mutações homozigóticas no gene *FOXN1* em humanos também são muito prejudiciais, causando desde problemas na pele e **alopecia congênita** até deficiência do sistema imunológico, o que torna essas pessoas muito suscetíveis a infecções.

Existem algumas linhagens de camundongos que são utilizadas para estudar diferentes doenças neuromusculares. Uma delas é a linhagem Dmd^{mdx} para **Distrofia Muscular de Duchenne**, que tem uma mutação espontânea do tipo **mutação sem sentido** no gene *Dmd* **murino**, o que faz com que não haja a produção da proteína **distrofina** no músculo esquelético desses animais. Entretanto, diferentemente dos humanos, nos camundongos Dmd^{mdx} o quadro clínico é muito leve e a sobrevida é bastante próxima da normal. Nos humanos, a Distrofia Muscular de Duchenne se manifesta em meninos desde os primeiros anos de vida e se caracteriza pela **degeneração** muscular progressiva, com perda da força muscular com uma série de consequências que comprometem a expectativa e a qualidade de vida.

Além dos modelos espontâneos de camundongo, também os modelos induzidos são muito úteis em pesquisas biomédicas. Uma das abordagens com esse tipo de modelo é o de validação de genes candidatos a diversas doenças. Um exemplo é o autismo, que é uma condição que pode ter origem monogênica ou **multifatorial**. Quando se tem um gene candidato para determinada forma de autismo, pode-se editar esse gene em camundongos e verificar se os animais apresentam características semelhantes às dos pacientes humanos. Um dos testes que pode ser feito para avaliar o comportamento social dos camundongos é o chamado teste de sociabilidade três câmaras (Figura 7). Nele, o camundongo central é observado pelos pesquisadores e suas ações são quantificadas. Entre os parâmetros avaliados, estão o tempo de exploração da gaiola, número de vezes e tempo gasto na interação com os outros camundongos etc.

Alopecia - também chamada de calvície, é a perda ou ausência de cabelo.

Mutação sem sentido - variante genética na qual um códon que levaria à inclusão de um aminoácido à cadeia polipeptídica é alterado, passando a sinalizar o término da tradução, isto é, torna-se um códon de parada (no RNA: UAG, UAA ou UGA).

Distrofina - proteína muscular que liga o citoesqueleto da célula do músculo a um complexo de proteínas de membrana que, por sua vez, interage com o colágeno extracelular.

Multifatorial - diz respeito a características cujas manifestações resultam de interações entre múltiplos fatores genéticos e ambientais.

Congênita - condição que, independentemente da sua causa, já se apresenta ao nascimento.

Distrofia Muscular de Duchenne - doença genética causada por variantes deletérias no gene da distrofina. Trata-se do tipo mais comum de distrofia muscular.

Murino - referente ao grupo *Murinae*, do qual fazem parte os camundongos.

Degeneração - deterioração de um tecido ou órgão.

Genética na Escola – ISSN: 1980-3540

Teste de sociabilidade três câmaras

Normalmente, os roedores preferem passar mais tempo com outro roedor e investigam um novo indivíduo mais do que um familiar (novidade social). Com base nessa tendência, este teste é útil para avaliar o comportamento social dos animais. Nele, a gaiola é dividida em três câmaras com aberturas entre elas. Após a habituação à câmara vazia (a do meio), o animal em análise encontra outros indivíduos nas outras câmaras. O tempo gasto cheirando cada animal, o tempo gasto em cada câmara e o número de entradas são registrados e analisados.

É aplicável, por exemplo, a estudos envolvendo autismo, efeito de medicamentos na sociabilidade, comportamento parental e interesse em novidades sociais.

Figura 7.
Teste de sociabilidade.

Cão (*Canis lupus familiaris*)

Por que utilizar o melhor amigo do homem em pesquisas? Existem pelo menos 450 doenças que surgem espontaneamente em cães e 80% delas também afetam humanos. Além disso, em muitos casos, a manifestação dessas doenças nos cães é comparável ao que ocorre em humanos - considerando sua complexidade social, comportamental, e do ambiente em que estão inseridos - e, assim sendo, o **potencial translacional** do modelo canino para o estudo de doenças humanas é maior que de outros modelos que são mais utilizados como, por exemplo, camundongos e *zebrafish*.

Potencial translacional - potencial de gerar informações ou ferramentas aplicáveis à saúde humana.

A resposta imune dos cães tem maior similaridade com a do humano do que outros modelos. Além disso, o tamanho maior dos cães torna-os fisiologicamente mais próximos dos humanos. Isso contribui, por exemplo, para mimetizar melhor a **biodistribuição** de fármacos em um tratamento experimental. Afinal, a biodistribuição tende a ser facilitada no corpo de animais menores. Muitos medicamentos testados em pequenos animais podem não alcançar seus alvos moleculares em humanos com a mesma eficácia, já que somos muito maiores, o que é uma barreira translacional.

Biodistribuição - passagem de um fármaco da corrente sanguínea para os tecidos. É afetada por fatores fisiológicos e propriedades físico-químicas do próprio fármaco.

Genética na Escola – ISSN: 1980-3540

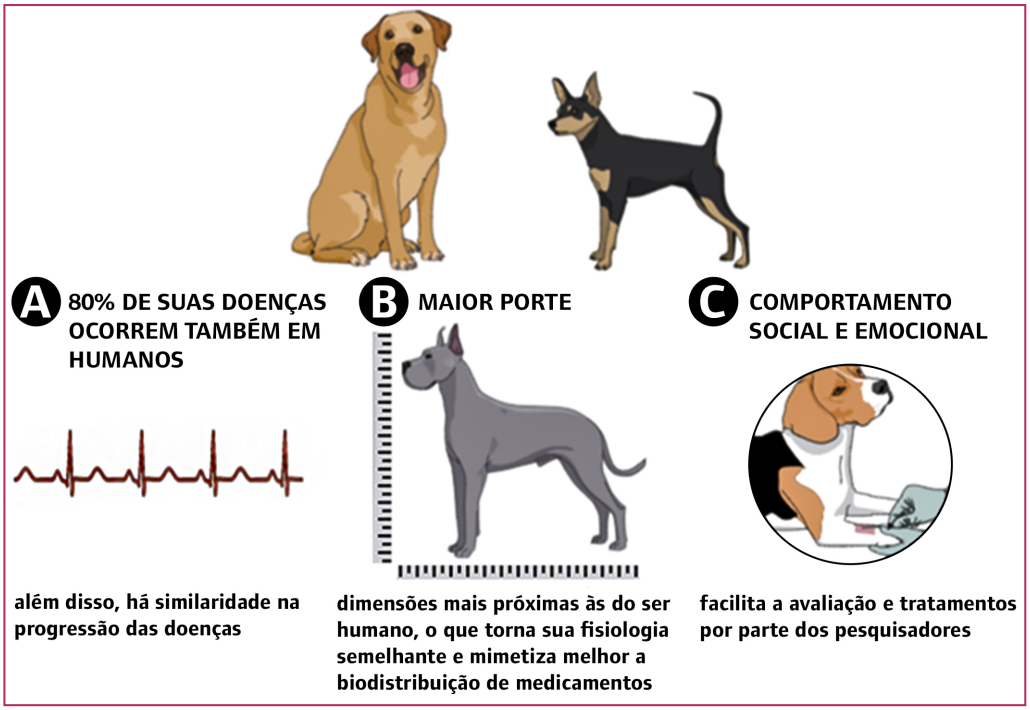


Figura 8.
Alguns benefícios de se utilizar o cão como modelo animal.

A variação genética entre cães de raças diferentes é maior que entre os da mesma raça. Assim sendo, cada raça de cão possui uma alta homogeneidade de fenótipo, em grande parte por causa do processo de **seleção artificial** conduzida por humanos para gerar animais com características desejadas em aparência e comportamento. No entanto, esse processo de seleção imposto por criadores acabou gerando também uma alta frequência de doenças genéticas em cães ditos "de raça/*pedigree*". Isso acontece porque a consanguinidade propicia uma maior probabilidade de ocorrer em homozigose alelos recessivos potencialmente patogênicos, ou seja, cães ditos "de raça" são mais propensos a certas doenças (as autossômicas recessivas) que os sem raça definida (ditos "vira-latas"). Isso abre oportunidades na pesquisa científica para elucidar a relação entre genes e doenças.

Doenças são causadas por alterações genéticas, condições ambientais, ou por uma interação entre variantes genéticas e o ambiente. Os cães domésticos compartilham, em grande parte, do mesmo ambiente que os seres humanos e, em consequência, disso podem desenvolver diversas patologias que nos afetam, já que estão expostos a microrganismos, poluentes, toxinas e outras variáveis ambientais. O mesmo não ocorre com animais mantidos em **biotérios** especializados com ambiente rigidamente controlado (quanto à temperatura, dieta, luminosidade, condições sanitárias etc.). Embora em muitos contextos científicos o ambiente controlado seja positivo para o sucesso do experimento, é importante avaliar também uma intervenção em um ambiente mais comparável ao do ser humano e, nesse caso, o cão doméstico tende a reproduzir melhor o que acontece conosco.

Biotérios - ambientes em que são mantidos animais em condições adequadas à utilização em experimentos científicos ou produção biotecnológica (ex.: de vacinas e soros).

Seleção artificial - processo pelo qual humanos selecionam animais ou plantas com características desejáveis para efetuar cruzamentos, com o objetivo de obter organismos com fenótipos desejados.

A sociabilidade e o comportamento emotivo dos cães facilitam a avaliação clínica das condições de saúde e intervenções terapêuticas, pois eles expressam mais claramente dor e incômodo. Essas características também permitem que os próprios donos consigam relatar a melhora ou a piora dos cães, o que também pode facilitar a detecção de efeitos colaterais e efetividade de tratamentos a longo prazo.

Apesar das vantagens apresentadas, os estudos com cães não são tão frequentes e as principais razões são a dificuldade de se obter o número necessário de indivíduos com a condição estudada (o que é importante para que os resultados sejam estatisticamente significativos) e o alto custo de manutenção de

um canil - desde a alimentação, contratação de pessoal especializado (como médicos veterinários e cuidadores), à alocação de espaço físico adequado e limpeza.

Um bom exemplo de modelo canino para o estudo de uma doença humana é a Distrofia Muscular em Golden Retrievers (GRMD), que ocorre por mutação espontânea no gene da distrofina. Ao contrário do camundongo *Dmd*^{mdx}, que tem um quadro leve, os cães GRMD têm uma progressão clínica comparável com a Distrofia Muscular de Duchenne em humanos. Em ambos os casos (cão e humano), os afetados têm perda progressiva de massa muscular e a doença culmina na morte por insuficiência respiratória ou cardíaca em idade prematura. Por isso, esse modelo canino mimetiza melhor tal condição humana que o modelo *Dmd*^{mdx} murino (Figura 9).

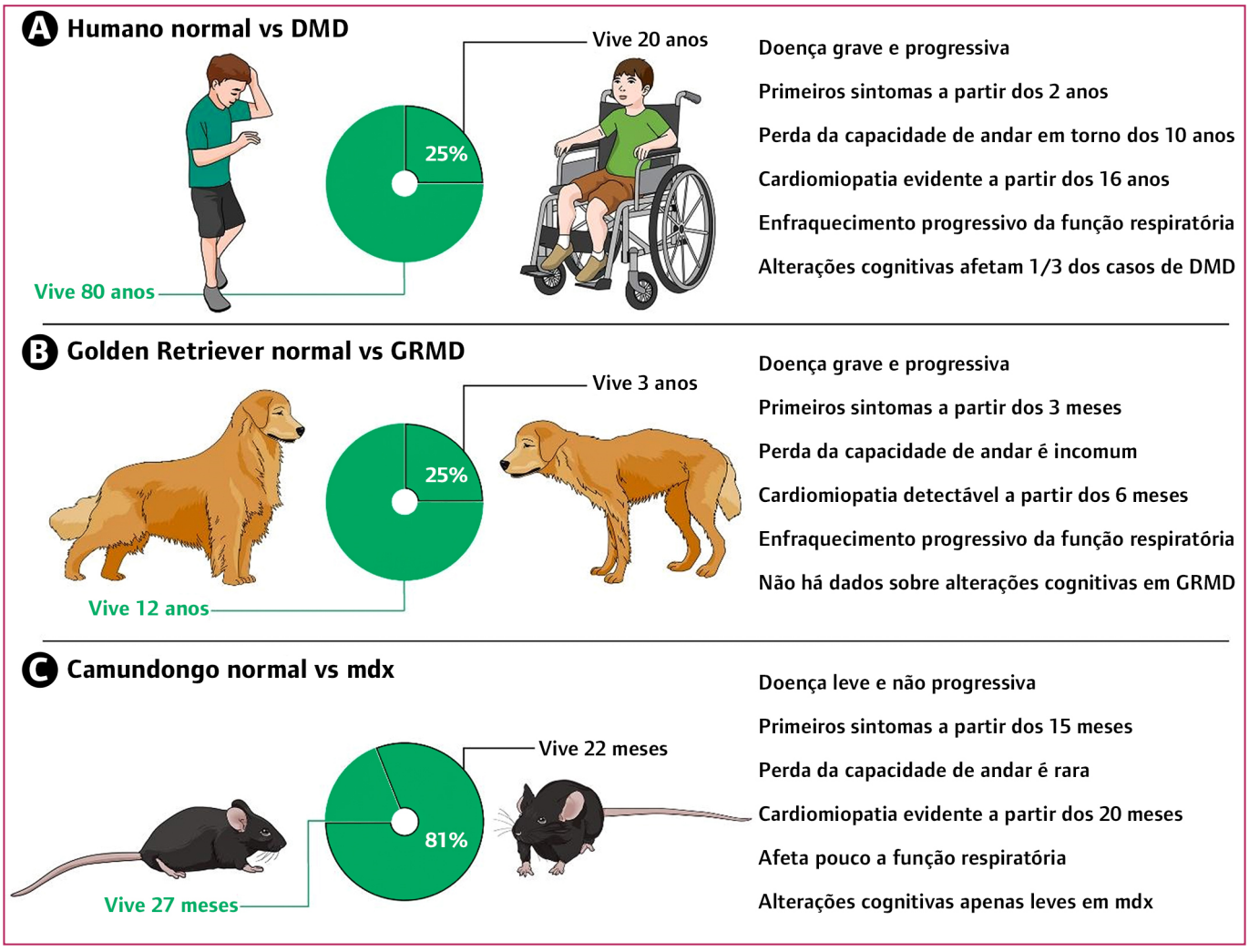


Figura 9.
Análise comparativa entre modelos de Distrofia Muscular de Duchenne.

Recentemente, percebeu-se que uma variante genética rara no gene *JAGGED1*, quando presente no modelo GRMD, é protetiva para a progressão da doença, isto é, animais com variante patogênica no gene da distrofina tinham uma melhor qualidade de vida (mais semelhante à dos cães selvagens) se apresentassem também a variante em *JAGGED1*. Os pesquisadores agora procuram estudar o gene *JAGGED1* para tentar encontrar soluções técnicas terapêuticas para a Distrofia Muscular de Duchenne em humanos. Esse tipo de descoberta pode ajudar no desenvolvimento de novas terapias para esta e outras doenças, tendo como alvo genes que podem impedir que o indivíduo desenvolva a patologia, os chamados "genes protetores".

Os cães também são um interessante modelo para avaliar tratamentos de câncer. A partir de um estudo de **necropsias** de 2000 cães descobriu-se que 23% do total e 45% daqueles acima de 10 anos morreram de câncer. Outro estudo, realizado pelos grupos de Mayana Zatz e Oswaldo Keith Okamoto, da Universidade de São Paulo, avaliou o potencial do vírus Zika para combater células tumorais do sistema nervoso central e os cães foram um dos modelos utilizados. Sabia-se até então que esse tipo de vírus causa microcefalia em bebês durante a gestação por atacar **células progenitoras neuronais**, mas que não há prejuízo neuronal em indivíduos adultos infectados. O resultado deste estudo mostrou que o vírus Zika é capaz de reduzir tumores cerebrais, o que pode ser uma possível terapia para cânceres no cérebro.

Como escolher um modelo animal?

Cada modelo animal tem suas particularidades (Figura 10). Uma escolha criteriosa do modelo animal é um ponto fundamental do planejamento da pesquisa. A decisão por um modelo específico depende dos objetivos do projeto e deve ser tomada com base em argumentos científicos e verificação de características intrínsecas das opções de espécies ou linhagens disponíveis. Um modelo inadequado pode implicar em resultados pouco confiáveis ou um menor potencial translacional.

Não há um melhor modelo animal para pesquisas envolvendo doenças genéticas humanas. Há modelos mais adequados a depender do tipo de doença, objetivo do estudo e infraestrutura disponível (espaço físico, dinheiro e equipamentos). E não se pode presumir que uma espécie é melhor por ser geneticamente mais semelhante aos humanos, ainda que esse possa ser um bom ponto de partida caso as evidências na literatura científica ainda sejam escassas para fundamentar uma escolha, por isso, é preciso verificar se alguma pesquisa anterior indicou que determinada espécie poderia ser inadequada para o tipo de estudo pretendido.

Sabemos que naturalmente existem diferenças entre humanos e modelos animais e, a depender do modelo, uma mutação em um gene pode ser patogênica em humanos, mas ter consequências mínimas em animais no contexto clínico, isto é, alterações fenotípicas imperceptíveis ou mesmo inexistentes. A escolha do melhor modelo depende de que pergunta queremos responder e de que maneira podemos extrapolar e aplicar os resultados obtidos à espécie humana.

Necrópsias - exame de cadáver para determinar a causa da morte.

