Mariana Soares Fogo

Estudo de expressão gênica em células neurais derivadas de células-tronco pluripotentes induzidas de indivíduos com Transtorno do Espectro Autista

Gene expression study in neural cells derived from induced pluripotent stem cells of individuals with Autism Spectrum Disorder

São Paulo

2018

Mariana Soares Fogo

Estudo de expressão gênica em células neurais derivadas de células-tronco pluripotentes induzidas de indivíduos com Transtorno do Espectro Autista

Gene expression study in neural cells derived from induced pluripotent stem cells of individuals with Autism Spectrum Disorder

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Biologia, na Área de Genética.

Orientadora: Dra. Andréa Laurato Sertié
Coorientadora: Dra. Karina Griesi Oliveira

São Paulo

2018

Fogo, Mariana Soares
Estudo de expressão gênica em células neurais derivadas de células-tronco pluripotentes induzidas de indivíduos com Transtorno do Espectro Autista
Número de páginas: 87

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Transtorno do Espectro Autista 2. Transcriptoma 3. RNAseq I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).	Prof(a). Dr(a).
Prof(a). Dr(a).	Prof(a). Dr(a). Orientador(a)

Y que la técnica es un arma, y que quien sienta que el mundo no es perfecto como debiera ser, debe luchar porque el arma de la técnica sea puesta al servicio de la sociedad, y por eso rescatar antes a la sociedad para que toda la técnica sirva a la mayor cantidad posible de seres humanos, y para que podamos construir la sociedad del futuro (...).

Ernesto Che Guevara

(discurso feito no Encontro Internacional de Estudantes de Arquitetura, 1963)

AGRADECIMENTOS

Concluir esse trabalho não foi tarefa simples. Nele coloquei muito esforço e dedicação e com ele precisei aprender a aceitar minhas próprias limitações, lidar com as frustrações cotidianas e entender as limitações e contradições que o fazer científico carrega. Fazer essa caminhada sozinha não teria sido fácil, que bom que pude contar com pessoas que, de perto ou de longe e cada uma a sua maneira, me alegraram nos momentos de desânimo, festejaram comigo os momentos de alegria e me ajudaram direta ou indiretamente a avançar no trabalho. Meus mais sinceros agradecimentos vão a cada uma dessas pessoas.

A meus pais, Neide e Haroldo, agradeço por acreditarem em mim mesmo quando eu mesma tinha dúvidas e pelas palavras (nem sempre) doces, mas que sempre me fizeram seguir em frente. Ao meu irmão Daniel, agradeço pela parceria, pelas conversas e por compartilhar comigo sonhos e luta.

À Tati e à Ju, companheiras de lab que se tornaram grandes amigas, agradeço pelas eternas conversas sobre Ciência, projetos e experimentos, pelo companheirismo de sempre e por aguentarem minhas intermináveis reclamações e lamentações. Muito do que eu sei hoje, aprendi e continuo aprendendo com vocês desde a época em que a gente era o time da cultura no lab da Irene. Também gostaria de agradecer a todos do lab onde fiz minha IC, em especial à Irene, por me ensinar a dar os primeiros passos como pesquisadora, por toda a paciência e pelo acolhimento (inclusive hoje em dia quando eu resolvo aparecer no lab para conversar ou comer alguma coisa).

Agradeço também aos amigos do lab 200, com quem dividi pequenas angústias e alegrias fazendo experimentos nos horários mais bizarros, por toda ajuda dentro e fora do laboratório. Aos amigos do IIEP, agradeço pela companhia. Gostaria de agradecer especialmente à Sandra e à Juliana pelas discussões, ensinamentos e pela amizade. Fez toda a diferença ter vocês comigo todos os dias... A Ju sempre tão otimista me dizendo que tudo ia dar certo e sempre tendo toda a paciência do mundo para me explicar as coisas e ouvir minhas ideias, e a Sandrita sempre pronta pra me ajudar em tudo que eu precisasse, me ouvir reclamando (e reclamar junto) e fazer companhia no lab e nos rolês da vida.

Agradeço aos amigos da bio pela companhia, pela força e pelas eternas discussões sobre Ciência que me fizeram encontrar respostas para uma série de questionamentos que me acompanharam ao longo do mestrado. Agradeço a meus amigos do colégio, que caminham comigo há pelo menos 15 anos e nunca me deixaram desanimar. Agradeço também aos amigos do Levante, que me ensinaram que eu nunca estou sozinha e que a vida é mais fácil quando a gente sonha junto.

À Rita, à Karina e à Andrea, agradeço pela oportunidade de fazer parte do grupo de pesquisa e estudar essa condição tão intrigante que é o TEA. À Karina, que me acompanhou mais de perto, agradeço pela orientação, pelas discussões e por me fazer enxergar coisas que sozinha eu não conseguia. À Andrea, agradeço pela paciência e toda a ajuda com os experimentos do estudo de caso. Gostaria de agradecer também a todos os colaboradores do Genoma, do IIEP e da Química que contribuíram com meu trabalho, em especial à Suzana e ao Julio, que tiveram a paciência de me ensinar um pouco sobre bioinformática e programação.

Esse trabalho carrega nele um pouquinho de cada um de vocês.

Por fim, agradeço à FAPESP, Capes e CNPq pelo auxílio financeiro, fundamental para o desenvolvimento e conclusão desse trabalho no preocupante cenário de cortes orçamentários que vivenciamos atualmente.

ÍNDICE

RESUMO 1
ABSTRACT 2
1. INTRODUÇÃO..... 3
2. OBJETIVOS 15
3. MATERIAL E MÉTODOS..... 16
4. RESULTADOS 30
5. DISCUSSÃO..... 53
6. CONCLUSÕES 60
7. ESTUDO DE CASO 61
8. BIBLIOGRAFIA..... 76
9. ANEXOS..... 83

RESUMO

O Transtorno do Espectro Autista (TEA) é um transtorno neuropsiquiátrico grave caracterizado por comprometimento da capacidade de interação social e comunicação e pela presença de interesses, atividades e comportamentos repetitivos e restritos. O TEA apresenta alta heterogeneidade genética e pode ser enquadrado em diferentes modelos de herança, o que torna difícil apontar os fatores genéticos associados ao transtorno e compreender de que forma eles interagem e ocasionam, em última instância, sua manifestação clínica. Ainda, fenômenos como penetrância incompleta e discordância entre gêmeos monozigóticos são frequentemente observados em famílias com indivíduos afetados, tornando ainda mais complexo o entendimento do envolvimento de cada um dos fatores genéticos na etiologia do TEA. Uma forma de contornar esses problemas é a utilização de abordagens transcriptômicas no estudo desse transtorno. Nesse contexto, a busca por variantes patogênicas deixa de ser o principal foco do estudo e a atenção volta-se para a investigação de vias de sinalização e processos biológicos que estejam desregulados durante o processo de diferenciação neuronal e que sejam compartilhados pelos indivíduos afetados. Usando como principal ferramenta o sequenciamento do transcriptoma de células progenitoras neurais (NPC) e neurônios derivados de células-tronco pluripotentes induzidas (iPSC), identificamos uma série de alterações transcricionais importantes entre pacientes e controles, sugerindo que, a despeito da grande heterogeneidade genética associada ao transtorno, essas alterações se refletem em processos biológicos comuns. Nossos resultados são consistentes com os obtidos em trabalhos anteriores realizados com outros modelos de estudo e as linhagens celulares geradas a partir de nosso protocolo parecem refletir de forma fidedigna o perfil de expressão do cérebro fetal. Ainda, em um caso particular incluído neste estudo, investigamos as consequências transcricionais e funcionais de uma duplicação na região 17p13.3 - alteração previamente implicada em outros casos de autismo - nas células neuronais deste paciente, contribuindo para a compreensão do papel desta duplicação na patofisiologia do TEA. Dessa forma, nosso trabalho reforça a validade do uso de células neurais derivadas de iPSC e de abordagens transcricionais para o estudo de transtornos do neurodesenvolvimento.

ABSTRACT

Autism Spectrum Disorder (ASD) is a severe neuropsychiatry disorder characterized by impaired social interaction and communication, restricted interests and repetitive behaviors. ASD is highly genetically heterogeneous and can be caused by different inheritance models, which hampers the identification of the genetic factors associated to the disorder. Incomplete penetrance and discordance between monozygotic twins are often observed in families of affected individuals, making it even more complex to understand the involvement of genetic factors in ASD. Applying transcriptomic approaches to study this disorder is an interesting way to overcome these problems. In this context, searching for pathogenic variants is no longer the study's main focus. Instead, one aims to investigate signaling pathways and biological process that are deregulated during neuronal differentiation and shared by affected individuals. Based on the transcriptome sequencing of neural progenitor cells (NPC) and neurons derived from induced pluripotent stem cells (iPSC), we were able to identify a series of important transcriptional alterations between patients and controls, which suggest that despite the great genetic heterogeneity associated to ASD, those alterations are reflected in common biological processes. Our results are consistent with those obtained in previous studies performed with other models and the lineages generated by our protocol seem to reliably reflect fetal brain expression profile. Moreover, in a particular case included in this study, we investigated transcriptional and functional consequences of a duplication located at 17p13.3 - an alteration previously found in other autism cases – in the neuronal cells of this patient, contributing towards a better comprehension of the role of this duplication in ASD pathophysiology. Hence, our work reinforces the validity of the use of neural cell derived from iPSC and transcriptional approaches in the studies of neurodevelopmental disorders.

1. INTRODUÇÃO

1.1 Aspectos clínicos e epidemiológicos do Transtorno do Espectro Autista

O Transtorno do Espectro Autista (TEA) é um transtorno neuropsiquiátrico grave caracterizado por comprometimento da capacidade de interação social e comunicação e pela presença de interesses, atividades e comportamentos repetitivos e restritos (American Psychiatric Association, 2013). Essas manifestações clínicas, quando precocemente apresentadas, são essenciais para que o diagnóstico do transtorno seja realizado (Jones, Gliga, Bedford, Charman, & Johnson, 2014). O grau de comprometimento que os indivíduos apresentam para cada um desses sintomas é bastante variado, indo desde formas mais leves, em que os pacientes têm uma vida independente, até aquelas que são bastante graves e estão, em muitos casos, associadas a quadros de deficiência intelectual (Griesi-Oliveira & Sertié, 2017). Além disso, outras manifestações clínicas não diagnósticas podem ainda ser observadas, tendo impacto negativo no cotidiano dos pacientes e de suas famílias, tais como comportamento agressivo, distúrbios do sono e da alimentação, problemas gastrointestinais, hiperatividade e ansiedade (Gorrindo et al., 2012; Krakowiak, Goodlin-Jones, Hertz-Picciotto, Croen, & Hansen, 2008; American Psychiatric Association, 1994).

Ainda que já tenham sido descritas estruturas do sistema nervoso central anatomicamente alteradas em indivíduos afetados pelo TEA, como o cerebelo, a amígdala e o córtex frontal (Donovan & Basson, 2017), essas alterações não são observadas em todos os pacientes. Também não foram ainda identificados quaisquer biomarcadores moleculares que possam ser usados no diagnóstico do TEA. Em função disso, existe uma grande dificuldade para se desenvolver ferramentas diagnósticas para este transtorno, e assim o diagnóstico pode ser definido apenas por meio de avaliações comportamentais, testes cognitivos e histórico clínico do paciente. Da mesma forma, não existem medicamentos usados para tratar os sintomas principais do transtorno; as terapias medicamentosas disponíveis tratam apenas as manifestações clínicas não diagnósticas citadas.

Estudos recentes apontam que o TEA tem prevalência de aproximadamente 1/68 indivíduos na população e é mais comum em meninos do que em meninas, apresentando um desvio sexual de 4:1 (Christensen et al., 2016). Em comparação com análises anteriores, esses dados indicam um aumento na prevalência do transtorno ao longo dos anos, o que pode ser reflexo de um aumento real no número

de indivíduos afetados na população devido, por exemplo, à maior exposição a fatores ambientais de risco. Entretanto, parte desse aumento provavelmente se deve a mudanças em critérios diagnósticos, melhorias em metodologias de análises epidemiológicas e aumento do conhecimento sobre o TEA tanto pelo público leigo quanto por profissionais da saúde (Baxter et al., 2015; Fombonne, 2009; Wing & Potter, 2002).

1.2 Aspectos genéticos do Transtorno do Espectro Autista

Sabe-se que tanto fatores ambientais, tais como infecções e exposição a agentes tóxicos (Bilbo, Block, Bolton, Hanamsagar, & Tran, 2018), quanto fatores genéticos estão envolvidos na manifestação do TEA. No entanto, diversas evidências nos permitem afirmar com segurança a grande relevância dos fatores genéticos na etiologia do transtorno, tais como: a maior concordância observada em gêmeos monozigóticos em comparação à observada em gêmeos dizigóticos, levando a estimativas de herdabilidade que variam entre 38 e 90%; o alto risco de recorrência do transtorno dentro de uma mesma família; a maior propensão que irmãos e pais de indivíduos afetados têm de apresentar características cognitivas e comportamentais qualitativamente similares àquelas observadas em pacientes; a associação com doenças monogênicas e o aumento do risco de manifestação do transtorno associado a mutações e alterações cromossômicas em diversos *loci* (Sandin, Lichtenstein, Larsson, Hultman, & Reichenberg, 2014; Ronald & Hoekstra, 2011; Hallmayer et al., 2011; Abrahams & Geschwind, 2008; Trottier, Srivastava, & Walker, 1999).

Por meio do uso tanto de abordagens mais tradicionais, como a cariotipagem e o estudo de genes candidatos, quanto de abordagens mais modernas, como o *array* e o sequenciamento de nova geração, uma série de genes e regiões cromossômicas de risco para o TEA já foram identificadas. No banco de dados SFARI (Simons Foundation Austism Research Initiave – https://gene.sfari.org/), que agrupa todas as informações sobre *loci* de risco para o TEA já publicadas, é possível notar que, até o momento, quase 500 genes apresentam evidências de associação com o TEA. É importante ressaltar que nenhum desses genes sozinho consegue explicar mais do que 1% do total de casos de TEA e a causa genética ainda é desconhecida em aproximadamente 75% dos casos (Bourgeron, 2015), evidenciando uma alta heterogeneidade genética associada ao transtorno.

A identificação de genes e regiões cromossômicas por meio dessas abordagens e sua associação com a manifestação clínica do transtorno só é possível se as alterações ou variantes (rearranjos cromossômicos, CNVs (*copy number*

variants), SNVs (*single nucleotide variants*) e pequenas deleções ou duplicações) contidas nesses *loci* forem raras de efeito médio ou alto, ou seja, se o efeito provocado por elas for forte o suficiente para que uma variante sozinha (ou poucas variantes em conjunto) consiga desestabilizar o desenvolvimento neuronal e levar à manifestação do fenótipo. Entretanto, considerando que o TEA é um transtorno comum na população, que estudos familiais indicam que o risco de recorrência é diretamente proporcional à porcentagem do genoma compartilhado com um indivíduo afetado (Risch et al., 2014; Sandin et al., 2014) e que os indivíduos afetados apresentam adaptabilidade reprodutiva reduzida, evidenciada pelo comprometimento da capacidade de interação social inerente ao transtorno, é possível especular que variantes comuns de baixo efeito também tenham papel importante em sua etiologia. A ideia subjacente a essa hipótese é a de que, como variantes de baixo efeito sofrem pressão seletiva menor, elas acabam se tornando comuns na população e se acumulando nos indivíduos. O fenótipo se manifestaria então quando um indivíduo acumulasse uma carga suficientemente grande dessas variantes, capaz de provocar efeitos danosos no desenvolvimento neuronal. De fato, variantes raras de efeito médio e alto em conjunto são responsáveis por causar o transtorno em 10-26% dos casos, enquanto estima-se que SNVs comuns estejam envolvidas na manifestação do transtorno em 17-60% dos casos (Bourgeron, 2015; Gaugler et al., 2014; Klei et al., 2012), evidenciando a grande importância dessas variantes e reforçando a alta heterogeneidade genética associada ao TEA. Contudo, nenhuma dessas variantes de baixo efeito foi identificada até o momento, uma vez que cada uma está associada a um risco muito pequeno e nenhum estudo de associação global já realizado teve poder estatístico suficiente para apontar alguma que fosse indiscutivelmente associada ao TEA (Huguet, Benabou, & Bourgeron, 2016; Bourgeron, 2015).

Assim, atualmente acredita-se que tanto variantes raras quanto variantes comuns expliquem a arquitetura genética do TEA e que ele é, portanto, um transtorno complexo que pode ser enquadrado em diferentes modelos de herança (figura 1.1). Nesse contexto, parte dos casos deve ser causada por uma única variante *de novo* de efeito alto (representando um modelo de herança monogênico) ou por algumas variantes raras de efeito médio (modelo de herança oligogênico). Em um segundo conjunto de casos, os indivíduos podem estar carregando uma carga média de variantes comuns de risco e manifestar o fenótipo apenas com o acúmulo de poucas variantes de efeito médio. Ainda, é possível que alguns dos indivíduos apresentem uma carga muito alta de variantes comuns de risco e apenas isso seja suficiente para que eles sejam afetados. Essas duas últimas situações se aproximariam mais de um modelo de herança poligênico. Dessa forma, cada indivíduo deve carregar um

conjunto próprio de variantes de risco para o TEA e o que determina, em ultima instância, se ele será afetado ou não é a quantidade de variantes raras e comuns presentes em seu genoma e como se dá a interação entre elas.

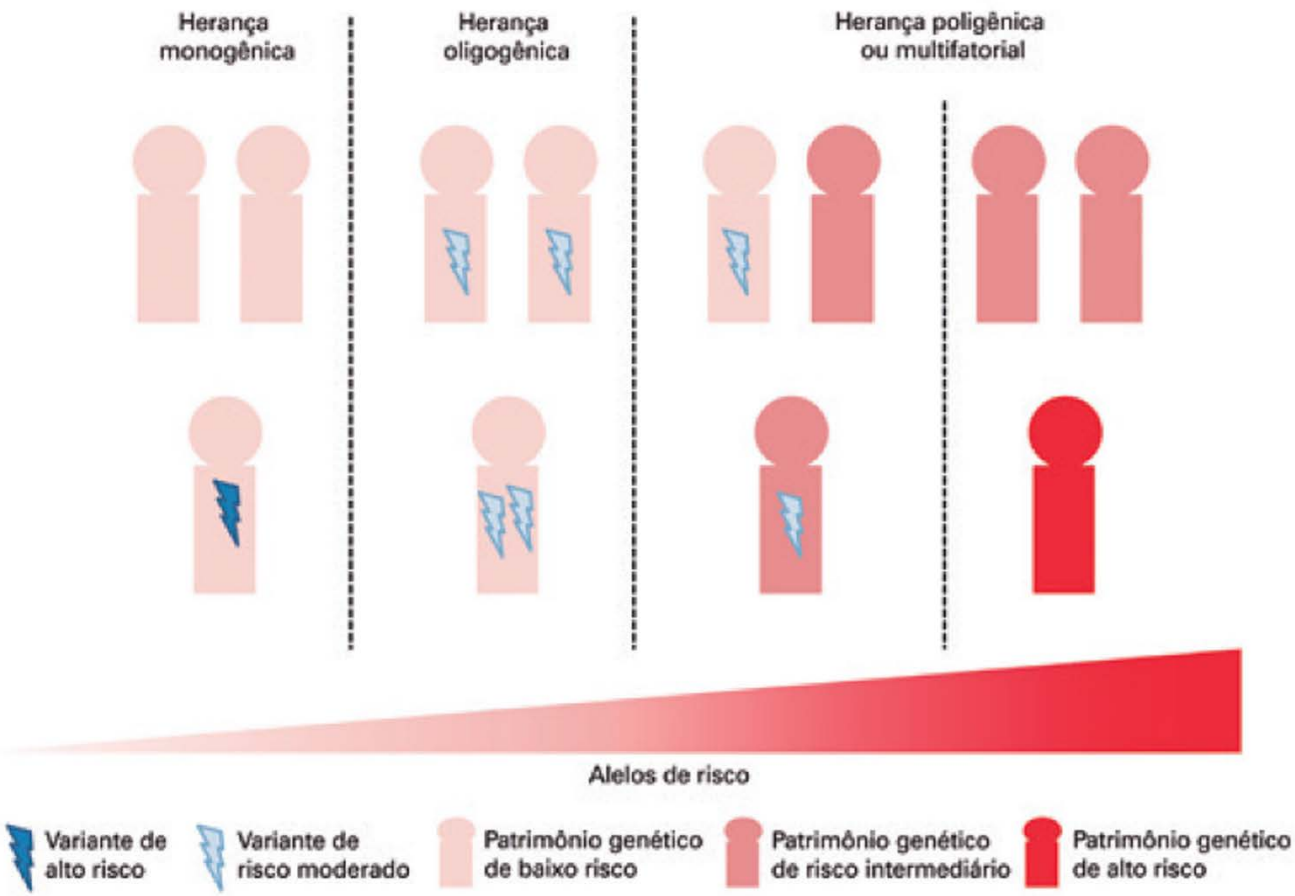

Figura 1.1: Interação entre variantes raras e comuns na manifestação do TEA. O TEA é um transtorno complexo que pode se encaixar em diferentes modelos de herança. Em alguns casos, uma única variante de novo de alto efeito ou algumas variantes raras de efeito médio são suficientes para sua manifestação, representando modelos de herança mono e oligogênicos, respectivamente (esquerda e centro). Ainda, de acordo com um modelo poligênico, uma carga média de variantes comuns de risco somada a poucas variantes de efeito médio ou uma carga muito alta de variantes comuns de risco podem levar à manifestação do transtorno (direita). Retirado de Griesi-Oliveira & Sertié, 2017.

1.3 Desafios e possibilidades no estudo do Transtorno do Espectro Autista

Um grande obstáculo no estudo de doenças que acometem o sistema nervoso central sempre foi a obtenção de tecido para condução de experimentos celulares e moleculares. No caso do TEA, no qual o cérebro em desenvolvimento é o órgão-chave para a compreensão da doença, a impossibilidade de retirada de tecido por meio de biopsias e a dificuldade de se obter amostras *post-mortem* foram por algum tempo um

grande entrave no avanço de estudos nessa área. Entretanto, o desenvolvimento das células-tronco pluripotentes induzidas (iPSC, do inglês *induced pluripotent stem cells*), resultado de um processo em que células somáticas são reprogramadas a um estado pluripotente a partir da superexpressão de genes específicos (Takahashi et al., 2007; Yu et al., 2007), tornou possível a superação desse obstáculo. Após a reprogramação, as iPSC podem ser rediferenciadas em tipos celulares provenientes de qualquer um dos três folhetos embrionários, incluindo neurônios. A grande vantagem dessa técnica no estudo do TEA é a obtenção de modelos neuronais já predispostos à doença, uma vez que os neurônios obtidos apresentam o genoma do indivíduo doador. Dessa forma, é possível usar as iPSC como modelo na identificação de alterações na expressão gênica, morfologia e função neuronal de pacientes e no estudo da neurogênese, particularmente importante em doenças de neurodesenvolvimento como o TEA (Lin et al., 2012).

Trabalhos realizados tendo células progenitoras neurais (NPC, do inglês *neural progenitor cells*) e neurônios derivados de iPSC como modelo mostram que essa é de fato uma abordagem promissora para o estudo de transtornos que acometem o sistema nervoso. Griesi-Oliveira e colaboradores (2014) identificaram uma translocação equilibrada que causou o rompimento do gene TRPC6 (que codifica para um canal de cátion independente de voltagem e permeável a Ca^{2+}) em um paciente com TEA e reprogramaram células de polpa de dente a um estado pluripotente a fim de entender as consequências funcionais do rompimento desse gene durante o processo de diferenciação neuronal. Usando essa estratégia, o grupo foi capaz de mostrar que o influxo de Ca^{2+} estava alterado nas NPC do paciente e que seus neurônios apresentavam neuritos mais curtos, menor arborização, densidade reduzida de espinhos dendríticos, menor número de sinapses e correntes de Na^{+} comprometidas em comparação às células derivadas dos controles. Nesse trabalho, também foi verificado o efeito de duas drogas diferentes sobre o fenótipo dos neurônios desse paciente, a hiperforina (a qual age ativando o canal TRPC6) e o IGF, e foram obtidos resultados positivos de resgate do fenótipo nas duas situações. Mais recentemente, Marchetto e colaboradores (2016), usando NPC e neurônios diferenciados a partir de iPSC derivadas de indivíduos afetados pelo TEA apresentando macrocefalia, mostraram que as células progenitoras dos pacientes apresentavam proliferação acelerada, enquanto alterações na neurogênese e redução na sinaptogênese foram observadas nos neurônios desses indivíduos. O grupo verificou ainda que as alterações observadas nos neurônios levaram a problemas na formação de redes neuronais, fenótipo que foi resgatado por meio do uso de IGF-1. Dessa forma, esses estudos mostram que a obtenção de células neurais a partir de

reprogramação celular é de fato uma abordagem viável tanto para o estudo dos mecanismos celulares e moleculares subjacentes à doença quanto para a triagem de drogas que apontem para potenciais terapias.

Um segundo fator que representa um grande desafio no estudo do TEA é a própria heterogeneidade genética associada ao transtorno. Como já explicitado, o desenvolvimento de técnicas mais sensíveis e que possibilitam uma análise ampla do genoma (ou do exoma) proporcionou um grande aumento no número de genes de risco identificados, porém, dada a complexidade da arquitetura genética do TEA, ainda há grande dificuldade de se identificar a causa genética em uma proporção importante dos casos. Nesse contexto, estudos de expressão gênica global surgem como uma estratégia interessante para contornar esse problema, conforme descrito a seguir.

1.4 Estudos de expressão gênica global

Os estudos de expressão gênica global surgem como uma alternativa para contornar o problema que a alta heterogeneidade genética e a dificuldade de identificar variantes de risco de baixo efeito trazem para o estudo do TEA, uma vez que o foco da análise migra do nível genômico, no qual os pacientes são altamente heterogêneos, para o nível transcriptômico, no qual é possível que se encontre uma maior homogeneidade entre os pacientes em termos de vias de sinalização e processos biológicos desregulados em função de alterações na expressão gênica (Han et al., 2014). Nesse contexto, os estudos de expressão gênica global representam também uma estratégia relevante para a identificação de marcadores moleculares que possam ser empregados no desenvolvimento de ferramentas diagnósticas, contribuindo para tornar o diagnóstico do TEA menos subjetivo do que ele é atualmente.

Para a realização desses estudos, são utilizadas técnicas que permitem a análise global da expressão gênica de um tecido em um dado momento, como o *microarray* e o RNAseq (estudo de expressão baseado no sequenciamento de nova geração) (Han et al., 2014). Enquanto o *microarray* quantifica a hibridação de cDNA a oligos específicos (pequenas sequências de DNA representando genes pré-determinados) para definir o nível de expressão dos genes que estão representados na lâmina, o RNAseq se baseia no sequenciamento do cDNA e alinhamento dos *reads* sequenciados a um genoma (previamente sequenciado ou montado a partir dos dados do RNAseq), sendo a expressão de determinado gene obtida a partir do número de *reads* que se alinham a ele (Zhao, Fung-Leung, Bittner, Ngo, & Liu, 2014; Govindarajan, Duraiyan, Kaliyappan, & Palanisamy, 2012); ambas as abordagens

podem ser usadas para comparar as diferenças de expressão gênica entre duas amostras (figuras 1.2 e 1.3). As diferenças entre essas duas técnicas fazem com que o *microarray* venha sendo progressivamente substituído pelo RNAseq, uma vez que problemas inerentes à primeira, como o ruído de fundo das hibridações e a limitação da análise aos genes representados pelos oligos da lâmina, são eliminados na última. Além disso, o RNAseq torna possível a identificação de novos genes e isoformas, a quantificação de diferentes isoformas de um gene isoladamente e a avaliação de expressão alelo-específica, abrindo novas possibilidades para os estudos de expressão (Zhao, Fung-Leung, Bittner, Ngo, & Liu, 2014; Malone & Oliver, 2011; Wang, Gerstein, & Snyder, 2009).

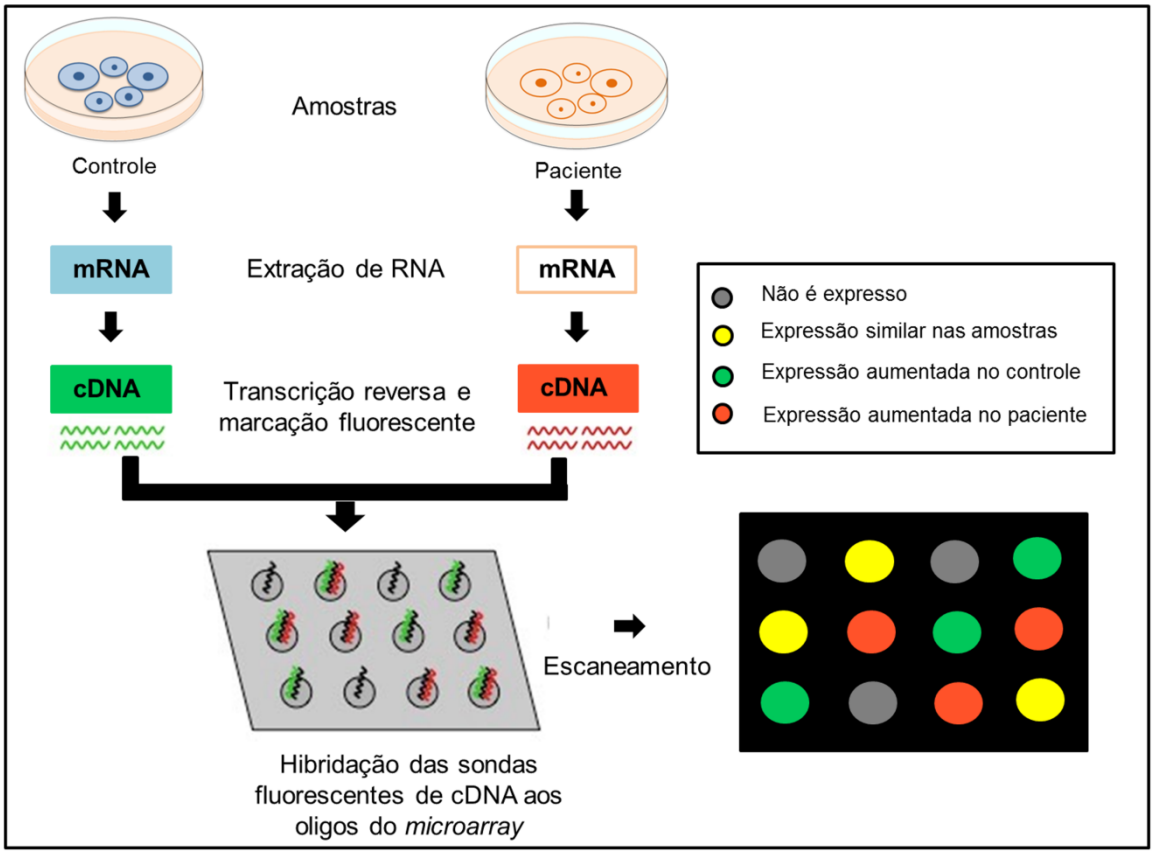

Figura 1.2: Princípios do *microarray*. O *microarray* é baseado na comparação entre a hibridação de cDNA com marcação fluorescente, proveniente de amostras distintas, a pequenas sequências de DNA que representam genes específicos. Nessa técnica, a diferença de expressão gênica entre as amostras é obtida por meio da quantidade relativa de fluorescência emitida em cada um dos poços onde os oligos se encontram.

