Universidade de São Paulo

Instituto de Matemática e Estatística - IME

Computational Systems Biology Laboratory - CSBL

**Transcriptômica de autismo e deficiências intelectuais: uma
abordagem de biologia de sistemas para identificar especificidades
e convergências**

Orientanda: Bruna Garbes Gonçalves Pinto

Orientador: Helder Takashi Imoto Nakaya

São Paulo

2020

BRUNA GARBES GONÇALVES PINTO

**Transcriptômica de autismo e deficiências intelectuais: uma
abordagem de biologia de sistemas para identificar
especificidades e convergências**

Dissertação apresentada ao Programa Interunidades de Pós-graduação em Bioinformática para obtenção do título de Mestre em Ciências pelo Instituto de Matemática e Estatística da Universidade de São Paulo.

Orientador: Prof. Dr.^o. Helder Takashi I. Nakaya

SÃO PAULO - SP

2020

AGRADECIMENTOS

Há um provérbio africano que diz que se você quer ir rápido, vá sozinho, mas se você quer ir longe, vá acompanhado. Durante o período em que desenvolvi meu projeto de mestrado tive a oportunidade de ser acompanhada por pessoas especiais que certamente me ajudaram a realizar esse trabalho das mais variadas formas possíveis. As primeiras pessoas que me acompanharam nessa jornada foram meus pais que não pouparam esforços para garantir que eu seguisse meus estudos. Sem eles eu não conseguiria ter sido a primeira pessoa da minha família a fazer graduação, e agora mestrado, nas melhores universidades deste país. Para com eles eu tenho eterna gratidão.

Também gostaria de agradecer ao meu orientador, Dr. Helder Nakaya, por ter me recebido em seu laboratório e por me deixar fazer parte desse time tão vibrante e talentoso que é o CSBL. A oportunidade, paciência e confiança que você teve em mim certamente me levaram muito mais longe do que eu mesma havia imaginado.

Aos meus colegas de laboratório que tanto me ajudaram e que tiveram paciência em responder a todos os meus questionamentos. Vocês não tem ideia de quão brilhantes vocês são!

Aos meus colegas da Morada da Vila por compartilhar as dores e as alegrias de ser um aluno de pós-graduação neste país.

À Fundação Instituto de Pesquisas Farmacêuticas - FIPFARMA projeto nº 87/Fundação Bill e Melinda Gates pelo financiamento, concessão da bolsa e incentivo à pesquisa, bem como à Universidade de São Paulo, por toda a infraestrutura e corpo docente de altíssima excelência que estiveram à minha disposição durante todo esse tempo. Alguns dos meus melhores dias foram dentro desta universidade e sempre a guardarei com muito carinho em meu coração.

Ao Instituto de Matemática e Estatística, por me ensinar tantas coisas. Foi lá que fiz meu primeiro curso de introdução à programação e eu não tinha ideia do quanto isso mudaria a minha vida para melhor.

À minha co-orientadora, Dr^a Karina Griesi, pela minha primeira oportunidade na área de bioinformática quando foi minha orientadora durante a minha iniciação científica e por me apresentar ao universo da genética dos transtornos

neuropsiquiátricos, uma área do conhecimento que é muito desafiadora, mas que me motiva a continuar e acreditar na ciência.

Ao meu companheiro Márcio, por sempre estar ao meu lado e por me ajudar em simplesmente tudo que eu preciso. Você é um ser humano incrível e eu tenho muita sorte em ter você ao meu lado.

E por último mas não menos importante, ao meu irmão Wadson, pela sua alegria e coragem em lidar com uma doença tão difícil quanto é a esquizofrenia. Tenho muito orgulho de ser sua irmã e é para você a quem dedico este trabalho.

“Dizem que a vida é para quem sabe viver, mas ninguém nasce pronto. A vida é para quem é corajoso o suficiente para se arriscar e humilde o bastante para aprender.”

- Clarice Lispector

PINTO, B.G.G. **Transcriptômica de autismo e deficiências intelectuais: uma abordagem de biologia de sistemas para identificar especificidades e convergências.** Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo. São Paulo. 2021