Células progenitoras neuronais - células que se diferenciam nas células maduras do sistema nervoso (neurônios, astrócitos e oligodendrócitos).

	zebrafish	camundongo	cão
Prolificidade	200-300 ovos/semana	5-10 filhotes/ninhada	muito variável
Tempo de gestação	-	21 dias	58-68 dias
Maturidade sexual	3 meses pós-fertilização	6-8 semanas pós-parto	6-9 meses
Expetativa de vida	média de 3,5 anos	média de 2 anos	10-13 anos
Comprimento do adulto	1,8-3,8 cm	8,2-8,7 cm (sem cauda)	15-110 cm
Custo	$	$$	$$$
Potencial translacional	+	++	+++

Figura 10.
Comparações entre *zebrafish*, camundongo e cão.

Em casos de divergência funcional entre animais e humanos, pode-se procurar aperfeiçoar o uso do modelo a fim de responder às questões centrais da pesquisa. Um exemplo é o caso do camundongo duplo mutante $Dmd^{mdx}/Large^{myd}$ desenvolvido em 2013 pelo grupo de pesquisa coordenado por Mariz Vainzof da Universidade de São Paulo. Enquanto o camundongo Dmd^{mdx} apresenta as mesmas características genéticas e bioquímicas que pacientes humanos com Distrofia Muscular de Duchenne, a fraqueza muscular é um componente que não está presente de forma muito evidenciada, ou seja, tal modelo não manifesta todos os aspectos funcionais da doença. Então os animais Dmd^{mdx} foram cruzados com animais com mutação (também espontânea) no gene *Large1* (relacionado a uma outra forma de distrofia muscular). A linhagem obtida, $Dmd^{mdx}/Large^{myd}$, apresenta grave e progressiva fraqueza muscular, semelhante ao que acontece em humanos, associada a deficiência de distrofina no músculo, de forma que é um modelo para estudo das distrofias musculares em geral.

Há um debate a respeito da usabilidade e confiabilidade dos animais para modelagem de algumas características humanas. Parte dos críticos argumenta que, mesmo em casos em que há grande similaridade genética, algumas características podem não ser tão similares. Sobre isso, cabe dizer que a **reprodutibilidade** de estudos nos quais são usados modelos animais, que são organismos complexos, depende de um rigoroso controle de variáveis que podem influenciar nos resultados. Essas variáveis podem ser o ***background* genético** de uma linhagem, tipo de alimentação oferecida, controle do ciclo de dia/noite, exposição a fatores estressantes, como barulho etc. Os animais de experimentação, sempre que possível, devem ainda ser mantidos em biotérios cuja estrutura e responsáveis pelo seu cuidado sejam adequadamente preparados para seu uso em pesquisa. Dessa forma, reduz-se a chance de observações experimentais contraditórias.

Vale a pena lembrar que toda **pesquisa clínica** realizada em humanos para validação de qualquer medicamento é precedida pela experimentação em modelos animais. A quantidade de trabalhos científicos em todo o mundo em que foram usados modelos animais em seu desenvolvimento para uma infinidade de aplicações e todos os avanços que propiciaram à sociedade, como medicamentos, vacinas, técnicas etc, não deixam dúvidas de que são importantes ferramentas quando usados corretamente.

Aspectos legais da experimentação animal no Brasil

No Brasil, todos os projetos de ensino ou pesquisa que envolvam animais vertebrados devem passar por uma apreciação ética prévia da CEUA (Comissão de Ética no Uso de Animais) da instituição na qual os animais serão utilizados. Trata-se de um procedimento obrigatório. As CEUAs são credenciadas e reguladas por um órgão hierarquicamente superior, o CONCEA (Conselho Nacional de Controle de Experimentação Animal), que está vinculado ao Ministério de Ciência, Tecnologia e Inovações do governo federal. Somente após a aprovação do projeto pela CEUA podem ser iniciados experimentos envolvendo animais. A apreciação ética da CEUA envolve a verificação de que o plano de trabalho proposto pelos pesquisadores está de acordo com legislação vigente, sobretudo a Lei nº. 11.794/2008, conhecida como **Lei Arouca**.

De maneira geral, a apreciação ética envolve as seguintes verificações:

1. A real necessidade da realização do experimento, tendo em vista os benefícios que poderia trazer à sociedade:
   - Vale a pena investigar/solucionar o **problema de pesquisa** em questão?
   - Foi verificado se alguém já solucionou ou estudou esse problema anteriormente?
   - Quais benefícios a pesquisa trará?
   - Que parcela da população e que segmentos da economia serão beneficiados?

Reprodutibilidade - característica de um experimento científico passível de ser reproduzido novamente apresentando resultados equiparáveis aos de execuções anteriores. Trata-se de uma condição importante para que os experimentos científicos tenham credibilidade, pois é preciso evitar conclusões errôneas a partir de efeitos aleatórios, erros de análise ou mesmo manipulações inadequadas.

Lei Arouca - é assim conhecida por ser de autoria do Deputado Sergio Arouca. Link de acesso a essa legislação: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/l11794.htm

***Background* genético** - "fundo genético"; composição genética de uma linhagem ou espécie (todos os alelos do genoma), excluindo-se o(s) gene(s) de interesse do estudo.

Problema de pesquisa - questão para a qual se buscam respostas em uma pesquisa científica.

Pesquisa clínica - testes realizados em humanos, após etapas prévias em modelos animais, para averiguar a segurança e a eficácia de uma intervenção (vacinas e medicamentos).

Genética na Escola – ISSN: 1980-3540

2. Se o projeto de pesquisa segue o princípio dos 3R's, que será apresentado mais adiante;

3. Se há pessoas na equipe capacitadas tecnicamente para os procedimentos propostos.

Além disso, a Lei Arouca proíbe a utilização de animais em atividades educacionais de primeiro e segundo graus, ficando restrita somente a estabelecimentos de ensino superior ou de educação profissional técnica da área biomédica.

A Lei Arouca também prevê sanções legais às pessoas físicas (pesquisadores) e jurídicas (instituições) que transgredirem essa legislação. As sanções variam entre advertência, multa, interdição, impedimento temporário ou definitivo do exercício da atividade com animais e suspensão de financiamentos provenientes de fontes oficiais de crédito e fomento científico.

Os princípios dos 3R's

Os Princípios dos 3Rs compõem uma norma técnica proposta em 1959 por William Russell e Rex Burch e amplamente difundida entre os países para uso de animais em pesquisas. Baseia-se em três pilares: "reduzir, substituir e aprimorar/refinar" (do inglês: *reduction*, *replacement* and *refinement*) (Figura 11). É imperativa, de acordo com esses princípios, a reflexão de sempre tentar reduzir o número de animais por procedimento experimental, substituir sempre que possível o uso de animais e aprimorar métodos para minimizar o desconforto animal.

Para minimizar o desconforto animal, utilizam-se, por exemplo, instalações limpas e adequadas (em espaço, temperatura, com disponibilidade de água, alimento, etc), além de sedativos, analgésicos ou anestésicos durante eventuais procedimentos que poderiam causar dor. Também é preciso que os pesquisadores que os manipulam tenham passado por treinamento prévio de capacitação e recomendável que não utilizem cosméticos que possam ter odores durante os procedimentos com animais, porque isso causa desconforto.

Uma opção bastante factível de substituição de animais, especificamente para fins de ensino, são vídeos e técnicas modernas de apresentação 3D em computador, pois permitem a demonstração de fenômenos fisiológicos. Além disso, o uso de bonecos realísticos para treinamento de estudantes da área de saúde (por exemplo, na realização de suturas) são aliados muito interessantes na redução do uso desnecessário de animais nas aulas práticas.

Do ponto de vista da pesquisa científica, além das melhorias que têm ocorrido no desenvolvimento de melhores modelos animais, um grande esforço está sendo dedicado à busca de abordagens alternativas ou mesmo prévias ao uso de animais: **culturas celulares**, abordagens ***in vitro*** e o uso de **células-tronco** para diferenciação no tecido de interesse. Quando nos referimos a "abordagens prévias", quer dizer que resultados pouco promissores nesses ensaios preliminares podem levar à desistência de se seguir avaliando tal hipótese, excluindo a necessidade de se testar em animais. No entanto, cabe dizer que nenhum processo vivo ocorre isoladamente. Portanto, em algumas situações, os estudos envolvendo animais tendem a mimetizar melhor a complexidade dos processos biológicos de organismos vivos.

Culturas celulares - são conjuntos de células desenvolvidas sob condições controladas em ambientes artificiais.

In vitro - expressão latina que designa experimentos da área biológica que têm lugar fora de sistemas vivos como, por exemplo, em tubos de ensaio.

Células-tronco - células que conseguem se diferenciar e se transformar em diversos tipos celulares.

Para saber mais

ANDERSEN M. L., WINTER L. M. F. Animal models in biological and biomedical research - experimental and ethical concerns. *AnAcadBras Cienc.* v. 91, 2019. doi:10.1590/0001-3765201720170238.

CARLI, G. J., SOUZA, T. A. J., PEREIRA, T. C. A revolucionária técnica de edição genética "CRISPR". *Genética na Escola.* v. 12, n. 2, p. 114-123, 2017.

SOUZA, T. A. J., Pereira, T. C. O impacto na sociedade da tecnologia de edição gênica com base no sistema CRISPR-Cas9. *Genética na escola.* v. 12, n. 2, p. 124-131, 2017.

Modelos animais na pesquisa em DMD #Fala, Mariz! - Com Mariz Vainzof: https://youtu.be/y5u-k33rYlrs

Genética na Escola – ISSN: 1980-3540

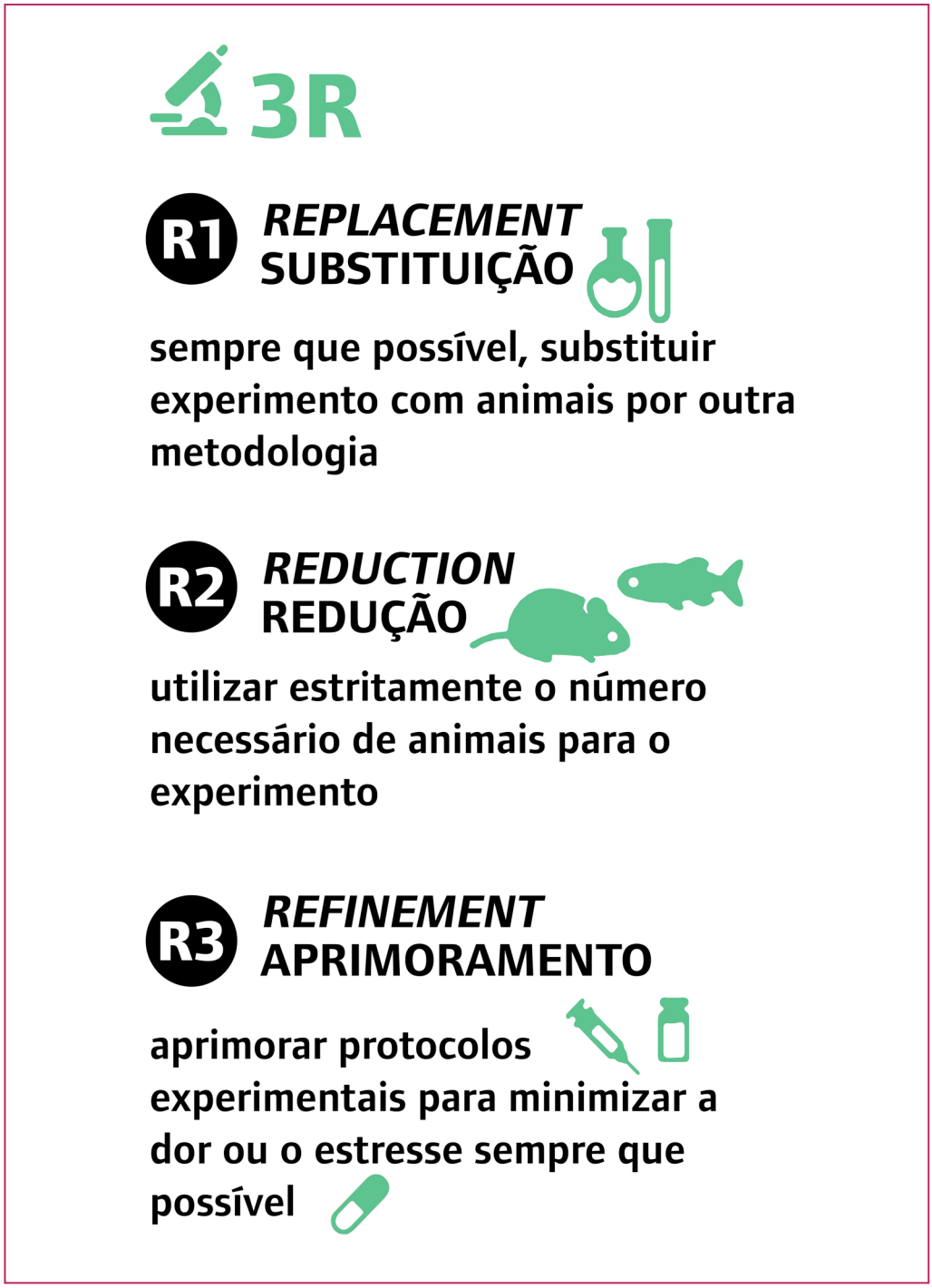


Figura 11.
A norma técnica dos 3R's.

Ilustrações

As figuras 1 e 3 a 11 foram produzidas pelos autores utilizando recursos *Mind the Graph* (https://www.mindthegraph.com/), respeitando as condições de seus termos de uso.

Financiamento

Os autores são gratos pelo apoio das agências de fomento FAPESP, CAPES e CNPq, que financiam seus projetos (CEPID e INCT) e bolsas de pós-graduação.

Capítulo 4.3

CRISPR-Cas9: conceitos fundamentais e aplicações em saúde humana

CARVALHO, L. M. L; SOUZA, VASCONCELOS, F. T. G. R. CRISPR-Cas9: conceitos fundamentais e aplicações em saúde humana

(em vias de submissão)

CRISPR-Cas9: conceitos fundamentais e aplicações em saúde humana

CRISPR-Cas9: fundamental concepts and applications in human health

CRISPR-Cas9: conceptos fundamentales y aplicaciones en la salud humana

Laura Machado Lara Carvalho[*]; Felipe Tadeu Galante Rocha de Vasconcelos[*]

*Centro de Estudos do Genoma Humano e Células-tronco

Resumo

CRISPR-Cas9 é uma ferramenta molecular com inúmeras aplicações, incluindo o desenvolvimento de modelos animais e celulares de doenças, terapia gênica, xenotransplantes, diagnóstico, controle de vetores, desenvolvimento de vacinas recombinantes e identificação e validação de alvos terapêuticos. Neste manuscrito, revisamos mecanismos fundamentais e aplicações em ciências biomédicas, com enfoque em CRISPR-Cas9. Relacionamos o desenvolvimento da técnica CRISPR-Cas9 com o avanço de conhecimentos advindos da pesquisa básica, visto que essa metodologia é inspirada em um mecanismo natural de procariotos. Além da enzima Cas9 de *Streptococcus pyogenes*, também apresentamos outras enzimas Cas naturais e artificiais e suas particularidades. Por fim, discutimos sobre a eventual possibilidade de edição de genes humanos com finalidades não terapêuticas. O avanço das tecnologias baseadas em CRISPR tem aumentado sua versatilidade e especificidade e esperamos, portanto, que nas próximas décadas tenham impactos positivos na solução de diversas questões de saúde.

Palavras-chave: CRISPR-Cas9, edição genética, saúde humana, sgRNA, enzimas Cas

Abstract

CRISPR-Cas9 is a molecular tool with numerous applications, including the development of animal and cellular disease models, gene therapy, xenotransplants, diagnosis, disease vector control, development of recombinant vaccines, and identification and validation of therapeutic targets. In this manuscript, we reviewed fundamental mechanisms and applications in biomedical sciences, focusing on CRISPR-Cas9. We connected the development of the CRISPR-Cas9 technique with the advancement of knowledge from basic research, as this methodology is inspired by a natural mechanism. In addition to the Cas9 enzyme from *Streptococcus pyogenes*, we also presented other natural and artificial Cas9 enzymes and their particularities. Finally, we discussed the possibility of editing human genes for non-therapeutic purposes. The advancement of CRISPR-based

technologies has increased their versatility and specificity, and we therefore hope that in the coming decades they will have positive impacts on solving various health issues.

Keywords: CRISPR-Cas9, genetic editing, human health, sgRNA, Cas enzymes

Resumen

CRISPR-Cas9 es una herramienta molecular con innumerables aplicaciones, por ejemplo, para desarrollo de modelos animales y celulares de enfermedades, terapia génica, xenotransplantes, diagnóstico, control de vectores, desarrollo de vacunas recombinantes, identificación y validación de dianas terapéuticas. En este manuscrito, revisamos mecanismos fundamentales y aplicaciones en ciencias biomédicas, enfocándonos en CRISPR-Cas9. Relacionamos el desarrollo de la técnica CRISPR-Cas9 con el avance de conocimientos procedentes de la investigación básica, ya que esta metodología está inspirada en un mecanismo natural de procariotas. Además de la enzima Cas9 de *Streptococcus pyogenes*, también presentamos otras enzimas Cas naturales y artificiales y sus particularidades. Finalmente, discutimos la eventual posibilidad de edición de genes humanos con objetivos no terapéuticos. El avance de las tecnologías basadas en CRISPR está aumentando su versatilidad y especificidad. Esperamos, por lo tanto, que en las próximas décadas estas tecnologías tengan impacto positivo en la solución de diversas cuestiones de salud.

Palabras clave: CRISPR-Cas9, edición genética, salud humana, sgRNA, enzimas Cas

Introdução

Em 2020, a microbiologista francesa Emmanuelle Charpentier e a bioquímica norte-americana Jennifer Doudna foram conjuntamente laureadas com o Prêmio Nobel de Química pelo desenvolvimento de uma técnica precisa de edição de DNA: CRISPR-Cas9 (ISAACSON, 2021; LI; LI, 2020). Desde 2012, ano da publicação do manuscrito que lhes rendeu o Nobel (JINEK et al., 2012), o número de publicações envolvendo CRISPR cresceu extraordinariamente (**Figura 1A**). Tamanho entusiasmo com o tema tem motivo: o potencial de se utilizar tal técnica com uma ampla gama de objetivos de interesse para a humanidade, incluindo aplicações em saúde humana.