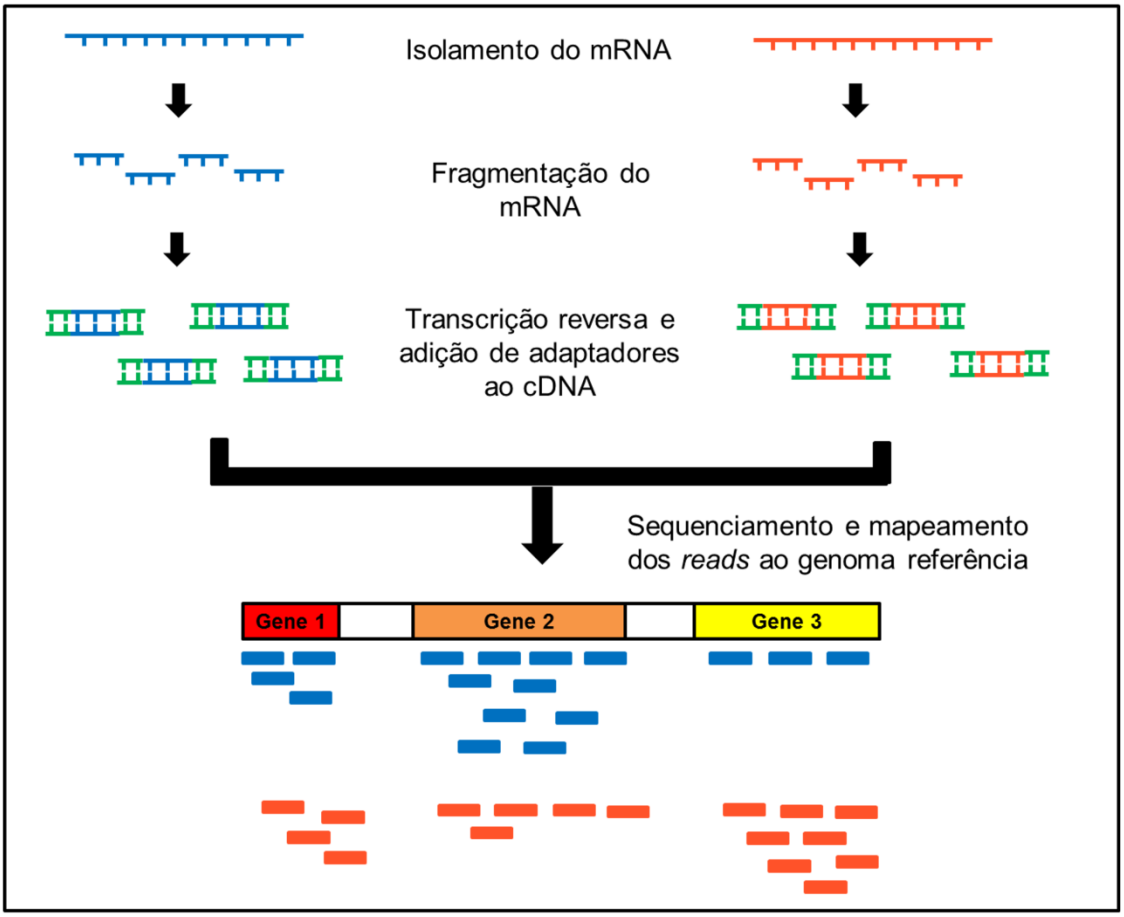

Figura 1.3: Princípios do RNAseq. Nessa técnica, o mRNA é isolado, fragmentado e usado como molde para a síntese de cDNA. Às móleculas de cDNA, são adicionados adaptadores, necessários para a realização do sequenciamento de nova geração. Após o sequenciamento, os *reads* gerados são alinhados contra um genoma referência e a diferença de expressão gênica entre duas amostras pode ser obtida pela diferença no número de *reads* que se alinham aos genes.

Até o momento, uma série de trabalhos utilizando uma dessas técnicas e diversos modelos de estudos para o TEA, que variam desde diferentes tipos celulares até amostras de cérebro *post-mortem*, já foram publicados e sugerem o comprometimento de vias biológicas específicas nos pacientes (Ansel, Rosenzweig, Zisman, Melamed & Gesundheit, 2017).

Por meio do uso de *microarrays*, Kong e colaboradores (2013) usaram células de sangue periférico para analisar o perfil de expressão gênica de 20 indivíduos afetados pelo TEA pareados com seus irmãos não afetados. Controlando-se os possíveis efeitos de idade, sexo e família, foram identificados 189 genes diferencialmente expressos entre os dois grupos. Dentre eles, genes envolvidos em vias ribossômicas, spliceossômicas e mitocondriais foram encontrados com expressão aumentada entre os pacientes, enquanto genes envolvidos em vias neuroreceptor-

ligante, resposta imune e sinalização de cálcio tiveram expressão reduzida nesse grupo.

Griesi-Oliveira e colaboradores (2013) realizaram estudos de *microarray* a partir do mRNA extraído de células-tronco de dentes humanos decíduos esfoliados (SHED, do inglês *stem cells from human exfoliated deciduous teeth*), um tipo celular proveniente do mesmo folheto embrionário que o cérebro, e identificaram 683 genes diferencialmente expressos entre pacientes e controles. Desse total, aproximadamente metade dos genes tem expressão no cérebro e estão envolvidos em mecanismo previamente associados ao TEA, como síntese proteica, processamento de mRNA, neurogênese, adesão celular e montagem do citoesqueleto. Os genes identificados como sendo diferencialmente expressos foram distribuídos em 68 vias de sinalização canônicas diferentes, sendo as vias relacionadas à síntese proteica, organização do citoesqueleto e formação de neuritos destacadas e apontadas pelos autores como especialmente relevantes para o TEA.

Voineagu e colaboradores (2011), por sua vez, usando amostras de tecidos *post-mortem* de duas áreas corticais distintas do cérebro de 19 pacientes e 17 controles, realizaram análises de expressão gênica diferencial e de módulos de coexpressão gênica (isto é, grupos de genes cuja expressão é altamente correlacionada) entre esses grupos. Dentre os genes diferencialmente expressos entre pacientes e controles, aqueles com expressão diminuída nos pacientes estavam enriquecidos para categorias relacionadas à função sináptica, enquanto os com expressão aumentada mostraram enriquecimento para categorias relacionadas à resposta imune e inflamatória. A análise de módulos de coexpressão, por sua vez, apontou a existência de dois módulos correlacionados ao status da doença. Consistente com o observado na análise de expressão gênica diferencial, um desses módulos foi caracterizado pelo enriquecimento para genes envolvidos em função sináptica, transporte vesicular e projeção neural, enquanto o outro apresentou uma grande quantidade de genes relacionados a resposta imunológica e marcadores gliais. Análises adicionais mostraram ainda que o módulo neuronal estava enriquecido com genes de risco para o TEA e o módulo imune não, sugerindo que as alterações na expressão gênica observadas no primeiro têm base genética, enquanto as identificadas no segundo provavelmente se devem a fenômenos secundários ou a fatores ambientais.

A partir desse mesmo modelo, mas usando um número maior de amostras de tecido cortical (57 provenientes de controles e 47 de pacientes), Gupta e colaboradores (2014) identificaram três módulos de coexpressão diferencialmente regulados entre os dois grupos. Dentre esses módulos, um estava enriquecido para

genes envolvidos em resposta imunológica e ativação M2 da micróglia enquanto os dois restantes estavam envolvidos em transmissão sináptica, estando um deles enriquecido para genes de atividade de canais iônicos relacionados a GABA e o outro para genes relacionados à sinalização por peptídeos e hormônios. Os autores acreditam que, ao aumentarem o número amostral em relação ao estudo de Voineagu e colaboradores (2011), foram capazes de criar módulos mais específicos e apontar com maior precisão os genes diferencialmente expressos entre os grupos.

Mais recentemente, foram publicados estudos de expressão utilizando NPC e neurônios derivados de iPSC como modelo. Mariani e colaboradores (2015) ao realizarem análises de redes de coexpressão gênica em organóides cerebrais derivados de iPSC de pacientes com TEA apresentando macrocefalia e de pais não afetados, mostraram que módulos enriquecidos para genes envolvidos em regulação transcricional de proliferação e destino celular, diferenciação neuronal e transmissão sináptica estavam superexpressos nos indivíduos afetados. Buscando validar funcionalmente essas análises, o grupo observou a ocorrência de alterações nas células dos pacientes que sugeriam aumento na maturação neuronal, supercrescimento sináptico e superprodução de NPC e neurônios inibitórios, gerando um desbalanço entre o número de células excitatórias e inibitórias. Ainda, o grupo conseguiu demostrar que a superprodução de neurônios inibitórios é decorrente da superexpressão de um fator de transcrição específico, denominado FOXG1. Marchetto e colaboradores (2016), por sua vez, usando NPC e neurônios derivados de iPSC de um grupo de pacientes com TEA apresentando macrocefalia (variando entre suave e extrema) e de controles, mostraram que as NPC dos indivíduos afetados apresentam aumento na proliferação e alterações na diferenciação em relação às NPC dos controles, enquanto seus neurônios apresentam neurogênese alterada e sinaptogênese reduzida, levando a problemas nas redes neuronais formadas. Os dados do RNAseq obtidos nesse trabalho corroboram o fenótipo celular observado, uma vez que, dentre os genes diferencialmente expressos identificados nas NPC, muitos dos que apresentam expressão aumentada entre os pacientes estão envolvidos no desenvolvimento do cérebro. Já nos neurônios, os genes com expressão diminuída estão enriquecidos para as categorias cílios e axonema, consistente com as alterações sinápticas observadas nessas células. Além disso, o grupo identificou 35 genes com expressão alterada entre pacientes e controles ao longo do processo de diferenciação de NPC para neurônios, muitos dos quais estão envolvidos na formação de canais de cálcio dependentes de voltagem.

Apesar das pequenas diferenças existentes entre os estudos, é possível notar uma grande consistência nos resultados apresentados por eles, mostrando que as

alterações genéticas observadas nos indivíduos acometidos pelo transtorno, ainda que distintas, devem convergir em vias de sinalização e processos biológicos comuns e reforçando a ideia de que a análise global do transcriptoma é de fato uma abordagem bastante promissora para o estudo do TEA. Além disso, estudos que fazem esse tipo de análise usando células neurais derivadas de iPSC, ainda bastante escassos, podem acrescentar dados valiosos ao corpo de conhecimento já construído, uma vez que esse modelo permite não só que se estude mais profundamente as vias moleculares já previamente identificadas e as refine, mas também que se monitore a célula desde o estágio em que está indiferenciada até se tornar um neurônio, tornando possível a análise dessas e de outras vias que possam estar alteradas durante toda a diferenciação neuronal.

1.5 Estudos de desvio de expressão alélica

Considerando que variantes que têm alto impacto no sucesso reprodutivo dos indivíduos tendem a ser raras na população, uma abordagem bastante popular no estudo do TEA é a busca por variantes *de novo* que possam estar associadas à manifestação de seus sintomas clínicos. Esse tipo de abordagem, apesar de importante na identificação de genes e *loci* relevantes, acaba por excluir quaisquer informações sobre variantes herdadas, as quais também devem ter papel na etiologia da doença, considerando-se, por exemplo, a agregação familiar observada em alguns casos. Nesse sentido, a expressão preferencial de um dos alelos de um gene em detrimento do outro (isto é, o desvio de expressão alélica) surge como um mecanismo que poderia explicar alguns dos fenômenos observados no TEA e em outras doenças neuropsiquiátricas, tais como penetrância incompleta e expressividade variável.

Nesse contexto, os dados de expressão gênica de um dado tipo celular provenientes de diferentes indivíduos, quando obtidos por meio de RNAseq, além de nos fornecer informações sobre as diferenças de expressão gênica associadas a uma determinada condição, podem ainda indicar a ocorrência de expressão preferencial de um dos alelos de um gene. O sequenciamento do mRNA, resultante do emprego dessa técnica, torna possível apontar a contribuição de cada alelo, quando em heterozigose, para o total de transcritos de um gene e, consequentemente, verificar se existe desvio de expressão para um deles.

Até pouco tempo atrás, apesar da expressão monoalélica em decorrência de *imprinting*, inativação do cromossomo X e eventos aleatórios específicos já ter sido bem documentada para determinados genes, ainda não estava claro para a maior parte do genoma se os dois alelos de um gene eram simultaneamente utilizados como

molde para a transcrição e, em caso positivo, se ambos contribuíam de forma similar para o total de transcritos. A fim de investigar essa questão, Deng e colaboradores (2014) realizaram o RNAseq de células embrionárias individuais de camundongos e mostraram que entre 12 e 24% dos genes tem expressão monoalélica em um dado momento e que o alelo ativo em uma célula pode não ser o mesmo que está ativo nas células vizinhas, apontando a estocasticidade do processo. Os genes que apresentam expressão bialélica são, em geral, constitutivos. Ainda, o grupo obteve resultados similares em fibroblastos e hepatócitos murinos, demonstrando que o mesmo processo ocorre em células maduras. Em 2015, Borel e colaboradores, fazendo uso de RNAseq em fibroblastos humanos individuais, obtiveram resultados semelhantes, sugerindo que o mesmo fenômeno deve estar ocorrendo também em células humanas.

O trabalho de Lin e colaboradores (2012) também demonstrou a ocorrência de desvio de expressão alélica em neurônios humanos. Usando um *pool* de células neuronais derivadas de iPSC, o grupo identificou 801 genes apresentando desvio de expressão alélica. O grupo verificou ainda que o desvio observado não se devia a origem parental do alelo, também sugerindo que eventos estocásticos devam estar envolvidos nesse processo. Além disso, dentre os genes identificados, havia um pequeno enriquecimento para genes previamente associados à esquizofrenia e ao TEA.

Ainda, Gregg e colaboradores (2010) verificaram, no cérebro de camundongos adultos, a ocorrência de desvio de expressão alélica de acordo com o sexo do indivíduo e a origem parental do alelo. Alguns dos genes identificados pelo grupo são candidatos para transtornos neurológicos ou tem função relevante no comportamento.

Esses estudos demonstram que o desvio de expressão alélica é um mecanismo comum (ocorrendo inclusive em neurônios) que pode ter relação com o desvio sexual ou a penetrância incompleta observados em transtornos que afetam o sistema nervoso, reforçando a hipótese de que ele pode ter um papel relevante na etiologia do TEA e que, portanto, investigações mais profundas neste sentido podem ser uma abordagem promissora para o estudo dessa doença.

2. OBJETIVOS

Os estudos de expressão gênica global conduzidos até o momento demonstram a existência de alterações em processos relevantes para o neurodesenvolvimento que são compartilhadas pelos indivíduos afetados pelo TEA. Trabalhos que fazem esse tipo de análise usando células neurais derivadas de iPSC ainda são bastante escassos, mas são importantes para a compreensão do transtorno uma vez que possibilitam o monitoramento da expressão gênica em células provenientes de indivíduos afetados durante diversos estágios da diferenciação neuronal. Os dois estudos realizados até o momento a partir desse modelo são baseados em coortes de indivíduos afetados pelo TEA que apresentam macrocefalia, o que indica a necessidade de condução desses experimentos em outros grupos de pacientes. Além disso, o estudo do transcriptoma pela técnica de RNAseq permite a análise de parâmetros ainda pouco explorados no TEA, como por exemplo, o possível papel do mecanismo de desvio de expressão alélica na etiologia do transtorno.

Posto isso, a partir da realização do RNAseq em NPC e neurônios derivados de iPSC de seis indivíduos afetados pelo TEA e seis controles, os objetivos desse trabalho são: (1) identificar vias de sinalização e processos biológicos que estejam alterados entre pacientes e controles por meio de análises de expressão gênica diferencial e de redes de coexpressão gênica em células neurais e (2) verificar se genes candidatos ao TEA agrupados no banco de dados SFARI apresentam desvio de expressão alélica nesses mesmos tipos celulares. Ainda, durante a caracterização genética de nossos pacientes, identificamos uma duplicação na região 17p13.3 afetando seis genes em um dos indivíduos. Os impactos transcricionais e funcionais da duplicação foram avaliados em NPC e neurônios desse paciente e os resultados desse estudo de caso são descritos no capítulo 7 da presente dissertação.

3. MATERIAL E MÉTODOS

3.1 Pacientes

Os pacientes e familiares que fazem parte deste estudo foram atendidos no Centro de Pesquisa sobre o Genoma Humano e Células-Tronco (CEGH-CEL), IB/USP. Estes pacientes foram encaminhados ao nosso serviço após serem diagnosticados por profissionais especializados com base nos critérios estabelecidos pelo DSM-IV (American Psychiatric Association, 1994). Nossos principais colaboradores são: Dr. Estevão Vadasz e equipe, do Serviço de Psiquiatria da Infância e da Adolescência (Faculdade de Medicina / Hospital das Clínicas da Universidade de São Paulo), e Dr. Wanderley Domingues, diretor do Centro Pró-Autista.

Em nosso atendimento, levantamos os dados relativos à gestação da criança, seu desenvolvimento nos primeiros anos de vida e histórico familiar, a fim de identificar outros indivíduos na família portadores de TEA ou outros transtornos psiquiátricos. Visando o conhecimento da sintomatologia dos pacientes, bem como o grau de comprometimento destes com relação às áreas de interação social, comunicação e comportamentos repetitivos, nosso grupo utiliza o questionário CARS (Childhood Autism Rating Scale) e o teste neuropsicológico WISC-IQ, para avaliação do coeficiente de inteligência (tabela 3.1), os quais foram aplicados pela psicóloga Dra. Elaine Zachi.

Para a realização desse estudo, foram selecionados seis pacientes não-sindrômicos do sexo masculino. Buscando formar uma amostra o mais homogênea possível frente a variabilidade do transtorno, procuramos compor nossa amostra exclusivamente por pacientes de alto funcionamento (isto é, indivíduos que não apresentam comprometimento da capacidade cognitiva). No entanto, um paciente formalmente avaliado por teste neuropsicológico apenas após o início do estudo, mostrou um QI<70, portanto sendo classificado como um paciente de baixo funcionamento (tabela 3.1). Na época da coleta do material biológico (dentes decíduos para o estabelecimento de linhagens celulares e sangue para a realização de testes moleculares específicos), os pacientes apresentavam idades variando entre 6 e 12 anos. Como controle, utilizamos o material de seis indivíduos do sexo masculino não afetados pelo TEA, também com idades variando entre 6 e 12 anos.

Todos os pacientes foram submetidos ao teste molecular para diagnóstico diferencial da Síndrome do X-Frágil (executado pela Mestra Vanessa Van Optiz, do CEGH-CEL). Mutações no gene FMR1 não foram encontradas em nenhum dos

pacientes. Além disso, foram realizados o sequenciamento do exoma (com o apoio da equipe do CEGH-CEL) e o teste de *array* CGH (em colaboração com o grupo da Prof. Dra. Carla Rosemberg) para cada um dos pacientes, a fim de posteriormente se realizar uma investigação mais ampla no genoma desses indivíduos (anexo 1).

Tabela 3.1: resultados dos testes CARS e WISC-IQ obtidos pelos pacientes incluídos em nossa amostra. Os indivíduos com índice <70 no WISC-IQ são considerados autistas de baixo funcionamento.

Registro	CARS	WISC-IQ
F2613	NA	NA
F2688	30.5	110
F2735	32	97
F6136	NA	NA
F6281	37.5	60
F7511	30.5	95

3.2 Aspectos éticos

Os projetos “Novas abordagens no estudo das causas genéticas dos transtornos do espectro autista”, que propõe o uso de células-tronco dos pacientes autistas para a pesquisa, e "Análise genômica e funcional dos transtornos do espectro autista: perspectivas de novas ferramentas de diagnóstico precoce" foram aprovados pelo Comitê de Ética em Pesquisa do Instituto de Biociências - USP (protocolos n° 049/2006 e 1.133.486, respectivamente).

3.3 Estabelecimento das culturas de células-tronco de dentes humanos decíduos esfoliados (SHED), geração de iPSC e diferenciação em NPC

Os procedimentos experimentais que envolvem o estabelecimento das SHED (*stem cells of human exfoliated deciduous teeth*), a geração de iPSC a partir de sua reprogramação e as etapas iniciais da diferenciação neuronal, originando as NPC, serão brevemente descritos aqui, uma vez que foram previamente realizados por nosso grupo e não foram repetidos durante o desenvolvimento desse estudo.

Dentes decíduos de pacientes e controles tiveram a polpa removida da coroa. As polpas foram então digeridas em uma solução de 0,25% de tripsina (Invitrogen) a 37°C por 30 minutos. Após a digestão, as polpas foram cultivadas em placas de seis poços com meio DMEM/F12 (Invitrogen) suplementado com 15% de soro fetal bovino (FBS - Hyclone) e 100M/ml estreptomicina (Biofluids) e incubadas a 37°C em 5% CO_2.

O meio foi trocado uma vez por semana até que as células-tronco (SHED) fossem liberadas da polpa (Costa et al., 2008).

As SHED foram então induzidas a expressar *SOX2, c-Myc, OCT4* e *KLF4* por meio da transdução retroviral desses genes, previamente descritos como importantes para a reprogramação de células diferenciadas a um estado pluripotente. Após a transdução, as SHED foram cultivadas sobre fibroblastos embrionários murinos, os quais se acredita que liberem fatores importantes para o processo de reprogramação e manutenção do estado pluripotente. Após aproximadamente duas semanas, as colônias de iPSC foram identificadas e transferidas para placas cobertas com matrigel (BD Biosciences) e cultivadas em meio mTeSR (Stem Cell Technologies). Essa reprogramação celular foi feita de acordo com o protocolo estabelecido por Takahashi e colaboradores (2007). Foram gerados de dois a três clones de iPSC para cada indivíduo, que usamos como replicatas biológicas nos experimentos.

A diferenciação das NPC a partir das iPSC foi realizada de acordo com o protocolo estabelecido por Marchetto e colaboradores (2010). Assim, quatro dias após o plaqueamento das iPSC, o meio mTeSR foi substituído por DMEM/F12 (Cellgro) suplementado com N2 e B27 (meio NB) a uma concentração final de 0,5X e 1μM de dorsomorfina. Após dois dias sob essas condições de cultivo, as colônias foram transferidas para placas de baixa aderência, formando corpos embrióides (EB, do inglês *embryoid bodies*), que foram cultivados com meio 0,5X NB e 1μM de dorsomorfina por mais dois dias sob agitação a 100rpm. No terceiro dia, a dorsomorfina foi substituída por 20ng/ml de FGF e 20ng/ml de EGF e o meio foi trocado a cada dois dias, até se completar uma semana de cultivo dos EBs. No sétimo dia, os EBs foram dissociados com o auxílio da enzima accutase (Millipore). As células obtidas após a dissociação foram plaqueadas sobre matrigel e mantidas em meio 0,5X NB acrescido de 20ng/ml de FGF e EGF, ocorrendo então a formação de rosetas. Após um período entre dois e três dias, as rosetas foram coletadas manualmente e transferidas para uma nova placa coberta por matrigel. Três dias após a transferência, as células foram removidas e replaqueadas em placas cobertas com 10μg/ml de poli-ornitina (Sigma) e 5 μg/ml de laminina (Invitrogen) e mantidas em meio 0,5X NB suplementado com 20ng/ml de EGF e FGF, onde formaram uma população homogênea de NPC. As células obtidas foram então congeladas com o meio CryoStor CS10 (StemCell Technologies) e armazenadas em nitrogênio líquido para serem utilizadas em experimentos posteriores.

3.4 Transdução das NPC com o vetor Synapsin::EGFP

Parte das NPC geradas foi transformada com o vetor Synapsin::EGFP (construção contendo o promotor do gene Synapsin dirigindo a transcrição de GFP – *green fluorescent protein*) por meio de transdução lentiviral. Como o gene que codifica para a sinapsina é transcrito apenas em neurônios, as células diferenciadas em neurônios passam a expressar GFP, o que pode ser usado como estratégia para isolar essa subpopulação da cultura. Para realizar a transdução, é necessário que o vetor seja previamente empacotado no envelope viral, o que é feito usando células HEK293T.

Para isso, células HEK293T foram cultivadas em placas de 150 mm com meio DMEM *high glucose* suplementado com 10% de FBS a 37°C em 5% CO_2. Quando as células atingiram a confluência necessária, foi realizada a transfecção usando PEI, o vetor Synapsin::EGFP e as sequências de DNA que codificam para proteínas estruturais do lentivírus (MDL, RSV e VSVG). A transfecção foi feita com PEI a uma quantidade final quatro vezes maior que a quantidade total de DNA. Para uma placa de 150 mm, foi preparada uma solução com 1ml de DMEM *high glucose*, 110µg de PEI e 27,5µg de DNA (12,2µg de Synapsin::EGFP, 8,1µg de MDL, 3,1µg de RSV e 4,1µg de VSVG). A solução foi incubada por cinco minutos, a fim de que o PEI se associasse às sequências de DNA, e logo em seguida distribuída na placa. As células foram então incubadas com a solução por seis horas, ao final das quais o meio foi removido e trocado novamente por meio DMEM *high glucose* suplementado com 10% de FBS. Três dias após a transfecção, o sobrenadante foi coletado, filtrado e centrifugado a 19.400 rpm por 2h a 12°C. Após a centrifugação, o sobrenadante foi removido, adicionamos 100µl DMEM *high glucose* ao pellet e o incubamos *overnight* a 4°C para que os vírus fossem lentamente ressuspendidos durante esse intervalo. À ressuspensão de vírus, adicionamos meio suficiente para completar 700µl, que foi então usado para a transdução de uma placa de 35 mm contendo NPC com confluência entre 50 e 70%.

As NPC a serem transduzidas, cultivadas em meio NB suplementado com 20ng/ml de EGF e FGF, tiveram o meio NB completamente removido e foram incubadas entre cinco e seis horas com a ressuspensão viral a 37°C em 5% CO_2. Após a incubação, a ressuspensão foi substituída novamente por meio NB suplementado. As células obtidas foram congeladas com o meio CryoStor e armazenadas em nitrogênio líquido para serem utilizadas posteriormente.

3.5 Diferenciação das NPC em neurônios

As NPC foram descongeladas em placas cobertas com 10µg/ml de poli-ornitina e 5µg/ml de laminina, onde foram mantidas em meio NB suplementado com 20ng/ml de EGF e FGF até terem confluência suficiente para serem transferidas para as placas de diferenciação e recongeladas para manutenção do estoque. Para a diferenciação neuronal, as NPC foram plaqueadas em baixa densidade (250 mil células para uma placa de 60 mm) em placas cobertas com 20µg/ml de poli-ornitina e 10µg/ml de laminina. Após o plaqueamento, as células foram mantidas em meio 0,5X NB suplementado com EGF e FGF por aproximadamente 24h, ao final das quais o meio foi substituído por meio 1X NB suplementado com 1 µM de ácido retinóico, tendo então início a diferenciação em neurônios.

As células em diferenciação foram mantidas no meio suplementado com ácido retinóico por quatro semanas, tendo metade do volume de meio trocado três vezes por semana para que a indução da diferenciação fosse mantida durante todo o tempo. Ao final desse período, obtivemos uma população apresentando neurônios maduros, já emitindo prolongamentos e formando conexões uns com os outros.

3.6 Imunofluorescência

As NPC e os neurônios foram plaqueados sobre lamínulas em placas de 24 poços a uma quantidade final de 100 mil células/poço e mantidos em meio NB (suplementado com EGF e FGF ou ácido retinóico) em condições adequadas. Aproximadamente 24 (para as NPC) ou 48h (para os neurônios) após o plaqueamento, o meio foi retirado dos poços e as células foram fixadas com paraformaldeído (PFA) por 10 minutos. Os poços foram então lavados com PBS 1X por 3 minutos em agitação. A lavagem foi repetida por três vezes e, após a última, as células foram mantidas em solução de bloqueio (5% de soro de burro e 0,1% de Triton em PBS 1X) durante 40 minutos. Depois do bloqueio, as células foram incubadas por aproximadamente 16h com anticorpos primários específicos em agitação a 4°C. Os anticorpos primários usados foram anti-SOX2 (1:100, Millipore), anti-Nestina (1:250, Millipore), anti-CTIP2 (1:100, Abcam) e anti-MAP2 (1:500, Millipore). Ao final desse período, os poços foram novamente lavados por três vezes com PBS 1X, as células mantidas em solução de bloqueio por 15 minutos e, em seguida, incubadas com anticorpos secundários apropriados por 1h. Os anticorpos secundários usados foram anti-mouse e anti-rabbit (1:250, LifeTechnologies). Os poços foram novamente lavados em PBS 1X e, após a última lavagem, as lâminas foram preparadas em Vecta Shield com DAPI. As imagens foram feitas no microscópio confocal IX51 Olympus.

3.7 Separação dos neurônios GFP+

Após as quatro semanas de diferenciação, as células das linhagens transformadas com o vetor Synapsin::EGFP foram submetidas a separação no citômetro FACSAria. Para isso, as células foram incubadas com a enzima TrypLE Express (Life Technologies) a 37°C por 20 minutos, coletadas em PBS com 1% de FBS e 0,5% de EDTA e centrifugadas a 1.500 rpm por quatro minutos. Após a centrifugação, as células foram ressuspendidas em meio NB com 1% de FBS e 0,5% de EDTA e passadas para tubos de citometria com tampa de célula-filtro (para que as células fossem bem dissociadas) e levadas em seguida ao citômetro.