RESUMO

Os transtornos do espectro autista (TEA) e a deficiência intelectual (DI) podem ser encontrados de forma isolada ou em combinação. Juntos, eles afetam cerca de 3% da população e, exceto quando associados à síndromes, ambos não possuem biomarcadores. Geneticamente, um dos fatores que dificulta a identificação de possíveis biomarcadores é o aspecto heterogêneo que essas doenças possuem. Isto é, diferentes genes podem estar alterados em diferentes pacientes, sendo ainda desconhecido quais seriam exatamente esses genes. Entretanto, se a presença de variantes genéticas responsáveis pela exibição desses fenótipos é bastante variada, por outro lado, é possível que tais alterações em níveis funcionais, e.g., transcrição gênica, sejam mais homogêneas entre os pacientes. Com efeito, vários estudos independentes, incluindo um de nosso grupo recém submetido, têm consistentemente encontrado uma rede de genes co-expressos desregulada, cuja função está relacionada a sinapse de neurônios de pacientes com TEA, indicando que o perfil de expressão desses genes poderia ser utilizado como biomarcador para o TEA. Todavia, ainda não se tem conhecimento se há outros perfis de desregulação dessa e de outras redes de co-expressão que possam ser utilizados para diferenciar DI e TEA. Deste modo, com o propósito avaliar as semelhanças e diferenças nos perfis de expressão dessas duas condições, o presente projeto pretende comparar os perfis de expressão gênica de células neuronais de indivíduos com TEA isolado, TEA+DI e DI isolado, visando identificar a existência de módulos de co-expressão cuja desregulação é exclusiva de uma ou outra condição. Para isso usaremos dados de transcriptoma públicos oriundos de tecido post-mortem. Um dos potenciais deste estudo é a contribuição para o desenvolvimento de ferramentas diagnósticas para estas duas condições, favorecendo assim uma intervenção terapêutica precoce e direcionada.

Palavras-chave: transtornos do espectro autista, deficiência intelectual, cérebro *post-mortem*, células tronco pluripotentes-induzidas, transcriptômica, análise de redes, módulos, CEMiTool.

ABSTRACT

Found isolated from each other or in combination, autism spectrum disorders (ASD) and intellectual deficiency (ID), affect around 3% of the population and, except when associated with a syndrome, both of them do not have a molecular biomarker. From a genetic point of view, one of the factors that makes the identification of possible biomarkers difficult, is the genetic heterogeneity associated with these diseases, i.e., different genes can be altered in different patients, being its total identification still unexplored. Though the presence of genetic variants associated with these phenotypes is still quite variable, it is possible that such alterations might result in functional alterations, e.g., at the genetic transcription level, that are more consistent among patients. Indeed, many independent studies, including one recently submitted to publication by our group, have consistently found a co-expression gene network as dysregulated in neurons from ASD patients, whose function is related to synapse, suggesting that the expression profile of these genes could be used as a biomarker for ASD. Nonetheless, there is still no knowledge regarding other possible dysregulation profiles related to this network or to any other co-expression gene network that could be used to differentiate ID and ASD. In this sense, we aim to evaluate the similarities and differences in the expression profiles of these two conditions. To achieve this goal the present study intends to compare the gene expression profile of neuronal cells from patients with ASD only, ASD + ID and ID only, to identify the existence of co-expressed gene networks, also called modules, whose dysregulation might be both exclusive or common to these conditions, by using publicly available transcriptomic datasets. In this case, we are particularly interested in datasets from post-mortem brain tissue

Keywords: autism spectrum disorders, intellectual deficiency, post-mortem brain, induced pluripotent stem cells, transcriptomics, network analysis, modules, CEMiTool.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - Critérios de inclusão e exclusão estabelecidos em cada etapa das revisões sistemáticas

FIGURA 2: Esquema representando as etapas do processo de normalização quantílica aplicada aos dados de microarray.

FIGURA 3: Número de trabalhos de acordo com a inclusão ou não nas fases do processo de revisão sistemática

FIGURA 4: A) Diagrama de Venn representando a sobreposição das amostras únicas entre os estudos de autismo idiopático. B) Número de amostras de acordo com a informação a respeito da presença de DI entre os pacientes com autismo idiopático.

FIGURA 5: A) Regiões cerebrais amostradas em cada um dos 8 estudos selecionados para a análise de construção de redes de co-expressão. B) Número de genes e número de módulos por estudo usados na construção das redes.

FIGURA 6: Rede representando as 100 sobreposições mais significativas que os módulos associados com o status da doença fazem com todos os estudos analisados.

FIGURA 7: Sobreposições entre cada um dos módulos associados aos transtornos avaliados em nosso trabalho com os módulos restantes de todos os estudos analisados.

FIGURA 8: Grafo representando as sobreposições dos módulos do bloco 1.

FIGURA 9: Grafo representando as sobreposições dos módulos do bloco 2.

FIGURA 10: Enriquecimento funcional dos módulos dos blocos 1 e 2.

FIGURA 11: Sobreposição dos módulos desregulados gerados a partir dos trabalhos feitos com TEA idiopáticos e DI sindrômica com os módulos obtidos por Gandal e colegas (2018).

LISTA DE TABELAS

TABELA 1: Representação das condições experimentais dos estudos de TEA idiopático utilizadas para avaliar desregulação de expressão gênica em TEA e DI considerando as informações clínicas a respeito de DI entre as amostras dos pacientes com TEA.