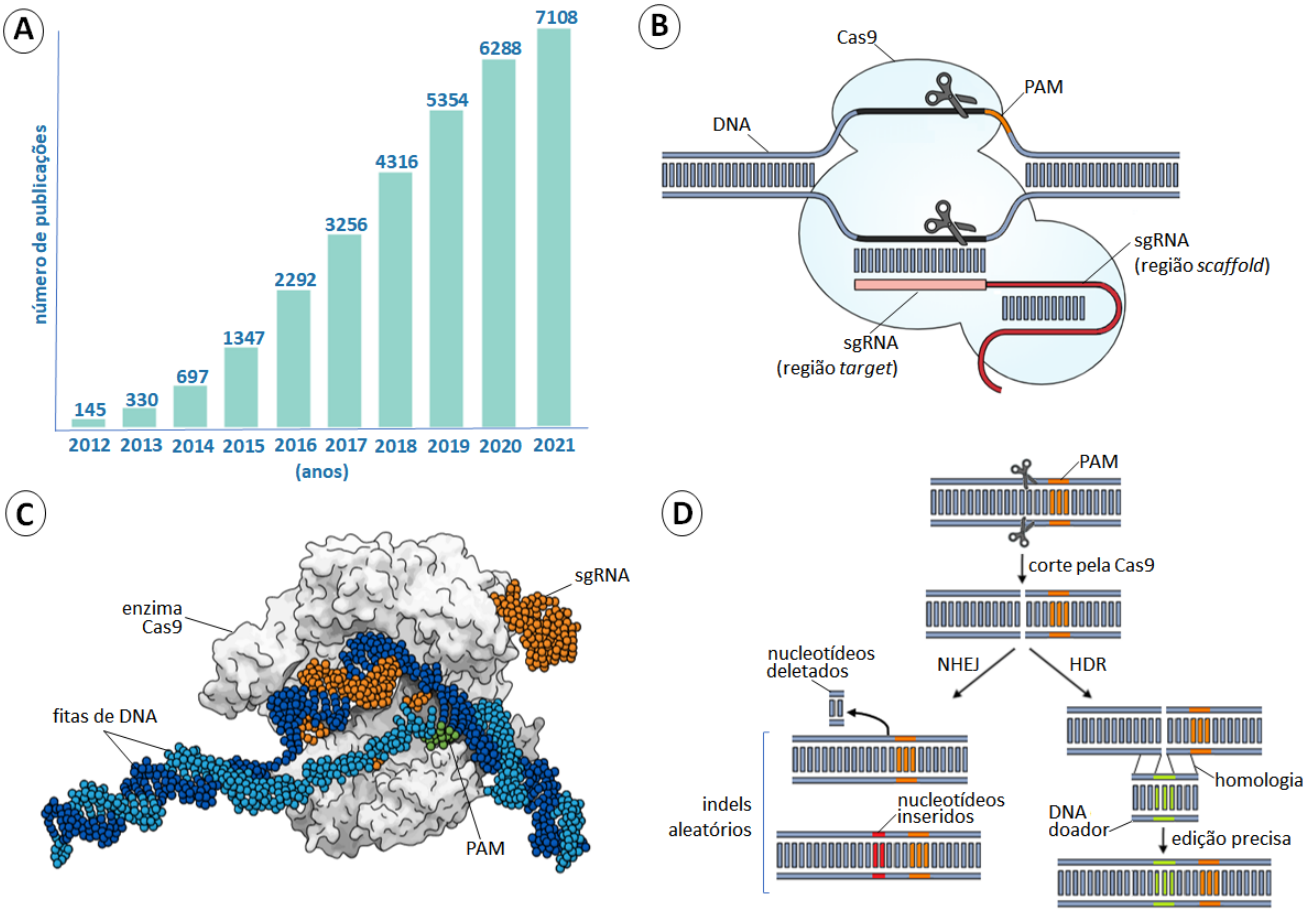


Figura 1 – CRISPR-Cas9: seu impacto na ciência e mecanismos fundamentais.
A – Crescimento do número de publicações em revistas indexadas ao PubMed/MEDLINE que mencionaram o termo "CRISPR" entre os anos 2012 e 2021. Busca realizada em 05/11/2022. **B** – Mecanismo geral do corte de DNA pela Cas9 guiada por um sgRNA. **C** – Visão tridimensional do complexo sgRNA + Cas9 atuando para cortar o DNA alvo. **D** – Mecanismos de reparo de DNA que têm como consequência a edição de sua sequência.

No âmbito das ciências biomédicas, CRISPR poderia ser utilizado, por exemplo, para criar modelos animais e celulares para estudo de doenças humanas (JACINTO et al., 2020), correção de mutações específicas em pacientes (terapia gênica) (GONÇALVES; PAIVA, 2017; ZHANG, 2021), viabilização de xenotransplantes (NAEIMI KARAROUDI et al., 2018), identificação e validação de alvos terapêuticos para intervenção farmacológica (L. LI et al., 2019; FELLMANN et al., 2017), controle de vetores de doenças (LI; AKBARI; WHITE, 2018), desenvolvimento de vacinas (ZHANG et al., 2020) e testes diagnósticos (JOLANY VANGAH et al., 2020).

Mecanismo geral de CRISPR-Cas9 para edição genética

Na técnica de CRISPR, utiliza-se uma nuclease Cas – muito comumente a Cas9 de *Streptococcus pyogenes*, SpCas9 – e um RNA (chamado de sgRNA - *single guide* RNA –, RNA guia ou simplesmente gRNA), que se liga de forma complementar ao DNA que se pretende editar. Quando se utiliza a Cas9, a sequência-alvo deve estar localizada imediatamente antes de um NGG - em que N é qualquer nucleotídeo seguido de duas guaninas (GG) –, sendo essa região denominada PAM (*Protospacer Adjacent Motif*) (**Figuras 1B e 1C**) (PEREIRA, 2016; ZHANG; WEN; GUO, 2014). Cabe dizer, porém, que a sequência PAM e sua localização tendem a ser diferentes ao se utilizar outros tipos de enzima Cas (MOJICA; MONTOLIU, 2016).

Ao reconhecer o PAM e haver complementariedade de sequência, o complexo sgRNA-Cas corta o DNA (PEREIRA, 2016; ZHANG; WEN; GUO, 2014). O sgRNA tem duas porções: (1) a porção *target*, que é complementar à região do DNA alvo que se pretende editar (específica) e (2) a *scaffold,* que é constante e permite o reconhecimento pela Cas9 (UDDIN; PARTSCHT; NAHAR, 2020) (**Figura 1B**). O DNA alvo pode ser de diversos tipos celulares e organismos.

Após o corte do DNA pela Cas, a célula eucariótica tenta repará-lo por dois mecanismos genéticos endógenos principais relacionados a quebras bifilamentares: junção de extremidades não homólogas (NHEJ - *Nonhomologous end-joining*) e recombinação homóloga (HDR - *homology-directed repair*) (DOUDNA; CHARPENTIER, 2014; MOJICA; MONTOLIU, 2016). O reparo por NHEJ é o mais frequente em células de mamíferos e ocorre em qualquer etapa do ciclo celular (PEREIRA, 2016). Nele, há a união das extremidades quebradas por ação da DNA ligase IV (GRIFFITHS (GRIFFITHS; ET AL., 2013), mas esse reparo é frequentemente associado a erros que causam inserções e deleções (indels) (**Figura 1D**), o que muitas vezes resulta em alterações no quadro de leitura (MOJICA; MONTOLIU, 2016; PEREIRA, 2016) e essa propensão é particularmente interessante quando se pretende provocar um *knockout* gênico (PEREIRA, 2016).

Já o reparo por HDR depende de uma estrutura de DNA *template* que apresente homologia com o DNA danificado (GRIFFITHS et al., 2013; PEREIRA, 2016). Naturalmente, a HDR ocorre em células em divisão (sobretudo nas fases S e G2) (LIU et al., 2019a) e enzimas especializadas aparam as extremidades quebradas e, em seguida, as cromátides-irmãs disponíveis são utilizadas como molde para uma síntese reparadora. Ao contrário do reparo por NHEJ, o reparo por HDR é um mecanismo praticamente sem erro (GRIFFITHS et al., 2013). Dessa forma, ao se adicionar um *template* de DNA (ou DNA doador) - que pode ser de uma (*single-strand* DNA, ssDNA) ou duas (*double-strand* DNA - dsDNA) fitas - em um contexto de edição por CRISPR, a célula pode utilizá-lo como molde (assim como usa as cromátides-irmãs em condições não artificiais) para um reparo por HDR (MOJICA; MONTOLIU, 2016). Isso é interessante no contexto de edição gênica pela possibilidade de o DNA doador ser desenhado de forma a ter extremidades homólogas à região cortada pela Cas, mas conter em seu interior certas diferenças que levam a uma edição precisa (**Figura 1D**). Dessa forma, são possibilidades na edição quando ocorre o HDR: deleção,

substituição alélica, introdução de sequências exógenas de DNA (*knock-in*) (MOJICA; MONTOLIU, 2016; PEREIRA, 2016) ou mesmo a inclusão de um códon de parada (*knockout* gênico) (MALLETT et al., 2019).

O grande interesse no mecanismo de reparo HDR é sua precisão em contraste com NHEJ (que gera indels aleatórias) (MOJICA; MONTOLIU, 2016). No entanto, na prática, a eficiência do reparo por HDR é bastante inferior à por NHEJ (LIU et al., 2019a; OKAMOTO et al., 2019), que é mais frequente mesmo na presença de um DNA doador (PEREIRA, 2016). Por esse motivo, há um esforço de diversos grupos de pesquisa para superar essa dificuldade técnica do HDR, seja aumentando sua frequência ou mesmo inibindo NHEJ (LIU et al., 2019a). As propostas metodológicas presentes na literatura envolvem, por exemplo, intervenções na expressão gênica da DNA ligase IV e de outras proteínas envolvidas em NHEJ, com o uso de RNAi (RNA de interferência) (PEREIRA, 2016).

A inativação - ou *knockout* gênico - é tipicamente alcançada por mutações que resultam em perda da atividade do produto final do gene e pode ser alcançada tanto por NHEJ quanto por HDR. O reparo por HDR com DNA doador é mais preciso, mas apresenta dificuldades técnicas já mencionadas anteriormente e, em função disso, NHEJ é muitas vezes o modelo de nocauteamento de escolha. No entanto, ainda que NHEJ seja biologicamente mais frequente, o que traz facilidades técnicas, esse modelo gera mutações aleatórias, sendo preciso escolher células/organismos em que a alteração provocada altere a matriz de leitura tradicional (PEREIRA, 2016).

Formas de *delivery*

A Cas9 pode ser entregue de três formas: como proteína, como mRNA ou como plasmídeo (**Figura 2A**), sendo que estes dois últimos contêm a sequência para produção intracelular da proteína, que é a forma que efetivamente corta o DNA alvo. Já o sgRNA pode ser entregue como RNA ou plasmídeo contendo a sequência para sua produção (EOH; GU, 2019; PEREIRA, 2016). A entrega da proteína Cas9, que pode formar diretamente o complexo ribonucleoproteína (RNP) com o sgRNA, é mais direta. A Cas9 em forma de RNA tem degradação mais rápida que a forma proteica e os plasmídeos podem se integrar ao DNA genômico da célula receptora, o que não é desejado. Comercialmente, as Cas9 em suas três formas de apresentação costumam ter uma sequência NLS (*nuclear localization signal*) acoplada, que facilita seu direcionamento ao núcleo (EOH; GU, 2019).

A entrega da Cas9 e do sgRNA pode envolver, por exemplo, métodos físicos, como a eletroporação (que perturba a membrana causando maior permeabilidade) e a microinjeção (que permite a edição de células únicas e é especialmente interessante para embriões). Além dos métodos físicos, há os vetores virais e não virais (**Figura 2B**). Dos vetores não virais, a forma mais usada são os lipossomos catiônicos, mas há também as nanopartículas (amino)lipídicas e vesículas extracelulares como exossomos. A aplicação de biomateriais como vetores é bastante promissora por sua versatilidade. Nanopartículas podem ser personalizadas para, por exemplo, melhorar a

especificidade por determinado tecido, bem como facilitar o transporte para o núcleo celular. Além disso, em contextos de terapia gênica, a biocompatibilidade que pode ser fornecida por esses tipos de *delivery* é uma alternativa interessante aos vetores virais, com os quais a maior preocupação é a potencial imunogenicidade (EOH; GU, 2019).

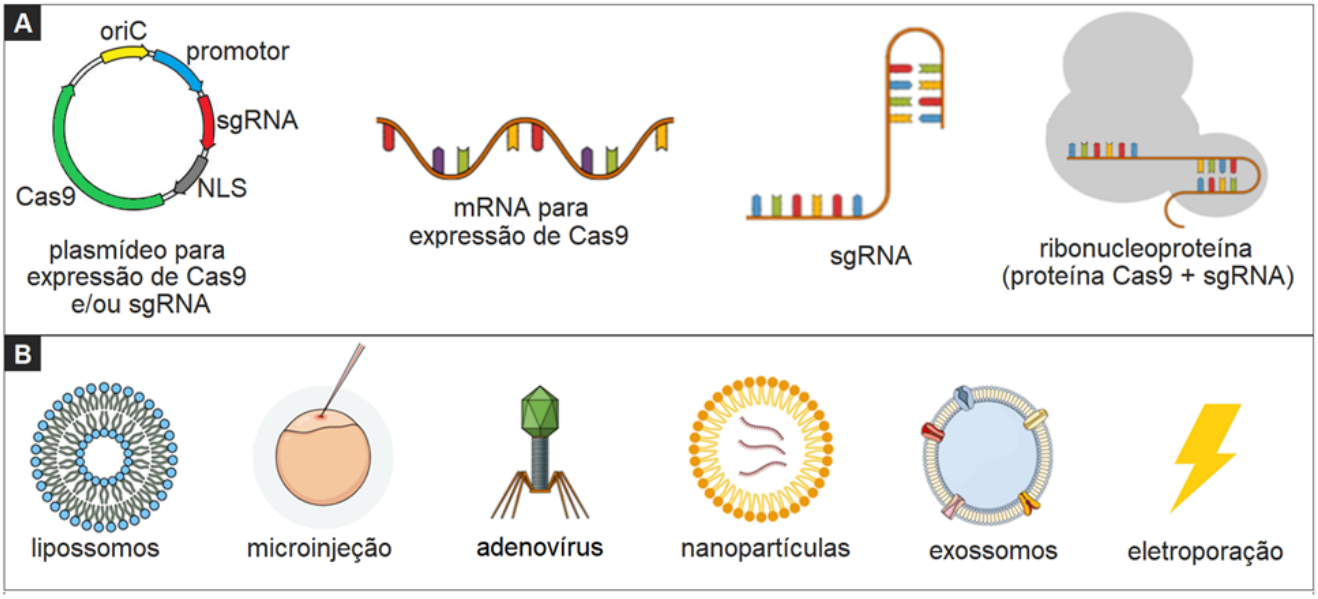


Figura 2 – Estratégias para entrega do sistema CRISPR-Cas9.
A – Formas de apresentação da Cas9 e do sgRNA. OriC – origem de replicação; NLS - *nuclear localization signal*. **B** – Sistemas de entrega da Cas9 e do sgRNA à célula (formas de *delivery*). A microinjeção é representada com embrião de peixe.

O laboratório imitou a natureza

O sistema CRISPR-Cas9 de edição genética foi inspirado em um mecanismo procarioto de clivagem de DNAs exógenos invasores. O entendimento desse mecanismo natural começou com um trabalho desenvolvido por Francisco Mojica (LANDER, 2016), quem identificou no genoma da arquea *Haloferax mediterranei* a existência de uma região com características peculiares: a presença de sequências palindrômicas de 30pb que se repetem, sendo intercaladas por sequências de aproximadamente 36pb que ele chamou de “espaçadoras” e que não aparentavam ter alguma regularidade (**Figura 3A**) (MOJICA et al., 1995). Em 2000, Mojica havia encontrado esse *locus*, ao qual chamou de CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats* – Repetições Palindrômicas Curtas Regularmentee Interespaçadas), em 20 procariotos diferentes e filogeneticamente distantes. Em 2003, percebeu que as sequências espaçadoras eram compatíveis com sequências de vírus e plasmídeos. Com isso, ele propôs que as sequencias espaçadoras do *locus* CRISPR teriam origem em DNAs exógenos arquivados para funcionar como uma memória imunológica, fazendo parte de um sistema imunológico adaptativo para defesa em invasões futuras (MOJICA et al., 2005). Essa organização que alterna sequências palindrômicas e espaçadoras está próxima de genes Cas (CRISPR-*associated*), que produzem proteínas que regulam todo o sistema (LANDER, 2016).

Pode-se resumir o sistema CRISPR naturalmente presente em procariotos da seguinte maneira: após a entrada de organismos invasores, trechos do DNA exógeno são incorporados. Essas sequências exógenas são as que Mojica chamou de espaçadoras e se intercalam com as sequências palindrômicas (**Figura 3A**). A partir do *locus* CRISPR, é transcrito um grande RNA que passa por um processamento, gerando pequenos RNAs, chamados de crRNAs (**cr**ispr **RNA**s), nos quais está presente uma sequência baseada no DNA exógeno e também na transcrição da porção repetitiva. A porção crRNA originada da sequência repetitiva interage com um outro RNA, o tracrRNA (**tra**nsactivating **cr**ispr **RNA**), e ambos formam um complexo com a Cas9. O crRNA reconhece o alvo e o tracrRNA age como um ativador (**Figura 3B**).