Para a separação dos neurônios, selecionamos a subpopulação de células apresentando corpo celular pequeno e alta intensidade de fluorescência verde. A definição desses parâmetros foi necessária porque notamos que a transcrição de GFP não é tão específica como imaginávamos a princípio e estava ocorrendo também em células de corpo celular maior que não apresentavam morfologia neuronal. Os neurônios recuperados foram submetidos a extração de RNA ou replaqueados para a realização de outros experimentos.

3.8 Extração do RNA total

Após as quatro semanas de diferenciação dos neurônios, as células foram dissociadas com o uso da enzima Tryple Express e o RNA foi extraído com uso do kit NucleoSpin RNA (Macherey-Nagel), de acordo com o protocolo do fabricante. O mesmo foi feito com os neurônios GFP+ depois de serem submetidos à separação pelo citômetro. A qualidade do RNA foi avaliada por eletroforese capilar no equipamento Bioanalyzer (Agilent Technologies) e a quantificação foi feita no espectrofotômetro Nanodrop (ThermoFisher).

3.9 Preparação das bibliotecas para sequenciamento do transcriptoma

A preparação das bibliotecas para sequenciamento do transcriptoma foi realizada com o uso do kit Truseq RNA Sample Preparation (Illumina), seguindo protocolo do fabricante. Após preparação, as bibliotecas foram analisadas quanto ao tamanho e qualidade dos fragmentos por eletroforese capilar no equipamento Bioanalyzer (Agilent Technologies) e quantificadas por meio de PCR em tempo real para injeção de quantidade ideal no sequenciador, com o uso do kit LIBRARY QUANTIFICATION PRIMER PREMIX AND QPCR MASTERMIX for Roche LightCycler® (Kapa Biosystems).

As bibliotecas foram então corridas no sequenciador Hiseq 2500, objetivando-se um total de 30 milhões de *reads* por amostra, lidos nos dois sentidos. Para a realização da corrida, tivemos o apoio técnico da equipe do CEGH-CEL.

3.10 Processamento dos dados brutos do sequenciamento do RNA

Após o sequenciamento, os *reads* foram demultiplexados e os arquivos FASTQ, que contém tanto a sequência dos *reads* quanto a informação de qualidade de cada uma das bases sequenciadas, foram gerados. A qualidade de cada biblioteca foi analisada pelo FASTQC e as amostras que não obtiveram um mínimo de 20 milhões de *reads* sequenciados foram eliminadas (tabela 4.1). Os *reads* foram então filtrados no Trimmomatic (Bolger, Lohse, & Usadel, 2014), um programa que remove bases de baixa qualidade e descarta *reads* muito curtos. Essa etapa é importante para diminuir os erros gerados durante o sequenciamento e aumentar a eficiência do alinhamento, tanto por meio da redução de seu tempo de execução quanto pelo reaproveitamento de *reads* que seriam descartados devido ao elevado número de bases mal-pareadas contra o genoma. Após a filtragem no Trimmomatic, os *reads* foram alinhados contra o genoma humano (hg19) com o uso do TopHat v2.1.0. Após o alinhamento, selecionamos 20 milhões de *reads* aleatórios em cada amostra, cuja contagem foi sumarizada por gene com o uso do pacote HTSeq (Anders, Pyl & Huber, 2015), dando origem a uma matriz contendo o número de *reads* observado para cada gene em cada uma das amostras sequenciadas. O número de *reads* foi então normalizado pelo tamanho total da biblioteca e pelo tamanho de cada gene específico, gerando os valores FPKM (*fragments per kilobase million*) de cada gene, os quais foram usados para definir os genes transcricionalmente ativos. Os genes apresentando FPKM ≥ 1 em pelo menos metade das amostras foram mantidos na análise, enquanto o restante foi excluído. Esse passo se faz necessário uma vez que se acredita que transcritos pouco abundantes devem ser resultado de ruídos técnicos ou biológicos, não de atividade transcricional (Hart, Komori, LaMere, Podshivalova, & Salomon, 2013).

O processamento dos dados brutos foi realizado em colaboração com o grupo do Prof. Dr. Eduardo Moraes Rego Reis, do Instituto de Química – USP.

3.11 Deconvolução e normalização dos dados de expressão

A partir dessa etapa, os dados do sequenciamento das NPC, que haviam sido gerados e processados anteriormente pelo nosso grupo, foram incluídos nas análises.

Usamos um *pipeline* desenvolvido pelo grupo da Dra. Irina Voineagu, da University of New South Wales, para estimar a proporção neuronal de cada uma de nossas amostras. Ao realizar as análises de componente principal para NPC e neurônios separadamente, verificamos que as amostras se agrupavam de acordo com a proporção neuronal em ambos os tipos celulares e que os módulos de coexpressão gênica (ver seção 3.13) estavam fortemente correlacionados com essas proporções, indicando a necessidade de normalização dos dados (figura 3.2).

A normalização foi feita por meio do RUVseq, um pacote do R usado para remover variações indesejadas em dados de RNAseq (Risso, Ngai, Speed, & Dudoit, 2014), também em colaboração com o grupo da Dra. Irina Voineagu. Para estimar essas variações, o RUVseq utiliza dois parâmetros que precisam ser definidos pelo usuário: um conjunto de genes controle negativo (isto é, que não sejam influenciados pela covariável de interesse) e um número k de fatores de variação indesejada. As etapas seguidas para a definição das listas de genes controle e dos valores k estão resumidas na figura 3.1. Brevemente, para gerar a lista de genes controle, identificamos os genes não diferencialmente expressos entre pacientes e controles nas NPC ou nos neurônios por meio da realização de uma análise de expressão gênica diferencial preliminar com o uso do pacote DESeq2 (Love, Huber, & Anders, 2014). Baseado nos p-valores de diferença de expressão obtidos por esta análise, geramos diferentes listas de genes controle, variando-se o p-valor usado para corte da lista entre 0,1 e 0,9. Cada uma dessas listas foi então sobreposta a uma lista de genes constitutivos previamente compilada (Eisenberg & Levanon, 2013), dando origem às listas finais de genes controle. Para cada lista de genes controle usada no RUVseq, variamos também o valor do fator k da análise de 1 a 5. A lista final de genes controle e o valor de k a ser utilizado para normalizar os dados de expressão foram escolhidos com base na análise de três parâmetros: (1) análise de componente principal: verificamos se houve uma tendência a se desfazer o efeito de agrupamento por proporção neuronal das amostras; (2) análise de clusters hierárquicos: verificamos se no cluster hierárquico houve tendência a se desfazer o agrupamento por proporção neuronal, e uma tendência de agrupamento das réplicas amostrais; (3) análise dos módulos de coexpressão gênica: verificamos se os módulos de coexpressão gênica gerados pelo pacote WGCNA (ver seção 3.13) deixavam de ter correlação significativa com a proporção neuronal. Assim, para as NPC, os parâmetros escolhidos foram p-valor para corte da lista dos genes não diferencialmente expressos ≥ 0.3 e fator k=2, enquanto para os neurônios, o p-valor escolhido foi de ≥ 0.4 e k=4 (figura 3.2). As tabelas com os *counts* normalizados pelo RUVseq foram então utilizadas nas análises subsequentes.

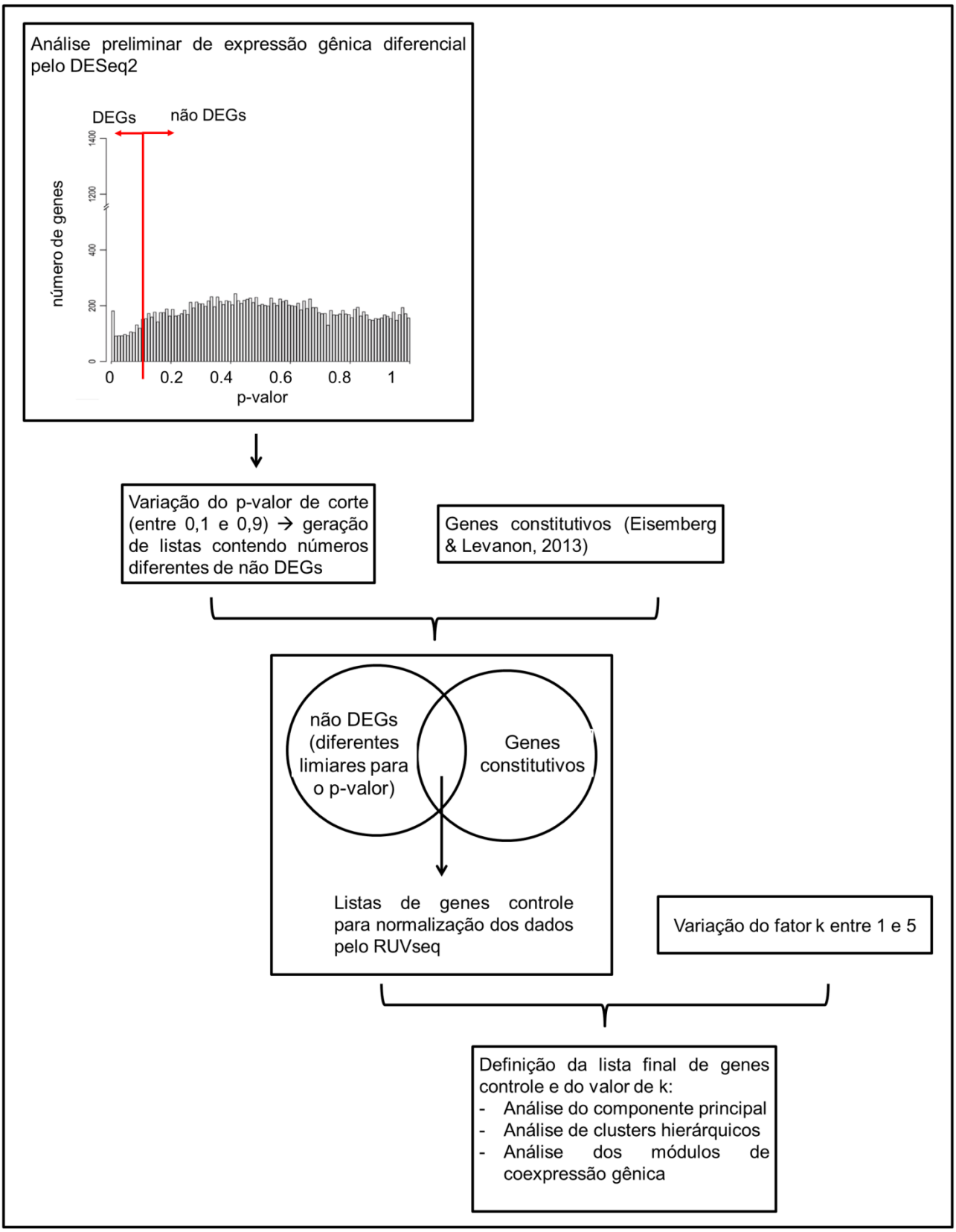

Figura 3.1: Determinação dos parâmetros para normalização dos dados de expressão gênica. O fluxograma descreve os diferentes testes realizados para a definição dos parâmetros que melhor desfaziam o efeito da proporção neuronal sobre nossos dados.

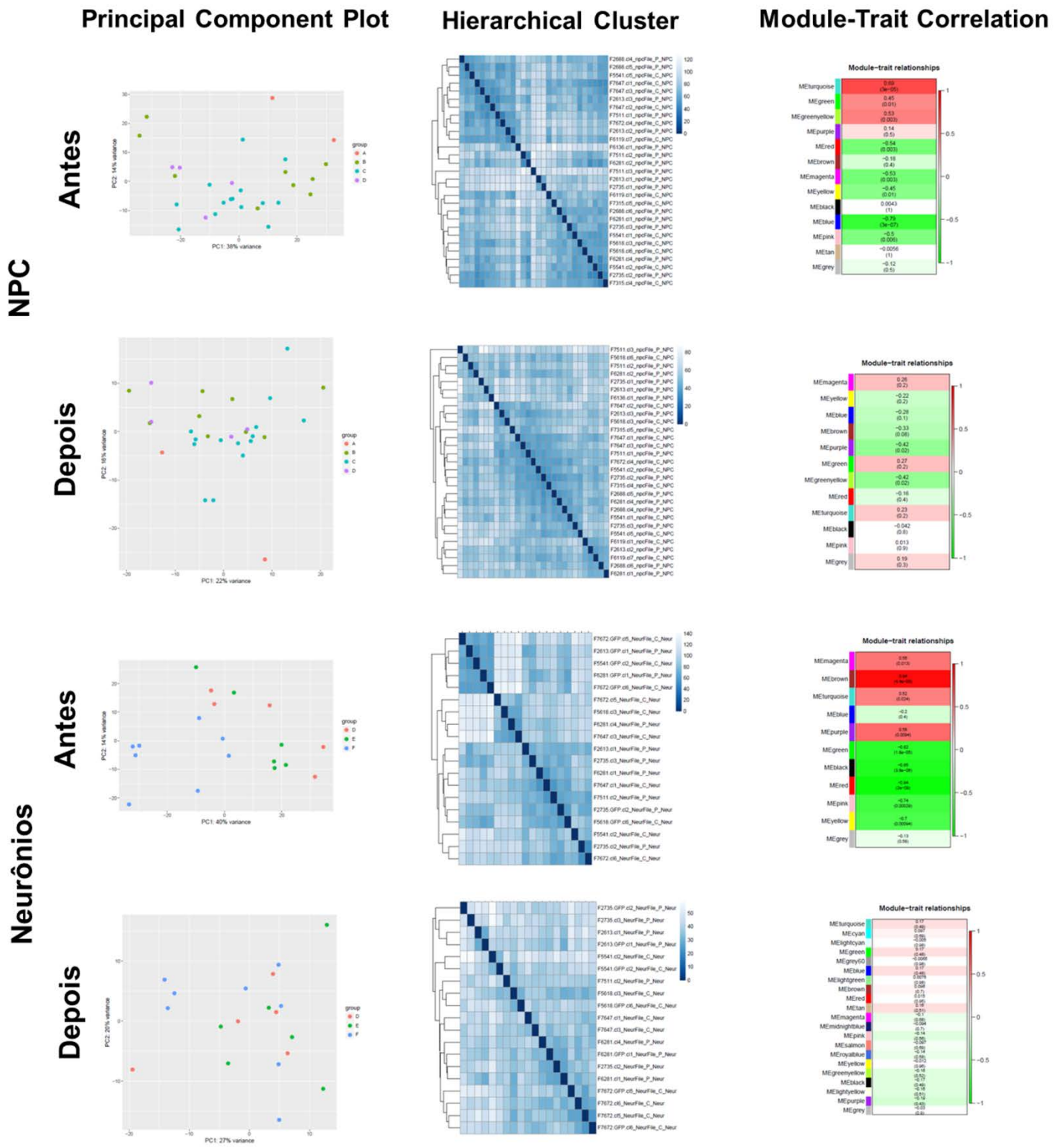

Figura 3.2: Normalização dos dados de expressão gênica. Utilizamos três diferentes abordagens para avaliar a normalização de nossos dados de expressão: análise de componente principal, clusterização hierárquica e análise de módulos de coexpressão gênica. Nos gráficos de componente principal, as diferentes cores indicam diferentes proporções neuronais; na tabela de módulos de coexpressão, a escala de cores representa correlações positivas (vermelho) ou negativas (verde) entre o nível de expressão do componente principal de cada módulo e a proporção neuronal das amostras. As análises realizadas antes da normalização dos dados mostram que o agrupamento das amostras e os módulos de coexpressão

identificados estão fortemente correlacionados à proporção neuronal das amostras. Após a normalização (realizada com o uso do pacote RUVseq a partir dos parâmetros especificados no texto), o efeito da proporção neuronal é desfeito e as amostras mostram uma tendência ao agrupamento por réplicas biológicas.

3.12 Classificação das amostras pelo BrainSpan

Para avaliar a qual idade embrionária as amostras de NPC e neurônios geradas a partir do nosso protocolo correspondem, comparamos o transcriptoma de cada uma delas aos transcriptomas de 259 amostras de cérebro fetal disponíveis no banco BrainSpan (www.brainspan.org) com idades variando entre 8 e 25 semanas pós-concepção. Para isso, calculamos a correlação de Spearman entre os níveis de expressão logaritmizados de nossas amostras e os das amostras provenientes do banco e ranqueamos essas últimas de acordo com os valores de correlação obtidos para cada uma das amostras de NPC ou neurônios. Em seguida, para cada amostra do BrainSpan, somamos os valores de classificação (*ranking*) para as NPC e os neurônios separadamente e verificamos quais delas apresentavam os menores valores após o somatório, ou seja, as maiores correlações gerais com esses tipos celulares.

3.13 Análise de expressão gênica diferencial e módulos de coexpressão gênica

Após a normalização, as análises de expressão gênica diferencial foram realizadas por meio do pacote DESeq2, considerando-se como genes diferencialmente expressos aqueles com $p<0,1$. As análises de coexpressão gênica foram realizadas por meio do pacote WGCNA (Langfelder & Horvath, 2008). Os módulos, gerados separadamente para NPC e neurônios, foram obtidos por meio da função blockwiseModule com os seguintes parâmetros: power=11 (NPC) ou 16 (neurônios), networkType = "signed", minModuleSize=50 (NPC) ou 150 (neurônios), mergeCutHeight=0.15, verbose=6, minKMEtoStay=0.5, nThreads=24 e maxBlockSize=2000. O módulo M0-grey foi configurado de forma a receber todos os genes que não foram agrupados em nenhum dos outros módulos. Para cada um dos módulos, o nível de expressão gênica de cada amostra foi resumido em um valor *eingengene*, usado para avaliar a correlação entre módulo e *status* da doença, lote de preparação da biblioteca ou proporção neuronal das amostras. Os valores de p calculados pelo WGCNA para estas correlações foram posteriormente submetidos à correção de Bonferroni. A função userListEnrichment do WGCNA (Langfelder, Luo, Oldham, & Horvath, 2011) foi usada para avaliar a sobreposição entre os módulos

gerados a partir de nossos dados e aqueles identificados em outros estudos de expressão gênica global usando células ou tecidos derivados de indivíduos com TEA (Mariani et al., 2015; Gupta et al., 2014; Voineagu et al., 2011). As análises no WGCNA também foram realizadas em colaboração com o grupo da Dra. Irina Voineagu.

A anotação funcional dos DEGs e dos módulos de coexpressão foram realizadas por meio de dois programas: DAVID (Database for Annotation, Visualization and Integrated Discovery - https://david.ncifcrf.gov/) e IPA (Ingenuity Pathway Analysis).

3.14 Validação funcional dos resultados

Ensaio de proliferação celular

Para o ensaio de proliferação celular, as NPC foram descongeladas em placas de 60 mm e cultivadas até atingirem alta confluência. As células foram então removidas das placas com o uso de Tryple, ressuspendidas em PBS 1X e centrifugadas a 400g por 10 minutos. Após a centrifugação, o sobrenadante foi descartado e as células ressuspendidas em 1 ml de CFSE diluído em PBS 1X a uma concentração final de 1µM/ml e incubadas a 37°C por 10 minutos. À ressuspensão, foram adicionados 9 ml de PBS 1X com 5% de FBS e as células foram novamente centrifugadas a 400g durante 10 minutos. O sobrenadante foi descartado e as células foram, uma última vez, ressuspendidas em 10 ml de PBS 1X com 5% de FBS, incubadas a 37°C por 15 minutos e centrifugadas a 400g durante 10 minutos. Após esse último procedimento, as células foram ressupendidas em meio NB suplementado com EGF e FGF e transferidas para placas de 6 poços a uma quantidade final de 150 mil células/poço. Aproximadamente 24, 48 e 72h após o plaqueamento, as células foram removidas de dois poços, ressuspendidas em PBS 1X com 1% de FBS e levadas ao citômetro para leitura de fluorescência. A taxa de proliferação celular foi determinada a partir da razão entre a fluorescência emitida nos pontos 48 ou 72h pós-plaqueamento e a emitida no ponto 24h. As análises estatísticas foram feitas no programa R, com o pacote lme4 (Bates, Mächler, Bolker, & Walker, 2015).

Análise de morfologia neuronal

Ao final de quatro semanas de diferenciação de linhagens transformadas com o vetor Synapsin::EGFP, as células foram dissociadas e submetidas à separação no

citômetro. Recuperamos a fração positiva para fluorescência verde e plaqueamos 20 mil neurônios GFP+ com 50 mil células provenientes de uma população heterogênea de neurônios e células da glia não transformadas com o vetor Synapsin::EGFP em *chamber slides* de oito poços. Esse plaqueamento foi feito de duas formas distintas: neurônios GFP+ combinados com células do mesmo indivíduo obtidas a partir de linhagens sem o vetor (condição *non-mixed*) ou neurônios GFP+ combinados com células sem o vetor de um indivíduo do grupo oposto (neurônios de um paciente com células de um controle, e vice-versa – condição *mixed*). Aproximadamente 72h após o plaqueamento, as células foram fixadas e preparadas para imunofluorescência (ver seção 3.6 para uma descrição mais detalhada do procedimento). O anticorpo primário usado nesse experimento foi anti-GFP (1:500, Millipore) e o secundário foi anti-mouse (1:250, LifeTechnologies).

O imageamento das células foi feito no microscópio IN Cell Analyzer 2200 (GE Healthcare Life Sciences) em aumento de 20x. A análise da morfologia neuronal foi realizada por meio do *plugin* NeurphologyJ, disponível no programa ImageJ. As medidas obtidas foram normalizadas pelo número de células na imagem e as análises estatísticas foram conduzidas no programa R, com auxílio do pacote geepack (Højsgaard, Halekoh, & Yan, 2006).

3.15 Processamento dos dados do exoma e análise de desvio de expressão alélica

Como já citado, todos os pacientes incluídos no presente estudo tiveram seus exomas sequenciados e os dados processados previamente. Brevemente, o sequenciamento completo do exoma desses indivíduos foi feito a partir do DNA obtido de células de sangue periférico. A preparação das bibliotecas e a captura dos fragmentos do exoma foram realizadas por meio do kit Illumina TruSeq, e o sequenciamento foi feito em Illumina HiSeq 2500, onde obtivemos *reads paired-end* de tamanho máximo de 100pb. Uma cobertura acima de 10 *reads* foi alcançada para aproximadamente 60% das bases cobertas pelo exoma de cada paciente. Os *reads* gerados após o sequenciamento foram alinhados ao genoma referência (hg19) pelo Burrows-Wheller Aligner (bwa-mem) (Li & Durbin, 2009). O alinhamento gerado foi ordenado e teve suas duplicatas de PCR marcadas com a ferramenta Picard (http://broadinstitute.github.io/picard). As etapas seguintes de recalibração das qualidades do alinhamento e a chamada de variantes foi feita pelo Genome Analysis Toolkit (GATK, versão 3.7, McKenna et al., 2010), enquanto o ANNOVAR (Wang, Li, &

Hakonarson, 2010) e um *script* em Perl foram usados para a anotação dessas variantes, com a finalidade de filtrar aquelas mais relevantes para o nosso trabalho.

Os dados já processados do transcriptoma dos pacientes também foram submetidos à chamada e anotação de variantes pelo GATK e ANNOVAR, respectivamente. Essas etapas foram realizadas com a colaboração do Dr. Guilherme Lopes Yamamoto e da Dra. Suzana Ezquina.

A fim de identificar genes com desvio de expressão alélica, identificamos as variantes informativas nos dados do exoma de cada paciente (isto é, aquelas em heterozigose, cuja proporção da expressão de cada alelo permanecesse entre 30 e 70%) e, em seguida, verificamos seu balanço alélico nos dados do transcriptoma: se o alelo com maior expressão respondesse por mais de 80% do total de *reads* em pelo menos dois clones de um mesmo indivíduo, então se considerava que, nesse gene, estava ocorrendo desvio de expressão alélica. Os genes apresentando menos de 10 *reads* no exoma ou no transcriptoma foram excluídos das análises.

Realizamos as análises separadamente nas NPC e nos neurônios de cada paciente, gerando uma lista de genes apresentando desvio de expressão alélica em NPC e outra em neurônios para cada indivíduo. A fim de criar uma lista consenso de genes apresentando desvio de expressão alélica para cada um dos tipos celulares, mantivemos em nossas duas listas finais (NPC e neurônios) apenas os genes que apresentaram desvio de expressão em pelo menos quatro pacientes.

As listas finais foram submetidas à análise no IPA, onde se verificou o enriquecimento para genes agrupados no SFARI e vias canônicas relevantes no neurodesenvolvimento.

4. RESULTADOS

4.1 Amostras disponíveis para as análises

As análises do transcriptoma realizadas durante o desenvolvimento desse projeto foram feitas a partir de NPC e neurônios derivados de iPSC de pacientes e controles. Para cada um dos doze indivíduos incluídos na amostra, entre dois e três clones foram diferenciados de iPSC para NPC e tiveram seus transcriptomas sequenciados previamente pelo nosso grupo. Parte desses clones foi ainda induzida à diferenciação terminal, gerando populações heterogêneas de neurônios maduros e células da glia. Essas amostras tiveram seus transcriptomas sequenciados como populações mistas ou, no caso das linhagens transduzidas com o vetor Synapsin::EGFP, foram enriquecidas para neurônios antes do sequenciamento por meio da seleção de células GFP+ por citometria. Após o sequenciamento, as amostras com menos de 20 milhões de *reads* totais sequenciados e aquelas apresentando proporção neuronal inferior a 50% nos neurônios e superior a 55% nas NPC foram removidas das análises realizadas posteriormente (tabela 4.1). Ao final dessas filtragens, restaram 19 amostras neuronais e 29 de NPC.

4.2 Caracterização das NPC e dos neurônios derivados de iPSC

As NPC obtidas a partir da diferenciação das iPSC foram mantidas em monocamada e apresentaram alta expressão dos marcadores de células-tronco neurais SOX2 e Nestina (figura 4.1A). Essas células foram ainda induzidas à diferenciação neuronal, gerando populações heterogêneas de neurônios maduros, nas quais foi verificada a expressão dos marcadores neuronais MAP2 e CTIP2 (figura 4.1B). De forma geral, a capacidade de diferenciação em NPC e neurônios foi similar entre as linhagens derivadas de pacientes e controles.

A fim de refinar a caracterização dessas células, utilizamos os dados do RNAseq gerados a partir delas para analisar as alterações na expressão gênica que ocorrem ao longo da diferenciação neuronal. Para isso, utilizamos os dados de todas as amostras (pacientes e controles) para realizar a análise de expressão gênica diferencial entre NPC e neurônios e, em seguida, a anotação funcional dos genes diferencialmente expressos (DEGs) obtidos. A análise mostrou a presença de 10.922 DEGs entre esses dois tipos celulares, 5.453 cuja expressão diminui e 5.469 cuja expressão aumenta conforme se dá a diferenciação de NPC para neurônios. A

anotação funcional revelou que há um enriquecimento para termos relacionados a ciclo celular e tradução quando se analisa apenas os genes *downregulated*, enquanto os genes *upregulated* estão enriquecidos para termos relacionados a desenvolvimento neuronal e transmissão sináptica (figura 4.1C). Esses resultados reforçam a validade de nosso modelo de estudo, uma vez que a expressão gênica das células se comporta exatamente de acordo com o esperado durante o processo de diferenciação neuronal, no qual se observa a diminuição da capacidade de proliferação celular e a aquisição progressiva da morfofisiologia neuronal.

Ainda, para avaliar a qual fase do desenvolvimento do cérebro humano as NPC e neurônios gerados a partir de nosso protocolo correspondem, comparamos os transcriptomas desses tipos celulares aos dados disponíveis no BrainSpan, um banco contendo transcriptomas de cérebros humanos *post-mortem* que abrangem desde o desenvolvimento embrionário até a idade adulta (Kang et al., 2011). A análise de correlação mostrou que nossas NPC refletem de forma consistente o perfil de expressão de cérebros de embriões no início do desenvolvimento (8-9 semanas pós-concepção), enquanto essa relação é um pouco menos marcada nos neurônios e eles são comparáveis a amostras variando entre 9 e 16 semanas pós-concepção (figura 4.1D), evidenciando a ocorrência de progressão temporal ao longo da diferenciação dessas células.

Tabela 4.1: amostras que foram submetidas ao sequenciamento do transcriptoma. As amostras apresentando número total de reads menor do que 20 milhões ou proporção neuronal inferior a 50% nos neurônios e superior a 55% nas NPC foram descartadas. P = paciente, C = controle.