TABELA 2: Descrição dos 8 estudos selecionados para a análise de redes de coexpressão.

TABELA 3: Informações clínicas a respeito do número de amostras de pacientes com TEA que também possuem DI.

TABELA 4: Parâmetros específicos utilizados na construção das redes de coexpressão por estudo.

TABELA 5: Associação com o status do transtorno e com a presença de DI nos estudos de TEA idiopático.

TABELA 6: Termos relacionados aos tipos celulares do banco Cell Marker enriquecidos para os genes dos módulos desregulados em TEA e DI sintomático.

Apêndice I

TABELA 1- Prevalência de autismo nas síndromes genéticas abordadas

TABELA 2- Termos de interesse utilizado na busca sistemática entre os trabalhos de transcriptômica de cérebro *post-mortem* para cada uma das doenças.

TABELA 3- Descrição do número de indivíduos controles e pacientes utilizados nos 8 estudos selecionados

Apêndice II

FIGURA 1: Reprodução do artigo publicado em terceira autoria na revista científica *Molecular Psychiatry* entitulado : “Transcriptomic of iPSC- derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder”

FIGURA 2: Reprodução do artigo publicado em primeira autoria na revista científica *The Journal of Infectious Diseases* entitulado “ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19.”

INTRODUÇÃO	3
1.1 Transtornos do espectro autista	3
1.3 Conhecimento atual sobre marcadores genéticos dos transtornos do espectro autista e da deficiência intelectual	5
1.4 Estudos de expressão gênica global relacionados ao autismo e à deficiência Intelectual	7
1.4.1 Análises em cérebro post-mortem de pacientes com TEA e DI	7
1.4.2 Análises em células neuronais derivadas de células-tronco pluripotentes induzidas de pacientes com TEA e DI	10
1.5 Biologia de sistemas	11
JUSTIFICATIVA	16
OBJETIVOS	17
Objetivo geral	17
Objetivos específicos	17
MATERIAIS E MÉTODOS	18
Revisão sistemática dos estudos em transcriptômica dos transtornos do espectro autista e das deficiências intelectuais	18
Preparação dos dados para análise e geração dos módulos pela ferramenta Cemitool	20
Análises de associação com as doenças originalmente estudadas e com as informações clínicas de deficiência intelectual	22
Análises de sobreposição de módulos entre os estudos	23
Análises de enriquecimento funcional	24
Sobreposição com o trabalho de Gandal e colegas (2018)	25
RESULTADOS	26
Estudos selecionados para a análise de redes de co-expressão	26
Voineagu et. al. (2011) - GSE28521	27
Gupta et. al (2014)	28
Chow et. al (2011) - GSE28475	28
Garbett et. al (2008)	29
Lin et al (2016) - GSE75303	30
Aldinger et al (2020) - GSE128380	31
Lockstone et al (2017) - GSE5390	32
Olmos-Serrano et al. (2016) - GSE59630	33
Caracterização das amostras	34
Geração das redes de co-expressão	36
Análises de associação entre expressão do módulo e presença do transtorno	38
Análises de sobreposição de módulos dos diferentes estudos	41
Análises de enriquecimento funcional	48

Validação com estudos robustos de transcriptômica	53
DISCUSSÃO E CONCLUSÃO	55
Apêndice I - Tabelas	61
Apêndice II - Trabalhos Publicados	64
Apêndice III - Projetos Adicionais	66
Apêndice IV - Pacotes utilizados	68
REFERÊNCIAS	72

INTRODUÇÃO

1.1 Transtornos do espectro autista

O autismo, ou os transtornos do espectro autista (TEA e, em inglês, ASD), é um conjunto de transtornos do neurodesenvolvimento que acomete o indivíduo logo em seus primeiros anos de vida, devendo seus sintomas estarem presentes até os três anos de idade ¹⁻⁴. Os sintomas do TEA são estabelecidos por um conjunto de alterações fenotípicas comportamentais, que incluem déficits no comportamento recíproco sócio-comunicativo, além da presença de comportamentos repetitivos e estereotipados.

Apesar dos traços fenotípicos serem característicos do TEA e preservados ao longo da vida do indivíduo afetado, há um consenso na literatura de que esses traços se apresentam de forma ampla e contínua, tanto na população geral quanto nos indivíduos afetados. Isso implica em uma noção de espectro, no qual o fenótipo observado assume graus, desde severos a mais leves de comprometimento funcional do indivíduo ^{1,2}. Estudos epidemiológicos como o da Rede Americana de Monitoramento de Deficiências do Desenvolvimento e Autismo (*The Autism and Developmental Disabilities Monitoring, ADDM*) do Centro de Controle e Prevenção de Doenças dos EUA (*Centers for Disease Control and Prevention, CDC*) estimam que 1,5% da população mundial possui autismo. Fatores geográficos ou socioeconômicos não estão associados à incidência desta doença, porém é claro o seu viés sexual: para cada menina diagnosticada com TEA, 4 meninos recebem o diagnóstico ^{5,6}.