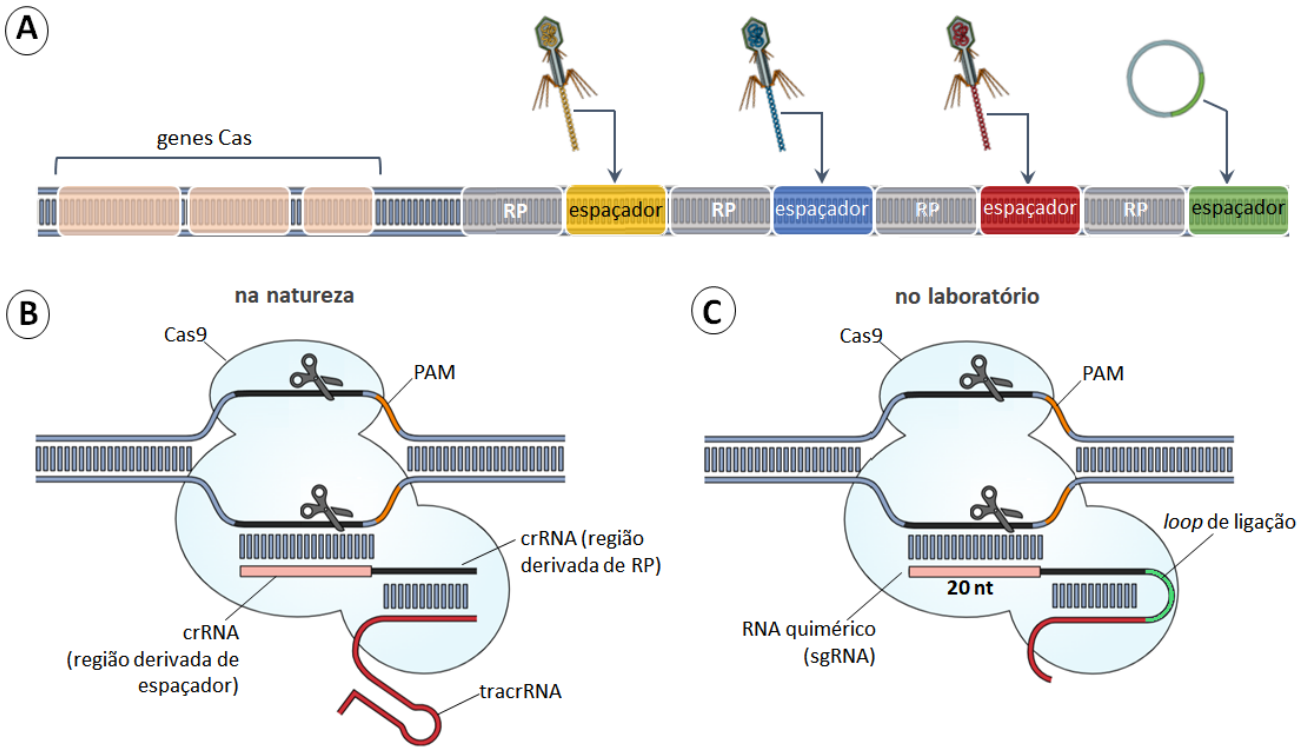


Figura 3 – O sistema de edição é baseado em um mecanismo natural de procariotos.
A – Organização simplificada do *locus* CRISPR em procariotos. **B** e **C** – comparação entre o mecanismo natural coordenado pelo *locus* CRISPR (*S. pyogenes*), baseado em dois RNAs (crRNA e tracrRNA) e o sistema utilizado em laboratório, com um RNA quimérico (sgRNA). RP – repetições palindrômicas; nt – nucleotídeos.

Em 2012, foi publicado pela *Science* o trabalho que deu fama a Jennifer Doudna e Emmanuelle Charpentier (JINEK et al., 2012). Nele, foram capazes de clivar sequências de DNA – tanto circular quanto linear - em pontos programados, utilizando SpCas9 recombinante e RNAs produzidos *in vitro*. Demonstraram que crRNA, tracrRNA e Cas9 eram necessários para a clivagem. Além disso, simplificaram a ferramenta molecular ao unir covalentemente (através de um *tetraloop* GAAA) as duas moléculas de RNA necessárias no mecanismo natural, formando um só RNA quimérico (sgRNA – **Figura 3C**). Dessa forma, o sgRNA que se utiliza de maneira artificial tem uma porção constante (*scaffold*) e, para direcioná-lo ao alvo pretendido, é preciso variar apenas a região *target*, composta por 20 nucleotídeos (**Figuras 1B e 3C**).

Outros tipos de enzimas Cas

Até o momento, apresentamos mecanismos associados à spCas9, que foi a inicialmente caracterizada e ainda hoje é muito utilizada. No entanto, essa enzima tem como alvo sequências flanqueadas por um PAM NGG, o que limita o rol de regiões que podem ser editadas. Foram identificadas enzimas ortólogas em outras espécies de bactérias que apresentam sequências PAM diferentes de NGG, aumentando a versatilidade da técnica. Algumas dessas proteínas são menores, o que também facilita seu acesso ao núcleo celular. Além disso, há versões modificadas dessas enzimas para atender a objetivos experimentais específicos. A **Tabela 1** apresenta alguns exemplos de enzimas Cas naturais e artificiais.

Tabela 1: Exemplos de Cas9 utilizadas em laboratório e suas características.

		Tamanho (aa)	PAM / PFS	*target* do sgRNA (pb)	Particularidades
naturais	SpCas9	1368	NGG	20	corte ~3pb 5' do PAM; pontas cegas
	FnCas9	1629	NGG	20	corte ~3pb 5' do PAM; pontas cegas
	SaCas9	1053	NNGRRT	21	corte ~3pb 5' do PAM; pontas cegas
	NmCas9	1082	NNNNGATT	24	corte ~3pb 5' do PAM; pontas cegas
	St1Cas9	1121	NNAGAAW	20	corte ~3pb 5' do PAM; pontas cegas
	St3Cas9	1409	NGGNG	20	corte ~3pb 5' do PAM; pontas cegas
	CjCas9	984	NNNNRYAC	22	corte ~3pb 5' do PAM; pontas cegas
	AsCpf1	1307	TTTV	24	corte 19/24pb 3' do PAM; pontas coesivas
	LbCpf1	1228	TTTV	24	corte 19/24pb 3' do PAM; pontas coesivas
	LshCas13a	1389	A, U ou C	14-28	tem como alvo RNA, assim como Cas13a de advindas de outras bactérias
artificiais	dCas	1368 (D10A, H840A)	NGG	20	não corta; serve de transportadora de outras moléculas
	HiFi SpCas9	1368 (R691A)	NGG	20	*high fidelity* (menos *off-targets*)
	HypaCas9	1368 (N692A, M694A, Q695A, H698A)	NGG	20	*high fidelity* (menos *off-targets*)
	xCas9-3.7	1368 (A262T, R324L, S409I, E480K, E543D, M694I, E1219V)	PAM: NGN ou GAW	20	compatibilidade com diferentes sequências PAM e *high fidelity* (menos *off-targets*)

PAM: *Protospacer Adjacent Motifs*; PFS: *Protospacer Flanking Sequence*, também chamado de rPAM, que é aplicável a sistemas que têm RNA como alvo. As variantes apresentadas nas enzimas artificiais têm como referência a SpCas9. Fontes: Adli, 2018; Brocken et al., 2018; Chen et al., 2019; Huang et al., 2022; Kick et al., 2022; Komor et al., 2017; Leenay & Beisel, 2017; Lim et al., 2018; Z. Liu et al., 2019; Vakulskas et al., 2018; J. Wang et al., 2020

Uma Cas artificial com inúmeras aplicações é a dCas9, do inglês *dead* (morta), que tem inativada a capacidade catalítica, não sendo capaz de cortar a dupla fita, mas mantém a capacidade de acoplar-se a um DNA alvo guiada por um sgRNA. Dessa forma, estando acoplada a outras

proteínas, a dCas9 acaba tendo função de transporte. Estando ligada a um ativador ou repressor transcricional, a dCas9 consegue regular a atividade de genes alvo. A dCas9 pode ser fusionada, por exemplo, à VP16 (fator de transcrição de *Herpesvirus simplex* tipo 1), para aumentar a transcrição de um gene. Para inibição da transcrição, o repressor KRAB (domínio de repressão transcricional *Krüppel-associated box*) costuma ser muito utilizado. Já o acoplamento ao GFP (proteína verde fluorescente) ou ao RFP (proteína vermelha fluorescente) permite marcar DNA (CASAS-MOLLANO et al., 2020; KARLSON et al., 2021; PEREIRA, 2016) **(Figura 4).** A dCas9 também pode carrear uma desaminase de citosina que atue em DNA (como APOBEC1) para catalisar a desaminação da citosina (C) convertendo-a em uracila (U). Após a conversão de C em U, a replicação do DNA ou mesmo a maquinaria de reparo de DNA converte esse par em A - T. Dessa forma, tem-se a edição precisa de um só par de bases (KUEH; HEROLD, 2016).

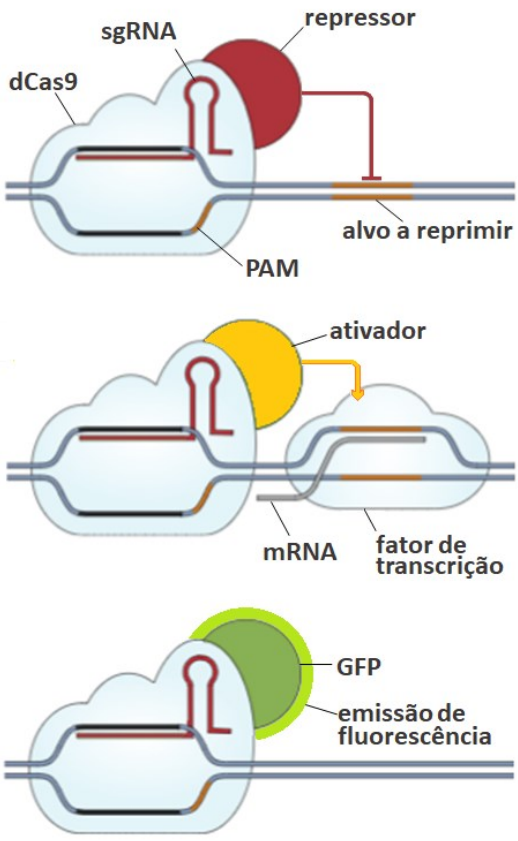


Figura 4 – Aplicações da dCas9 (*dead* Cas9)
A dCas9 é uma enzima alterada artificialmente para não ter função catalítica, mas manter sua capacidade de, guiada por um sgRNA, se aproximar de uma região alvo. Ela pode carrear outras moléculas que atuem, por exemplo, em repressão (**A**) ou em ativação (**B**) transcricional, além da possibilidade de marcação fluorescente (**C**).

Exemplos de aplicações voltadas para a saúde humana

A edição genética por CRISPR-Cas9 é uma ferramenta poderosa com inúmeras aplicações em ciências biomédicas. Uma dessas possibilidades é o desenvolvimento de modelos de doenças humanas para estudo de mecanismos fisiopatológicos, testes de drogas (TORRES-RUIZ; RODRIGUEZ-PERALES, 2017; VASCONCELOS et al., 2022) e investigação funcional de genes candidatos (PEREZ et al., 2019; SALAZAR-SILVA et al., 2021; SCHUMANN et al., 2021).

Animais mutantes são há muito tempo utilizados como modelos em pesquisas (CASTRACANE; HENSON, 2006; VAINZOF et al., 2008), porém a ocorrência e identificação de mutações espontâneas de interesse é rara (DAVISSON et al., 2012). Dessa forma, a edição de genes em modelos animais é uma alternativa bastante conveniente. Com o advento da tecnologia CRISPR-Cas9, tecnicamente mais fácil, barata e específica que as metodologias anteriores (PEREIRA, 2016), o desenvolvimento de modelos animais para doenças genéticas humanas tem sido beneficiado (TORRES-RUIZ; RODRIGUEZ-PERALES, 2017). São exemplos de doenças humanas com modelos animais gerados por CRISPR-Cas9: porco para doença de Huntington (YAN et al., 2018), camundongo para síndrome de insensibilidade androgênica (WU et al., 2022), macaco-cinomolgo para progeria de Hutchinson-Gilford (WANG et al., 2020) e zebrafish para síndrome de Phelan-McDermid (LIU et al., 2018).

Também a obtenção de alguns tipos de amostras com mutações causais de pacientes vivos pode ser dificultada pela invasividade da coleta, o que torna o sistema CRISPR-Cas9 uma alternativa atrativa. Estratégias baseadas em CRISPR-Cas9 permitem recriar experimentalmente em células cultivadas as mutações de pacientes (GUERNET; GRUMOLATO, 2017). Outra técnica que permite contornar a invasividade de coletas é a reprogramação celular, em que células diferenciadas, por exemplo de sangue periférico, se tornam pluripotentes (iPSC - *induced pluripotent stem cells*), podendo ser diferenciadas em diversos tipos celulares de interesse para o tipo doença, mantendo as características genômicas do doador das células, que pode ser inclusive um paciente portador de mutação causal. Em estudos envolvendo iPSC, o sistema CRISPR-Cas9 pode ser utilizado para (1) gerar mutações causais para desenvolver modelos de doença; (2) corrigir mutações causais para obter provas experimentais de uma reconstituição de fenótipo saudável a partir da correção. Além disso, CRISPR-Cas9 permite criar controles isogênicos (mesmo genoma exceto pela presença da mutação causal) (BEN JEHUDA; SHEMER; BINAH, 2018).

O desequilíbrio entre a oferta e a demanda de órgãos humanos para transplante são um problema que impacta na sobrevivência de muitas pessoas atualmente (RYCZEK et al., 2021) e, nesse sentido, uma promessa bastante aclamada pela comunidade científica e pela mídia tem sido o desenvolvimento de animais doadores de órgãos para humanos: os xenotransplantes. A ideia é utilizar a tecnologia CRISPR-Cas9 para editar genes de porcos e impedir com isso a produção de epítopos que causam rejeição aguda. E essa ferramenta tem permitido editar genes relacionados às principais incompatibilidades imunológicas entre as duas espécies (STEWART, 2022). Em 2021, rins de porcos geneticamente modificados foram transplantados para duas pessoas com morte cerebral; não houve rejeição a esses órgãos, que funcionaram normalmente enquanto os indivíduos foram mantidos em ventilação mecânica. No início de 2022, um primeiro paciente recebeu um coração de porco (REARDON, 2022) e inicialmente não houve rejeição. Além disso, foi relatado que o paciente estava bem após a cirurgia. Porém, sobreviveu por dois meses apenas, com morte por razões pouco claras (“Pig-to-human transplants take a leap toward reality”, 2022).

Outra possibilidade é a terapia gênica, isto é, a correção de mutações com finalidade terapêutica. CRISPR-Cas9 tem sido considerada uma ferramenta promissora para o tratamento de doenças genéticas humanas, incluindo doenças hematológicas, neurológicas e câncer (ZHANG, 2021). Os estudos mais avançados atualmente são voltados sobretudo para doenças eritrocitárias monogênicas, como β-talassemia, doença falciforme e deficiência de piruvato quinase; eles consistem em editar *ex vivo* células-tronco hematopoiéticas CD34+ advindas da medula óssea dos próprios pacientes e, em seguida, as células editadas são reintroduzidas. Os resultados de testes clínicos até o momento têm sido considerados promissores (QUINTANA-BUSTAMANTE et al., 2022).

Há também um destaque na terapia gênica aplicada a doenças hereditárias da retina, uma vez que é uma região anatomicamente acessível e capaz de ser monitorada *in vivo.* A aplicação do sistema CRISPR costuma ser por injeção sub-retiniana ou intravítrea e há vários estudos em fase pré-clínica e clínica em andamento, por exemplo para as doenças amaurose congênita de Leber, acromatopsia, retinitis pigmentosa e retinosquise ligada ao X (CHO et al., 2017; HERNÁNDEZ-JUÁREZ; RODRÍGUEZ-URIBE; BOROOAH, 2021).

Alguns dos desafios da terapia gênica são a edição fora do alvo (*off-targets*) e a necessidade de sistemas *delivery* que cheguem aos tipos celulares de interesse com eficiência e especificidade. Embora se tenha feito grandes esforços para limitar a possibilidade de ocorrência de *off-targets*, incrementando a afinidade das enzimas Cas, essa intercorrência ainda pode acontecer, o que pode ter impacto em outros genes de funções importantes. Além disso, ainda que existam muitos tipos diferentes de *delivery*, é difícil alcançar alguns tipos celulares (ZHANG, 2021). Em casos de pacientes com distrofias musculares, por exemplo, como alcançar de maneira eficiente e específica as células musculares, que são amplamente distribuídas anatomicamente?

CRISPR-Cas9 também pode ser útil na identificação e validação de alvos terapêuticos. Uma possibilidade é aplicar a edição *knockout* a diferentes genes e observar para cada um a consequência fenotípica da perda de função em um modelo de doença. Recentemente, ao se aplicar essa metodologia a células de um tipo de câncer de pulmão, se observou que a perda de função de uma enzima da via de biossíntese de pirimidinas (diidroorotato desidrogenase - DHODH) tem ação antitumoral. Posteriormente, também a aplicação de um fármaco (brequinar), que atua inibindo essa enzima, teve essa mesma ação nas células desse câncer de pulmão (LI et al., 2019b).

No controle de vetores de doenças, pode-se manipular insetos geneticamente para que haja transmissão de sequências que interferem no seu desenvolvimento ou prejudiquem sua capacidade reprodutiva, o que foi proposto por exemplo para reduzir a transmissão da malária e da leishmaniose (NATEGHI ROSTAMI, 2020).