Amostra	Condição	Tipo celular	Total de reads	Reads mapeados	Proporção neuronal	Amostra	Condição	Tipo celular	Total de reads	Reads mapeados	Proporção neuronal
F2613 cl1	P	NPC	36M	30M	0.44	**F7647 cl3**	C	NPC	87M	78M	0.48
F2613 cl2	P	NPC	149M	136M	0.44	**F7672 cl4**	C	NPC	77M	63M	0.44
F2613 cl3	P	NPC	86M	79M	0.47	**F7672 cl5**	C	NPC	134M	123M	0.71
F2688 cl4	P	NPC	77M	68M	0.47	**F7672 cl6**	C	NPC	175M	162M	0.56
F2688 cl5	P	NPC	98M	90M	0.47	**F2613 cl1**	P	Neu	55M	42M	0.82
F2688 cl6	P	NPC	41M	37M	0.47	**F2613.GFP cl1**	P	Neu	50M	37M	1
F2735 cl1	P	NPC	40M	34M	0.45	**F2688 cl4**	P	Neu	46M	36M	0.49
F2735 cl2	P	NPC	29M	26M	0.49	**F2688 cl6**	P	Neu	99M	76M	0.4
F2735 cl3	P	NPC	30M	28M	0.51	**F2735 cl2**	P	Neu	83M	60M	0.64
F6136 cl1	P	NPC	93M	74M	0.34	**F2735 cl3**	P	Neu	68M	51M	0.88
F6136 cl4	P	NPC	140M	128M	0.60	**F2735.GFP cl2**	P	Neu	54M	40M	0.71
F6281 cl1	P	NPC	46M	39M	0.49	**F6136 cl4**	P	Neu	50M	39M	0.5
F6281 cl2	P	NPC	83M	75M	0.50	**F6281 cl1**	P	Neu	63M	43M	0.64
F6281 cl4	P	NPC	28M	26M	0.49	**F6281 cl4**	P	Neu	46M	36M	0.6
F7511 cl1	P	NPC	30M	26M	0.46	**F6281.GFP cl1**	P	Neu	29M	20M	0.94
F7511 cl2	P	NPC	137M	124M	0.47	**F7511 cl2**	P	Neu	27M	19M	0.59
F7511 cl3	P	NPC	29M	26M	0.35	**F7511.GFP cl2**	P	Neu	19M	15M	

Amostra	Condição	Tipo celular	Total de reads	Reads mapeados	Proporção neuronal	Amostra	Condição	Tipo celular	Total de reads	Reads mapeados	Proporção neuronal
F5541 cl1	C	NPC	49M	45M	0.55	**F5541 cl1**	C	Neu	19M	14M	
F5541 cl2	C	NPC	21M	19M	0.56	**F5541 cl2**	C	Neu	73M	54M	0.53
F5541 cl5	C	NPC	28M	24M	0.43	**F5541.GFP cl1**	C	Neu	16M	12M	
F5618 cl3	C	NPC	58M	54M	0.46	**F5541.GFP cl2**	C	Neu	48M	35M	0.9
F5618 cl4	C	NPC	19M	16M		**F5618 cl3**	C	Neu	62M	49M	0.63
F5618 cl6	C	NPC	20M	19M	0.46	**F5618 cl6**	C	Neu	26M	20M	0.38
F6119 cl1	C	NPC	70M	66M	0.44	**F5618.GFP cl6**	C	Neu	110M	79M	0.81
F6119 cl4	C	NPC	14M	13M		**F7647 cl1**	C	Neu	64M	47M	0.72
F6119 cl7	C	NPC	22M	18M	0.44	**F7647 cl3**	C	Neu	143M	115M	0.57
F7315 cl4	C	NPC	135M	124M	0.51	**F7672 cl5**	C	Neu	85M	68M	0.62
F7315 cl5	C	NPC	49M	45M	0.51	**F7672 cl6**	C	Neu	67M	49M	0.57
F7647 cl1	C	NPC	22M	20M	0.47	**F7672.GFP cl5**	C	Neu	40M	30M	1
F7647 cl2	C	NPC	35M	32M	0.47	**F7672.GFP cl6**	C	Neu	32M	22M	0.95

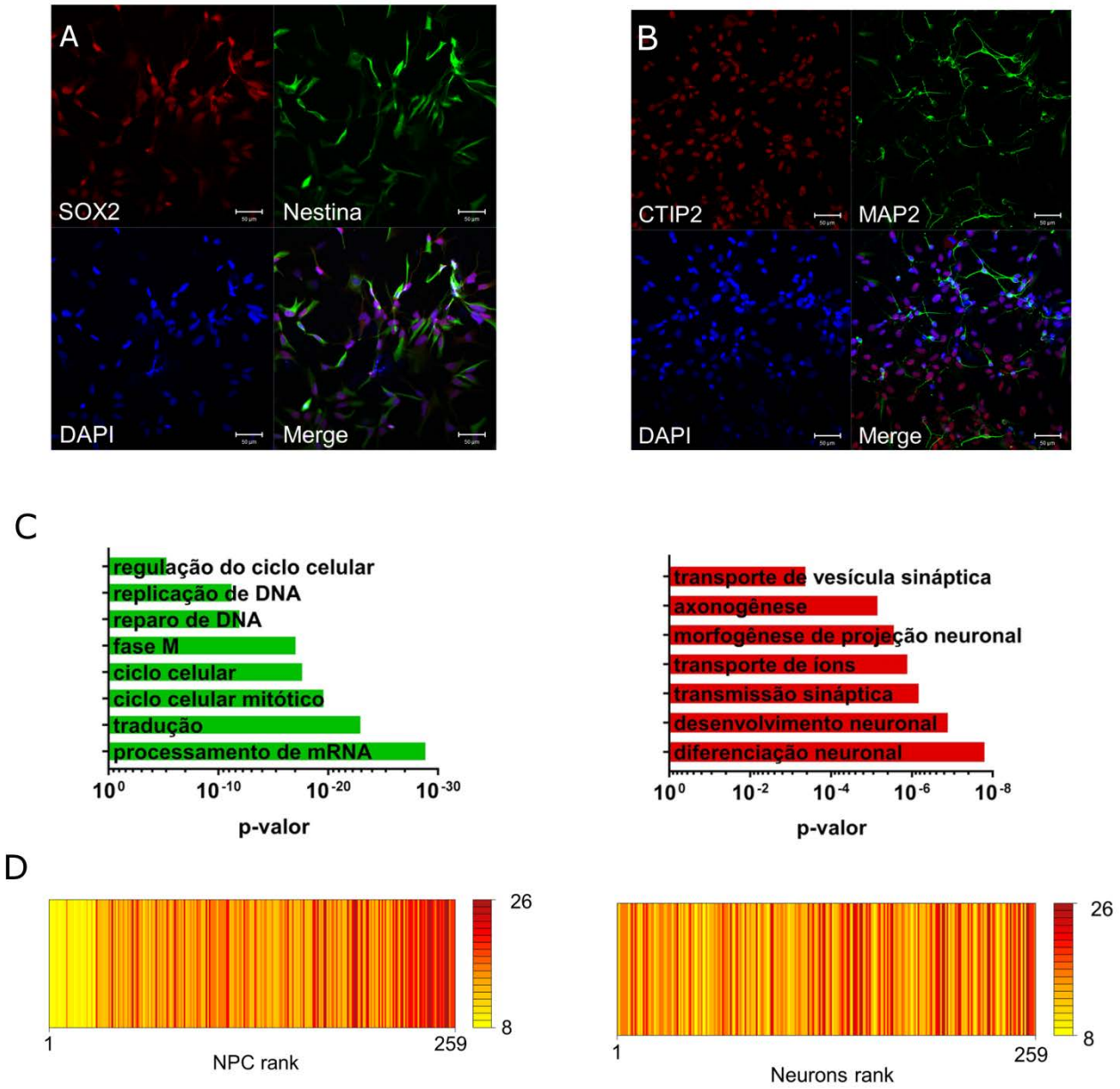

Figura 4.1: Caracterização celular. (A e B) As NPC e os neurônios usados nesse estudo expressam marcadores de células-tronco neurais (SOX2 e Nestina) e de neurônios maduros (CTIP2 e MAP2), respectivamente. (C) O gráfico mostra o enriquecimento para as funções destacadas dentre os genes que apresentam redução na expressão (em verde) ou aumento de expressão (em vermelho) ao longo da diferenciação de NPC para neurônios. (D) Usando os dados do BrainSpan, calculamos a correlação entre os transcriptomas de nossas amostras e os transcriptomas provenientes de cérebros em diferentes fases do desenvolvimento embrionário. Cada linha vertical nos gráficos representa uma amostra do banco posicionada de acordo com a somatória dos *rankings* obtidos a partir da correlação que ela apresentou com cada uma de nossas amostras (quanto maior o índice de correlação, melhor ranqueada a amostra foi e, portanto, a somatória dos índices tende a ser menor). Verificamos que as NPC têm o perfil de expressão correspondente ao início do desenvolvimento embrionário (8-9 semanas pós-concepção; à esquerda), enquanto os neurônios correspondem a fases variando entre 9 e 16

semanas pós-concepção (à direita). As barras laterais com escala colorimétrica indicam a semana pós-concepção que cada cor representa.

4.3 Estudo de expressão gênica diferencial

Genes diferencialmente expressos estão envolvidos em processos importantes para o neurodesenvolvimento e a função neuronal

A fim de identificar vias de sinalização e processos biológicos desregulados durante o desenvolvimento neuronal em indivíduos afetados pelo TEA, os transcriptomas de pacientes e controles foram comparados em dois momentos diferentes: no início da diferenciação, por meio das amostras de NPC, e ao final dela, por meio das amostras de neurônios. Para isso, o pacote DESeq2 foi utilizado para a realização das análises de expressão gênica diferencial entre os dois grupos e as ferramentas DAVID e IPA para a obtenção das anotações funcionais dos DEGs identificados.

Nas NPC, a análise de expressão gênica diferencial mostrou a presença de 543 DEGs entre pacientes e controles (alguns dos quais têm sido relacionados à manifestação do TEA no banco de dados SFARI e são apresentados na tabela 4.2). A anotação funcional dessa lista revelou que os genes com expressão aumentada nos pacientes estão enriquecidos para vias canônicas e funções celulares relacionadas a síntese/degradação de proteínas, fosforilação oxidativa e ciclo celular (figura 4.2A e B), processos importantes no início do desenvolvimento neuronal. Observamos ainda um enriquecimento para genes expressos em células Cajal-Retzius e no córtex de cérebro fetal, concordando com o fato das NPC estarem presentes no início do desenvolvimento, e para genes expressos em plaquetas (figura 4.2C). Ainda, analisamos a similaridade entre nossa lista de DEGs e listas identificadas em outros estudos e verificamos a ocorrência de sobreposição significativa com DEGs identificados em um estudo de expressão gênica anterior do nosso grupo, realizado com SHED (figura 4.2D). Nos genes com expressão diminuída nos pacientes, por sua vez, se observa um enriquecimento para funções relacionados a cílio e uma sobreposição significativa com genes identificados em um estudo anterior usando células neuronais derivadas de iPSC como modelo (Mariani et al., 2015; figura 4.2A e D).

Tabela 4.2: DEGs identificados entre NPC de pacientes e controles que apresentam evidências de estarem relacionados ao TEA.

Gene	Fold Change	SFARI score
CCT4	1.14	3
EIF4E	1.20	4
FEZF2	-1.33	4

GIGYF1	-1.17	3
MARK1	1.22	4
OTX1	-1.31	4
UBE3C	1.14	3
UPF3B	-1.15	2S

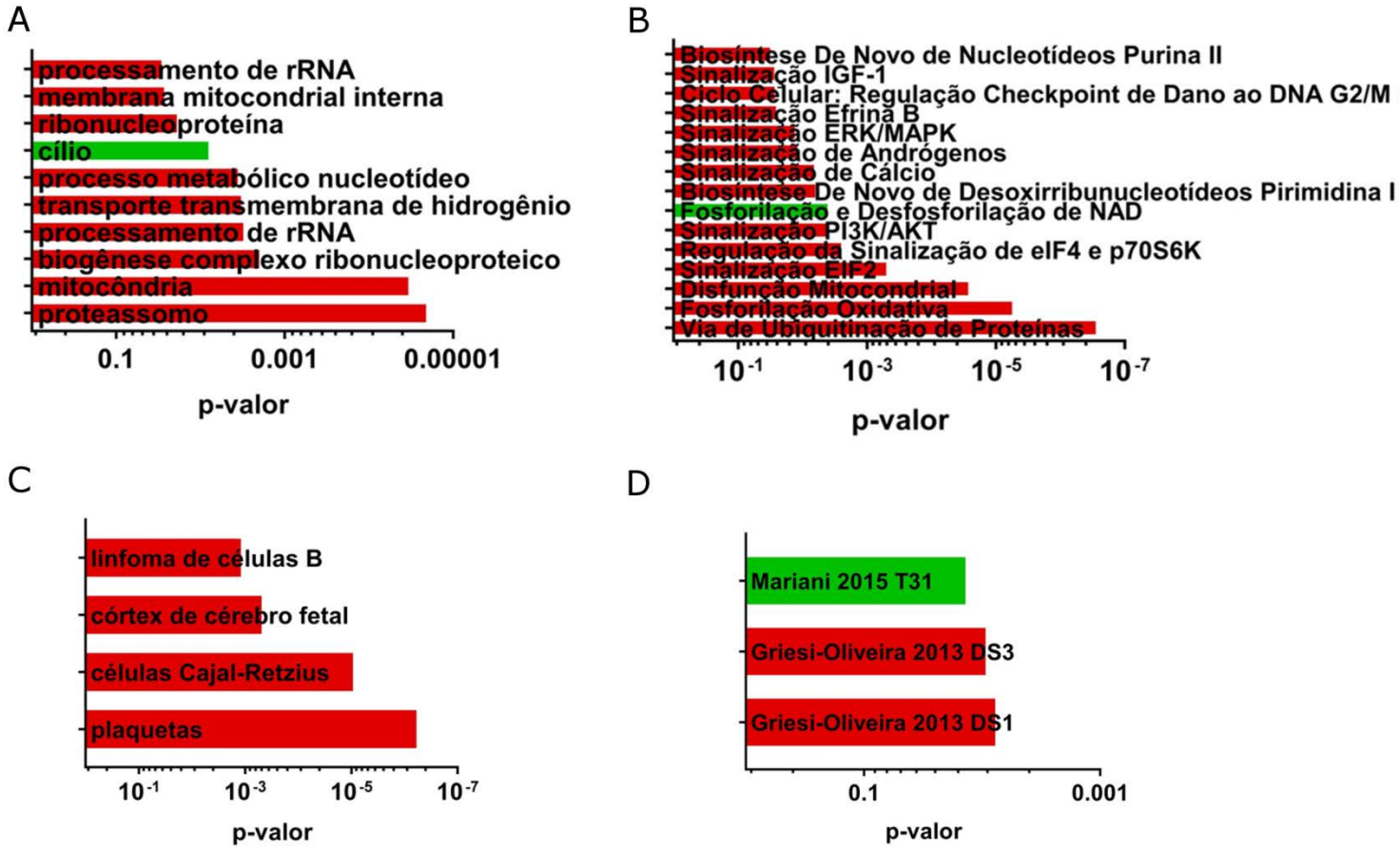

Figura 4.2: Os DEGs identificados nas NPC têm função relevante no neurodesenvolvimento. Os gráficos mostram funções/termos Gene Onthology (A), vias canônicas (B), tecidos (C) e conjuntos de DEGs identificados em outros estudos de expressão gênica global (D) que estão enriquecidos por genes com expressão aumentada (vermelho) ou diminuída (verde) nas NPC dos pacientes.

Nos neurônios, por sua vez, identificamos 1.390 DEGs entre pacientes e controles (os que já são associados ao TEA são destacados na tabela 4.3). A anotação funcional dos genes com expressão aumentada nos pacientes mostra um enriquecimento para vias canônicas e termos importantes no neurodesenvolvimento e na função neuronal, tais como transmissão sináptica, desenvolvimento neuronal, transporte de neurotransmissores, canais dependentes de voltagem e transporte de íons (figura 4.3A e B). Corroborando esses resultados, uma proporção importante desses genes é expressa no cérebro, no hipocampo e

na amígdala (figura 4.3C). Observamos também sobreposições significativas entre nossa lista e genes reunidos no banco de dados SFARI ou DEGs identificados em trabalhos anteriores usando diferentes modelos de estudo para o TEA (Mariani et al., 2015; Voineagu et al., 2011; figura 4.3D), sugerindo a relevância desses genes para a patofisiologia do TEA. Os genes com expressão diminuída nos pacientes, por sua vez, estão enriquecidos para funções relacionadas ao reparo de DNA e organização do citoesqueleto de actina (figura 4.3A e B), mas não apresentam sobreposição com genes do SFARI ou DEGs identificados em estudos anteriores.

Tabela 4.3: DEGs identificados entre neurônios de pacientes e controles que apresentam evidências de estarem relacionados ao TEA.

Gene	Fold Change	SFARI score
ADCY5	1.23	4
AGAP1	1.16	4
ANK2	1.16	1
ATP2B2	1.35	3
AUTS2	-1.13	3
BCL11A	-1.31	2S
CACNA1E	1.55	3
CACNA1I	1.46	4
CAPRIN1	-1.08	3
CC2D1A	1.17	3
CDH8	1.26	4
CLN8	1.15	4
CNTNAP5	1.28	4
CTTNBP2	1.20	3
DENR	-1.09	3
DNER	1.17	4
DNMT3A	-1.11	3
GABRB3	1.17	2
GALNT14	1.35	4
GPC4	-1.25	3
GRIK5	1.25	3
HECTD4	1.11	4
HECW2	1.25	3
HLA-A	1.31	4
KCNQ3	1.27	3
LMX1B	-1.51	3
MAOA	1.43	4

MEF2C	-1.25	4S
NRXN2	1.16	4
NTNG1	1.34	4S
PARK2	1.23	3
PPP1R1B	-1.42	4
RERE	-1.17	4S
RIMS1	1.23	2
SCFD2	-1.21	4
SHANK1	1.27	3
SHANK3	1.20	1S
SLC12A5	1.58	3
SLC4A10	1.25	4
SLC6A1	1.31	2
SLC9A9	1.35	4
SPAST	-1.09	2
STXBP1	1.18	3S
SYN1	1.26	4
TOP3B	-1.21	4
TSPAN7	1.18	4
UNC13A	1.37	4S

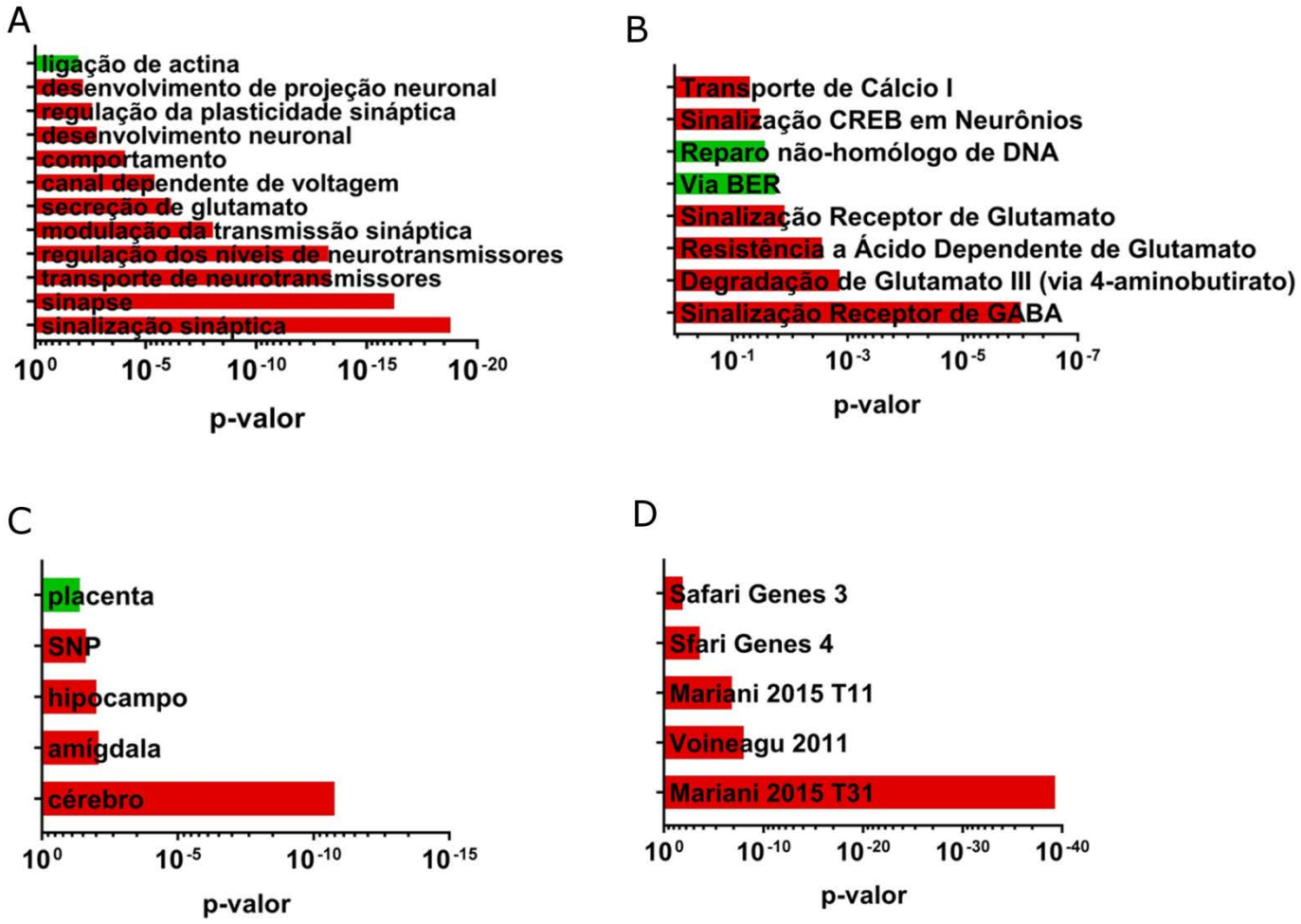

Figura 4.3: Os DEGs identificados nos neurônios são relevantes para a função neuronal. Os gráficos mostramfunções/termos Gene Onthology (A), vias canônicas (B), tecidos (C) e genes encontrados no SFARI ou conjuntos de DEGs identificados em outros estudos de expressão gênica global (D) que estão enriquecidos por genes com expressão aumentada (vermelho) ou diminuída (verde) nos pacientes.

As diferenças de expressão existentes entre pacientes e controles também são observadas em níveis de organização mais complexos

Considerando que as anotações funcionais dos DEGs em NPC e neurônios indicam a ocorrência de alterações em processos biológicos importantes para o neurodesenvolvimento e função neuronal, decidimos verificar se essas diferenças de expressão também são observadas em um contexto sistêmico por meio da análise de grupos de genes coexpressos. Para isso, usamos o pacote WGCNA para identificar módulos de coexpressão gênica em NPC e neurônios e analisar quais desses módulos tem correlação significativa com a doença. As anotações funcionais dos módulos foram feitas por meio do DAVID e do IPA.

Usando o conjunto completo de dados do transcriptoma das NPC (pacientes e controles), identificamos onze módulos de coexpressão gênica nessas células (tabela 4.4). Dentre eles, apenas o módulo M10-blue, com expressão aumentada nos pacientes e enriquecido para genes relacionados a tradução, ciclo celular e síntese de ATP, apresentou diferença significativa de expressão entre os dois grupos. De fato, diversos genes que compõem esse módulo foram identificados como DEGs nas NPC e o módulo se mostra enriquecido por genes expressos em cérebro fetal, células Cajal-Retzius e plaquetas (tabela 4.4).

Como o pacote WGCNA permite não apenas identificar módulos de coexpressão gênica, mas também comparar módulos identificados em diferentes trabalhos, avaliamos a sobreposição entre M10-blue e módulos descritos em estudos de TEA anteriores (Mariani et al, 2015; Gupta et al, 2014; Voineagu et al, 2011). Verificamos uma alta sobreposição entre M10-blue e mod4 (p=1,7E-36), um módulo relacionado à tradução identificado por Gupta e colaboradores, e uma sobreposição menor, mas ainda significativa, entre ele e o módulo red identificado por Mariani e colaboradores (p=0,009). Uma caracterização mais detalhada desse módulo correlacionado ao *status* da doença, assim como dos outros dez módulos identificados nas NPC, pode ser encontrada na tabela 4.4.

Assim como o que foi realizado nas NPC, utilizamos o conjunto completo de dados dos neurônios e identificamos 20 módulos de coexpressão gênica, três dos quais apresentam correlação significativa com o *status* da doença (tabela 4.5). O primeiro deles, M1-turquoise, módulo com expressão aumentada nos pacientes, é caracterizado por apresentar uma alta proporção de DEGs (51,2%) e enriquecimento para genes que são expressos em tecidos neurais (cérebro, amígdala e hipocampo) e estão envolvidos em processos essenciais para a função neuronal, tais como transmissão sináptica e transporte de íons e de neurotransmissores. Ainda, a comparação entre nossas análises e resultados obtidos em estudos anteriores revelou a ocorrência de sobreposições altas entre M1-turquoise e módulos neuronais correlacionados ao *status* da doença apontados nesses trabalhos: m12 identificado por Voineagu e colaboradores (p=8E-23), mod1 e mod6 identificados por Gupta e colaboradores (p=1,4E-26 e 3,6E-5, respectivamente) e brown e blue identificados por Mariani e colaboradores (p=2,3E-43 e 1,35E-5, respectivamente).

Os outros dois módulos correlacionados ao TEA nos neurônios, M16-pink e M18-purple, apresentam expressão reduzida nos pacientes e também são caracterizados pela presença de proporções altas de DEGs (aproximadamente 40% em ambos os casos). M16-pink está enriquecido para genes relacionados a citoesqueleto e apresenta sobreposição com os módulos green (p=8,6E-8) e mod5 (p=0,003), identificados por Mariani et al. e Gupta et al., respectivamente, e correlacionados ao status da doença nesses estudos. M18-purple, por sua vez, está enriquecido para genes relacionados a

síntese proteica e processamento de RNA não codificante, mas não apresenta sobreposição com módulos correlacionados à doença em outros estudos. Informações mais detalhadas sobre esses e os outros 17 módulos identificados nos neurônios estão reunidas na tabela 4.5.

Tabela 4.4: análise dos módulos de coexpressão gênica identificados a partir das amostras de NPC.

Módulo	p-valor TEA corrigido	p-valor prop neuronal corrigido	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M1-turquoise	0,22	1	1186	7,3	Splicing alternativo, regulação transcricional, organização de microtúbulos, cílios	testículos, cérebro	3,0	Mariani: midnightblue (3E-21), **blue** (1E-9); Gupta: mod2 (2E-4), **mod1** (9E-3)
M2-yellow	0,22	1	971	7,9	metabolismo/processamento de RNAs não codificantes, ribossomos, tradução, organização mitocondrial, transcrição	sangue de cordão umbilical, medula óssea, bexiga urinária, placenta	1,0	Mariani: turquoise (3E-27), red (4E-26); Gupta: mod8 (8E-4), mod9 (0,02)
M3-green	0,99	1	733	4,6	-	-	0	Gupta: mod12 (4,8E-11)
M4-red	1	1	544	0	organização de matriz extracelular, desenvolvimento do sistema esquelético	baço	3,3	Mariani: purple (1,4E-23), royalblue (4E-11); Gupta: **mod5** (5E-3), mod7 (6E-3)
M5-black	1	1	577	3,5	proteína de desenvolvimento	-	2,6	Mariani: greenyellow (2E-3)
M6-pink	0,44	1	535	9,5	regionalização do cérebro, desenvolvimento do sistema nervoso, cílios	rim fetal, cérebro fetal	2,6	Mariani: lightgreen (1E-17), black (2,7E-9); Gupta: mod7 (7E-3)
M7-magenta	1	1	562	0,7	fosfoproteína, cAMP, fator de liberação de guanina	cérebro fetal, testículos, cérebro	3,9	Mariani: turquoise (9,7E-10), **blue** (5E-6), black (1E-4); Gupta: mod2 (7E-3)

Módulo	p-valor TEA corrigido	p-valor prop neuronal corrigido	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M8-purple	0,22	0,22	484	7,4	organização mitocondrial, fosforilação oxidativa, tradução, metabolismo de nucleosídeos, actina e miosina	pulmão, células Cajal-Retzius, pele, linfoma células B, plaquetas, córtex de cérebro fetal, sangue de cordão umbilical, placenta, músculo	1,2	Mariani: **green** (5E-14), red (8E-9), **yellow** (9E-4); Gupta: mod4 (8E-27), mod10 (2E-9), **mod5** (6,4E-5)
M9-greenyellow	1	0,22	372	1,9	organização de matriz extracelular, adesão celular, angiogênese	fígado, pâncreas, placenta	1,9	Mariani: greenyellow (1,4E-10), **yellow** (1E-9), purple (4E-9), **green** (3E-6), royalblue (3E-3); Gupta: mod11 (2,8E-9), **mod5** (6E-7), mod4 (0,01)
M10-blue	**0,044**	1	265	20,4	tradução, estabelecimento de localização de proteínas, metabolismo de nucleosídeos, síntese de ATP, ciclo celular mitótico, regulação de ubiquitinação, interação vírus-hospedeiro	músculo, córtex de cérebro fetal, pele, adenocarcinoma de cólon, câncer de colo do útero, linfoma células B, fígado, epitélio, células Cajal-Retzius, hipófise, pulmão, próstata, plaquetas, bexiga urinária, pâncreas, linfa, medula óssea, sangue de cordão umbilical, câncer de mama, placenta, útero, rim, cólon, osso, células B	1,9	Mariani: red (9E-3); Gupta: mod4 (1,7E-36)

Módulo	p-valor TEA corrigido	p-valor prop neuronal corrigido	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M11-brown	1	0,88	223	2,2	resposta a proteínas mal formadas, retículo endoplasmático	fígado	2,7	Mariani: **green** (7,8E-7), purple (4E-3), greenyellow (7E-3); Gupta: **mod5** (0,01)

*consideramos os genes reunidos no banco de dados e classificados até o score 4.

**Voineagu et al., 2011; Gupta et al., 2014; Mariani et al., 2015. Os módulos correlacionados ao *status* da doença nesses três estudos estão destacados em negrito na tabela.

Tabela 4.5: análise dos módulos de coexpressão gênica identificados a partir das amostras neuronais.

Módulo	p-valor TEA corrigido	p-valor proporção neuronal	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M1-turquoise	**0,022**	0,49	605	51,2	sinapse, transporte de íons, canal iônico, transmissão de impulso nervoso, processos neurológicos	cérebro, amígdala, hipocampo	5,1	Mariani: **brown** (2,3E-43), **blue** (1,35E-5), **tan** (0,001); Gupta: **mod1** (1,4E-26), mod2 (1,7E-18), **mod6** (3,6E-5); Voineagu: **m12** (8E-23)
M2-tan	1	0,51	500	16,4	biosíntese de colesterol, transmissão sináptica, transporte de íons	cérebro	8	Mariani: **brown** (3,8E-28), **blue** (6,9E-10); Gupta: **mod1** (1E-14), mod2 (2,8E-13), **mod6** (3E-3); Voineagu: **m12** (2E-4)
M3-cyan	0,8	0,69	431	16,7	-	-	4,2	-
M4-midnightblue	1	0,7	298	1,7	desenvolvimento do esqueleto, organização de matriz extracelular, adesão focal, via de sinalização TGF-beta, interação ECM-receptor	plasma, fibroblasto, folículo capilar	3,4	Mariani: **green** (4,6E-05)
M5-lightcyan	1	0,98	275	3,6	regulação transcricional, splicing alternativo	-	2,9	-
M6-grey60	1	0,98	307	4,9	-	-	3,3	Mariani: **yellow** (1E-7), **green** (6E-3); Gupta: **mod5** (0,01)

Módulo	p-valor TEA corrigido	p-valor proporção neuronal	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M7-lightgreen	1	0,98	196	10,7	cílio	testículos	2,6	Mariani: grey60 (4E-13)
M8-lightyellow	0,44	0,51	280	16,1	regulação transcricional	-	2,1	Mariani: red (8E-4)
M9-royalblue	0,26	0,58	254	18,5	processamento de RNA, tradução, spliceossomo	sangue de cordão umbilical, bexiga urinária	2,0	Mariani: turquoise (2,6E-16), red (2,7E-05); Gupta: mod4 (2E-5), mod9 (5E-5)
M10-blue	1	0,48	142	1,4	-	-	2,8	-
M11-brown	1	0,7	167	0	ciclo celular, divisão celular	epitélio, medula óssea, linfa, testículos	1,2	Mariani: turquoise (6,9E-81)
M12-yellow	0,38	0,96	202	25,2	regulação transcricional	-	3,0	-
M13-green	1	0,48	221	1,4	-	-	5,0	-
M14-red	1	0,95	272	0	retículo endoplasmático	rim fetal, cérebro	7,0	Mariani: **brown** (0,04)
M15-black	1	0,49	200	0,5	processamento de RNA, spliceossomo	pulmão	2,5	Mariani: black (0,002); Gupta: mod9 (0,04)
M16-pink	**0,05**	0,56	138	40,6	ligação à actina, citoesqueleto	placenta, plaquetas	0,0	Mariani: greenyellow (3,9E-9), **green** (8,6E-8), red (1E-5); Gupta: **mod5** (0,003)
M17-magenta	1	0,68	147	0,7	regulação transcricional	-	2,7	Mariani: greenyellow (0,001); Gupta: **mod5** (1,2E-5)

Módulo	p-valor TEA corrigido	p-valor proporção neuronal	Nº total de genes	% de DEGs	Termo GO	Tecido	% genes SFARI*	Sobreposição com módulos de outros estudos**
M18-purple	**0,048**	0,43	200	40,5	biosíntese de proteínas, processamento de ncRNA	epitélio	2,0	Mariani: red (8,6E-6)
M19-greenyellow	1	0,52	95	2,1	ribossomo, tradução, biosíntese de proteínas	pele	1,1	Gupta: mod4 (1E-27)
M20-salmon	1	0,69	162	9,3	Homeobox, ligação ao DNA	-	4,9	-

*consideramos os genes reunidos no banco de dados e classificados até o score 4.

**Voineagu et al., 2011; Gupta et al., 2014; Mariani et al., 2015. Os módulos correlacionados ao *status* da doença nesses três estudos estão destacados em negrito na tabela.

4.4 Estudos funcionais

As NPC derivadas dos pacientes apresentam uma tendência à diminuição da proliferação

As análises de anotação funcional tanto dos DEGs quanto do módulo associado ao *status* da doença mostraram o enriquecimento de algumas vias de sinalização e processos biológicos relacionados a ciclo celular. Por essa razão, realizamos um ensaio de proliferação celular por CSFE a fim de validar funcionalmente nossos resultados. As células plaqueadas foram coletadas em três dias consecutivos e submetidas à citometria de fluxo imediatamente após as coletas.

Os dados obtidos a partir desse ensaio indicam uma redução (ainda que não significativa) da proliferação das células derivadas de pacientes. Essa tendência é observada quando se considera tanto o intervalo entre 24 e 48h pós-plaqueamento quanto o intervalo entre 24 e 72h pós-plaqueamento (figura 4.4).

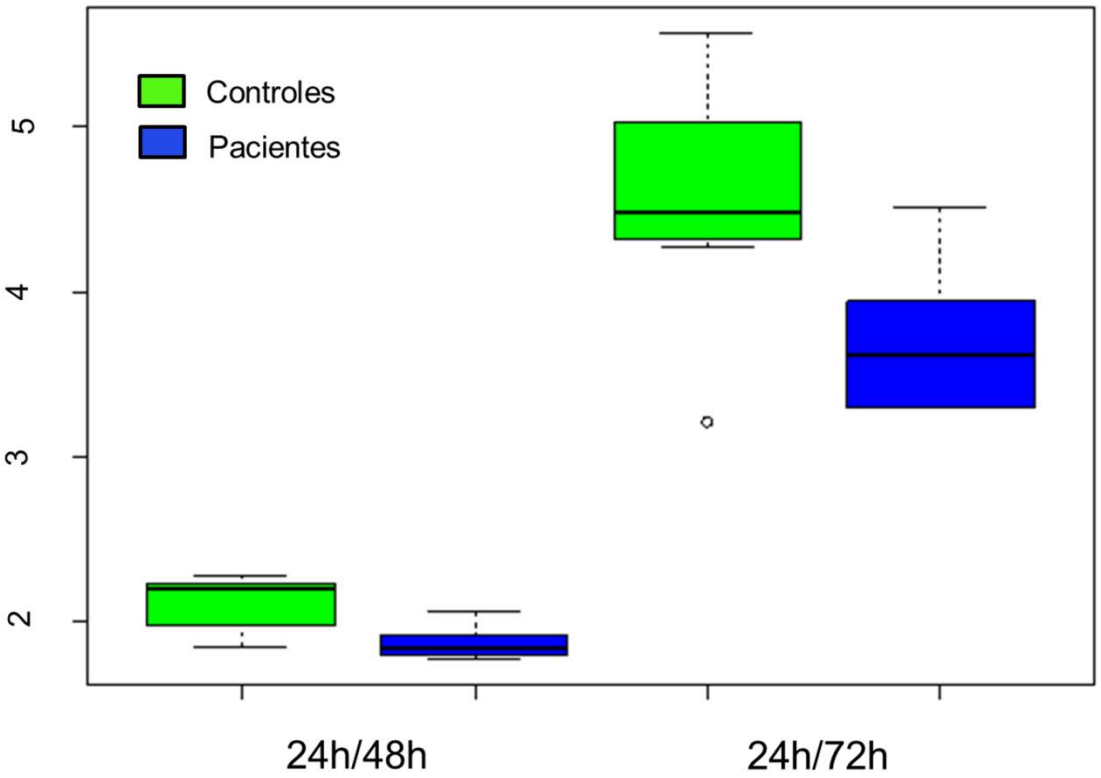

Figura 4.4: A proliferação celular é diminuída nas NPC derivadas dos pacientes. O gráfico mostra a razão de emissão de fluorescência entre 24h e 48h (esquerda) ou 24h e 72h (direita) pós-plaqueamento. Foram usadas amostras de cinco controles e quatro pacientes (entre um e dois clones de cada indivíduo) para a realização desse experimento.

Os neurônios derivados de pacientes e controles apresentam diferenças morfológicas importantes

A anotação funcional da lista de DEGs identificada nos neurônios mostra que processos biológicos como desenvolvimento neuronal, desenvolvimento de projeção neuronal e organização do citoesqueleto de actina estão alterados nos indivíduos afetados pelo TEA. A fim de validar funcionalmente esses resultados, analisamos as diferenças morfológicas entre neurônios derivados de pacientes e controles. A morfologia dos neurônios GFP+ de um indivíduo co-cultivados com células não marcadas desse mesmo indivíduo (condição *non-mixed*) ou com células não marcadas provenientes de um indivíduo do grupo oposto (um paciente + um controle, condição *mixed*) foi avaliada após a realização de imunofluorescência para a proteína GFP. Dessa forma, garantimos a medida apenas das células do indivíduo de interesse. As imagens obtidas foram analisadas por meio do NeurphologyJ (um *plugin* do programa ImageJ) quanto à área média do corpo celular (soma), comprimento médio dos neuritos, número de neuritos por célula e arborização. Esses parâmetros foram normalizados em relação ao número de células em cada imagem.

A comparação entre as células de pacientes e controles plaqueadas isoladamente (condição *non-mixed*) mostra uma redução significativa na arborização e no comprimento médio dos neuritos nos pacientes ($p=0,015$ e $p=0,007$, respectivamente), enquanto a área média do soma se encontra significativamente aumentada nesses indivíduos ($p=0,029$; figura 4.5). Quando se considera as medidas das células dos pacientes *non-mixed* e as das células dos mesmos pacientes nas amostras em co-cultura com células de controles (condição *mixed*), é possível notar que a área média do soma é significativamente reduzida ($p=0,02$) e o comprimento médio dos neuritos sofre um aumento ($p<0,001$), mostrando que o co-cultivo com células derivadas de indivíduos não afetados tem impacto positivo em parte das alterações identificadas nas células dos pacientes. O resgate de fenótipo observado foi total para o comprimento médio dos neuritos e parcial para a área média do soma, uma vez que os neurônios derivados de pacientes co-cultivados com células controle ainda apresentam corpo celular significativamente maior que os neurônios controle ($p<0,001$; figura 4.5). Não foram observadas diferenças significativas entre os neurônios derivados de controles plaqueados isoladamente e aqueles co-cultivados com células de pacientes (figura 4.5).

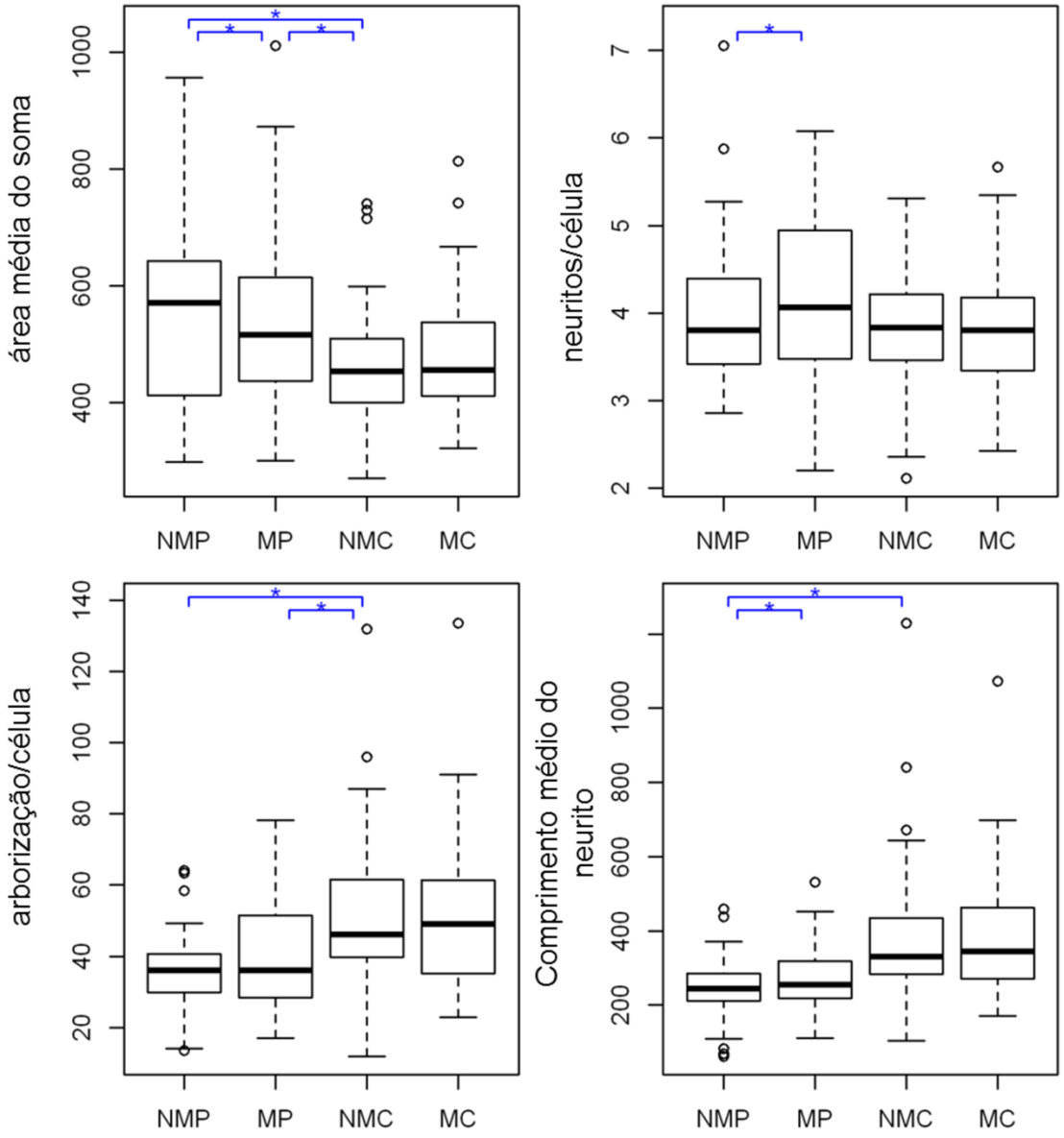

Figura 4.5: Diferenças morfológicas entre neurônios derivados de pacientes e controles. Os gráficos mostram as medidas obtidas para quatro parâmetros diferentes relacionados à morfologia neuronal. Foram usadas amostras provenientes de dois pacientes e três controles para a realização desse experimento. NM: *non-mixed*, M: *mixed*, P: paciente e C: controle.

4.5 Análises de desvio de expressão alélica

Os dados de expressão gênica obtidos a partir do RNAseq de um dado tipo celular provenientes de diferentes indivíduos, além de nos fornecer informações sobre as diferenças de expressão gênica associadas a uma determinada condição, podem ainda indicar a ocorrência de expressão preferencial de um dos alelos de um gene. No caso de doenças como o TEA, em que se observa agregação familiar, o desvio de expressão alélica pode ser um mecanismo relevante para explicar, por exemplo, a penetrância incompleta observada em casos familiais. Considerando isso e o fato de que só é possível

diferenciar os alelos de um mesmo gene caso eles estejam em heterozigose, identificamos, no exoma dos pacientes, variantes em heterozigose e verificamos seu balanço alélico nos dados do transcriptoma, buscando por indícios da ocorrência de expressão preferencial de um dos alelos.

Em nossas análises, identificamos uma média, por indivíduo, de 461 variantes nas quais é possível observar desvio de expressão alélica nas NPC e 544 nos neurônios, distribuídas em 311 e 373 genes, respectivamente. Curiosamente, a maior parte dessas variantes apresenta desvio de expressão para o alelo referência; do total de variantes identificadas nos seis indivíduos, apenas aproximadamente 23% nas NPC e 25% nos neurônios têm expressão desviada para o alelo alternativo. Observamos ainda que em média, por paciente, 147 variantes nas NPC e 187 nos neurônios apresentam desvio para o alelo paterno, enquanto uma média de 135 variantes nas NPC e 166 nos neurônios apresentam desvio para o alelo materno.

As listas de genes que apresentaram desvio de expressão alélica em quatro ou mais pacientes nas NPC (90 genes; tabela 4.6) e nos neurônios (74 genes; tabela 4.7) foram submetidas a análises de enriquecimento para genes do SFARI e anotação funcional. Apenas nas NPC observamos genes do SFARI (*DOCK1* e *SGSH*) apresentando desvio de expressão e, consequentemente, nenhuma das duas listas mostrou enriquecimento para genes agrupados nesse banco na análise realizada no IPA. Verificamos ainda que a lista obtida a partir das amostras de NPC apresenta sobreposição com uma via canônica relevante para o neurodesenvolvimento, a via de sinalização Notch (genes *DTX2* e *NOTCH2*; p=0,01), enquanto a que foi obtida a partir das amostras neuronais está enriquecida para a via de reparo de DNA por recombinação homóloga (gene *RAD52*; p=0,04).

Tabela 4.6: genes apresentando desvio de expressão alélica nas amostras de NPC de quatro indivíduos ou mais.

LINC01128	LOC642846	FRG1BP	PLEC
SLC35E2B	GXYLT1	COL6A2	WASH1
CDK11A,CDK11B	WASF3	PI4KA	PTENP1
NBPF3	KPNA3	BMS1P20	CNTNAP3B
PPIEL	KCTD12	BCR	FRG1JP
NBPF1	MEG3	MN1	NBPF10
CROCCP2	TBC1D2B	RRP7BP	FAM35DP
CROCC	LOC646214	SDHAP1	PWAR5
PDE4DIP	IPW	NCBP2-AS2	NOMO1
NOTCH2	ASB7	NAP1L5	MEIS3P1
NBPF8,NBPF9	WASH3P	TMEM128	SGSH

LINC01347	NME4	RPL9	NLRP2
LGALS8	ACSF3	FRG1	SNHG17
SEC22B	VAMP2	PRIM2	SEPT11
LOC441666	ELAC2	NT5DC1	ZDHHC11
AGAP4	MAP2K3	PEG10	PRICKLE4
LOC439994	ARID3A	ZFAND2A	POM121
DOCK1	UHRF1	TMEM106B	RARRES2
HOTS	ZNF813	CHCHD2	MMP16
H19	ZDBF2	DTX2	PDXDC2P
DDX12P	ATL2	PVRIG2P,STAG3L5P-PVRIG2P-PILRB	SDHAP2
TDG	WASH2P	SLC35B4	
LOC100288778	CCDC74A	PCM1	

Tabela 4.7: genes apresentando desvio de expressão alélica nas amostras neuronais de quatro indivíduos ou mais.

SLC35E2B	RAD52	CYP2D7	FRG1JP
CDK11A,CDK11B	NCAPD2	KLHL24	SLC35E2B
FAM131C	LOC642846	MUC20	PWAR5
NBPF1	LOC100240735	NCBP2-AS2	USP6
CROCCP2	FLJ12825	CRIPAK	VAMP2
MST1P2	WASF3	TMEM128	RPSAP58
THRAP3	KPNA3	RPL9	NLRP2
ZNF697	IPW	NAP1L5	SNHG17
NOTCH2	TBC1D2B	FRG1	SDHAP2
NBPF8,NBPF9	WASH3P	TPPP	SDHAP1
PDE4DIP	ZNF597	PCDHA8	LINC00265
SEC22B	NOMO1	PCDHB9	POM121
FAM231D,LOC388692	MAP2K3	PRIM2	LINC01000
C1orf226	ARID3A	TMEM106B	WASH1
LGALS8	PGPEP1	PEG10	SAPCD2
LINC01347	ANKRD36	PVRIG2P,STAG3L5P-PVRIG2P-PILRB	CROCC
SPRN	WASH2P	PCM1	CBX3
HOTS	FRG1BP	CNTNAP3	
H19	BCR	CNTNAP3B	

5. DISCUSSÃO

O TEA é um transtorno neuropsiquiátrico que apresenta arquitetura genética complexa. Considerando a dificuldade de apontar as variantes associadas à manifestação do fenótipo em cada paciente individualmente, usamos a identificação de processos transcricionalmente desregulados que são comuns aos indivíduos afetados como estratégia para o estudo desse transtorno. Para isso, realizamos o sequenciamento do transcriptoma de NPC e neurônios derivados de iPSC de seis indivíduos do sexo masculino afetados pelo TEA e seis controles pareados por sexo e idade. Ainda que os pacientes incluídos em nossa amostra sejam aparentemente heterogêneos em relação às alterações genômicas associadas à manifestação do fenótipo, as anotações funcionais dos genes diferencialmente expressos identificados nas NPC e nos neurônios revelam que essas alterações convergem em vias canônicas e funções celulares específicas.

Nas NPC, a lista de DEGs entre pacientes e controles mostra um enriquecimento para vias e funções relacionadas a respiração celular, síntese e degradação de proteínas e proliferação/ciclo celular. Problemas no controle dessas funções já foram previamente relacionados a alterações no neurodesenvolvimento.

As mitocôndrias estão envolvidas em uma série de processos que são essenciais para a função neuronal, tais como produção de ATP, homeostase de Ca^{2+}, regulação da apoptose, liberação de neurotransmissores e plasticidade sináptica (Frye & Rossignol, 2011). De fato, diversas evidências apontam o papel da disfunção mitocondrial na etiologia de transtornos neurológicos e psiquiátricos distintos (Manji et al., 2012; Trushina & McMurray, 2007; Prabakaran et al., 2004; Konradi et al., 2004; Dotti et al., 1993). No caso específico do TEA, a meta-análise de três estudos que investigaram a proporção de indivíduos concomitantemente afetados pelo TEA e doença mitocondrial mostra que a prevalência dessa última é de 5% dentre os autistas, muito maior do que a de 0,01% observada na população geral (Rossignol & Frye, 2012). A prevalência de valores anormais de biomarcadores da função mitocondrial em sangue periférico é ainda maior do que a de doenças mitocondriais (Rossignol & Frye, 2012) sugerindo a associação entre TEA e disfunção mitocondrial. Estudos realizados a partir de neuroimageamento, culturas *in vitro* e cérebros *post-mortem* reforçam a alta prevalência de disfunções mitocondriais entre os indivíduos afetados pelo TEA (Rossignol & Frye, 2012). Apesar de ainda não estar claro se as disfunções observadas nos pacientes estão envolvidas na etiologia do transtorno ou se elas são subproduto da desregulação de outros processos celulares, se acredita que elas têm papel relevante nas comorbidades acometendo diferentes órgãos e sistemas que são frequentemente observadas nos indivíduos afetados pelo TEA. A identificação de funções e vias canônicas relacionadas a fosforilação oxidativa, disfunção

mitocondrial e fosforilação/desfosforilação de NAD sugere que o grupo de pacientes incluídos em nosso estudo apresentam desregulação da função mitocondrial, ainda que doenças mitocondriais não tenham sido diagnosticadas nesses indivíduos.

Uma série de genes previamente associados ao TEA codifica proteínas envolvidas no controle da tradução, mecanismo que, no contexto neuronal, parece ser importante para a função sináptica (Santini & Klann, 2014). Parte dessas proteínas são importantes na via de sinalização mTOR, altamente associada ao TEA e responsável pelo controle da tradução, degradação de proteínas e autofagia (Faus-Garriga, Novoa, & Ozaita, 2017; Magdalon, Sánchez-Sánchez, Griesi-Oliveira & Sertié, 2017; Suzuki et al., 2015; Costa-Mattioli & Monteggia, 2013; Zoncu, Sabatini, & Efeyan, 2011). Curiosamente, diversas vias canônicas identificadas como enriquecidas pelos DEGs encontados nas NPC têm papel relevante na via mTOR, seja por modular sua ativação (como é o caso das sinalizações de IGF-1, ERK/MAPK e PI3K/AKT) ou por participar mais diretamente da síntese proteica (no caso das sinalizações eIF4 e p70S6K), apontando sua desregulação em nossos pacientes.

A disfunção mitocondrial e a desregulação da via mTOR também podem ter papel relevante em alterações na proliferação e ciclo celular, uma vez que esses processos dependem de tradução proteica e de um aporte energético apropriado para ocorrerem normalmente. Ainda, diferentes estudos apontam a importância da via ERK/MAPK na proliferação celular ao demonstrarem que a deleção de genes que codificam proteínas ERK causa uma série de alterações no cérebro de camundongos, incluindo problemas na proliferação de células progenitoras (Pucilowska et al., 2015; Pucilowska, Puzerey, Karlo, Galan, & Landreth, 2012; Samuels et al., 2008). Considerando a identificação de alterações transcricionais nessas vias em nosso estudo, procuramos validá-las funcionalmente por meio da avaliação da proliferação das NPC dos pacientes. Embora não significativa, detectamos uma tendência à redução da proliferação nesses indivíduos, sugerindo que as alterações observadas nas NPC em nível transcricional têm impacto negativo sobre processos celulares importantes. Em parte dos estudos realizados em camundongos (Pucilowska, Puzerey, Karlo, Galan, & Landreth, 2012; Samuels et al., 2008), os animais apresentaram alterações no tamanho do cérebro, enquanto reduções no tamanho do cérebro não foram observadas em nenhum dos indivíduos incluídos em nosso estudo. Acreditamos que a desregulação da via ERK/MAPK (possivelmente em conjunto com a desregulação de mTOR e vias relacionadas à função mitocondrial) observada na anotação funcional dos DEGs possa ter influência sobre as alterações identificadas, porém a manifestação do fenótipo se dá de forma mais branda porque essas alterações ocorrem apenas em nível transcricional, enquanto nos camundongos foram provocadas alterações genômicas importantes.

É importante ressaltar que a lista de DEGs identificada nas NPC está enriquecida de genes expressos em córtex de cérebro fetal e células Cajal-Retzius, mostrando relevância desses genes no córtex em desenvolvimento. Ainda, observamos um enriquecimento para genes expressos em plaquetas, células que, apesar de não terem função direta no desenvolvimento do sistema nervoso, são de fácil acesso e podem ser importantes no desenvolvimento de ferramentas diagnósticas.

Nos neurônios, a lista de DEGs entre pacientes e controles mostra um enriquecimento para vias e funções essenciais para o neurodesenvolvimento e função neuronal, tais como sinapse, transporte de neurotransmissores, desenvolvimento de projeção neuronal, canais dependentes de voltagem, organização do citoesqueleto de actina e sinalização de glutamato e GABA. Corroborando esses resultados, uma proporção significativa dos DEGs é expressa no cérebro, no hipocampo e na amígdala. A relevância desses processos biológicos na etiologia do TEA fica clara a medida em que uma série de genes de risco para o transtorno tem envolvimento em seu controle e manutenção (Lin, Frei, Kilander, Shen, & Blatt, 2016). Ainda, demonstramos que essas desregulações transcricionais se refletem em alterações celulares a partir da avaliação da morfologia neuronal. Nossas análises indicam que os neurônios derivados de indivíduos afetados pelo TEA apresentam alteração no tamanho do corpo celular, menor arborização e neuritos mais curtos em relação aos neurônios derivados dos controles, alterações já observadas tanto em pacientes sindrômicos quanto não sindrômicos (Griesi-Oliveira et al., 2014; Doers et al., 2014; Sheridan et al., 2011; Marchetto et al., 2010). Curiosamente, quando esses mesmos neurônios são co-cultivados com células derivadas de controles, observamos um aumento no comprimento médio dos neuritos e uma redução na área do corpo celular, sugerindo que o fenótipo pode ser ao menos parcialmente resgatado se o desenvolvimento dos neurônios se der em microambientes apropriados. De fato, a importância da interação entre astrócitos e neurônios para o desenvolvimento neuronal adequado já foi previamente demonstrada em modelos de Síndrome de Rett e, mais recentemente, TEA (Russo et al., 2017; Williams et al., 2014). Em concordância com nossos resultados, neurônios derivados de indivíduos afetados apresentaram melhorias nos parâmetros morfofisiológicos avaliados quando co-cultivados com astrócitos controle em ambos os trabalhos. De forma recíproca, esses estudos mostram que os astrócitos derivados de pacientes têm influência negativa sobre os neurônios controle, levando a desregulações morfofisiológicas nesses últimos quando as células são co-cultivadas, resultado não obtido em nossas análises. Nos dois estudos anteriores, os grupos usaram protocolos específicos para geração de astrócitos a partir das iPSC, obtendo populações enriquecidas para esse tipo celular, enquanto o protocolo de diferenciação terminal empregado em nossas culturas celulares gera populações heterogêneas de neurônios e células da glia, as quais são compostas por uma

proporção de astrócitos que não conhecemos ao certo. Ainda, considerando que durante o desenvolvimento do sistema nervoso central de vertebrados, os neurônios são gerados antes das células da glia e que o mesmo padrão é observado em culturas de células (Miller & Gauthier, 2007), é possível que nossas linhagens diferenciadas durante um intervalo de tempo relativamente curto contenham quantidades pequenas de astrócitos, tornando difícil a comparação entre os estudos anteriores e o nosso. Com base nisso, algumas hipóteses podem ser levantadas para explicar os resultados discordantes entre nosso estudo e os outros dois trabalhos citados. Primeiramente, é possível que uma pequena quantidade de astrócitos controle seja suficiente para reverter o fenótipo dos neurônios dos autistas, mas uma pequena quantidade de astrócitos autistas não seja suficiente para induzir alterações nos neurônios dos controles, talvez por uma maior resiliência destes últimos à sinalização inadequada pelos astrócitos de pacientes devido a um *background* genético de baixa susceptibilidade ao TEA. Outra possibilidade, considerando que provavelmente temos uma quantidade menor de astrócitos ou temos astrócitos ainda imaturos em comparação aos estudos citados, é que a resposta observada seja mais relacionada a interação entre os neurônios de pacientes e controles. Assim, de forma similar à hipótese levantada para os astrócitos, é possível que a sinalização dos neurônios controle seja suficiente para resgatar o fenótipo dos neurônios autistas, mas os neurônios controle são mais resilientes e não sofrem alteração ao interagir com os neurônios autistas. Finalmente, outra hipótese ainda é a de que a diferenciação de células progenitoras em astrócitos é retardada nos indivíduos autistas e, assim, as culturas neuronais dos pacientes teriam menos astrócitos que as culturas dos controles ao final do nosso protocolo de quatro semanas, influenciando menos na resposta morfofisiológica dos neurônios dos controles. É importante ressaltar também que em doenças de arquitetura genética complexa como o TEA, em que na maior parte das vezes não se conhece as variantes patogênicas e o *background* genético dos indivíduos afetados, é possível que experimentos similares apresentem resultados diversos pelo fato das células derivadas de pacientes responderem de formas diferentes aos estímulos.

A fim de verificar se as alterações identificadas a partir dos DEGs também são preservadas em um contexto sistêmico, realizamos análises de módulos de coexpressão gênica. Nas NPC, apenas o módulo M10-blue, enriquecido para os DEGs identificados nesse tipo celular e, consequentemente, para funções relacionadas a biossíntese proteica, função mitocondrial e ciclo celular apresentaram correlação com o *status* da doença. Devido à grande relevância desses processos celulares, proporções significativas de genes que compõem esse módulo são expressas em uma ampla variedade de tecidos, dentre eles córtex de cérebro fetal, células Cajal-Retzius e plaquetas, observados também na análise de expressão gênica diferencial. É interessante destacar também a expressão

desses genes em pele, que, assim como as plaquetas, é um tecido de fácil acesso, podendo ter papel relevante em ferramentas diagnósticas a serem desenvolvidas. No entanto, para avançar nesse sentido, é necessário que se verifique primeiramente se o padrão de expressão gênica nesses tecidos é similar ao observado em células e tecidos neuronais. A alta sobreposição identificada entre M10-blue e mod4, módulo enriquecido para genes envolvidos em tradução e não correlacionado ao *status* da doença em cérebro *post-mortem* (Gupta et al., 2014), sugere que esse módulo é preservado ao longo da diferenciação, porém só se encontra desregulado nos pacientes em estágios precoces do neurodesenvolvimento.