Métodos de recombinação são usados como estratégia no desenvolvimento de algumas vacinas. O objetivo de se desenvolver vacinas baseadas em proteínas recombinantes é obter um

produto eficaz e com menos reações adversas. As proteínas a serem incluídas nesse tipo de produto são cuidadosamente estudadas e escolhidas (UNNIKRISHNAN; RAPPUOLI; SERRUTO, 2012). No entanto, considerando que os métodos convencionais de recombinação são demorados, laboriosos, mais caros e menos eficientes, CRISPR acaba sendo uma ferramenta interessante para esse tipo de aplicação (NAEEM et al., 2022; TANG et al., 2021).

Há diferentes estratégias de diagnóstico de doenças infecciosas e não infecciosas em que CRISPR pode ser utilizado (KIM; JI; KOH, 2021; GUK et al., 2017), por exemplo, relataram um método de hibridização mediada por CRISPR-dCas9 para a detectar *Staphylococcus aureus* resistente ao antibiótico meticilina, que é um desafio em termos de tratamento. Eles aproveitaram a ligação específica do DNA alvo a um dCas9-sgRNA ligado a *beads* magnéticas. Com um imã, recuperaram o que ficou ligado e coraram com um marcador fluorescente intercalante de DNA (SYBR Green I), o que levou à emissão de fluorescência. Isto é, amostras que emitiam o sinal fluorescente eram positivas para *S. aureus* resistente a meticilina e as demais negativas. Há também diversas propostas em diagnóstico utilizando a enzimas Cas13, que tem como alvo RNA, com aplicações na detecção de cepas específicas de alguns vírus (como zika e dengue), distinção de bactérias patogênicas, genotipagem humana e diagnóstico de câncer (GOOTENBERG et al., 2017).

Bebês à la carte?

A maioria das terapias em estudo envolvendo CRISPR envolvem abordagem *ex vivo*, isto é, células de pacientes são coletadas, editadas e reimplantadas (HIRAKAWA et al., 2020; LI et al., 2020). No entanto, algumas doenças congênitas ou de início precoce afetam gravemente o paciente (RANISCH, 2020), o que é um argumento utilizado em favor da edição de gametas e embriões, o que é proibido em grande número de países.

Em 2018, o cientista chinês He Jiankui anunciou o nascimento de gêmeas humanas que tiveram o gene *CCR5* editado pré-implantação (ORMOND et al., 2019), pois o receptor codificado por esse gene é necessário para que o HIV infecte o indivíduos (LEDERMAN et al., 2006). Tal notícia gerou grande repercussão mundial. Tratava-se de um experimento com casais em que os homens eram portadores do HIV, objetivando evitar a transmissão para seus filhos. Esse experimento não pode ser considerado uma intervenção terapêutica justificável, visto que há técnicas de lavagem de sêmen em contexto de reprodução assistida para evitar a transmissão em casos de casais sorodiscordantes (RAPOSO, 2019) e essas técnicas são seguras (ZAFER et al., 2016). O experimento de Jianku passou a ser encarado como “melhoramento genético”, algo preocupante pois não prioriza o bem-estar dos indivíduos envolvidos. A China não aporta base legal para o experimento realizado por Jiankui e a comunidade científica tem sido bastante crítica a ele (RAPOSO, 2019). As revistas *Nature* e *Science* se recusaram a publicar seu artigo (FURTADO, 2019). Cabe ainda mencionar que intervenções em linhagem germinativa e embriões são

desaprovadas pela Declaração Universal sobre o Genoma Humano e os Direitos Humanos (1997) e Convenção de Oviedo (1999) (RANISCH, 2020).

Tecnicamente, o que podemos dizer é que, no atual estágio de desenvolvimento da técnica de CRISPR, ainda não há segurança para a edição de embriões humanos. Apesar de tida como específica, CRISPR pode editar erroneamente outras regiões do genoma que não a de interesse (*off-targets*), o que pode acarretar doenças (ex.: câncer) (ROSSANT, 2018; ORMOND et al., 2019). Em casos em que um ou ambos os pais são sabidamente portadores de uma variante genética patogênica, o diagnóstico genético pré-implantacional, no qual são selecionados embriões não portadores para implantação em útero, é uma alternativa tecnicamente mais segura que a edição por CRISPR.

Para características muito priorizadas por eugenistas dos séculos XIX e XX - como inteligência e cor da pele - o potencial de CRISPR, ou mesmo do diagnóstico genético pré-implantacional, ainda é bastante limitado. Afinal, essas características estão relacionadas a diversos genes e também a fatores ambientais. Interferir molecularmente em características assim ainda não é possível. Eticamente, esse tipo de proposta funcionaria como aprimoramento genético motivado pelo desejo consumidor de ter filhos com características pré-estabelecidas, o que no passado foi chamado por Sergio Pena de “bebês à la carte” (PENA, 2009) e aqui ilustramos metaforicamente com a **Figura 5**. A modulação de características como inteligência, cor da pele, potencial atlético e cor de olhos, ainda que tecnicamente distante, como eventual proposta poderia acirrar as possibilidades de discriminação genética, tanto das pessoas que passaram por edição quanto das não “melhoradas”.

Considerações finais

Todos os seres vivos, grandes ou pequenos, têm suas “cartas na manga” para sobreviver. Os procariotos têm um sofisticado mecanismo molecular de memória imunológica para defesa contra DNAs invasores. Já a curiosidade humana permitiu-nos entender diversos processos naturais e criar a partir deles procedimentos que resolvem problemas cotidianos e até mesmo previnem e curam doenças.

É fascinante perceber que a ciência básica nos fez entender mecanismos de procariotos que permitiram à humanidade uma ferramenta extremamente versátil. Tamanhas são as expectativas sobre o de uso das ferramentas CRISPR que foram apenas oito anos entre a publicação do artigo que rendeu fama a Doudna e Charpentier (JINEK et al., 2012) e o prêmio Nobel que receberam juntas. De qualquer maneira, é preciso reconhecer que muitos outros pesquisadores também estiveram envolvidos na construção desse conhecimento e desenvolvimento da técnica, o que provavelmente começou nos anos 90 com Francisco Mojica (LANDER, 2016).

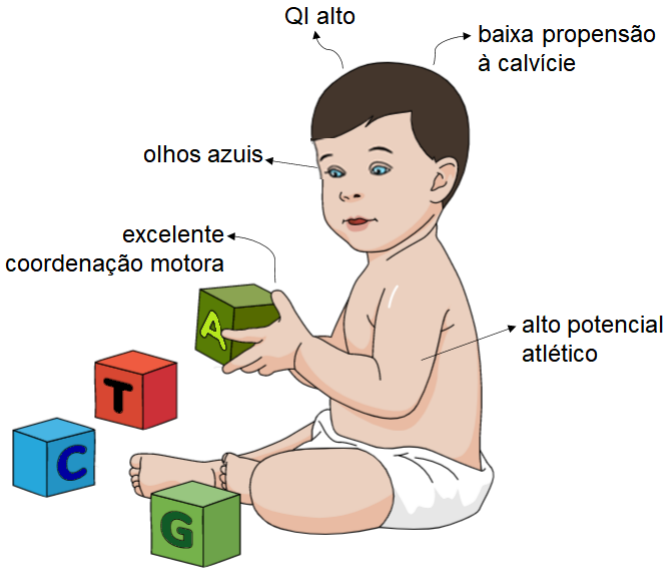


Figura 5 - Representação metafórica da obtenção de seres humanos com características desejadas a partir da manipulação de genes não relacionados a doenças, o que ainda é tecnicamente impossível e também envolve conflitos éticos.

Muito ainda precisa ser discutido sobre os limites que a sociedade poderia estabelecer ao uso das ferramentas CRISPR, principalmente no que diz respeito à edição de genes humanos. Como toda poderosa tecnologia, há questões que precisam ser enfrentadas e discutidas, tendo sido aqui destacadas: a segurança e a possibilidade (ainda remota) de uso dessa ferramenta com objetivos neoeugenistas. De qualquer maneira, as expectativas do uso de CRISPR em prol da saúde humana são fartas, conforme discutido neste manuscrito. Nesse sentido, esperamos ver nas próximas décadas várias soluções voltadas para sobrevivência e qualidade de vida.

Referências

ADLI, M. The CRISPR tool kit for genome editing and beyond. **Nature communications**, v. 9, n. 1, p. 1911, maio 2018.

BEN JEHUDA, R.; SHEMER, Y.; BINAH, O. Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9. **Stem Cell Reviews and Reports**, v. 14, n. 3, p. 323–336, 6 jun. 2018.

BROCKEN, D. J. W.; TARK-DAME, M.; DAME, R. T. dCas9: A Versatile Tool for Epigenome Editing. **Current Issues in Molecular Biology**, p. 15–32, 2018.

CASAS-MOLLANO, J. A. et al. CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes. **The CRISPR Journal**, v. 3, n. 5, p. 350–364, 1 out. 2020.

CASTRACANE, V. D.; HENSON, M. C. The Obese (ob/ob) Mouse and the Discovery of Leptin. In: **Leptin**. Boston, MA: Springer US, 2006. p. 1–9.

CHEN, W. et al. Molecular basis for the PAM expansion and fidelity enhancement of an evolved

Cas9 nuclease. **PLOS Biology**, v. 17, n. 10, p. e3000496, 11 out. 2019.

CHO, G. Y. et al. CRISPR in the Retina: Evaluation of Future Potential. In: **Precision Medicine, CRISPR, and Genome Engineering**. [s.l.] Springer, 2017. p. 147–155.

DAVISSON, M. T. et al. Discovery Genetics: The History and Future of Spontaneous Mutation Research. **Current Protocols in Mouse Biology**, v. 2, n. 2, p. 103–118, jun. 2012.

DOUDNA, J. A.; CHARPENTIER, E. The new frontier of genome engineering with CRISPR-Cas9. **Science**, v. 346, n. 6213, p. 1258096, 28 nov. 2014.

EOH, J.; GU, L. Biomaterials as vectors for the delivery of CRISPR–Cas9. **Biomaterials Science**, v. 7, n. 4, p. 1240–1261, 2019.

FELLMANN, C. et al. Cornerstones of CRISPR-Cas in drug discovery and therapy. **Nature reviews. Drug discovery**, v. 16, n. 2, p. 89–100, fev. 2017.

GONÇALVES, G. A. R.; PAIVA, R. DE M. A. Gene therapy: advances, challenges and perspectives. **Einstein (São Paulo)**, v. 15, n. 3, p. 369–375, set. 2017.

GOOTENBERG, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. **Science**, v. 356, n. 6336, p. 438–442, 28 abr. 2017.

GRIFFITHS, A. J. F.; ET AL. **Introdução à genética.** . 10. ed. Rio de Janeiro: Guanabara Koogan, 2013.

GUERNET, A.; GRUMOLATO, L. CRISPR/Cas9 editing of the genome for cancer modeling. **Methods**, v. 121–122, p. 130–137, maio 2017.

GUK, K. et al. A facile, rapid and sensitive detection of MRSA using a CRISPR-mediated DNA FISH method, antibody-like dCas9/sgRNA complex. **Biosensors and Bioelectronics**, v. 95, p. 67–71, set. 2017.

HERNÁNDEZ-JUÁREZ, J.; RODRÍGUEZ-URIBE, G.; BOROOAH, S. Toward the Treatment of Inherited Diseases of the Retina Using CRISPR-Based Gene Editing. **Frontiers in Medicine**, v. 8, 1 out. 2021.

HUANG, X. et al. Recent Advances in Improving Gene-Editing Specificity through CRISPR–Cas9 Nuclease Engineering. **Cells**, v. 11, n. 14, p. 2186, 13 jul. 2022.

ISAACSON, W. O prêmio Nobel. In: **A decodificadora**. 1. ed. Rio de Janeiro: Intrínseca, 2021. p. 492–500.

JACINTO, F.; LINK, W.; FERREIRA, B. CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. **J Cell Mol Med**, v. 24, n. 7, p. 3766–3778, 2020.

JINEK, M. et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial

Immunity. **Science**, v. 337, n. 6096, p. 816–821, 17 ago. 2012.

JOLANY VANGAH, S. et al. CRISPR-Based Diagnosis of Infectious and Noninfectious Diseases. **Biological Procedures Online**, v. 22, n. 1, p. 22, 14 dez. 2020.

KARLSON, C. K. S. et al. CRISPR/dCas9-Based Systems: Mechanisms and Applications in Plant Sciences. **Plants**, v. 10, n. 10, p. 2055, 29 set. 2021.

KICK, L. M. et al. Structure and mechanism of the RNA dependent RNase Cas13a from Rhodobacter capsulatus. **Communications Biology**, v. 5, n. 1, p. 71, 20 dez. 2022.

KIM, S.; JI, S.; KOH, H. R. CRISPR as a Diagnostic Tool. **Biomolecules**, v. 11, n. 8, p. 1162, 6 ago. 2021.

KOMOR, A. C.; BADRAN, A. H.; LIU, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. **Cell**, v. 168, n. 1–2, p. 20–36, jan. 2017.

KUEH, A.; HEROLD, M. The new editor—targeted genome engineering in the absence of homology-directed repair. **Cell Death Discovery**, v. 2, n. 1, p. 16042, 13 dez. 2016.

LANDER, E. S. The Heroes of CRISPR. **Cell**, v. 164, n. 1–2, p. 18–28, jan. 2016.

LEDERMAN, M. M. et al. Biology of CCR5 and Its Role in HIV Infection and Treatment. **JAMA**, v. 296, n. 7, p. 815, 16 ago. 2006.

LEENAY, R. T.; BEISEL, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. **Journal of Molecular Biology**, v. 429, n. 2, p. 177–191, jan. 2017.

LI, L. et al. Identification of DHODH as a therapeutic target in small cell lung cancer. **Science Translational Medicine**, v. 11, n. 517, p. eaaw7852, 6 nov. 2019a.

LI, L. et al. Identification of DHODH as a therapeutic target in small cell lung cancer. **Science Translational Medicine**, v. 11, n. 517, 6 nov. 2019b.

LI, M.; AKBARI, O. S.; WHITE, B. J. Highly Efficient Site-Specific Mutagenesis in Malaria Mosquitoes Using CRISPR. **G3: Genes|Genomes|Genetics**, v. 8, n. 2, p. 653–658, fev. 2018.

LI, Z.-Q.; LI, C.-H. **CRISPR/Cas9 from bench to bedside: what clinicians need to know before application?Military Medical Research**, dez. 2020.

LIM, K.; YOON, C.; YOKOTA, T. Applications of CRISPR/Cas9 for the Treatment of Duchenne Muscular Dystrophy. **Journal of Personalized Medicine**, v. 8, n. 4, p. 38, 24 nov. 2018.

LIU, C. et al. CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. **Molecular Autism**, v. 9, n. 1, p. 23, 2 dez. 2018.

LIU, M. et al. Methodologies for Improving HDR Efficiency. **Frontiers in Genetics**, v. 9, 7 jan. 2019a.

LIU, Z. et al. Expanded targeting scope and enhanced base editing efficiency in rabbit using optimized xCas9(3.7). **Cellular and Molecular Life Sciences**, v. 76, n. 20, p. 4155–4164, 27 out. 2019b.

MALLETT, D. R. et al. Efficient and modular CRISPR-Cas9 vector system for *Physcomitrella patens*. **Plant Direct**, v. 3, n. 9, p. e00168, set. 2019.

MOJICA, F. J. M. et al. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. **Molecular Microbiology**, v. 17, n. 1, p. 85–93, jul. 1995.

MOJICA, F. J. M. et al. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. **Journal of Molecular Evolution**, v. 60, n. 2, p. 174–182, fev. 2005.

MOJICA, F. J. M.; MONTOLIU, L. On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. **Trends in Microbiology**, v. 24, n. 10, p. 811–820, out. 2016.

NAEEM, M. et al. CRISPR/Cas System Toward the Development of Next-Generation Recombinant Vaccines: Current Scenario and Future Prospects. **Arabian Journal for Science and Engineering**, 23 set. 2022.

NAEIMI KARAROUDI, M. et al. Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 Gene Editing Technique in Xenotransplantation. **Frontiers in immunology**, v. 9, p. 1711, 5 set. 2018.

NATEGHI ROSTAMI, M. CRISPR/Cas9 gene drive technology to control transmission of vector-borne parasitic infections. **Parasite Immunology**, v. 42, n. 9, 12 set. 2020.

OKAMOTO, S. et al. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. **Scientific Reports**, v. 9, n. 1, p. 4811, 18 dez. 2019.

PENA, S. D. **Igualmente diferentes**. Belo Horizonte: Editora UFMG, 2009.

PEREIRA, T. C. **Introdução à técnica de CRISPR**. 1. ed. Ribeirão Preto: SBG, 2016.

PEREZ, Y. et al. Mutations in the microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. **Brain**, v. 142, n. 3, p. 574–585, 1 mar. 2019.

Pig-to-human transplants take a leap toward reality. **Nature Medicine**, v. 28, n. 3, p. 423–423, 21 mar. 2022.

QUINTANA-BUSTAMANTE, O. et al. Gene Editing for Inherited Red Blood Cell Diseases.

Frontiers in Physiology, v. 13, 28 mar. 2022.

REARDON, S. First pig-to-human heart transplant: what can scientists learn? **Nature**, v. 601, n. 7893, p. 305–306, 20 jan. 2022.

RYCZEK, N. et al. CRISPR/Cas Technology in Pig-to-Human Xenotransplantation Research. **International Journal of Molecular Sciences**, v. 22, n. 6, p. 3196, 21 mar. 2021.

SALAZAR-SILVA, R. et al. NCOA3 identified as a new candidate to explain autosomal dominant progressive hearing loss. **Human Molecular Genetics**, v. 29, n. 22, p. 3691–3705, 21 jan. 2021.

SCHUMANN, T. et al. Deletion of the diabetes candidate gene Slc16a13 in mice attenuates diet-induced ectopic lipid accumulation and insulin resistance. **Communications Biology**, v. 4, n. 1, p. 826, 1 dez. 2021.

STEWART, Z. A. Xenotransplantation: The Contribution of CRISPR/Cas9 Gene Editing Technology. **Current Transplantation Reports**, 30 ago. 2022.