Nos neurônios, dos vinte módulos identificados, três apresentaram correlação com o *status* da doença. Dois desses módulos, M16-pink e M18-purple têm expressão diminuída nos pacientes e estão relacionados a organização do citoesqueleto e biossíntese de proteínas, respectivamente. O módulo M1-turquoise, por sua vez, tem expressão aumentada nos pacientes e se destaca por apresentar uma grande proporção de DEGs (51,2%), estando consequentemente relacionado a funções neuronais importantes, como sinapse, transporte de íons e processos neurológicos, e enriquecido por genes expressos no cérebro, no hipocampo e na amígdala. Esse módulo apresenta sobreposições altamente significativas com os módulos neuronais correlacionados ao *status* da doença identificados nos três trabalhos anteriores com os quais comparamos nossos dados (Mariani et al., 2015; Gupta et al., 2014; Voineagu et al., 2011), apontando a consistência entre os diferentes estudos mesmo com o uso de modelos distintos e sugerindo a relevância desse módulo na patofisiologia do TEA. Curiosamente, a desregulação desse módulo se dá em sentidos opostos nos diferentes modelos estudados, sendo caracterizado de forma geral por uma expressão reduzida em cérebros *post-mortem* de pacientes (Gupta et al., 2014; Voineagu et al., 2011) e aumentada em neurônios derivados de iPSC desses indivíduos (Mariani et al., 2015). Considerando que nossas amostras neuronais correspondem a fases variando entre 9 e 16 semanas pós-concepção, enquanto os cérebros *post-mortem* são obtidos de indivíduos em estágios posteriores, muitos deles já em idade adulta, acreditamos que as diferenças de expressão observadas entre esses dois modelos sejam decorrentes de mudanças transcricionais que ocorrem ao longo do desenvolvimento, ainda que a estrutura do módulo seja preservada. De fato, Parikshak e colaboradores (2013) caracterizaram a trajetória de expressão de módulos corticais construídos a partir de dados do BrainSpan e mostraram que aqueles relacionados a transmissão sináptica apresentam expressão reduzida em estágios precoces do desenvolvimento, mas sofrem um aumento ao longo do tempo. Em vista dos resultados obtidos a partir de cérebros *post-mortem* e células neuronais derivadas de iPSC de indivíduos afetados pelo TEA, acreditamos que o módulo neuronal correlacionado ao

status da doença apresenta expressão elevada nos pacientes em estágios precoces do neurodesenvolvimento comparada à expressão nos controles, mas essa expressão não sofre um acréscimo suficientemente grande ao longo do tempo que a iguale ao aumento de expressão verificado nos indivíduos não afetados, o que se apresenta como uma redução da expressão do módulo nos pacientes em estágios posteriores do neurodesenvolvimento.

Um dos módulos identificados por Parikshak e colaboradores (2013), nomeado M16, está altamente enriquecido para genes previamente relacionados ao TEA e genes que compõem o módulo m12 identificado no trabalho de Voineagu e colaboradores (2011), de forma que acreditamos que ele seja equivalente ao módulo neuronal preservado nos diferentes estudos analisados. Curiosamente, M16 está também enriquecido para genes carregando variantes *de novo* raras que produzem proteínas truncadas identificadas a partir da análise do exoma de indivíduos afetados pelo TEA em quatro estudos distintos, reforçando a relevância desse módulo para a patofisiologia do TEA e conectando as alterações genéticas observadas em pacientes a um circuito específico que pode ser funcionalmente explorado.

Utilizamos ainda os dados de expressão gerados a partir do sequenciamento do transcriptoma para verificar a ocorrência de desvio de expressão alélica nas células de nossos pacientes, mecanismo que pode ser relevante em fenômenos frequentemente observados no TEA, como penetrância incompleta, discordância entre gêmeos monozigóticos ou excesso de indivíduos do sexo masculino afetados. Nas NPC, dois genes de risco para o TEA (*DOCK1* e *SGSH*) apresentaram desvio de expressão de forma consistente entre os pacientes, ainda que variantes que causem prejuízos na expressão não tenham sido identificadas em nenhum dos indivíduos incluídos nesse estudo. Verificamos ainda que a via de sinalização Notch, que atua na inibição da diferenciação de progenitoras neurais, na especificação dos subtipos de progenitoras e, mais tarde, na maturação neuronal (Yoon & Gaiano, 2005), está enriquecida por genes que apresentam desvio de expressão nas NPC. Nos neurônios, observamos que a via de reparo de DNA por recombinação homóloga, essencial na manutenção da homeostase do sistema nervoso (Lee & McKinnon, 2007), apresenta enriquecimento para genes apresentando desvio de expressão. Dessa forma, variantes em *DOCK1*, *SGSH* e nos genes que compõem essas duas vias (*DTX2*, *NOTCH2* e *RAD52)* precisam ser analisadas com cuidado, em especial quando estão presentes em indivíduos da família não afetados, a fim de se entender sua contribuição na manifestação do TEA nos indivíduos afetados.

Embora os conjuntos de genes identificados com desvio de expressão alélica em NPC e neurônios estejam pouco enriquecidos para genes de risco para o TEA ou vias canônicas relevantes no neurodesenvolvimento, devemos considerar que isso pode ser

reflexo de nosso pequeno tamanho amostral, associado à opção de apontar apenas os genes apresentando desvio de expressão alélica de forma consistente (isto é, em quatro ou mais indivíduos). Esses fatores provavelmente limitaram o número de genes informativos, ou seja, aqueles em que tínhamos indivíduos com variantes em heterozigose. Dessa forma, acreditamos que uma ampliação amostral, que pode ser realizada com indivíduos não autistas e incluir também indivíduos do sexo feminino, pode nos proporcionar uma compreensão mais ampla do conjunto de genes apresentando desvio de expressão alélica em células neurais e, assim, possibilitar um entendimento mais adequado de se esse de fato é um mecanismo associado às questões de penetrância incompleta e desvio sexual observados no TEA.

6. CONCLUSÕES

Nossas análises apontam a ocorrência de desregulações transcricionais em um conjunto de genes importantes no neurodesenvolvimento, alterações também observadas em um contexto sistêmico quando são analisados módulos de coexpressão gênica. Os neurônios diferenciados a partir de nosso protocolo manifestam de forma consistente as alterações observadas nos cérebros de indivíduos afetados pelo TEA e em modelos celulares semelhantes aos empregados nesse estudo, sugerindo que a grande heterogeneidade genética associada ao transtorno se reflete em alterações de processos biológicos específicos, as quais são observadas mesmo em modelos de estudo e/ou subgrupos de pacientes diferentes. Verificamos também que nossas amostras neuronais parecem representar de forma mais próxima o perfil de expressão do cérebro fetal e a comparação entre nossos resultados e aqueles obtidos em estudos anteriores sugere que esse conjunto de genes associados à função sináptica apresenta desregulações opostas no período fetal em relação ao período pós-natal, apontando a relevância do uso de iPSC no estudo de transtornos do neurodesenvolvimento. Ainda, ao demonstrar a ocorrência de desregulações transcricionais associadas ao *status* da doença que não se mantem em estágios posteriores do neurodesenvolvimento, fomos capazes de apontar a importância do uso das NPC para entender de forma mais ampla os processos envolvidos na patofisiologia do TEA. Nesse último caso, a necessidade da utilização de reprogramação celular se torna ainda mais evidente devido à impossibilidade de obtenção de amostras de cérebro de indivíduos afetados que correspondam a estágios tão precoces do desenvolvimento.

Ao migrar o foco das análises do nível genômico, no qual os pacientes são altamente heterogêneos, para o transcriptômico, no qual é possível observar desregulações comuns entre esses indivíduos, obtivemos resultados que abrem caminho para uma melhor compreensão sobre os mecanismos envolvidos na manifestação do TEA e, eventualmente, das comorbidades associadas a ele. Em particular, a observação consistente da desregulação do módulo M1-turquoise em indivíduos afetados pelo TEA sugere sua importância para o desenvolvimento de ferramentas diagnósticas e de terapias medicamentosas voltadas para o tratamento dos principais sintomas do transtorno, focando em drogas que tenham como alvo genes desse módulo.

7. ESTUDO DE CASO

Introdução

O *locus* 17p13.3 apresenta grande instabilidade genômica e é caracterizado pela presença de diversas CNV (*copy number variants*) não recorrentes (Gu et al., 2015). Ainda que duplicações de diferentes tamanhos identificadas nessa região apresentem penetrância incompleta, elas têm sido associadas a transtornos do neurodesenvolvimento, como TEA e deficiência intelectual (Curry et al., 2013; Prasad et al., 2012; Capra et al., 2012; Bruno et al., 2010). Com base em análises de sobreposição dessas duplicações, Bruno e colaboradores (2010) identificaram uma região mínima dentro desse *locus* que parece estar fortemente associada à manifestação do TEA e contém apenas um gene, o *YWHAE.* Ainda, discute-se a possibilidade do *CRK*, gene adjacente ao *YWHAE*, também estar contribuindo com a manifestação do transtorno (Curry et al., 2013).

O gene *YWHAE* codifica a proteína 14-3-3epsilon, pertencente a uma família de proteínas cujos membros interagem com mais de 200 parceiros por meio de ligações a resíduos de serina e treonina fosforilados e estão envolvidos em uma série de eventos celulares, tais como regulação do ciclo celular e transdução de sinais (Cornell, Wachi, Zhukarev, & Toyo-oka, 2016; Berg, Holzmann, & Riess, 2003). Ainda, as proteínas dessa família são altamente expressas no cérebro, sugerindo sua importância na função neuronal (Berg, Holzmann & Riess, 2003). A relevância da proteína 14-3-3epsilon na proliferação, migração e diferenciação de células progenitoras neurais já foi demonstrada (Toyo-oka et al., 2014; Toyo-Oka et al., 2003). Mais recentemente, Cornell e colaboradores (2016) mostraram também o papel dessa proteína na morfogênese neuronal por meio da interação e estabilização de Dcx, uma proteína que se liga a microtúbulos. Além disso, Wachi e colaboradores (2017) observaram alterações em uma série de comportamentos neuropsiquiátricos (como atividade locomotora, memória e sociabilidade) em camundongos nocaute para o gene *YWHAE*, sugerindo o envolvimento de 14-3-3epsilon na patofisiologia de transtornos neurológicos.

O gene *CRK*, por sua vez, codifica a proteína Crk, pertencente a uma família de proteínas adaptadoras que são responsáveis por mediar a formação de complexos proteicos envolvidos na transdução de sinais em resposta a uma série de estímulos extracelulares (Feller, 2001). As proteínas da família Crk são expressas em muitos tecidos e participam de diversos processos biológicos, incluindo regulação da morfologia, migração, proliferação e diferenciação celular (Feller, 2001). No desenvolvimento do cérebro, as proteínas Crk estão envolvidas no controle da migração neuronal e formação das camadas do córtex por meio da ativação da via C3G/Rap1 e na promoção de tradução

proteica e, possivelmente, crescimento de dendritos e desenvolvimento de espinhos dendríticos por meio da ativação da via mTOR em resposta ao estímulo por Relina, proteína codificada por um gene cuja associação com o TEA já está bem estabelecida (figura 7.1A e B; Lee & D'Arcangelo, 2016; Park & Curran, 2008; Ballif et al., 2004).

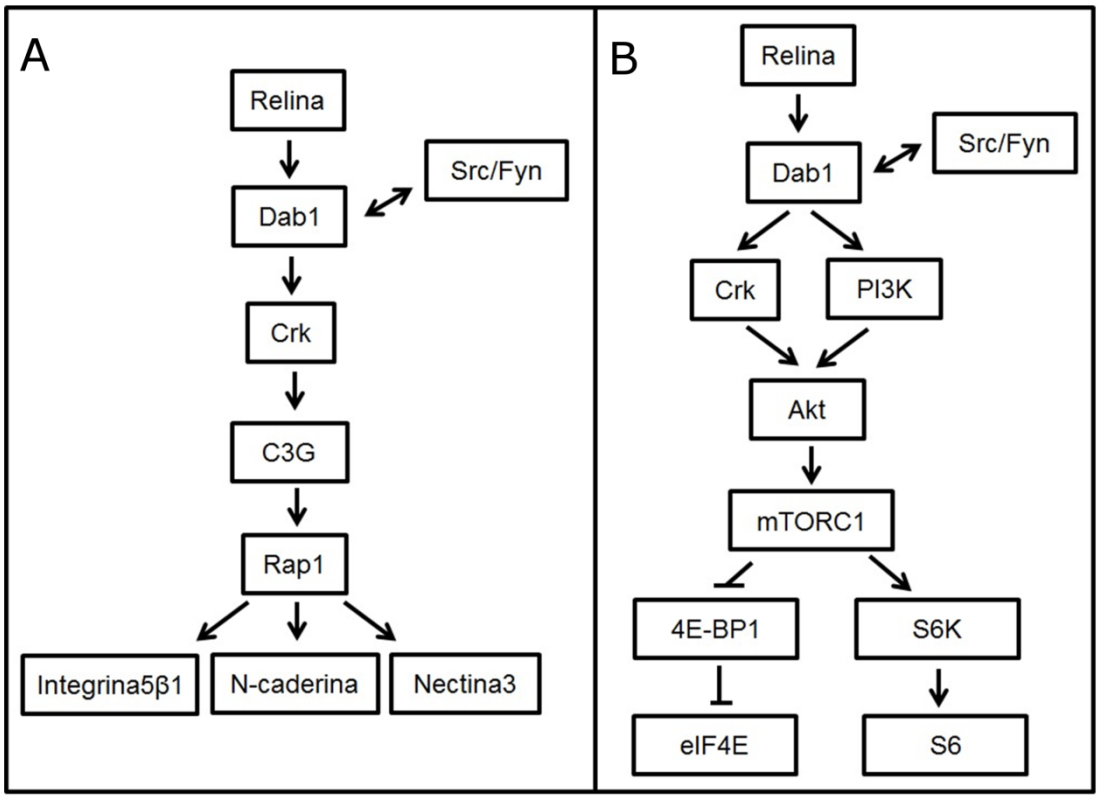

Figura 7.1: A proteína Crk participa da ativação de vias relevantes para o neurodesenvolvimento, tais como a C3G/Rap1, relacionada a migração neuronal, e Akt/mTOR, envolvida no controle de tradução.

A associação de ambos os genes ao TEA e deficiência intelectual é baseada em análises de sobreposição de duplicações e estudos funcionais conduzidos em linhagens celulares e organismos modelo que apontam sua importância no neurodesenvolvimento, mas até o momento nenhum estudo usando células de pacientes foi realizado. Nosso grupo identificou uma duplicação *de novo* de 345kb no *locus* 17p13.3 envolvendo seis genes, dentre eles *YWHAE* e *CRK*, em um indivíduo do sexo masculino diagnosticado com TEA. Assim, nesse estudo, analisamos as consequências funcionais dessa duplicação avaliando a contribuição dos genes *YWHAE* e *CRK* nas possíveis alterações moleculares em células progenitoras neurais (NPC) e neurônios derivados de células-tronco pluripotentes induzidas (iPSC). Considerando ainda a penetrância incompleta do fenótipo em indivíduos apresentando duplicações nessa região, verificamos a existência de outras variantes que pudessem estar contribuindo para a manifestação do quadro nesse paciente.

Material e métodos

Avaliação clínica e coleta de material biológico

O paciente (F2613), encaminhado ao CEGH-CEL (Centro de Pesquisa sobre o Genoma Humano e Células-Tronco) já previamente diagnosticado com TEA, foi submetido a uma anamnese detalhada, por meio da qual levantamos o histórico do paciente e de sua família e avaliamos seu grau de comprometimento em relação às áreas comportamentais e cognitivas afetadas pelo transtorno. Além disso, realizamos um exame físico cuidadoso a fim de identificar características dismórficas ou outras alterações nesse indivíduo.

Coletamos material genético do paciente e de seus pais para a realização de *array* CGH e sequenciamento do exoma. Obtivemos ainda dentes decíduos do paciente e de três indivíduos do sexo masculino não afetados pelo TEA, a partir dos quais foram estabelecidas as linhagens celulares usadas para a realização dos experimentos funcionais e sequenciamento do transcriptoma. Esse estudo foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Biociências – USP (protocolos n° 049/2006 e 1.133.486).

Array CGH

A presença de desequilíbrios genômicos submicroscópicos foi investigada pela técnica de hibridação genômica comparativa baseada em *array – array* CGH. Utilizamos a plataforma da Agilent 4X 180K (Agilent Technologies, Califórnia, USA). Esses oligoarrays contêm quatro áreas com aproximadamente 180.000 oligonucleotídeos de 60pb. Os procedimentos de purificação das amostras, hibridação e lavagem foram realizados conforme descrito pelo fabricante.

As imagens do *array*, obtidas com o Agilent High-Resolution Microarrayscanner, foram processadas e analisadas utilizando o pacote de programas Feature Extraction e Agilent Genomic Workbench (ambos da Agilent Technologies), usando o algoritmo estatístico ADM-2 e o limiar de sensibilidade 6,0. Apenas as alterações abrangendo três oligonucleotídeos consecutivos com razão $\log_2$ alterada foram consideradas pelo programa como possível alteração no número de cópias de determinado segmento genômico.

As alterações de número de cópias identificadas nos pacientes foram comparadas às variações documentadas no banco de dados *Database of Genomic Variants* (DGV), que compila as CNV presentes em amostras de indivíduos normais. Foram comparadas também às informações do *Database of Chromosomal Imbalances and Phenotype in Humans using Ensembl Resources* (DECIPHER), que documenta desequilíbrios genômicos identificados em indivíduos afetados por diferentes patologias.

Cultura e diferenciação de células

Dentes decíduos de pacientes e controles tiveram a polpa removida da coroa. As polpas foram então digeridas em uma solução de 0,25% de tripsina (Invitrogen) a 37°C por 30 minutos e cultivadas em placas de seis poços com meio DMEM/F12 (Invitrogen) suplementado com 15% de soro fetal bovino (FBS - Hyclone) e 100M/ml estreptomicina (Biofluids) e incubadas a 37°C em 5% CO_2. O meio foi trocado uma vez por semana até que as células-tronco (SHED) fossem liberadas da polpa (Costa et al., 2008).

As SHED foram induzidas a expressar os fatores de pluripotência *SOX2, c-Myc, OCT4* e *KLF4* por meio de transdução retroviral (Takahashi et al., 2007) e, após esse procedimento, cultivadas sobre fibroblastos embrionários murinos. Aproximadamente duas semanas após a transdução, as colônias de iPSC foram identificadas e transferidas para placas cobertas com matrigel (BD Biosciences) e cultivadas em meio mTeSR (Stem Cell Technologies).

A diferenciação das NPC a partir das iPSC foi realizada de acordo com o protocolo estabelecido por Marchetto e colaboradores (2010). Quatro dias após o plaqueamento das iPSC, o meio mTeSR foi substituído por DMEM/F12 (Cellgro) suplementado com N2 e B27 (meio NB) a uma concentração final de 0,5X e 10µM de dorsomorfina e, depois de dois dias, as colônias foram transferidas para placas de baixa aderência e mantidas sob agitação a 100rpm por mais dois dias. Durante esse processo, corpos embrióides (EBs) foram formados. No terceiro dia, a dorsomorfina foi substituída por 20ng/ml de FGF e 20ng/ml de EGF e o meio foi trocado a cada dois dias, até se completar uma semana de cultivo dos EBs. No sétimo dia, os EBs foram dissociados com o auxílio da enzima accutase (Millipore). As células obtidas após a dissociação foram plaqueadas sobre matrigel e mantidas em meio 0,5X NB suplementado com 20ng/ml de EGF e FGF até que rosetas fossem formadas. As rosetas foram então coletadas manualmente e transferidas para uma nova placa coberta por matrigel. Três dias após esse procedimento, as células foram novamente removidas e semeadas em placas cobertas com 10µg/ml de poli-ornitina (Sigma) e 5 µg/ml de laminina (Invitrogen) e mantidas em meio 0,5X NB suplementado com 20ng/ml de EGF e FGF, onde formaram uma população homogênea de NPC.

Parte das NPC foi transformada com o vetor Synapsin::EGFP (construção contendo o promotor do gene Synapsin dirigindo a transcrição de GFP) por transdução lentiviral. Primeiramente, o vetor foi empacotado no envelope viral após a transfecção de células HEK293T com PEI (110µg) associado ao Synapsin::EGFP (12,2µg) e sequências de DNA que codificam para proteínas estruturais do lentivírus (8,1µg de MDL, 3,1µg de RSV e 4,1µg de VSVG). As células, mantidas em placas de 150 mm com DMEM *high glucose*

suplementado com 10% de FBS, foram incubadas com a solução de transfecção por seis horas, ao final das quais o meio foi removido e substituído pelo mesmo meio. Três dias após a transfecção, o sobrenadante foi coletado, filtrado e centrifugado a 19.400 rpm por 2h a 12°C. O *pellet* formado foi ressuspendido por aproximadamente 16h em 100µl de DMEM *high glucose* a 4°C e, ao final desse intervalo, adicionamos à ressuspensão meio suficiente para completar 700µl, que foi então usada na transdução. As NPC a serem transduzidas foram incubadas entre cinco e seis horas com a ressuspensão viral a 37°C em 5% CO_2. Após a incubação, a ressuspensão foi substituída por meio NB suplementado com EGF e FGF.

Para a diferenciação neuronal, as NPC foram plaqueadas em baixa densidade (250 mil células para uma placa de 60 mm) em placas cobertas com 20µg/ml de poli-ornitina e 10µg/ml de laminina e mantidas em meio NB suplementado com EGF e FGF por aproximadamente 24h. Ao final desse intervalo, o meio foi substituído por meio NB suplementado com 1 µM de ácido retinóico, tendo então início a diferenciação neuronal. As células foram mantidas nesse meio por quatro semanas, ao final das quais se observou a formação de uma população heterogênea de neurônios e células da glia.

Após as quatro semanas de diferenciação neuronal, as células das linhagens transformadas com o vetor Synapsin::EGFP foram submetidas a separação no citômetro FACSAria. Para isso, as células foram dissociadas com a enzima TrypLE Express (Life Technologies) a 37°C por 20 minutos, coletadas em PBS com 1% de FBS e 0,5% de EDTA e centrifugadas a 1.500 rpm por quatro minutos. Após a centrifugação, as células foram ressuspendidas em meio NB com 1% de FBS e 0,5% de EDTA e passadas para tubos de citometria com tampa de célula-filtro. Para a separação dos neurônios, selecionamos a subpopulação de células apresentando corpo celular pequeno e alta intensidade de fluorescência verde.

Western blot

As NPC foram lisadas e coletadas em tampão RIPA (Thermo Fisher Scientific) com inibidores de protease e fosfatase (para preservação do extrato proteico) quando atingiram confluência entre 90 e 100%. Os neurônios, após separação da fração GFP+ no citômetro, também foram lisados e coletados em RIPA com inibidores. Os lisados foram centrifugados a 15000 rpm a 4°C por 15 minutos e a concentração proteica do sobrenadante foi determinada com o uso do kit Pierce BCA (Thermo Fisher Scientific). As amostras foram preparadas em tampão apropriado a uma quantidade final de 5 µg (para os neurônios) ou 10 µg (para as NPC) de proteína e, após serem desnaturadas, foram carregadas em géis de poliacrilamida de 10 ou 15% . As proteínas foram separadas no gel por eletroforese e,

em seguida, transferidas para membranas de nitrocelulose (GE Healthcare). As membranas foram mantidas em solução de bloqueio (5% de leite em TBST 1X) por aproximadamente 2h e incubadas com anticorpos primários específicos diluídos nessa mesma solução *overnight,* a 4°C e em agitação. Os anticorpos primários usados foram: anti-AKT (1:1000; Cell Signaling), anti-pAKT (S473, 1:1000; Cell Signaling #9271), anti-pS6 (S240/244, 1:2500; Cell Signaling #5364), anti-4E-BP1 (1:2000), anti-DCX (1:200; Santa Cruz Biotechnology #8066) e anti-β-actina (1:10000, Sigma). Após a incubação, as membranas foram lavadas três vezes com TBST 1X e incubadas com os anticorpos secundários apropriados (anti-mouse-HRP (1:2000 ou 1:10000, quando o anticorpo primário era anti-β-actina; Cell Signaling), anti-rabbit-HRP (1:2000; Cell Signaling) ou anti-goat-HRP (1:2000)). O sinal resultante da reação enzimática entre a HPR e o substrato quimioluminescente (Bio-Rad) foi detectado e fotografado por meio do Chemidoc Touch Imaging System (Bio-Rad). A intensidade das bandas foi quantificada com o uso do software ImageLab 5.2.1 (Bio-Rad) e normalizada pelos níveis de β-actina correspondentes. As análises estatísticas foram feitas no programa R v3.4.1 com o pacote lme4 (Bates et al., 2015).

Ensaio multiplex

A fim de avaliar de forma mais ampla alterações na via Akt/mTOR nas NPC do paciente, utilizamos os kits MIlliplex Map Akt/mTOR Total Magnetic Bead 11-Plex e MIlliplex Map Akt/mTOR Phosphoprotein Magnetic Bead 11-Plex (EDM Millipore, #48-612MAG e #48-611MAG, respectivamente), que permitem a análise dos níveis proteicos de uma série de proteínas totais e fosforiladas que compõem a via. Os extratos proteicos foram obtidos a partir de células que foram lisadas e coletadas na solução de lise fornecida nos kits, à qual adicionamos inibidores de protease e fosfatase. A preparação das amostras e o ensaio foram realizados segundo o protocolo do fabricante.

Sequenciamento, processamento e análise dos dados dos sequenciamentos de exoma e transcriptoma

O sequenciamento completo do exoma do paciente e de seus pais foi feito a partir do DNA obtido de células de sangue periférico. A preparação das bibliotecas e a captura do exoma foram realizadas por meio do kit Illumina TruSeq e o sequenciamento foi feito em Illumina HiSeq 2500, onde obtivemos *reads paired-end* de tamanho máximo de 100pb. Uma cobertura acima de 10 *reads* foi alcançada para aproximadamente 60% das bases cobertas pelo exoma de cada paciente. Os *reads* gerados após o sequenciamento foram

alinhados ao genoma referência (hg19) pelo Burrows-Wheller Aligner (bwa-mem) (Li & Durbin, 2009). O alinhamento gerado foi ordenado e teve suas duplicatas de PCR marcadas com a ferramenta Picard (http://broadinstitute.github.io/picard). As etapas seguintes de recalibração das qualidades do alinhamento e a chamada de variantes foi feita pelo Genome Analysis Toolkit (GATK, versão 3.7, McKenna et al., 2010), enquanto o ANNOVAR (Wang, Li & Hakonarson, 2010) e um *script* em Perl foram usados para a anotação dessas variantes, com a finalidade de filtrar aquelas mais relevantes para o nosso trabalho.

A fim de identificar variantes potencialmente patogênicas no paciente, o exoma desse indivíduo foi filtrado para as variantes raras *missense* e de perda de função (introdução de códon de parada prematuro (*stopgain*), alteração no quadro de leitura (*frameshift*) e alteração em sítios de *splicing*). Para isso, foram excluídas das análises as variantes com MAF>0,01 nos bancos de dados públicos *1000 Genomes Project* (http://www.internationalgenome.org/), *NHLBI GO Exome Sequencing Project* (http://evs.gs.washington.edu/EVS/) e ABraOM (http://abraom.ib.usp.br/), qualidade ruim no sequenciamento e/ou encontradas em genes hipervariáveis. As variantes *missense* foram ainda filtradas por patogenicidade de acordo com os bancos de predição SPRING e CADD (Kircher et al., 2014; Wu, Li, & Jiang, 2014); apenas aquelas classificadas nos dez primeiros *rankings* do SPRING e apresentando score ≥ 20 no CADD foram mantidas na análise.

O sequenciamento do transcriptoma foi feito a partir de NPC e neurônios derivados de iPSC do paciente e de seis controles do sexo masculino. A preparação das bibliotecas foi realizada com o uso do kit Truseq RNA Sample Preparation (Illumina), a análise do tamanho e qualidade dos fragmentos foi feita no equipamento Bioanalyzer (Agilent Technologies) e a quantificação das bibliotecas foi realizada por meio de PCR em tempo real com o uso do kit LIBRARY QUANTIFICATION PRIMER PREMIX AND QPCR MASTERMIX for Roche LightCycler® (Kapa Biosystems). O sequenciamento foi feito em Illumina HiSeq 2500, objetivando-se um total de 30 milhões de *reads* por amostra, lidos nos dois sentidos. A qualidade das bibliotecas foi analisada pelo FASTQC e o Trimmomatic (Bolger, Lohse, & Usadel, 2014) foi usado para remover bases de baixa qualidade e descartar *reads* muito curtos. Os *reads* foram alinhados contra o genoma humano (hg19) com o uso do TopHat v2.1.0. Após o alinhamento, selecionamos 20 milhões de *reads* aleatórios em cada amostra, os quais foram usados para a geração da matriz de contagem de *reads* no programa HTSeq (Anders, Pyl & Huber, 2015). O número de *reads* foi normalizado pelo tamanho total da biblioteca e pelo tamanho de cada gene específico, gerando um valor FPKM para cada gene. Mantivemos nas análises posteriores apenas os genes apresentando FPKM≥1 em pelo menos metade das amostras, uma vez que a

presença de transcritos pouco abundantes pode ser decorrente de ruídos técnicos ou biológicos e não de atividade transcricional propriamente dita. A normalização dos dados foi feita por meio do RUVseq (Risso, Ngai, Speed, & Dudoit, 2014). Utilizamos como genes controle negativo a sobreposição entre uma lista de genes constitutivos previamente compilada (Eisenberg & Levanon, 2013) e a lista de genes não diferencialmente expressos entre os controles e um grupo de seis indivíduos afetados pelo TEA (que inclui o paciente descrito nesse estudo de caso) obtida para as NPC ou para os neurônios. Para a identificação dos genes não diferencialmente expressos, realizamos uma análise de expressão gênica diferencial preliminar com o uso do pacote DESeq2 (Love, Huber & Anders, 2014). As análises finais de expressão gênica diferencial também foram realizadas por meio do pacote DESeq2 e consideramos como genes diferencialmente expressos aqueles com $p<0,1$.

Resultados

Histórico clínico

O paciente F2613, atualmente com 19 anos, nasceu a termo após uma gravidez sem complicações. A mãe e o pai, indivíduos não aparentados, tinham 36 e 55 anos, respectivamente, no momento da concepção. O paciente foi encaminhado ao nosso serviço diagnosticado como autista de alto funcionamento. Quando criança, tinha dificuldade em realizar contato visual, balançava muito as mãos (movimento de *flapping*) e apresentava comportamento auto e hetero agressivo. Aos 18 anos, quando foi reavaliado no CEGH-CEL, o paciente apresentou comportamento agressivo, vocabulário rico e boa dicção, apesar da dificuldade em manter sua linha de raciocínio. O paciente não apresenta mal formações ou dismorfismos.

Identificação da duplicação

Após análise dos resultados do *array* CGH, foi identificada no paciente uma duplicação intersticial de 345kb no braço curto do cromossomo 17, afetando a banda 17p13.3 (figura 7.2B) entre as posições genômicas 1.048.587-1.393.982 (Build 37/Hg19 do genoma humano). Essa duplicação envolve seis genes: *ABR*, *BHLHA9*, *TUSC5*, *YWHAE*, *CRK* e *MYO1C* (figura 7.2A).

A fim de determinar se essa CNV foi herdada, amostras de DNA dos pais também foram submetidas ao teste. A mesma duplicação não foi identificada em nenhum dos genitores (figura 7.2C e D), tendo ocorrido *de novo* no paciente.

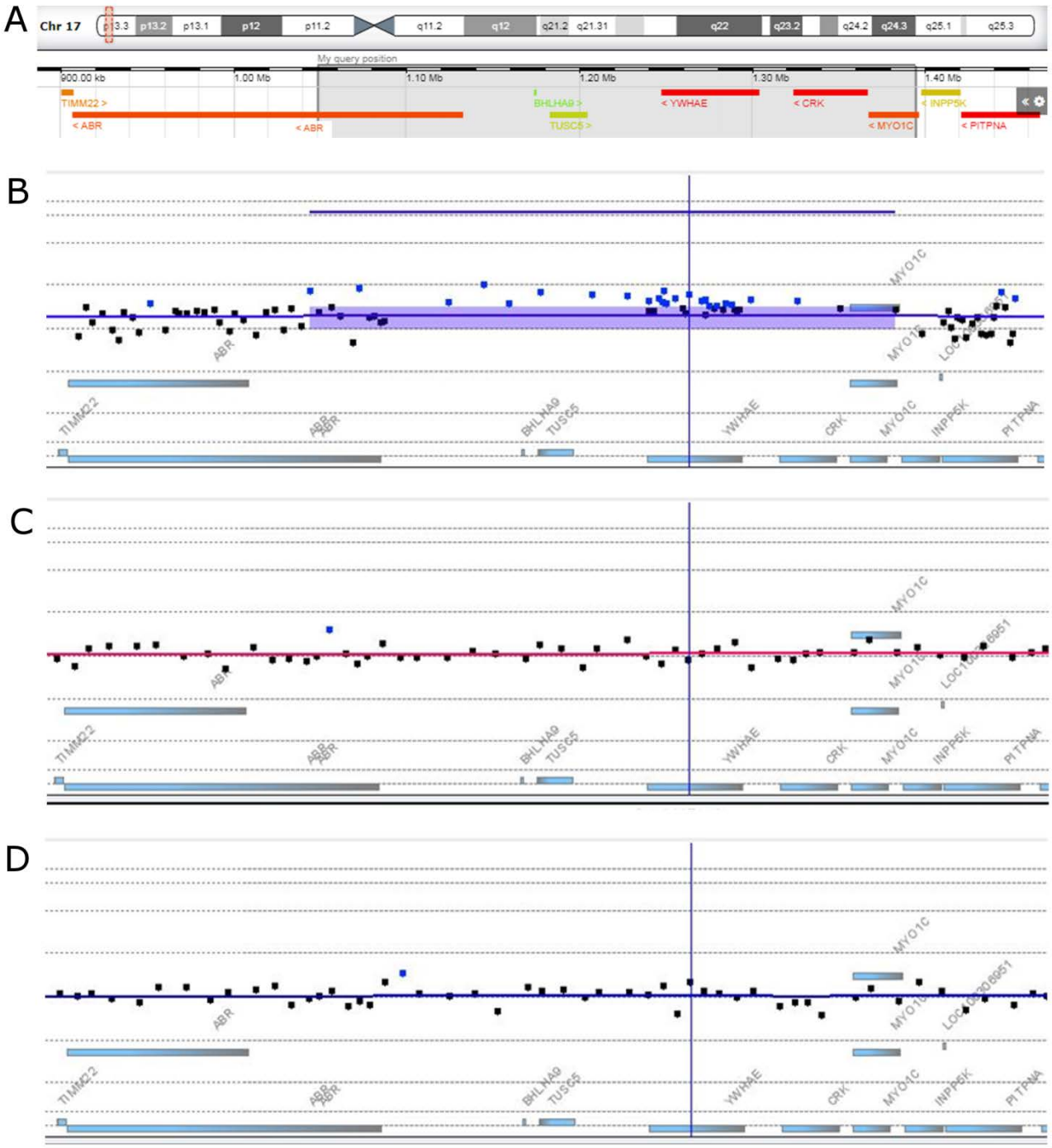

Figura 7.2: O paciente F2613 carrega uma duplicação no cromossomo 17. Região de 345kb envolvendo seis genes (A) que está duplicada no paciente (B), mas não é decorrente de herança materna (C) ou paterna (D).

Os genes YWHAE e CRK têm expressão aumentada em células progenitoras neurais e neurônios derivados do paciente

Considerando que a presença de uma CNV altera o número normal de cópias de um gene, é provável que a ocorrência de uma variante desse tipo (seja uma duplicação ou uma deleção) tenha implicações na expressão dos genes contidos nela. A fim de verificar

se a CNV identificada em nosso paciente leva a alterações na expressão dos genes duplicados, utilizamos os dados de expressão gênica global de NPC e neurônios obtidos pelo nosso grupo. Observamos que dois desses genes, *BHLHA9* e *TUSC5*, não são expressos nos tipos celulares analisados (FPKM<1). Dentre os quatro genes restantes, apenas *YWHAE* e *CRK* apresentaram expressão significativamente aumentada no paciente em relação aos controles. Esse aumento foi observado tanto em NPC quanto em neurônios (tabela 7.1). Com base nesses resultados, decidimos explorar as consequências funcionais da alteração da expressão gênica de *YWHAE* e *CRK*.

Tabela 7.1: expressão dos genes que estão localizados na duplicação do paciente F2613 em relação a indivíduos controle.

	NPC		**Neurônios**	
	Fold change	**p-valor**	***Fold change***	**p-valor**
ABR	1.03	0.98	-1.01	0.96
YWHAE	1.53	**2.93E-12**	1.23	**0.003**
CRK	1.55	**1.68E-08**	1.21	**0.014**
MYO1C	1.32	0.28	1.00	0.99

A ativação da via mTOR está alterada em células progenitoras neurais do paciente

A via mTOR atua em uma série de processos celulares importantes e é ativada por meio da ativação direta ou indireta da proteína mTOR por Akt, uma quinase que, por sua vez, pode ser ativada por Crk (figura 7.1B). Nesse contexto, considerando o aumento da expressão do gene que codifica Crk em nosso paciente, avaliamos a ativação de Akt e da via mTOR por meio da análise dos níveis proteicos e de ativação de componentes da via nas NPC desse indivíduo.

Os níveis proteicos de Akt, p-Akt e p-S6 mensurados por western blot estão significativamente aumentados no paciente em relação aos controles (figura 7.3A, p=0,018, p=0,029 e p<0,001, respectivamente), apontando um aumento da ativação da via mTOR no paciente. Os níveis de 4E-BP1 também mostram uma tendência a estarem aumentados nesse indivíduo (figura 7.3). O aumento de p-S6 foi confirmado por um segundo ensaio (multiplex), o qual também apontou uma tendência de aumento da proteína p-mTOR (figura 7.3B).

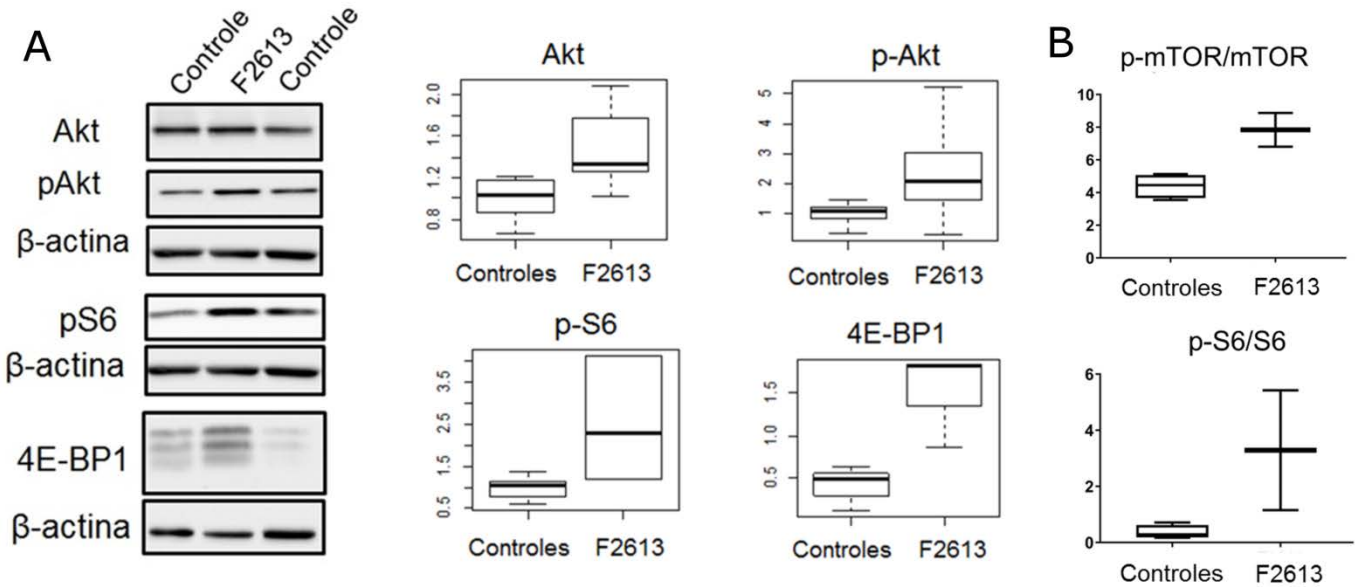

Figura 7.3: Componentes da via mTOR apresentam níveis proteicos aumentados nas NPC derivadas do paciente F2613. (A) Os imunoblots representativos mostram que Akt, p-Akt, p-S6 e 4E-BP1 estão aumentados nas células do paciente. Os gráficos mostram as análises dessas mesmas proteínas (normalizadas pelos níveis de β-actina) feitas a partir dos dados de três clones de NPC derivados do indivíduo F2613 e três controles diferentes obtidos em pelo menos dois experimentos para Akt (p=0,018), p-Akt (p=0,029) e p-S6 (p<0,001). (B) Os gráficos mostram os dados obtidos após a realização do ensaio multiplex, no qual foram utilizados os extratos proteicos de dois clones de NPC do indivíduo F2613 e três controles diferentes, e evidenciam uma tendência de aumento de p-S6 e mTOR.

Nossos dados de expressão gênica global indicam que os genes que codificam a proteína que ativa p-S6 (S6K) ou que codificam as proteínas totais analisadas (Akt e 4E-BP1) não apresentam expressão alterada no paciente (p=0.96, p=0.98, p=0.90, respectivamente), sugerindo que o aumento nos níveis proteicos observado não se dá em nível transcricional.

A proteína Dcx está diminuída em células progenitoras neurais e neurônios do paciente

Dcx é uma proteína que se liga a microtúbulos e os estabiliza, tendo papel importante na morfogênese de neurônios uma vez que está envolvida no controle do crescimento de neuritos (Cornell, Wachi, Zhukarev, & Toyo-oka, 2016). 14-3-3epsilon, codificada pelo gene *YWHAE*, interage com Dcx e a protege da ubiquitinação, diminuindo sua degradação (Cornell, Wachi, Zhukarev, & Toyo-oka, 2016). Dessa forma, um aumento nos níveis de 14-3-3epsilon poderia levar a uma um acúmulo de Dcx decorrente da redução de sua degradação. Considerando o aumento observado na expressão de *YWHAE*, decidimos analisar os níveis proteicos de Dcx nas NPC e nos neurônios do indivíduo F2613. Verificamos que, em ambos os tipos celulares, existe uma tendência a Dcx estar reduzido no paciente em relação aos controles (figura 7.4A e B).

Segundo os dados do estudo de transcriptoma realizado, o gene *DCX* não tem expressão alterada nas células neurais do paciente (p=0,88 em NPC e p=0,81 em neurônios), sugerindo que a alteração observada nos níveis dessa proteína não ocorre em nível transcricional.

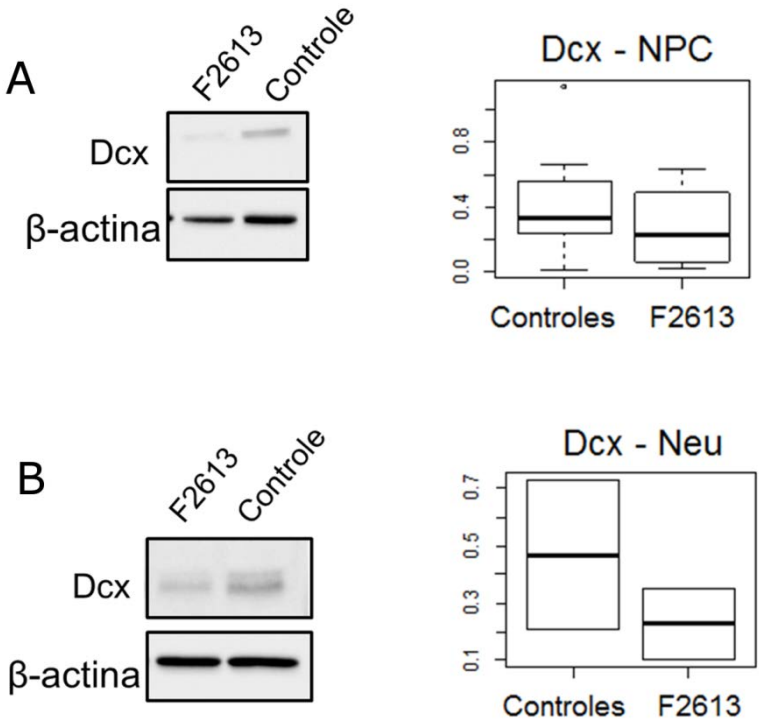

Figura 7.4: A proteína Dcx tem expressão diminuída nas células derivadas do paciente F2613. Os imunoblots representativos mostram que Dcx tende a estar reduzida nas NPC (A) e nos neurônios (B) do paciente em relação aos controles. Os gráficos mostram as análises dessa proteína (normalizada pelos níveis de β-actina) feitas a partir dos dados de três clones de NPC do indivíduo F2613 e de três controles diferentes (p=0,28) (A), e de dois clones de neurônios do paciente e de dois controles diferentes (B).

O indivíduo F2613 apresenta outras variantes raras em genes relevantes para o neurodesenvolvimento

Considerando a penetrância incompleta do fenótipo associado a duplicações na região 17p13.3, usamos os dados do exoma do paciente F2613 para verificar a ocorrência de SNV ou pequenas *indels* raras que pudessem estar contribuindo para a manifestação do fenótipo nesse indivíduo. Identificamos quatro variantes potencialmente patogênicas em genes relevantes para o neurodesenvolvimento e/ou previamente associadas ao TEA (tabela 7.2). A primeira delas é uma variante *stopgain* rara no gene *NGDN*, que codifica uma proteína reguladora da tradução durante o desenvolvimento do sistema nervoso (Jung, Lorenz, & Richter, 2006). Os dados do transcriptoma indicam uma redução significativa na expressão desse gene em NPC e neurônios do paciente F2613 (p=0,008 e

0,025, respectivamente), sugerindo que a variante identificada de fato apresenta impacto sobre sua transcrição. As três restantes, por sua vez, são variantes *missense* com forte indício de patogenicidade, segundo duas ferramentas de predição utilizadas, nos genes *PRKD1, CTNND2* e *CTTNBP2*, encontrados no SFARI e envolvidos em diversas funções importantes no desenvolvimento do sistema nervoso, tais como formação de junções aderentes e interação com o citoesqueleto (Sroka et al., 2016; Turner et al., 2015; Shih, Lee, Chen, & Hsueh, 2014; Hsueh, 2012).

Tabela 7.2: variantes potencialmente patogênicas identificadas na análise do exoma do indivíduo F2613.

Gene	Tipo de variante	SFARI score	SPRING rank*	CADD score**
NGDN	*stopgain*	-	-	-
PRKD1	*missense*	S	1	32
CTNND2	*missense*	2	6	24.4
CTTNBP2	*missense*	3	1	26.1

*De acordo com o método SPRING, variantes *missense* classificadas entre os dez primeiros rankings são consideradas como provavelmente patogênicas.
** De acordo com o método CADD, variantes com score acima de 20 estão dentro do conjunto dos 1% de variantes consideradas mais patogênicas do genoma.

Discussão

Duplicações na região 17p13.3, em especial aquelas incluindo os genes *YWHAE* e *CRK*, parecem estar associadas a manifestação do TEA, porém essa relação ainda não havia sido demonstrada em células neurais de indivíduos afetados. Identificamos uma duplicação de 345kb na região 17p13.3 em um paciente afetado pelo TEA e investigamos os impactos funcionais dessa CNV em NPC e neurônios derivados de iPSC desse indivíduo. Demonstramos que, dos seis genes encontrados na duplicação, apenas *CRK* e *YWHAE* apresentam expressão significativamente aumentada nas células do indivíduo F2613. O primeiro desses genes codifica a proteína Crk, que tem envolvimento na ativação da via mTOR, via altamente associada ao TEA e importante no controle de processos neuronais relevantes, tais como síntese/degradação de proteínas, crescimento de dendritos e desenvolvimento de espinhos dendríticos (Faus-Garriga, Novoa & Ozaita, 2017; Magdalon, Sánchez-Sánchez, Griesi-Oliveira & Sertié, 2017; Lee & D'Arcangelo, 2016; Suzuki et al., 2015). O aumento na expressão de Akt e mTOR, proteínas que estão envolvidas na ativação da via mTOR, e de 4E-BP1 e S6, relevantes no controle da tradução e importantes efetoras dessa via (Zoncu, Sabatini & Efeyan, 2011), sugere que a duplicação do gene *CRK* tem impacto funcional no neurodesenvolvimento uma vez que parece estar associada à superativação da via mTOR nas NPC de nosso paciente.

O gene *YWHAE* codifica a proteína 14-3-3epsilon, pertencente a uma família de proteínas cujos membros se ligam a diversos parceiros e podem alterar a interação desses últimos com outras proteínas (Cornell & Toyo-oka, 2017). Cornell e colaboradores (2016) demonstraram que 14-3-3epsilon interage diretamente com Dcx, protegendo-a de ubiquitinação. Como consequência dessa interação, a superexpressão de 14-3-3epsilon levou a diminuição da degradação de Dcx e observou-se seu acúmulo no extrato proteico (Cornell, Wachi, Zhukarev, & Toyo-oka, 2016). Entretanto, em nossas análises, o paciente apresentando expressão aumentada do gene *YWHAE* mostrou uma tendência à redução dos níveis de Dcx em NPC e neurônios.

Os dados obtidos por Cornell e colaboradores são resultantes da superexpressão de *Ywhae* em embriões de camundongo e linhagens estabelecidas de neuroblastoma, modelos que, apesar de serem bastante úteis no estudo de transtornos neuropsiquiátricos, não apresentam *background* genético já predisposto à manifestação do TEA como o indivíduo F2613. Dessa forma, uma possível explicação para a divergência de resultados entre os dois estudos é a maior complexidade de interações que estão ocorrendo durante o desenvolvimento do sistema nervoso do paciente, uma vez que ele apresenta alterações em outros genes relevantes para o neurodesenvolvimento. Um exemplo disso seria a atuação de uma mesma proteína nas vias em que *CRK* e *YWHAE*, os dois genes que apresentam claras alterações transcricionais, estão inseridos. Cornell e colaboradores (2016) relatam que a interação com 14-3-3epsilon é dependente da fosforilação de Dcx na treonina 42, resíduo fosforilado por ROCK (Amano et al., 2010). Paralelamente, ROCK tem papel relevante no controle da ativação da via mTOR, uma vez que ativa PTEN, proteína que regula negativamente Akt (Li et al., 2005). Quando PTEN é inibida, Akt é ativada e pode levar, em última instância, à ativação da via mTOR. Considerando que o aumento da expressão de *CRK* nas células do paciente parece estar levando à superativação de Akt, é possível que ROCK esteja sendo exacerbadamente recrutada para ativar PTEN como forma de compensar o efeito do aumento da ativação de Akt. O recrutamento excessivo por essa via pode estar levando a um prejuízo na interação entre ROCK e Dcx e, consequentemente, uma diminuição da fosforilação dessa última proteína. Como a interação entre 14-3-3epsilon e Dcx e a consequente proteção dessa última contra ubiquitinação dependem da fosforilação promovida por ROCK, uma possível interferência na função da quinase decorrente do aumento da expressão de *CRK* pode estar levando ao aumento da degradação de Dcx observado nas células de nosso paciente e provocando alterações importantes na dinâmica de crescimento de neuritos. Alternativamente, ainda que a transcrição de *YWHAE* esteja aumentada nas células do paciente, é possível que os níveis da proteína 14-3-3epsilon estejam reduzidos, o que, infelizmente, não pudemos

avaliar. Nesse caso, a redução nos níveis de Dcx observada no paciente seria resultante da disponibilidade diminuída de 14-3-3epsilon.

Considerando ainda a penetrância incompleta do fenótipo associado a essa duplicação, identificamos variantes raras nos genes *NGDN, PRKD1, CTNND2* e *CTTNBP2* que podem estar atuando como *hits* adicionais no indivíduo F2613. No *NGDN*, a variante *stopgain* parece alterar a disponibilidade do transcrito, o que é evidenciado pela diminuição de sua expressão em nosso estudo por RNAseq, No entanto, é necessário que se realize análises funcionais para avaliar as consequências dessa e das outras três variantes no fenótipo celular.

Nesse estudo, demonstramos pela primeira vez o impacto funcional da duplicação na região 17p13.3 em células neurais derivadas de um indivíduo afetado pelo TEA. Ainda que sejam necessários experimentos adicionais para responder algumas questões em aberto (como a verificação dos níveis de 14-3-3epsilon nas células do paciente e a avaliação do envolvimento de ROCK nas alterações observadas), nossos resultados mostram a ocorrência de alterações importantes em vias e processos relevantes para o neurodesenvolvimento nesse paciente, as quais são condizentes com o aumento de expressão de *CRK e YWHAE*, reforçando o protagonismo desses genes no fenótipo decorrente de duplicações em 17p13.3.

8. BIBLIOGRAFIA

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews. Genetics*, *9*(5), 341–355. https://doi.org/10.1038/nrg2346

Amano, M., Tsumura, Y., Taki, K., Harada, H., Mori, K., Nishioka, T., … Kaibuchi, K. (2010). A proteomic approach for comprehensively screening substrates of protein kinases such as rho-kinase. *PLoS ONE*, *5*(1), 1–9. https://doi.org/10.1371/journal.pone.0008704

American Psychiatric Association. (1994). Diagnostic and Statiscal Manual of Mental Disorders. *4thed*. Washington (DC): American Psychiatric Association.

American Psychiatric Association. (2013). Diagnostic and Statiscal Manual of Mental Disorders. 5*thed*. Washington (DC): American Psychiatric Association.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Ansel, A., Rosenzweig, J. P., Zisman, P. D., Melamed, M., & Gesundheit, B. (2017). Variation in gene expression in autism spectrum disorders: An extensive review of transcriptomic studies. *Frontiers in Neuroscience*, *10*(JAN). https://doi.org/10.3389/fnins.2016.00601

Ballif, B.A., Arnaud, L., Arthur, W.T., Guris, D., Imamoto, A. & Cooper, J.A. (2004) Activation of a Dab1/CrkL/C3G/Rap1 pathway in Reelin-stimulated neurons. *Curr Biol*., 14(7):606-10.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, *45*(3), 601–613. https://doi.org/10.1017/S003329171400172X

Berg, D., Holzmann, C., & Riess, O. (2003). 14-3-3 Proteins in the Nervous System. *Nature Reviews Neuroscience*, *4*(9), 752–762. https://doi.org/10.1038/nrn1197

Bilbo, S. D., Block, C. L., Bolton, J. L., Hanamsagar, R., & Tran, P. K. (2018). Beyond infection - Maternal immune activation by environmental factors, microglial development, and relevance for autism spectrum disorders. *Experimental Neurology*, *299*, 241–251. https://doi.org/10.1016/j.expneurol.2017.07.002

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borel, C., Ferreira, P. G., Santoni, F., Delaneau, O., Fort, A., Popadin, K. Y., … Antonarakis, S. E. (2015). Biased allelic expression in human primary fibroblast single cells. *The American Journal of Human Genetics,* 96, 70–80. https://doi.org/10.1016/j.ajhg.2014.12.001

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, *16*(9), 551–563. https://doi.org/10.1038/nrn3992

Bruno, D. L., Anderlid, B. M., Lindstrand, A., Van Ravenswaaij-Arts, C., Ganesamoorthy, D., Lundin, J., … Schoumans, J. (2010). Further molecular and clinical delineation of co-locating 17p13.3 microdeletions and microduplications that show distinctive phenotypes. *Journal of Medical Genetics*, *47*(5), 299–311. https://doi.org/10.1136/jmg.2009.069906

Capra, V., Mirabelli-Badenier, M., Stagnaro, M., Rossi, A., Tassano, E., Gimelli, S., & Gimelli, G. (2012). Identification of a rare 17p13.3 duplication including the BHLHA9 and YWHAE genes in a family with developmental delay and behavioural problems. *BMC Medical Genetics*, *13*, 7–9. https://doi.org/10.1186/1471-2350-13-93

Christensen, D. L., Baio, J., Braun, K. V. N., Bilder, D., Charles, J., Constantino, J. N., …

Yeargin-Allsopp, M. (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Cornell, B., & Toyo-oka, K. (2017). 14-3-3 Proteins in Brain Development: Neurogenesis, Neuronal Migration and Neuromorphogenesis. *Frontiers in Molecular Neuroscience*, *10*(October), 1–17. https://doi.org/10.3389/fnmol.2017.00318

Cornell, B., Wachi, T., Zhukarev, V., & Toyo-oka, K. (2016). Regulation of neuronal morphogenesis by 14-3-3epsilon (*Ywhae*) via the microtubule binding protein, doublecortin. *Human Molecular Genetics*, *25*(20), ddw270. https://doi.org/10.1093/hmg/ddw270

Costa-Mattioli, M., & Monteggia, L. M. (2013). mTOR complexes in neurodevelopmental and neuropsychiatric disorders. *Nature Neuroscience*, *16*(11), 1537–1543. https://doi.org/10.1038/nn.3546

Costa, A., Bueno, D. F., Martins, M. T., Kerkis, I., Kerkis, A., Fanganiello, R. D., … Passos-Bueno, M. R. (2008). Reconstruction of Large Cranial Defects in Nonimmunosuppressed Experimental Design With Human Dental Pulp Stem Cells. *The Journal of Craniofacial Surgery*, *19*, 204–210.

Curry, C. J., Rosenfeld, J. A., Grant, E., Gripp, K. W., Anderson, C., Aylsworth, A. S., … Dobyns, W. B. (2013). The duplication 17p13.3 phenotype: Analysis of 21 families delineates developmental, behavioral and brain abnormalities, and rare variant phenotypes. *American Journal of Medical Genetics, Part A*, *161*(8), 1833–1852. https://doi.org/10.1002/ajmg.a.35996

Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science,* 343, 193-196. https://doi.org/10.1126/science.1245316

Doers, M. E., Musser, M. T., Nichol, R., Berndt, E. R., Baker, M., Gomez, T. M., … Bhattacharyya, A. (2014). iPSC-Derived Forebrain Neurons from FXS Individuals Show Defects in Initial Neurite Outgrowth. *Stem Cells and Development*, *23*(15), 1777–1787. https://doi.org/10.1089/scd.2014.0030

Donovan, A. P. A., & Basson, M. A. (2017). The neuroanatomy of autism - a developmental perspective. *Journal of Anatomy*, *230*(1), 4–15. https://doi.org/10.1111/joa.12542

Dotti, M. T., Manneschi, L., Malandrini, A., De Stefano, N., Caznerale, F., & Federico, A. (1993). Mitochondrial dysfunction in Rett syndrome. An ultrastructural and biochemical study. *Brain & Development*, *15*, 103–106. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8214327

Eisenberg, E., & Levanon, E. Y. (2013). Human housekeeping genes, revisited. *Trends in Genetics*, *29*(10), 569–574. https://doi.org/10.1016/j.tig.2013.05.010

Faus-Garriga, J., Novoa, I., & Ozaita, A. (2017). mTOR signaling in proteostasis and its relevance to autism spectrum disorders. *AIMS Biophysics*, *4*(1), 63–89. https://doi.org/10.3934/biophy.2017.1.63

Feller, S. M. (2001). Crk family adaptors - signalling complex formation and biological roles. *Oncogene*, *20*(44), 6348–6371. https://doi.org/10.1038/sj.onc.1204779

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591–598. https://doi.org/10.1203/PDR.0b013e31819e7203

Frye, R., & Rossignol, D. (2011). Mitochondrial Dysfunction Can Connect the Diverse Medical Symptoms Associated With Autism Spectrum Disorders. *Pediatr Res*, *69*(5), 41R–47R. https://doi.org/10.1203/PDR.0b013e318212f16b

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Gorrindo, P., Williams, K. C., Lee, E. B., Walker, L. S., McGrew, S. G., & Levitt, P. (2012). Gastrointestinal dysfunction in autism: Parental report, clinical evaluation, and associated factors. *Autism Research*, *5*(2), 101–108. https://doi.org/10.1002/aur.237

Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). Microarray and its

applications. *Journal of Pharmacy and Bioallied Sciences*, *4*(6), 310–312. https://doi.org/10.4103/0975-7406.100283

Gregg, C., Zhang, J., Butler, J. E., Haig, D. and Dulac, C. (2010). Sex-specific parent-of-origin allelic expression in the mouse brain. *Science,* 329, 682-685. https://doi.org/10.1126/science.1190831

Griesi-Oliveira K., Sunaga D. Y., Alvizi L., Vadasz E. & Passos-Bueno M. R. (2013). Stem cells as a good tool to investigate dysregulated biological systems in autism spectrum disorders. *Autism Res.,* 6(5), 354-361. https://doi.org/10.1002/aur.1296

Griesi-Oliveira, K., Acab, A., Gupta, a R., Sunaga, D. Y., Chailangkarn, T., Nicol, X., … Muotri, a R. (2014). Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons. *Molecular Psychiatry*, (March), 1–16. https://doi.org/10.1038/mp.2014.141

Griesi-Oliveira, K., & Sertié, A. L. (2017). Autism spectrum disorders: an updated guide for genetic counseling. *Einstein (São Paulo)*, *15*(2), 233–238. https://doi.org/10.1590/s1679-45082017rb4020

Gu, S., Yuan, B., Campbell, I. M., Beck, C. R., Carvalho, C. M. B., Nagamani, S. C. S., … Lupski, J. R. (2015). Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Human Molecular Genetics*, *24*(14), 4061–4077. https://doi.org/10.1093/hmg/ddv146

Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., … Arking, D. E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature Communications*, *5*, 1–8. https://doi.org/10.1038/ncomms6748

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., … Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism, *Arch Gen Psychiatry*, 68(11), 1095-1102.