TANG, N. et al. Application of CRISPR-Cas9 Editing for Virus Engineering and the Development of Recombinant Viral Vaccines. **The CRISPR Journal**, v. 4, n. 4, p. 477–490, 1 ago. 2021.

TORRES-RUIZ, R.; RODRIGUEZ-PERALES, S. CRISPR-Cas9 technology: applications and human disease modelling. **Briefings in Functional Genomics**, v. 16, n. 1, p. 4–12, jan. 2017.

UDDIN, B.; PARTSCHT, P.; NAHAR, T. **Genome Editing of Mammalian Cells Using CRISPR-Cas: From In Silico Designing to In-Culture Validation**. [s.l: s.n.].

UNNIKRISHNAN, M.; RAPPUOLI, R.; SERRUTO, D. Recombinant bacterial vaccines. **Current Opinion in Immunology**, v. 24, n. 3, p. 337–342, jun. 2012.

VAINZOF, M. et al. Animal Models for Genetic Neuromuscular Diseases. **Journal of Molecular Neuroscience**, v. 34, n. 3, p. 241–248, 18 mar. 2008.

VAKULSKAS, C. A. et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. **Nature Medicine**, v. 24, n. 8, p. 1216–1224, 6 ago. 2018.

VASCONCELOS, F. et al. Modelos animais no estudo de doenças genéticas humanas. **Genética na Escola**, v. 1, p. 26–40, 2022.

WANG, F. et al. Generation of a Hutchinson–Gilford progeria syndrome monkey model by base editing. **Protein & Cell**, v. 11, n. 11, p. 809–824, 29 nov. 2020.

WANG, J.; ZHANG, C.; FENG, B. The rapidly advancing Class 2 CRISPR-Cas technologies: A customizable toolbox for molecular manipulations. **Journal of Cellular and Molecular Medicine**, v. 24, n. 6, p. 3256–3270, mar. 2020.

WU, S. et al. Two Compact Cas9 Ortholog-Based Cytosine Base Editors Expand the DNA Targeting Scope and Applications In Vitro and In Vivo. **Frontiers in Cell and Developmental Biology**, v. 10, 1 mar. 2022.

YAN, S. et al. A Huntingtin Knockin Pig Model Recapitulates Features of Selective Neurodegeneration in Huntington's Disease. **Cell**, v. 173, n. 4, p. 989- 1002.e13, maio 2018.

ZAREI, A. et al. Creating cell and animal models of human disease by genome editing using CRISPR/Cas9. **The Journal of Gene Medicine**, v. 21, n. 4, p. e3082, 15 abr. 2019.

ZHANG, B. CRISPR/Cas gene therapy. **Journal of Cellular Physiology**, v. 236, n. 4, p. 2459–2481, 22 abr. 2021.

ZHANG, F.; WEN, Y.; GUO, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. **Human Molecular Genetics**, v. 23, n. R1, p. R40–R46, 15 set. 2014.

ZHANG, W.-W. et al. A second generation leishmanization vaccine with a markerless attenuated Leishmania major strain using CRISPR gene editing. **Nature Communications**, v. 11, n. 1, p. 3461, 10 dez. 2020.

Discussão geral

Nesta tese apresentamos uma investigação de SNVs, indels e CNVs que pudessem justificar o quadro clínico de 20 pacientes de uma coorte de obesidade sindrômica. Para uma das entidades genéticas identificadas, a síndrome de Xia-Gibbs, desenvolvemos modelos de iPSC (*induced pluripotent stem cells*) e *zebrafish* para estudar mecanismos fisiopatológicos. A seguir, discutimos as contribuições específicas, as perspectivas futuras e apresentamos as principais conclusões deste trabalho.

Contribuições no diagnóstico de obesidade sindrômica

Em anos anteriores, o grupo de pesquisa da Profa. Celia Koiffman - sobretudo através dos trabalhos da Dra. Carla D'Angelo (D'ANGELO et al., 2014, 2013, 2018; D'ANGELO; KOIFFMANN, 2012) - investigou casuísticas de obesidade sindrômica com um olhar para as variantes de número de cópias (CNVs; *copy number variations*). Nesta tese, miramos também nas variantes dos tipos de nucleotídeo único (SNVs; *single-nucleotide variants*) e indels (inserções e deleções). Trata-se do primeiro estudo a aplicar uma investigação genômica abrangente envolvendo uma análise em conjunto de CNVs, SNVs e indels para uma coorte de obesidade sindrômica. Especificamente para CNVs, a taxa diagnóstica (3/15; 20%) foi similar à encontrada em outros trabalhos envolvendo casuísticas maiores analisadas por microarranjos (D'ANGELO et al., 2018; VUILLAUME et al., 2014). Mas a investigação combinada de CNVs, SNVs e indels a partir de dados de NGS teve alto rendimento diagnóstico (7/15; ~47%) e demonstrou ser mais efetiva.

Outro achado relevante do nosso estudo foi que todas as variantes causais identificadas estavam em genes conhecidos de deficiência intelectual, sendo que apenas para dois deles (*MYT1L* e *EHMT1*) é reconhecida a associação com obesidade. Ao olhar para esses dados, é preciso ter em conta que a obesidade é uma característica comum na população, conforme dados epidemiológicos brasileiros apresentados na **Introdução Geral**. Além disso, alguns trabalhos indicam uma maior prevalência de obesidade em indivíduos com deficiência intelectual em comparação com a população geral (EMERSON et al., 2016; HSIEH; RIMMER; HELLER, 2014; KRAUSE et al., 2016; MELVILLE et al., 2008; STEDMAN; LELAND, 2010). Vários fatores podem justificar

essa maior prevalência de obesidade em pessoas com deficiência intelectual: menos atividade física, hábitos nutricionais ruins, distúrbios psiquiátricos e uso de medicamentos psiquiátricos. Com isso, estabelecemos que uma regra geral para a investigação de pacientes com obesidade sindrômica é que a etiologia da condição sindrômica pode ser independente da origem da obesidade.

Considerando que apenas dois dos genes envolvidos nos diagnósticos positivos da nossa casuística têm associação bem estabelecida com obesidade, uma análise baseada em painel de genes de obesidade sindrômica nos levaria a uma taxa diagnóstica muito inferior à que obtivemos. Dessa forma, deve-se priorizar na investigação genética de casos de obesidade sindrômica abordagens com análises genômicas globais, como microarranjos e sequenciamento de exoma / genoma.

Ao interrogar a possibilidade de cada variante causal ter associação com obesidade, nos chamou a atenção que, além do paciente 5 da nossa casuística, outros dois com a síndrome de Xia-Gibbs que chegaram ao nosso conhecimento também tinham comportamento hiperfágico. Como a associação entre a síndrome de Xia-Gibbs com hiperfagia e obesidade não está descrita na literatura e as fotos de pacientes presentes em relatos de casos publicados nos indicam que muitos não têm obesidade, entramos em contato com a *Xia-Gibbs Society* (organização de pais) e com o Dr. Richard Gibbs (*Baylor College of Medicine*), quem descreveu a síndrome. Ambos nos relataram que alguns pacientes têm obsessão por comida e problemas de saciedade. Além disso, no site da *Xia-Gibbs Society* (organização de pais - xia-gibbs.org/) há uma parte dedicada a apresentar os gostos de alguns pacientes, chamada de "*our favorite things*" e, para muitos deles, "comer" - de maneira geral ou pratos específicos – está entre as três atividades favoritas. Dessa forma, trabalhamos com a hipótese de que variantes em *AHDC1* poderiam interferir em vias de fome e saciedade, podendo causar quadros de obesidade sindrômica. Essa hipótese e o fato de não existirem modelos biológicos bem estabelecidos para estudo da síndrome de Xia-Gibbs e do papel de *AHDC1* no neurodesenvolvimento nos levaram a propor o estabelecimento de modelos com essas finalidades. A descrição do desenvolvimento desses modelos foi apresentada no **Capítulo 3**.

Apresentamos um artigo original descrevendo a investigação genética de uma casuística de obesidade sindrômica (**Capítulo 2.1**) e um artigo de revisão (**Capítulo 1.1**) sobre as principais formas sindrômicas de obesidade, com suas características clínicas, etiologia e diagnóstico molecular. Por não terem sido apresentados nesta tese em ordem cronológica, é possível notar que os achados do artigo original contribuíram na discussão

do artigo de revisão. A partir do artigo de revisão de obesidade sindrômica (**Capítulo 1.1**), também foi possível estabelecer algumas observações gerais sobre sua etiologia; embora a obesidade sindrômica tenha vias biológicas diversas, é possível identificar uma convergência para mecanismos que impactam no neurodesenvolvimento. Também foi possível categorizar esses mecanismos nas seguintes classes: (1) relacionados a *imprinting* genético (síndromes de Prader-Willi, pseudoparatireoidismo e Temple); (2) causados por mutações em genes diretamente envolvidos em regulação de atividade transcricional (síndromes de Kleefstra tipo 1, Börjeson-Forssman-Lehmann, Cohen, relacionada ao *MYT1L* e Luscan-Lumish); e (3) ciliopatias (síndromes de Alström, MORM e Carpenter tipo 1).

Esperamos que os nossos achados contribuam para o diagnóstico genético de indivíduos com obesidade sindrômica. Afinal, além do efeito psicológico positivo de se conseguir um diagnóstico, essa informação também impacta no tratamento dos pacientes e planejamento familiar. No entanto, cabe lembrar que a maioria dos casos de obesidade sindrômica não tem sua etiologia identificada (KAUR et al., 2017), o que deve ser motivo para encorajar o desenvolvimento de mais pesquisas clínicas e genéticas para essa condição, inclusive com casuísticas maiores.

Contribuições dos relatos de casos

Descrevemos o primeiro caso brasileiro da síndrome relacionada ao *MYT1L*, fizemos uma revisão de mais 21 casos relatados na literatura envolvendo SNVs e indels nesse gene e determinamos as frequências fenotípicas nesses casos (CARVALHO et al., 2021b - **Capítulo 2.2**). O fenótipo é variável, mas deficiência intelectual / atraso de desenvolvimento e atraso de fala estavam sempre presentes. Obesidade estava presente em 47,6% dos pacientes e sobrepeso em outros 23,8%. A maioria dos pacientes com obesidade ou sobrepeso apresentavam comportamento hiperfágico, sendo esse o provável mecanismo para ganho excessivo de peso nessa síndrome. Macrocefalia, estrabismo, alterações de comportamento e malformações de mãos / pés também eram características recorrentes.

Relatamos dois casos envolvendo o gene *MED13L*, sendo um isolado (no qual a obesidade está presente) e um familial (CARVALHO et al., 2021a - **Capítulo 2.3**). Até então, todos os casos descritos na literatura eram *de novo*, tendo sido o primeiro relato de

caso familial, em decorrência de suposto mosaicismo gonadal. Não há uma associação reconhecida dessa forma sindrômica de deficiência intelectual com obesidade.

Publicamos o segundo trabalho, e primeiro caso brasileiro, relacionando o gene *SCAF4* a uma síndrome mendeliana do neurodesenvolvimento (CARVALHO et al., 2022) . Até então, apenas o trabalho de Fliedner et al. (2020) havia feito essa relação. Esperamos que a descrição das características clínicas do nosso paciente possa contribuir para definir o espectro fenotípico da síndrome. De acordo com a nossa análise, níveis variados de deficiências cognitivas e de fala estão presentes em todos os casos descritos até o momento. Outras características descritas em alguns pacientes incluem epilepsia, malformações cardíacas e problemas de comportamento. Microcefalia na infância foi observada apenas no nosso paciente. Ainda que o número de casos descritos seja limitado, identificamos que 2/8 pacientes apresentaram IMC z-score superior a +2 (*Table 1*- CARVALHO et al., 2022 - **Capítulo 2.4**), de maneira que provavelmente não há uma associação dessa nova síndrome com obesidade.

As linhagens de iPSC estabelecidas para estudo da síndrome de Xia-Gibbs

A reprogramação de células adultas para obter iPSC, sua diferenciação em tecidos de interesse e identificação de vias biológicas alteradas é uma estratégia para estudar a fisiopatologia de doenças genéticas (BASSETT, 2017), visto que essas células mantêm particularidades genômicas dos indivíduos, incluindo a mutação causal (PERKEL, 2015). Trata-se de uma alternativa a coletas invasivas, como de partes do cérebro, o que seria eticamente proibitivo para pacientes vivos. A reprogramação a partir de diferentes tipos celulares já foi descrita. Nossa escolha pelas PBMC (*peripheral blood mononuclear cells*) foi com base em sua coleta minimamente invasiva (HOKAYEM et al., 2016; Mack et al., 2011) e conveniência.

O método de reprogramação proposto por Okita et al. (2011) tem sido bastante utilizado com bons rendimentos pelo grupo da Profa. Maria Rita Passos-Bueno, que foi colaboradora nessa parte do trabalho. Nós obtivemos quatro linhagens de iPSC, sendo três derivadas de pacientes com a síndrome de Xia-Gibbs, com diferentes mutações de perda de função em *AHDC1*, e uma de um indivíduo não afetado. Essas quatro linhagens apresentaram fenótipo de pluripotência, capacidade de diferenciação nos três folhetos germinativos e foram aprovadas em testes de qualidade, conforme descrito no **Capítulo**

3. Esses resultados permitem que as linhagens estabelecidas sejam utilizadas com confiabilidade em estudos funcionais futuros sobre a síndrome de Xia-Gibbs. Duas outras linhagens derivadas de indivíduos não afetados foram estabelecidas por grupos parceiros (da Profa. Maria Rita Passos-Bueno e da Profa. Lygia Pereira) e poderão completar o grupo controle nesses estudos.

Para estudo de condições clínicas em que há prejuízo de funções cognitivas, como é o caso da síndrome de Xia-Gibbs, é muito utilizada a diferenciação de iPSC em neurônios corticais (ARDHANAREESWARAN et al., 2017). O grupo da Profa. Maria Rita Passos-Bueno tem bem estabelecido um protocolo de diferenciação de iPSC em neurônios corticais por dupla inibição de SMAD seguida de neurogênese *in vitro*, conforme Shi e colaboradores (2012) e adaptações introduzidas na dissertação de Branco (2020). Pretendemos utilizar esse protocolo para obter neurônios corticais e, como passos seguintes para identificar em larga escala as vias biológicas alteradas, empregaremos tecnologias ômicas em abordagens caso-controle. Iniciaremos essa investigação com o RNAseq.

Considerando as características clínicas da síndrome de Xia-Gibbs, nós entendemos ser relevante priorizar a investigação de vias relacionadas com neurodiferenciação, plasticidade, função sináptica, mielinização, remodelamento de cromatina e outros genes relacionados a deficiência intelectual / neurodesenvolvimento. Além do RNAseq, outras abordagens poderão ser adotadas em futuros estudos funcionais para propor vias biológicas alteradas na síndrome de Xia-Gibbs. Pode-se, por exemplo, investigar por espectrometria de massas quais proteínas são diferentemente expressas (proteômica), por ATAC-Seq quais regiões da cromatina apresentam alterações de acessibilidade, ou ainda identificar por sequenciamento de bissulfito em escala genômica as alterações de metilação.

Outra possibilidade interessante é a avaliação eletrofisiológica dos neurônios corticais que pretendemos obter. O Centro de Estudos do Genoma Humano e Células-tronco fez recentemente a aquisição do *Multielectrode Array* (*Axion Byosistems*), aplicável ao estudo de células excitáveis, como é o caso de neurônios. Esse equipamento tem sistemas de matrizes de múltiplos eletrodos que capturam o potencial de campo ou atividade em toda uma população de células. Além de diferenças no padrão eletrofisiológico espontâneo entre caso e controle, podemos comparar também a reposta a diferentes neurotransmissores.

São várias as possibilidades de uso do modelo celular desenvolvido para investigação fisiopatológica da síndrome de Xia-Gibbs. Dessa forma, haverá desdobramentos em projetos desenvolvidos pelo nosso laboratório e também por colaborações, o que é facilitado pelo depósito que fizemos das linhagens de iPSC com mutações em *AHDC1* no hpscreg.eu.

O modelo de *zebrafish ahdc1*$^{+/-}$

Estudos funcionais envolvendo modelos animais têm potencial de contribuir no entendimento da relação entre mutação e fenótipo. Nós escolhemos o *zebrafish* pelas várias vantagens apresentadas por esse modelo: genoma com considerável similaridade com o humano, manutenção relativamente barata (em comparação com modelos mamíferos), rápido desenvolvimento, grande prole (SETH; STEMPLE; BARROSO, 2013), ovos visíveis a olho nu (o que facilita sua manipulação) (SPENCE et al., 2007), embriões transparentes e com desenvolvimento fora do corpo (SANTORIELLO; ZON, 2012; SPENCE et al., 2007). Adicionalmente, o *zebrafish* tem sido usado como modelo para estudar doenças do neurodesenvolvimento (DE ABREU et al., 2020; MESHALKINA et al., 2018; PEREZ et al., 2019; SAKAI; IJAZ; HOFFMAN, 2018) e o gene *ahdc1* está presente em uma só cópia em seu genoma (dado do ncbi.nlm.nih.gov/gene).

O *zebrafish* pode ser usado como modelo de doenças monogênicas ao se editar o gene de interesse. O fato de a fecundação ocorrer fora do corpo da fêmea nesse modelo torna mais fácil a obtenção de *zigotos* para injeção. Utilizamos a técnica de edição mediada por CRISPR-Cas9 para obter nosso modelo de *zebrafish* com mutação em *ahdc1*. Essa técnica de edição tem sido utilizada para gerar modelos de estudo para diversas doenças humanas, incluindo as do neurodesenvolvimento (LIU et al., 2018; MENEGHETTI et al., 2019; PEREZ et al., 2019). Nós verificamos que o protocolo proposto por Vejnar et al. (2016) é reproduzível e foi eficiente para a nossa proposta de gerar a mutação de perda de função no gene *ahdc1*. Nós também recomendamos os *checkpoints* metodológicos que utilizamos (teste de eficiência *in vitro* e edição do gene de pigmentação *slc45a2),* pois eles evitam o uso desnecessário de animais quando os reagentes de edição genética não estão adequados.