Han, G., Sun, J., Wang, J., Bai, Z., Song, F., & Lei, H. (2014). Genomics in Neurological Disorders. *Genomics, Proteomics and Bioinformatics*, *12*(4), 156–163. https://doi.org/10.1016/j.gpb.2014.07.002

Hart, T., Komori, H. K., LaMere, S., Podshivalova, K., & Salomon, D. R. (2013). Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-778

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R Package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*(2), 1–11. https://doi.org/10.18637/jss.v015.i02

Hsueh, Y. P. (2012). Neuron-specific regulation on F-actin cytoskeletons: The role of CTTNBP2 in dendritic spinogenesis and maintenance. *Communicative and Integrative Biology*, *5*(4), 334–336. https://doi.org/10.4161/cib.20364

Huguet, G., Benabou, M., & Bourgeron, T. (2016). The Genetics of Autism Spectrum Disorders. In P. Sassone-Corsi & Y. Christen (Eds.), *A Time for Metabolism and Hormones* (pp. 101–129). Springer. https://doi.org/10.1007/978-3-319-27069-2

Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience and Biobehavioral Reviews*, *39*, 1–33. https://doi.org/10.1016/j.neubiorev.2013.12.001

Jung, M.-Y., Lorenz, L., & Richter, J. D. (2006). Translational Control by Neuroguidin, a Eukaryotic Initiation Factor 4E and CPEB Binding Protein. *Molecular and Cellular Biology*, *26*(11), 4277–4287. https://doi.org/10.1128/MCB.02470-05

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., … Šestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, *478*, 483–489. https://doi.org/10.1038/nature10523

Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., … Devlin, B. (2012). Common genetic variants, acting additively, are a major source of risk for

autism. *Molecular Autism*, *3*(1), 1–13. https://doi.org/10.1186/2040-2392-3-9
Kong, S. W., Shimizu-Motohashi, Y., Campbell, M. G., Lee, I. H., Collins, C. D., Brewster, S. J., … Kunkel, L. M. (2013). Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics* 14(2), 143–152.
Konradi, C., Eaton, M., MacDonald, M. L., Walsh, J., Benes, F. M., & Heckers, S. (2004). Molecular evidence for mitochondrial dysfunction in bipolar disorder. *Arch. Gen. Psychiat.*, *61*, 300–308. https://doi.org/10.1001/archpsyc.61.3.300
Krakowiak, P., Goodlin-Jones, B., Hertz-Picciotto, I., Croen, L. A., & Hansen, R. L. (2008). Sleep problems in children with autism spectrum disorders, developmental delays, and typical development: A population-based study. *Journal of Sleep Research*, *17*(2), 197–206. https://doi.org/10.1111/j.1365-2869.2008.00650.x
Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*. https://doi.org/10.1186/1471-2105-9-559
Langfelder, P., Luo, R., Oldham, M. C., & Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Computational Biology*, *7*(1). https://doi.org/10.1371/journal.pcbi.1001057
Lee, G. H., & D'Arcangelo, G. (2016). New Insights into Reelin-Mediated Signaling Pathways. *Frontiers in Cellular Neuroscience*, *10*, 1–8. https://doi.org/10.3389/fncel.2016.00122
Lee, Y., & McKinnon, P. J. (2007). Responding to DNA double strand breaks in the nervous system. *Neuroscience*, *145*(4), 1365–1374. https://doi.org/10.1016/j.neuroscience.2006.07.026
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
Li, Z., Dong, X., Wang, Z., Liu, W., Deng, N., Ding, Y., … Wu, D. (2005). Regulation of PTEN by Rho small GTPases. *Nature Cell Biology*, *7*(4), 399–404. https://doi.org/10.1038/ncb1236
Lin, M., Hrabovsky, A., Pedrosa, E., Wang, T., Zheng, D., & Lachman, H. M. (2012). Allele-Biased Expression in Differentiating Human Neurons: Implications for Neuropsychiatric Disorders. *PLoS ONE*, *7*(8), 1–10. https://doi.org/10.1371/journal.pone.0044017
Lin, Y.-C., Frei, J. A., Kilander, M. B. C., Shen, W., & Blatt, G. J. (2016). A Subset of Autism-Associated Genes Regulate the Structural Stability of Neurons. *Frontiers in Cellular Neuroscience*, *10*, 1–35. https://doi.org/10.3389/fncel.2016.00263
Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8
Magdalon, J., Sánchez-Sánchez, S. M., Griesi-Oliveira, K., & Sertié, A. L. (2017). Dysfunctional mTORC1 signaling: A convergent mechanism between syndromic and nonsyndromic forms of autism spectrum disorder? *International Journal of Molecular Sciences*, *18*(3), 1–27. https://doi.org/10.3390/ijms18030659
Malone, J. H., & Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology*, *9*. https://doi.org/10.1186/1741-7007-9-34
Manji, H., Kato, T., Di Prospero, N. A., Ness, S., Beal, M. F., Krams, M., & Chen, G. (2012). Impaired mitochondrial function in psychiatric disorders. *Nature Reviews Neuroscience*, *13*, 293–307. https://doi.org/10.1038/nrn3229
Marchetto, M. C. N. M., Carromeu, C., Acab, A., Yu, D., Yeo, G. W., Mu, Y., … Muotri, A. R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell*, *143*(4), 527–539. https://doi.org/10.1016/j.cell.2010.10.016.A
Marchetto, M. C., Bellinson, H., Tian, Y., Freitas, B. C., Fu, C., Vadodaria, K. C., … Muotri, A. R. (2016). Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Molecular Psychiatry* 00, 1-16. https://doi.org/10.1038/mp.2016.95
Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., … Vaccarino, F. M. (2015). FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron

Differentiation in Autism Spectrum Disorders. *Cell*, *162*(2), 375–390. https://doi.org/10.1016/j.cell.2015.06.034

McKenna, A., Matthew, H., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110.20

Miller, F. D., & Gauthier, A. S. (2007). Timing Is Everything: Making Neurons versus Glia in the Developing Cortex. *Neuron*, *54*, 357–369. https://doi.org/10.1016/j.neuron.2007.04.019

Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., … Geschwind, D. H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, *155*(5), 1008–1021. https://doi.org/10.1016/j.cell.2013.10.031

Park, T., & Curran, T. (2008). Crk and CrkL play essential overlapping roles downstream of Dab1 in the Reelin pathway. *J Neurosci*, *28*(50), 13551–13562. https://doi.org/10.1523/JNEUROSCI.4323-08.2008.Crk

Prabakaran, S., Swatton, J., Ryan, M., Huffaker, S., Huang, J.-J., Griffin, J., … Bahn, S. (2004). Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Molecular Psychiatry*, *9*, 684–697. https://doi.org/10.1038/sj.mp.4001511

Prasad, A., Merico, D., Thiruvahindrapuram, B., Wei, J., Lionel, A. C., Sato, D., … Scherer, S. W. (2012). A Discovery Resource of Rare Copy Number Variations in Individuals with Autism Spectrum Disorder. *G3 Genes|Genomes|Genetics*, *2*(12), 1665–1685. https://doi.org/10.1534/g3.112.004689

Pucilowska, J., Puzerey, P. A., Karlo, J. C., Galan, R. F., & Landreth, G. E. (2012). Disrupted ERK Signaling during Cortical Development Leads to Abnormal Progenitor Proliferation, Neuronal and Network Excitability and Behavior, Modeling Human Neuro-Cardio-Facial-Cutaneous and Related Syndromes. *Journal of Neuroscience*, *32*(25), 8663–8677. https://doi.org/10.1523/JNEUROSCI.1107-12.2012

Pucilowska, J., Vithayathil, J., Tavares, E. J., Kelly, C., Karlo, J. C., & Landreth, G. E. (2015). The 16p11.2 Deletion Mouse Model of Autism Exhibits Altered Cortical Progenitor Proliferation and Brain Cytoarchitecture Linked to the ERK MAPK Pathway. *Journal of Neuroscience*, *35*(7), 3190–3200. https://doi.org/10.1523/JNEUROSCI.4864-13.2015

Risch, N., Hoffmann, T. J., Anderson, M., Croen, L. A., Grether, J. K., & Windham, G. C. (2014). Familial recurrence of autism spectrum disorder: Evaluating genetic and environmental contributions. *American Journal of Psychiatry*, *171*(11), 1206–1213. https://doi.org/10.1176/appi.ajp.2014.13101359

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*(9), 896–902. https://doi.org/10.1038/nbt.2931

Ronald, A. & Hoekstra, R.A. (2011). Autism Spectrum Disorders and Autistic Traits: A Decade of New Twin Studies. *Am J Med Genet Part B: Neuropsychiatric Genetics*, 156(3), 255–274. https://doi.org/10.1002/ajmg.b.31159

Rossignol, D. A., & Frye, R. E. (2012). Mitochondrial dysfunction in autism spectrum disorders: A systematic review and meta-analysis. *Molecular Psychiatry*, *17*(3), 290–314. https://doi.org/10.1038/mp.2010.136

Russo, F. B., Freitas, B. C., Pignatari, G. C., Fernandes, I. R., Sebat, J., Muotri, A. R., & Beltrão-Braga, P. C. B. (2017). Modeling the Interplay Between Neurons and Astrocytes in Autism Using Human Induced Pluripotent Stem Cells. *Biological Psychiatry*, *83*(7), 569–578. https://doi.org/10.1016/j.biopsych.2017.09.021

Samuels, I. S., Karlo, J. C., Faruzzi, A. N., Pickering, K., Herrup, K., Sweatt, J. D., … Landreth, G. E. (2008). Deletion of ERK2 Mitogen-Activated Protein Kinase Identifies Its Key Roles in Cortical Neurogenesis and Cognitive Function. *Journal of Neuroscience*, *28*(27), 6983–6995. https://doi.org/10.1523/JNEUROSCI.0679-08.2008

Sandin, S., Lichtenstein, P., Larsson, H., Hultman, C., & Reichenberg, A. (2014). The

familial risk of autism. *JAMA*, *311*(17), 1770–1777. https://doi.org/10.1001/jama.2014.4144.The

Santini, E., & Klann, E. (2014). Reciprocal signaling between translational control pathways and synaptic proteins in autism spectrum disorders. *Science Signaling*, *7*(349), 1–10. https://doi.org/10.1126/scisignal.2005832

Sheridan, S. D., Theriault, K. M., Reis, S. A., Zhou, F., Madison, J. M., Daheron, L., … Haggarty, S. J. (2011). Epigenetic characterization of the FMR1 gene and aberrant neurodevelopment in human induced pluripotent stem cell models of fragile X syndrome. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0026203

Shih, P.-Y., Lee, S.-P., Chen, Y.-K., & Hsueh, Y.-P. (2014). Cortactin-binding protein 2 increases microtubule stability and regulates dendritic arborization. *Journal of Cell Science*, *127*(16), 3521–3534. https://doi.org/10.1242/jcs.149476

Sroka, R., Van Lint, J., Katz, S.-F., Schneider, M. R., Kleger, A., Paschke, S., … Eiseler, T. (2016). Cortactin is a scaffolding platform for the E-cadherin adhesion complex and is regulated by protein kinase D1 phosphorylation. *Journal of Cell Science*, *129*(12), 2416–2429. https://doi.org/10.1242/jcs.184721

Suzuki, A. M., Griesi-Oliveira, K., De Oliveira Freitas Machado, C., Vadasz, E., Zachi, E. C., Passos-Bueno, M. R., & Sertie, A. L. (2015). Altered mTORC1 signaling in multipotent stem cells from nearly 25% of patients with nonsyndromic autism spectrum disorders. *Molecular Psychiatry*, *20*(5), 550–552. https://doi.org/10.1038/mp.2014.175

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Toyo-Oka, K., Shionoya, A., Gambello, M. J., Cardoso, C., Leventer, R., Ward, H. L., … Wynshaw-Boris, A. (2003). 14-3-3ε is important for neuronal migration by binding to NUDEL: A molecular explanation for Miller-Dieker syndrome. *Nature Genetics*, *34*(3), 274–285. https://doi.org/10.1038/ng1169

Toyo-oka, K., Wachi, T., Hunt, R. F., Baraban, S. C., Taya, S., Ramshaw, H., … Wynshaw-Boris, A. (2014). 14-3-3ε and ζ Regulate Neurogenesis and Differentiation of Neuronal Progenitor Cells in the Developing Brain. *Journal of Neuroscience*, *34*(36), 12168–12181. https://doi.org/10.1523/JNEUROSCI.2513-13.2014

Trottier, G., Srivastava, L., & Walker, C. D. (1999). Etiology of infantile autism: A review of recent advances in genetic and neurobiological research. *Journal of Psychiatry and Neuroscience*, *24*(2), 103–115.

Trushina, E., & McMurray, C. T. (2007). Oxidative stress and mitochondrial dysfunction in neurodegenerative diseases. *Neuroscience*, *145*, 1233–1248. https://doi.org/10.1016/j.neuroscience.2006.10.056

Turner, T. N., Sharma, K., Oh, E. C., Liu, Y. P., Collins, R. L., Sosa, M. X., … Chakravarti, A. (2015). Loss of δ-catenin function in severe autism. *Nature*, *520*(7545), 51–56. https://doi.org/10.1038/nature14186

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., … Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–386. https://doi.org/10.1038/nature10110

Wachi, T., Cornell, B., & Toyo-oka, K. (2017). Complete ablation of the 14-3-3epsilon protein results in multiple defects in neuropsychiatric behaviors. *Behavioural Brain Research*, *319*, 31–36. https://doi.org/10.1016/j.bbr.2016.11.016

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), 1–7. https://doi.org/10.1093/nar/gkq603

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq : a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63.

Williams, E. C., Zhong, X., Mohamed, A., Li, R., Liu, Y., Dong, Q., … Chang, Q. (2014). Mutant astrocytes differentiated from Rett syndrome patients-specific iPSCs have adverse effects on wildtype neurons. *Human Molecular Genetics*, *23*(11), 2968–2980. https://doi.org/10.1093/hmg/ddu008

Wing, L. & Potter, D. (2002). The epidemiology of autistic spectrum disorders: is the prevalence rising? *Mental Retardation and Developmental Disabilities Research Reviews,* 8(3), 151-61.

Wu, J., Li, Y., & Jiang, R. (2014). Integrating Multiple Genomic Data to Predict Disease-Causing Nonsynonymous Single Nucleotide Variants in Exome Sequencing Studies. *PLoS Genetics*, *10*(3). https://doi.org/10.1371/journal.pgen.1004237

Yoon, K., & Gaiano, N. (2005). Notch signaling in the mammalian central nervous system: Insights from mouse mutants. *Nature Neuroscience*, *8*(6), 709–715. https://doi.org/10.1038/nn1475

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., … Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science*, *318*(5858), 1917–1920. https://doi.org/10.1126/science.1151526

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, *9*(1), 1–13. https://doi.org/10.1371/journal.pone.0078644

Zoncu, R., Sabatini, D., & Efeyan, A. (2011). mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol*, *12*(1), 21–35. https://doi.org/10.1038/nrm3025.mTOR

9. ANEXOS

ANEXO 1

Caracterização genética dos pacientes

Pipeline para priorização de variantes baseado na integração de dados genômicos e transcriptômicos

A fim de identificar variantes potencialmente patogênicas nos pacientes, o exoma de cada indivíduo foi filtrado para as variantes *missense* e de perda de função (introdução de códon de parada prematuro (*stopgain*), alteração no quadro de leitura (*frameshift*) e alteração em sítios de *splicing*) raras, ou seja, aquelas com MAF (*minor allele frequency*) maior do que 0,01 nos bancos de dados públicos *1000 Genomes Project* (http://www.internationalgenome.org/), *NHLBI GO Exome Sequencing Project* (http://evs.gs.washington.edu/EVS/) e ABraOM (http://abraom.ib.usp.br/) foram excluídas da análise. Como controle interno do experimento, foram mantidas apenas as variantes com boa qualidade no sequenciamento e MAF ≤ 0,05 na corrida. Além disso, as variantes encontradas em genes hipervariáveis também foram excluídas de análises posteriores.

Após essas filtragens, verificamos se as variantes selecionadas eram expressas nas células neurais derivadas de iPSC. Para isso, consideramos como expressos apenas os genes com FPKM ≥ 1 em pelo menos mais da metade das amostras sequenciadas de neurônios (uma vez que a maior parte dos genes expressos em NPC também são expressos em neurônios e a abrangência de genes expressos é maior nos neurônios do que nas NPC). Ainda, para determinar se o exon em que a variante foi encontrada é expresso nessas células, verificamos se a mesma variante aparecia também no sequenciamento dos transcriptomas das NPC e/ou dos neurônios. As variantes *missense* localizadas em genes e exons representados no transcriptoma foram ainda filtradas por patogenicidade de acordo com os bancos de predição SPRING (*SNV Prioritization Via the Integration of Genomic Data* - Wu, Li & Jiang, 2014) e CADD (*Combined Annotation-Dependent Depletion* - Kircher et al., 2014). O SPRING integra seis *scores* de efeito funcional calculados por outros métodos (SIFT, PolyPhen2, LRT, MutationTaster, GERP e PhyloP) e cinco *scores* de associação derivados de diferentes dados genômicos (ontologia gênica, interações proteína-proteína, sequências proteicas, anotações de domínios proteicos e de vias gênicas) para calcular o q-valor de cada uma das SNV e, com base nele, classifica-las. Em estudos de simulação com dados de sequenciamento de exoma, o grupo que desenvolveu o método SPRING demonstrou que entre aproximadamente 45 e

50% das SNVs causativas estão classificadas nos dez primeiros *rankings* e, a partir dessa informação, definimos o critério de corte de variantes em nossas análises. O CADD, por sua vez, integra anotações referentes a funcionalidade, patogenicidade, severidade da doença, efeitos regulatórios medidos experimentalmente e associação de traço complexo em um único *score*, que é usado para classificar as variantes (SNV e pequenas *indels*) em genomas individuais. O *score* indica o quão deletéria uma variante deve ser no genoma (por exemplo, as variantes com *score* ≥ 10 têm predição de serem 10% das variantes mais deletérias, aquelas com *score* ≥ 20 indicam 1% das variantes mais deletérias e assim por diante), informação que também usamos para definir nosso critério de corte. Dessa forma, em nossas análises, mantivemos apenas as variantes classificadas dentre os dez primeiros *rankings* do SPRING e apresentando score ≥ 20 no CADD.

A fim de reduzir o número de variantes de risco e identificar aquelas com maior probabilidade de estarem envolvidas na manifestação do transtorno em determinado paciente, verificamos se as que foram selecionadas estavam localizadas em genes encontrados no SFARI ou apresentavam alterações transcricionais (expressão gênica diferencial e/ou desvio de expressão alélica) no paciente em questão. Genes com funções relevantes no neurodesenvolvimento e apresentando expressão diferencial tanto nas NPC quanto nos neurônios de um mesmo indivíduo foram considerados potencialmente patogênicos mesmo quando variantes raras não foram identificadas, uma vez que alterações transcricionais mantidas ao longo da diferenciação também pode ser um indício de patogenicidade. Esse processo de priorização de variantes é resumido na figura 9.1.

Ainda, a partir dos resultados obtidos no *array* CGH, identificamos os genes localizados em regiões deletadas ou duplicadas e verificamos a ocorrência de alterações transcricionais importantes, que também pudessem estar contribuindo com o quadro clínico. Ao final dessas análises, obtivemos um conjunto de genes que podem estar associados à manifestação do transtorno para cada um dos indivíduos.

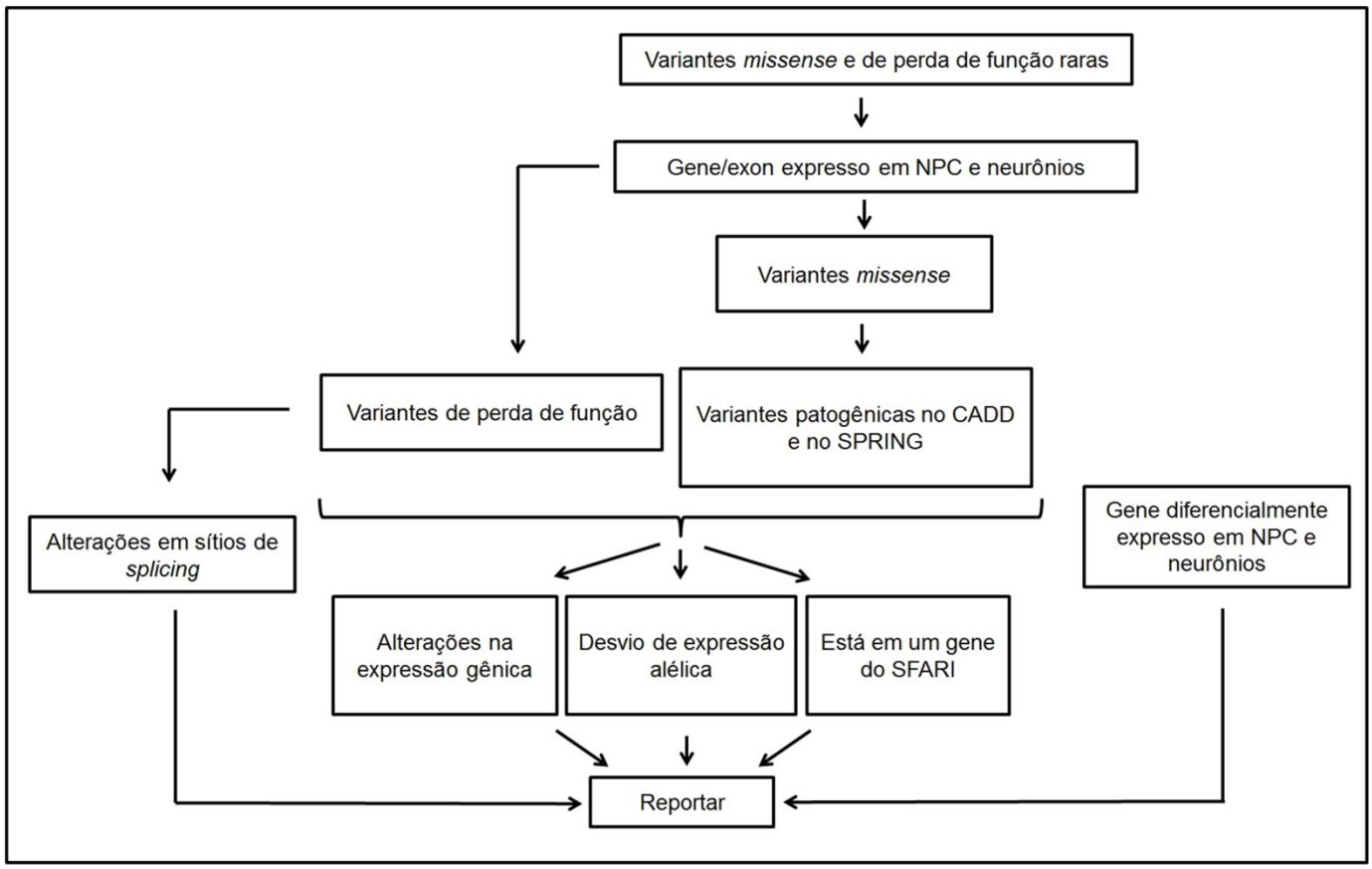

Figura 9.1: Esquema descrevendo a abordagem usada para identificar e priorizar variantes potencialmente patogênicas nos indivíduos afetados pelo TEA incluídos nesse estudo.

Variantes potencialmente patogênicas identificadas nos pacientes

A integração entre dados genômicos e transcriptômicos revelou a presença de genes portando variantes raras que apresentam expressão alterada, seja aumento ou diminuição da expressão gênica total ou a expressão preferencial de um dos alelos em detrimento do outro (tabela 9.1). Como exemplo dos resultados obtidos após essa análise, destacamos o caso do gene *NGDN*, que codifica uma proteína reguladora da tradução durante o desenvolvimento do sistema nervoso e, no paciente F2613, carrega uma variante *stopgain* e apresenta expressão reduzida em NPC e neurônios. As alterações observadas sugerem a ocorrência de decaimento do mRNA proveniente do alelo alternativo devido à inserção de um códon de parada prematuro. Também no indivíduo F2613, identificamos uma duplicação *de novo* de 345kb afetando a banda 17p13.3, região que tem sido relacionada à manifestação do TEA em uma série de estudos. Dois genes (*YWHAE* e *CRK*) dos seis localizados nessa CNV são funcionalmente relevantes para o neurodesenvolvimento e apresentam expressão aumentada tanto nas NPC quanto nos neurônios do paciente, sugerindo sua importância na manifestação do fenótipo nesse indivíduo.

No paciente F2688, identificamos uma variante de *splicing* no *CACNA1H*, gene que codifica uma das subunidades do canal de cálcio dependente de voltagem e é classificado como *score* 2 no SFARI. Observamos a ocorrência de desvio de expressão para o alelo portando essa variante nas NPC e nos neurônios desse indivíduo, indicando a ocorrência de alteração no sítio de *splicing* e expressão do íntron. Dessa forma, essa abordagem nos possibilitou descartar diversas variantes de *splicing* que não evidenciavam incorporação de íntrons ao transcrito.

Identificamos ainda uma variante *missense* rara no gene *GIGYF1*, previamente associado ao TEA e cujo produto proteico está envolvido na regulação da transdução de sinais. No paciente F6136, no qual ela foi identificada, observamos que o gene apresenta expressão diminuída em NPC, alteração que pode estar associada à alteração genômica verificada, porém essa relação é mais difícil de ser estabelecida nos casos em que pouco se sabe sobre o impacto da variante na proteína produzida.

Ao analisar genes com expressão desregulada em NPC e neurônios de um mesmo indivíduo, fomos capazes ainda de apontar genes funcionalmente relevantes para o desenvolvimento do sistema nervoso nos quais não foram identificadas variantes patogênicas após as filtragens do exoma. Exemplos disso são os casos do *DMD* e do *SNORD116*, já associados ao TEA e apresentando expressão diferencial nas NPC e nos neurônios do paciente F2735 e F7511, respectivamente (tabela 9.1). Nesse mesmo padrão, identificamos o gene *PCDHGB6* no paciente F6281, que, apesar de não ter sido associado ao TEA até o momento, é uma molécula de adesão específica do cérebro.

Os genes não destacados aqui, mas que também foram apontados como potencialmente envolvidos na manifestação do TEA após nossa análise integrada, podem ser encontrados na tabela 9.1.

Tabela 9.1: genes funcionalmente relevantes para o desenvolvimento e função neuronal apresentando alterações potencialmente patogênicas nos pacientes. As alterações foram identificadas por meio de análises integradas entre dados exômicos e transcriptômicos.

	Gene	Evidência
	YWHAE	dup 17p13.3 levando ao aumento na expressão em NPC e neurônios
	CRK	dup 17p13.3 levando ao aumento na expressão em NPC e neurônios
	NGDN	variante *stopgain* levando à redução da expressão em NPC e neurônios
	ZNF154	variante em sítio de *splicing*
F2613	*PRKD1*	variante *missense* em um gene do SFARI (score S)
	CTNND2	variante *missense* em um gene do SFARI (score 2)
	CTTNBP2	variante *missense* em um gene do SFARI (score 3)
	PEBP1	DEG em NPC e neurônios
	PCDHB10	DEG em NPC e neurônios

	SCAMP5	DEG em NPC e neurônios
F2688	*NPS*	dup 10q26.2 levando ao aumento da expressão em NPC
	CACNA1H	variante em sítio de *splicing* em um gene do SFARI (score 2)
	RELN	duas variantes *missense* em um gene do SFARI (score 1)
	SLC6A1	variante *missense* em um gene do SFARI (score 2)
	TERF2	variante *missense* em um gene do SFARI (score 3)
	NRP2	variante *missense* em um gene do SFARI (score 4)
F2735	*C9orf78*	variante em sítio de *splicing*; DEG em NPC e neurônios
	C1orf198	variante *stopgain* levando à redução da expressão em NPC e neurônios
	UTRN	variante *missense*; expressão aumentada em neurônios
	SH3BP2	variante *missense*; expressão diminuída em neurônios
	DMAP1	variante *missense*; expressão aumentada em neurônios
	ZNF398	variante *missense*; expressão aumentada em neurônios
	RIMS1	variante *missense* em um gene do SFARI (score 2)
	MED13	variante *missense* em um gene do SFARI (score 2)
	MYO9B	variante *missense* em um gene do SFARI (score 3)
	ETFB	variante *missense* em um gene do SFARI (score 3)
	CADPS2	variante *missense* em um gene do SFARI (score 4)
	SEMA3C	DEG em NPC e neurônios
	CACFD1	DEG em NPC e neurônios
	PCP4	DEG em NPC e neurônios
	DMD	DEG em NPC e neurônios (SFARI score S)
F6136	*GIGYF1*	variante *missense* em um gene do SFARI (score 3); expressão diminuída em NPC
	GRIK5	variante *missense* em um gene do SFARI (score 3)
	CEP135	variante *missense* em um gene do SFARI (score 3)
	SCN8A	variante *missense* em um gene do SFARI (score 3)
F6281	*CHD8*	variante *missense* em um gene do SFARI (score 1S)
	ASH1L	variante *missense* em um gene do SFARI (score 1)
	SRCAP	variante *missense* em um gene do SFARI (score 2)
	EHMT1	variante *missense* em um gene do SFARI (score 3S)
	PCDHGB6	DEG em NPC e neurônios
F7511	*CHD7*	variante *missense* em um gene do SFARI (score S)
	SCN9A	variante *missense* em um gene do SFARI (score 2)
	TRIO	variante *missense* em um gene do SFARI (score 2)
	AKAP9	variante *missense* em um gene do SFARI (score 3)
	CAPRIN1	variante *missense* em um gene do SFARI (score 3)
	PREX1	variante *missense* em um gene do SFARI (score 3)
	NRXN2	variante *missense* em um gene do SFARI (score 4)
	SNORD116	DEG em NPC e neurônios