O modelo *zebrafish ahdc1*$^{+/-}$ estabelecido é fértil e a prole apresentou proporções genotípicas mendelianas. Ele não tem alterações fenotípicas visíveis óbvias, embora ainda devamos realizar avaliação mais detalhada; cabe mencionar, porém, que os dismorfimos observáveis em pacientes humanos com a síndrome de Xia-Gibbs costumam ser leves. Para facilitar colaborações, depositamos o modelo desenvolvido no zfin.org.

Como próximos passos, pretendemos avaliar a expectativa de vida do modelo, sua fecundidade em comparação com cruzamentos entre selvagens e verificar se o homozigoto é viável. Outra pretensão é conseguir padronizar uma técnica de criopreservação de gametas em nosso laboratório, para que seja possível ter um *backup* da linhagem que desenvolvemos.

Considerando que muitos pacientes com a síndrome de Xia-Gibbs têm diagnóstico de autismo (JIANG et al., 2018) e que alguns podem apresentar hiperatividade (MUBUNGU et al., 2021), pretendemos realizar ensaios comportamentais. Para isso, foi estabelecida uma colaboração com a Dra. Alline de Campos, da USP de Ribeirão Preto, com experiência em ensaios de comportamento com *zebrafish*.

Sobre a hipótese de que *AHDC1* possa interferir em vias neurais de fome e saciedade, pretendemos utilizar o modelo de *zebrafish* em um ensaio que avalia o apetite por observação, conforme descrito por Falcinelli e colaboradores (2016). Aos oito dias pós-fertilização, larvas de *zebrafish* serão alimentadas por seis minutos com nauplios recém-eclodidos de *Artemia salina*. A atividade de ingestão alimentar de cada larva de *zebrafish* será monitorada por filmagem e quantificada. Somente após o ensaio será realizada a genotipagem para categorizar os animais como pertencentes ao grupo caso ou controle.

Assim como para o modelo celular, pretendemos realizar uma avaliação transcriptômica do modelo de *zebrafish*. Usaremos diferentes fases do desenvolvimento e compararemos animais *ahdc1*$^{+/-}$ com selvagens. Na obtenção das amostras de *zebrafish* para RNAseq, a dificuldade inicial é distinguir as larvas heterozigotas das selvagens. Por isso, padronizamos um protocolo de coleta (adaptado de DUPRET et al., 2018). Em suma, as larvas são cortadas em ponto padronizado (imediatamente inferior ao vitelo), sendo que a porção caudal é encaminhada para genotipagem e a porção na qual estão os tecidos encefálicos em desenvolvimento é armazenada. Após genotipagem, são feitos *pools* com as porções superiores de mesmo genótipo. Para animais adultos, podemos genotipar utilizando fragmento de nadadeira caudal e proceder com a coleta de encéfalos para o RNAseq.

Os conhecimentos da fisiopatologia da síndrome de Xia-Gibbs e o papel de *AHDC1* no neurodesenvolvimento podem ser aprofundados com modelos celulares e animais, além da possibilidade de propor potenciais alvos terapêuticos. Porém, a relevância do período pré-natal no neurodesenvolvimento aliado ao diagnóstico somente pós-natal são fatores limitantes na efetividade dos possíveis tratamentos farmacológicos. Ainda assim, talvez seja possível reverter certos déficits moleculares, eletrofisiológicos e comportamentais, principalmente em indivíduos mais jovens, em que ainda há maior plasticidade neural.

Sendo AHDC1 uma proteína intrinsecamente desordenada (IDP - *intrinsically disordered protein*), conforme proposto na discussão do **Capítulo 3**, sua estrutura é tão dinâmica que métodos convencionais de *design* de drogas baseados em estrutura não podem ser aplicados diretamente (RUAN et al., 2019), ou seja, a eventual proposta de opções terapêuticas deveria ser pensada tendo como alvo outras moléculas envolvidas nos mecanismos fisiopatológicos da síndrome. Portanto, é imperativo investigar os mecanismos subjacentes às mutações em *AHDC1* e nossos modelos podem ser ferramentas valiosas para esse propósito.

Conclusões

A etiologia da obesidade sindrômica tende a ser relacionada com o neurodesenvolvimento. Há uma ampla variabilidade genética e diversas vias biológicas podem estar implicadas, mas os mecanismos fisiopatológicos gerais podem ser separados três grupos principais: (1) relacionados a *imprinting* genético; (2) causados por mutações em genes diretamente envolvidos em regulação de atividade transcricional; e (3) ciliopatias.

Em investigações de casos de obesidade sindrômica, não se deve descartar a possibilidade de que a etiologia seja independente da obesidade. Por isso, métodos de investigação abrangente do genoma (como microarranjos e sequenciamento de exoma / genoma) devem ser priorizados em detrimento de painéis de genes de obesidade sindrômica.

Para estudo da síndrome de Xia-Gibbs, desenvolvemos uma linhagem de *zebrafish* com mutação de perda de função em *ahdc1*, além de três linhagens de iPSC derivadas de células de pacientes com essa síndrome. Tais modelos biológicos são os

primeiros para a síndrome de Xia-Gibbs e serão utilizados para investigar seus mecanismos fisiopatológicos, inclusive vislumbrando a proposição de alvos terapêuticos, bem como para avaliar a nossa hipótese de que mecanismos neurais de fome e saciedade possam estar alterados nos pacientes.

Referências

ARDHANAREESWARAN, K.; MARIANI, J.; COPPOLA, G.; ABYZOV, A.; VACCARINO, F. M. Human induced pluripotent stem cells for modelling neurodevelopmental disorders. **Nature Reviews Neurology**, v. 13, n. 5, p. 265–278, 2017. ISSN: 1759-4758. DOI: 10.1038/nrneurol.2017.45

BASSETT, A. R. Editing the genome of hiPSC with CRISPR/Cas9: disease models. **Mammalian Genome**, v. 28, n. 7–8, p. 348–364, 2017. ISSN: 0938-8990. DOI: 10.1007/s00335-017-9684-9

BENITO-KWIECINSKI, S.; LANCASTER, M. A. Brain Organoids: Human Neurodevelopment in a Dish. **Cold Spring Harbor Perspectives in Biology**, v. 12, n. 8, p. a035709, 2020. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a035709

BRANCO, E. V. **Caracterização genótipo-fenótipo em novas síndromes associadas ao Transtorno do Espectro Autista**. 2020. Universidade de São Paulo, São Paulo, 2020. DOI: 10.11606/D.41.2020.tde-20012020-101726

CARVALHO, L. M. L.; COSTA, S. S.; CAMPAGNARI, F.; KAUFMAN, A.; BERTOLA, D. R.; SILVA, I. T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. Two novel pathogenic variants in MED13L: one familial and one isolated case. **Journal of Intellectual Disability Research**, v. 2021, 2021a. DOI: https://doi.org/10.1111/jir.12891.

CARVALHO, L. M. L.; D'ANGELO, C. S.; MUSTACCHI, Z.; DA SILVA, I. T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. A novel MYT1L mutation in a boy with syndromic obesity: Case report and literature review. **Obesity Research & Clinical Practice**, 2021b. ISSN: 1871403X. DOI: 10.1016/j.orcp.2021.01.001

CARVALHO, L. M. L.; PINTO, C. F.; SCLIAR, M. O.; OTTO, P. A.; KREPISCHI, A. C. V.; ROSENBERG, C. SCAF4 -related syndromic intellectual disability. **American Journal of Medical Genetics Part A**, 2022. ISSN: 1552-4825. DOI: 10.1002/ajmg.a.63032

D'ANGELO, C. S.; VARELA, M. C.; DE CASTRO, C. IE; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; PEREZ, A. B. A.; KOIFFMANN, C. P. Investigation of selected genomic deletions and duplications in a cohort of 338 patients presenting with syndromic obesity by multiplex ligation-dependent probe amplification using synthetic probes. **Molecular Cytogenetics**, v. 7, n. 1, p. 75, 2014. ISSN: 1755-8166. DOI: 10.1186/s13039-014-0075-6

D'ANGELO, C. S. et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. **Molecular Cytogenetics**, v. 11, n. 1, 2018. ISSN:

17558166. DOI: 10.1186/s13039-018-0363-7

D'ANGELO, C. S.; KOHL, I.; VARELA, M. C.; DE CASTRO, C. I. E.; KIM, C. A.; BERTOLA, D. R.; LOURENÇO, C. M.; PEREZ, A. B. A.; KOIFFMANN, C. P. Obesity with associated developmental delay and/or learning disability in patients exhibiting additional features: Report of novel pathogenic copy number variants. **American Journal of Medical Genetics Part A**, v. 161, n. 3, p. 479–486, 2013. ISSN: 15524825. DOI: 10.1002/ajmg.a.35761

D'ANGELO, C. S.; KOIFFMANN, C. P. Copy Number Variants in Obesity-Related Syndromes: Review and Perspectives on Novel Molecular Approaches. **Journal of Obesity**, v. 2012, p. 1–15, 2012. ISSN: 2090-0708. DOI: 10.1155/2012/845480

DE ABREU, M. S.; GENARIO, R.; GIACOMINI, A. C. V. V.; DEMIN, K. A.; LAKSTYGAL, A. M.; AMSTISLAVSKAYA, T. G.; FONTANA, Barbara Dotto; PARKER, Matthew O.; KALUEFF, Allan V. Zebrafish as a Model of Neurodevelopmental Disorders. **Neuroscience**, 2020. ISSN: 03064522. DOI: 10.1016/j.neuroscience.2019.08.034

DUPRET, B.; VÖLKEL, P.; FOLLET, P.; LE BOURHIS, X.; ANGRAND, P. O. Combining genotypic and phenotypic analyses on single mutant zebrafish larvae. **MethodsX**, v. 5, p. 244–256, 2018. ISSN: 22150161. DOI: 10.1016/j.mex.2018.03.002

EMERSON, E.; ROBERTSON, J.; BAINES, S.; HATTON, C. Obesity in British children with and without intellectual disability: cohort study. **BMC Public Health**, v. 16, n. 1, p. 644, 2016. ISSN: 1471-2458. DOI: 10.1186/s12889-016-3309-1

FALCINELLI, S.; RODILES, A.; UNNIAPPAN, S.; PICCHIETTI, S.; GIOACCHINI, G.; MERRIFIELD, D. L.; CARNEVALI, O. Probiotic treatment reduces appetite and glucose level in the zebrafish model. **Scientific Reports**, v. 6, n. 1, p. 18061, 2016. ISSN: 2045-2322. DOI: 10.1038/srep18061

FLIEDNER, A. et al. Variants in SCAF4 Cause a Neurodevelopmental Disorder and Are Associated with Impaired mRNA Processing. **The American Journal of Human Genetics**, v. 107, n. 3, p. 544–554, 2020. ISSN: 00029297. DOI: 10.1016/j.ajhg.2020.06.019

FREEL, B. A.; SHEETS, J. N.; FRANCIS, K. R. iPSC modeling of rare pediatric disorders. **Journal of Neuroscience Methods**, v. 332, p. 108533, 2020. ISSN: 01650270. DOI: 10.1016/j.jneumeth.2019.108533

HSIEH, K.; RIMMER, J. H.; HELLER, T. Obesity and associated factors in adults with intellectual disability. **Journal of Intellectual Disability Research**, v. 58, n. 9, p. 851–863, 2014. ISSN: 09642633. DOI: 10.1111/jir.12100

JIANG, Y. et al. The phenotypic spectrum of Xia-Gibbs syndrome. **American Journal of Medical Genetics Part A**, v. 176, n. 6, p. 1315–1326, 2018. ISSN: 15524825. DOI: 10.1002/ajmg.a.38699

KAUR, Y.; DE SOUZA, R. J.; GIBSON, W. T.; MEYRE, D. A systematic review of genetic syndromes with obesity. **Obesity Reviews**, v. 18, n. 6, p. 603–634, 2017. ISSN: 14677881. DOI: 10.1111/obr.12531

KRAUSE, S.; WARE, R.; MCPHERSON, L.; LENNOX, N.; O'CALLAGHAN, M. Obesity in adolescents with intellectual disability: Prevalence and associated characteristics. **Obesity Research & Clinical Practice**, v. 10, n. 5, p. 520–530, 2016. ISSN: 1871403X. DOI: 10.1016/j.orcp.2015.10.006

LIU, C. X.; LI, C. Y.; HU, C. C.; WANG, Y.; LIN, J.; JIANG, Y. H.; LI, Q.; XU, X. CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. **Molecular Autism**, v. 9, n. 1, p. 23, 2018. ISSN: 2040-2392. DOI: 10.1186/s13229-018-0204-x

LOPEZ-RAMIREZ, M. A.; CALVO, C. F.; RISTORI, E.; THOMAS, J. L.; NICOLI, S. Isolation and Culture of Adult Zebrafish Brain-derived Neurospheres. **Journal of Visualized Experiments**, n. 108, 2016. ISSN: 1940-087X. DOI: 10.3791/53617

MELVILLE, C. A.; COOPER, S. A.; MORRISON, J.; ALLAN, L.; SMILEY, E.; WILLIAMSON, A. The Prevalence and Determinants of Obesity in Adults with Intellectual Disabilities. **Journal of Applied Research in Intellectual Disabilities**, v. 21, n. 5, p. 425–437, 2008. ISSN: 13602322. DOI: 10.1111/j.1468-3148.2007.00412.x

MENEGHETTI, G. et al. The epg5 knockout zebrafish line: a model to study Vici syndrome. **Autophagy**, v. 15, n. 8, p. 1438–1454, 2019. ISSN: 1554-8627. DOI: 10.1080/15548627.2019.1586247

MESHALKINA, D. A. et al. Zebrafish models of autism spectrum disorder. **Experimental Neurology**, v. 299, p. 207–216, 2018. ISSN: 00144886. DOI: 10.1016/j.expneurol.2017.02.004

MUBUNGU, G. et al. Clinical presentation and evolution of Xia-Gibbs syndrome due to p.Gly375ArgfsTer3 variant in a patient from DR Congo (Central Africa). **American Journal of Medical Genetics Part A**, v. 185, n. 3, p. 990–994, 2021. ISSN: 1552-4825. DOI: 10.1002/ajmg.a.62049

OKITA, K. et al. A more efficient method to generate integration-free human iPS cells. **Nature Methods**, v. 8, n. 5, p. 409–412, 2011. ISSN: 1548-7091. DOI: 10.1038/nmeth.1591

PEREZ, Y.; BAR-YAACOV, R.; KADIR, R.; WORMSER, O.; SHELEF, I.; BIRK, O. S.; FLUSSER, H.; BIRNBAUM, R. Y. Mutations in the microtubule-associated protein MAP11 (C7orf43) cause microcephaly in humans and zebrafish. **Brain**, v. 142, n. 3, p. 574–585, 2019. ISSN: 0006-8950. ISSN: 0036-8075. DOI: 10.1093/brain/awz004

PERKEL, J. M. Disease modeling with patient-specific iPS cells. **Science**, v. 347, n. 6227, p. 1271-1273, 2015. DOI: 10.1126/science.347.6227.1271

RUAN, H.; SUN, Q.; ZHANG, W.; LIU, Y.; LAI, L. Targeting intrinsically disordered proteins at the edge of chaos. **Drug Discovery Today**, v. 24, n. 1, p. 217–227, 2019. ISSN: 13596446. DOI: 10.1016/j.drudis.2018.09.017

SAKAI, C.; IJAZ, S.; HOFFMAN, E. J. Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. **Frontiers in Molecular Neuroscience**, v. 11, 2018. ISSN: 1662-5099. DOI: 10.3389/fnmol.2018.00294

SANTORIELLO, C.; ZON, L. I. Hooked! Modeling human disease in zebrafish. **Journal of Clinical Investigation**, v. 122, n. 7, p. 2337–2343, 2012. ISSN: 0021-9738. DOI: 10.1172/JCI60434

SETH, A.; STEMPLE, D. L.; BARROSO, I. The emerging use of zebrafish to model metabolic disease. **Disease Models & Mechanisms**, v. 6, n. 5, p. 1080–1088, 2013. ISSN: 1754-8403. DOI: 10.1242/dmm.011346

SHI, Y.; KIRWAN, P.; LIVESEY, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. **Nature Protocols**, v. 7, n. 10, p. 1836–1846, 2012. ISSN: 1754-2189. DOI: 10.1038/nprot.2012.116

SPENCE, R.; GERLACH, G.; LAWRENCE, C.; SMITH, C. The behaviour and ecology of the zebrafish, Danio rerio. **Biological Reviews**, v. 83, n. 1, p. 13–34, 2007. ISSN: 14647931. DOI: 10.1111/j.1469-185X.2007.00030.x

STEDMAN, K. V; LELAND, L. S. Obesity and intellectual disability in New Zealand. **Journal of intellectual & developmental disability**, England, v. 35, n. 2, p. 112–115, 2010. ISSN: 1469-9532 (Electronic). DOI: 10.3109/13668251003717928.

VEJNAR, C. E.; MORENO-MATEOS, M. A.; CIFUENTES, D.; BAZZINI, A. A.; GIRALDEZ, A. J. Optimized CRISPR–Cas9 System for Genome Editing in Zebrafish. **Cold Spring Harbor Protocols**, v. 2016, n. 10, p. pdb.prot086850, 2016. ISSN: 1940-3402. DOI: 10.1101/pdb.prot086850

VUILLAUME, M. L. et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. **American Journal of Medical Genetics Part A**, v. 164, n. 8, p. 1965–1975, 2014. ISSN: 15524825. DOI: 10.1002/ajmg.a.36587

XIA, F. et al. De Novo Truncating Mutations in AHDC1 in Individuals with Syndromic Expressive Language Delay, Hypotonia, and Sleep Apnea. **The American Journal of Human Genetics**, v. 94, n. 5, p. 784–789, 2014. ISSN: 00029297. DOI: 10.1016/j.ajhg.2014.04.006

YANG, N. Generation of pure GABAergic neurons by transcription factor programming. **Nature Methods**, v. 14, n. 6, p. 621–628, 2017. ISSN: 1548-7091. DOI: 10.1038/nmeth.4291

Anexos

Anexo 1 - *SnapShot: Directed Differentiation of Pluripotent Stem Cells.*

Anexo 2 – Dados dos pacientes com síndrome de Xia-Gibbs que tiveram células utilizadas para desenvolvimento de modelos de iPSC.

Anexo 3 - Outras produções durante o período do doutorado e distinções.

Anexo 1

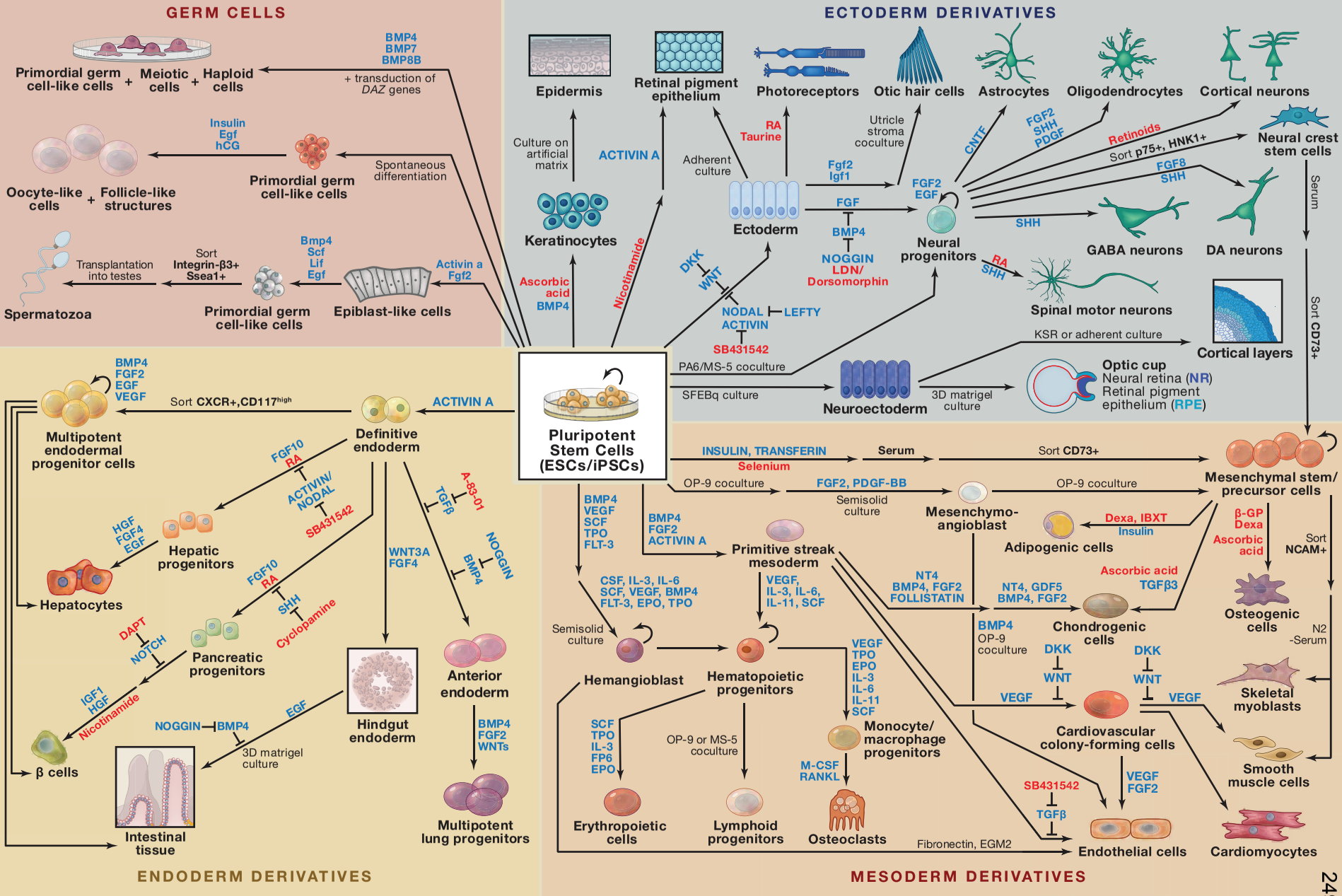
SnapShot: Directed Differentiation of Pluripotent Stem Cells

Cell

Luis A. Williams, Brandi N. Davis-Dusenbery, and Kevin C. Eggan
HHMI, Harvard University, Cambridge, MA 02138, USA

1174 Cell 149, May 25, 2012 ©2012 Elsevier Inc. DOI 10.1016/j.cell.2012.05.015

See online version for legend and references.



Anexo 2

Dados dos pacientes com síndrome de Xia-Gibbs que tiveram células utilizadas para desenvolvimento de modelos de iPSC

caso	idade	sexo	antropometria	cognição e comunicação	comportamento alimentar	comportamento social	outros dados clínicos	histórico familiar
P1	4 anos	M	IMC z-score: 5,70 (obesidade) ao nascimento: 3,35kg obs.: com 1,5 anos o perímetro cefálico era 54cm (+4,67 SD - macrocefalia)	atraso de desenvolvimento global; não fala	dificuldade de sucção quando bebê (usou tubo de alimentação) e pouco ganho de peso; tem hiperfagia; episódios de engasgo principalmente com líquidos	diagnóstico de autismo; atitudes repetitivas; agitação; automutilação; episódios de agressão	redução de movimentos fetais; hipotonia neonatal; hipogenitalismo; instabilidade térmica; ronca; boca pequena; ponte nasal baixa; orelhas protuberantes; olhos amendoados; pé torto; reflexo patelar exacerbado; disgenesia de corpo caloso; mega cisterna magna; sulcos acentuados entre giros cerebrais; redução de volume da substância branca periventricular e subcortical com dilatação irregular ex-vacuo dos ventrículos laterais e do 3º ventrículo	pais não consanguíneos; tem um irmão e dois meios-irmãos (nenhum afetado)
P2	8,1 anos	M	IMC z-score 0,26 (eutrofia) perímetro cefálico de 57,5cm aos 6 anos (+4.2 SD – macrocefalia)	importante comprometimento cognitivo e atraso de fala; não lê, não escreve e não faz contas; compreende apenas ordens simples	importante ansiedade por comida	em geral, é tranquilo, mas fica agitado em lugares abertos e com barulho	ao nascer, apresentou cianose, comunicação atrioventricular do tipo fossa oval com 4,7mm de fluxo esquerda-direita e átrio direito com discreta dilatação; aos três meses apresentou hidrocele bilateral, discreta hernia inguinal à direita, além de uma hérnia umbilical e fissura anal; atraso de marcos motores; tem hipotonia leve;	pais não consanguíneos; irmão por parte de pai não afetado

							abaulamento frontal; braquicefalia; pés planos		
P3	10 anos	M	IMC z-score 2,06 (obesidade) ao nascimento o perímetro cefálico era de 32cm (-1,71 SD)	boa articulação da fala e bom vocabulário, mas não lê, não escreve e não faz contas	compulsividade alimentar	movimentos repetitivos; inquieto; hiperfoco em marcas e instrumentos musicais; déficit de atenção e hiperatividade	hipotonia neonatal; atualmente tem problema de coordenação motora fina; não tem bom controle de esfíncteres; escoliose; tendões curtos; pequeno cisto aracnoide na fossa posterior; sono difícil	pais não consanguíneos; filho único	

P1

P2

P3

Anexo 3

Outras produções durante o período do doutorado e distinções

Artigos

SILVA, I. S.; FERREIRA, C. N.; COSTA, L. B. X.; SOTER, M. O.; CARVALHO, L. M. L.; ALBUQUERQUE, J. C.; SALES, M. F.; CANDIDO, A. L.; REIS, F. M.; VELOSO, A. A.; GOMES KB. Polycystic ovary syndrome: clinical and laboratory variables related to new phenotypes using machine-learning models. **Journal of Endocrinological Investigation**, v. 5, n. 3, p. :497-505, 2021. DOI: 10.1007/s40618-021-01672-8

VASCONCELOS, F. T. G. R.; BARBOSA, I. N.; CARVALHO, L.M. L.; SOUZA, L. S.; KREPISCHI, A. C. V. Revisitando o Dogma Central: a relação entre genes e proteínas. **Genética na Escola**. v. 16, p. 196-207, 2021. DOI: 10.55838/1980-3540.ge.2021.380

RODRIGUES, K. F.; PIETRANI, N. T.; CARVALHO, L. M. L.; BOSCO A. A.; SANDRIM, V. C.; FERREIRA, C. N.; GOMES K. B. Haptoglobin levels are influenced by Hp1-Hp2 polymorphism, obesity, inflammation, and hypertension in type 2 diabetes mellitus. **Endocrinología, Diabetes y Nutrición**, v. 66, n. 2, p. 99-107, 2019. DOI: 10.1016/j.endinu.2018.07.008

CARVALHO, L.M.L.; REIS, F. M.; CANDIDO A. L.; NUNES F. F. C.; FERREIRA C. N.; GOMES K. B. Polycystic Ovary Syndrome as a systemic disease with multiple molecular pathways: a narrative review. **Endocrine Regulations**, v. 52, p. 208-221, 2018. DOI: 10.2478/enr-2018-0026

Trabalhos apresentados em eventos

CARVALHO, L. M. L.; BRANCO, E. V.; MOREIRA, D. P.; KOBAYASHI, G. S.; SARAFIAN, R. D.; ARAUJO, F. T.; HSIA, G. S. P.; BERTOLLO, E. M. G.; BUCK, C. B.; PEREIRA, L. V.; KOIFFMANN, C. P.; PASSOS-BUENO, M. R. S. E.; MENDES, T. A. O.; KREPISCHI, A. C. V.; ROSENBERG, C. Development of cellular and animal models for the Xia-Gibbs syndrome. *In:* 67th Brazilian Congress of Genetics, Natal - RN, 2022

FIALHO, D. M. C., CARVALHO, L. M. L; KREPISCHI, A. C. V. Etiologia genética da deficiência intelectual sindrômica em indivíduos com genitores consanguíneos. *In:* 30º Simpósio Internacional de Iniciação Científica e Tecnológica da USP - SIICUSP, São Paulo – SP, 2022

FIALHO, D. M. C., CARVALHO, L. M. L; BERTOLA, D. R.; ROSENBERG, C.; KREPISCHI, A. C. V. Genetic etiology of syndromic intellectual disability in

individuals with consanguineous parents. *In:* GeneTime Conference 2022 (UFMG), Belo Horizonte – MG, 2022

FIALHO, D. M.; CARVALHO, L. M. L.; BERTOLA, D. R.; ROSENBERG, C.; KREPISCHI, A. C. V. Genetic etiology of syndromic intellectual disability in individuals with consanguineous parents. *In:* Mostra de Trabalhos acadêmicos da 25º da Semana Temática da Biologia do Instituto de Biociências da Universidade de São Paulo, São Paulo-SP, 2022

CARVALHO, L.M.L.; COSTA, S. S.; CAMPAGNARI, F; KAUFMAN, A.; BERTOLA, D. R.; DA SILVA, ISRAEL T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. NOVAS VARIANTES PATOGÊNICAS EM MED13L: RELATO DE UM CASO FAMILIAL E UM CASO ISOLADO. *In:* XXXII Congresso Brasileiro de Genética Médica, *online*, 2021.

CARVALHO L. M. L.; D'ANGELO, CARLA S.; VILLELA, D. C. M.; COSTA, S. S.; BERTOLA, D. R.; DA SILVA, ISRAEL T.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. Investigation of genetic variants in a Brazilian cohort of syndromic obesity. *In:* 66th Brazilian Congress of Genetics, *online*., 2021. v. 66.

CARVALHO, L. M. L.; COSTA, S. S.; Goloni-Bertollo, E. M.; GALBIATTI, A. L. S.; Pavarino, E.C.; KREPISCHI, A. C. V.; KOIFFMANN, C. P.; ROSENBERG, C. The Xia-gibbs syndrome caused by a novel variant in the *AHDC1* gene. *In:* 65th Brazilian Congress of Genetics, Águas de Lindóia - SP, 2019

GOMES K. B.; RODRIGUES, K. F.; PIETRANI, N. T.; CARVALHO L. M. L.; BOSCO, A. A.; SANDRIM, V. C.; FERREIRA, C. N. Os níveis de haptoglobina estão associados ao diabetes mellitus tipo 2 e são influenciados pelo polimorfismo hp1-hp2, obesidade e hipertensão. *In:* 33º Congresso Brasileiro de Endocrinologia e Metabologia, Belo Horizonte - MG, 2018.

Cursos ministrados

CARVALHO, L.M.L.; VASCONCELOS, F. T. G. R. CRISPR: do sal à edição genética. *In*: III Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2022. (Duração: 1 hora)

CARVALHO, L.M.L.; VASCONCELOS, F. T. G. R. Modelos animais no estudo de doenças genéticas humanas. *In*: III Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2022. (Duração: 1 hora)

CARVALHO, L.M.L. Redação científica. *In:* FMU - Centro Universitário das Faculdades Metropolitanas Unidas, 2022. (Duração: 3 horas)

CARVALHO, L.M.L. Redação científica. *In:* FMU - Centro Universitário das Faculdades Metropolitanas Unidas, 2021. (Duração: 3 horas)

CARVALHO, L.M.L.; VASCONCELOS, F. T. G. R. CRISPR: do sal à edição genética. *In*: II Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2021. (Duração: 1 hora)

CARVALHO, L.M.L.; BARROS, J. S. C(r)omo(s)somos? Citogenética Humana. *In*: II Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2021. (Duração: 3 horas)

CARVALHO, L.M.L.; VASCONCELOS, F. T. G. R. Modelos animais no estudo de doenças genéticas humanas. *In*: II Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2021. (Duração: 1 hora)

CARVALHO, L.M.L.; SOUZA, L. S.; VASCONCELOS, F. T. G. R.; RIBEIRO JUNIOR, A. F. Genética Humana. *In*: I Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2020. (Duração: 6 horas)

Organização de eventos

NETTO, L. E. S.; ARIAS, M. C.; FERRARI, M. F. R.; CARVALHO, L.M.L., et al. III Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2022

ARIAS, M. C.; FERRARI, M. F. R.; NETTO, L. E. S.; CARVALHO, L.M.L. et al. II Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2021

FERRARI, M. F. R.; ARIAS, M. C.; NETTO, L. E. S.; CARVALHO, L.M.L. et al. I Curso de Férias em Genética e Evolução do Instituto de Biociências da USP, 2020

MENDES, T. A. O.; KREPISCHI, A. C. V.; CARVALHO L. M. L.; VASCONCELOS, F. T. G. R.; BARROS, J. S.; PIRES, S. F. Curso Crispr-Cas9: do desenho à análise dos experimentos, 2020

Outros tipos de produções

CARVALHO, L. M. L.; GRIECO, A.; SOARES, M.; TRENCH, B.; DESSEN, E. Como o PEIXE PAULISTINHA ajuda os cientistas a entender genética (participação: apresentação e roteiro), 2022 (vídeo educativo publicado no canal Genoma USP do Youtube - https://youtu.be/bNoLuPrxOS8)

CARVALHO, L. M. L.; GRIECO, A.; SOARES, M.; TRENCH, B.; DESSEN, E. Ética na prática: animais na ciência (participação: apresentação e roteiro), 2022 (vídeo educativo publicado no canal Genoma USP do Youtube - https://youtu.be/gRDKRCxR4P4)

CARVALHO, L. M. L.; VASCONCELOS, F. T. G. R.; SOUZA, L. S.; GRIECO, A.; ALEGRIA, T. G. P.; SOARES, M.; TRENCH, B.; DAVINA, L.; DESSEN, E. O tamanho do gene tem relação com o tamanho da proteína? (participação: apresentação e roteiro), 2021 (vídeo educativo publicado no canal Genoma USP do Youtube - https://youtu.be/nnloCGes01c)

VASCONCELOS, F. T. G. R.; CARVALHO, L. M. L.; SOUZA, L. S.; GRIECO, A.; ALEGRIA, T. G. P.; SOARES, M.; TRENCH, B.; DAVINA, L.; DESSEN, E. Mais genes = mais complexo? 2021 (participação no roteiro - vídeo educativo publicado no canal Genoma USP do Youtube - https://youtu.be/vAvHpLPMMaE)

NASLAVSKY M.; MENDES, C. B.; CARVALHO, L.M.L et al. Governo corta 42% da verba do Ministério da Ciência e Tecnologia. 2019. (Comentário em programa de TV - Rede TVT)

Outras atividades

Representante discente titular na Comissão Coordenadora do Programa de Pós-graduação em Ciências Biológicas (Biologia Genética) – Instituto de Biociências da USP. Período: março/2019 a março/2022

Representante discente suplente na Comissão de Pesquisa do Instituto de Biociências da USP. Período: março/2021 a março/2022

Monitora da disciplina Genética Humana e Médica e Oncogenética, do curso de graduação em Medicina da USP. Carga horária semanal: 6 horas. Professora responsável: Maria Rita Passos-Bueno. Semestre: 2021/2

Monitora da disciplina Genética Humana, do curso de graduação em Odontologia da USP. Carga horária semanal: 6 horas. Professora Responsável: Ana Krepischi. Semestre: 2021/1

Distinções

Menção honrosa concedida pela Sociedade Brasileira de Genética no Prêmio Painel Pós-graduação pelo trabalho *Development of cellular and animal models for the Xia-Gibbs syndrome*. 67º Congresso Brasileiro de Genética. 2022

Aprovação em primeiro lugar na seleção do Programa de Doutorado em Biologia Genética do Instituto de Biociências – USP (2017/2).