Com relação à presença de comorbidades, sabe-se que outras desordens neuropsiquiátricas, como epilepsia, transtorno do déficit de atenção com hiperatividade (TDAH) e deficiência intelectual (DI) são comumente encontradas entre os pacientes, além de distúrbios do sono e de problemas gastrointestinais ¹. Com relação à deficiência intelectual, estima-se que ela esteja presente em cerca de um terço dos indivíduos com autismo ⁵.

Com exceção dos casos associados a síndromes bem caracterizadas por seus

sinais clínicos particulares, não há características fisiológicas ou anatômicas específicas para o TEA ¹. A ausência de biomarcadores para a maior parte dos casos de TEA deve-se principalmente ao fato de sua etiologia ser variada, envolvendo fatores genéticos, epigenéticos e ambientais. Como consequência, na maioria dos casos, o diagnóstico é clínico, isto é, feito somente com base no comportamento do indivíduo e em suas habilidades cognitivas. Mesmo que essas condições não possuam cura, é sabido que uma intervenção precoce possibilita um melhor prognóstico para os indivíduos afetados ⁷⁻⁹, trazendo alívio e conforto para a família do paciente que, uma vez estabelecida a identidade diagnóstica de seu filho, pode então se concentrar em informações mais específicas sobre a doença, como por exemplo, as opções de tratamento disponíveis.

1.2 Deficiência intelectual

Assim como o TEA, a deficiência intelectual (DI) também é considerada um transtorno do neurodesenvolvimento. Ela é definida por um nível intelectual abaixo da média (QI<70) e por prejuízos em comportamentos adaptativos, incluindo cuidados pessoais e autonomia, além de prejuízos em habilidades sociais e acadêmicas. Tipicamente manifesta-se antes dos 18 anos ⁴ e afeta 1,5% da população ^{5,6}.

O termo deficiência intelectual pode ser compreendido como um termo guarda-chuva, no qual diversas síndromes se incluem. Nos casos em que fatores genéticos são responsáveis pela sua etiologia, os modelos de herança variam, podendo incluir herança ligada ao cromossomo X, autossômica recessiva e dominante, além dos casos de alterações cromossômicas. As formas idiopáticas, *i.e.*, as formas nas quais o fator etiológico que leva à manifestação do fenótipo é desconhecido, também estão incluídas neste termo e, na verdade, correspondem à maior proporção dos casos ¹⁰.

A DI pode ocorrer tanto de forma isolada quanto em comorbidade com malformações congênitas, epilepsia e alterações do processamento sensorial ¹. A propósito, uma das comorbidades mais comuns entre os casos de DI é justamente o TEA ¹¹. Como resultado, o diagnóstico diferencial dessas duas condições é de difícil

estabelecimento. Mesmo com o desenvolvimento de algumas escalas clínicas como as descritas por Wilkins e colegas (2009) ¹¹ que tentam diferenciar os três grupos de pacientes (pacientes com DI e TEA isoladamente e pacientes com DI e TEA concomitantemente) ¹¹, elas ainda são aplicáveis somente em indivíduos adultos.

De modo análogo ao TEA, exceto nos casos nos quais as manifestações clínicas são conhecidas como, por exemplo, síndrome de Down (SD) e síndrome do X-frágil, não há para DI características fisiológicas ou anatômicas específicas que possam ser usadas como biomarcadores. Isso decorre do fato de que a deficiência intelectual também possui uma natureza complexa, na qual mecanismos genéticos, epigenéticos e ambientais exercem um papel importante em sua etiologia. Além disso, sob o ponto de vista genético, a DI também é uma doença geneticamente heterogênea, pois seu componente genético pode ser composto por mutações em diferentes genes que levam à manifestação do mesmo fenótipo.

Ainda que não haja cura para DI e TEA, sabe-se que uma intervenção precoce possibilita um melhor prognóstico para os indivíduos afetados, pois permite a diferenciação dos pacientes que possuem TEA e DI dos que possuem isoladamente uma dessas condições, proporcionando assim, um atendimento que considera as necessidades distintas desses grupos ^{11,12}.

1.3 Conhecimento atual sobre marcadores genéticos dos transtornos do espectro autista e da deficiência intelectual

Tanto TEA quanto DI são doenças complexas. Isto é, diferentes modelos de herança estão associados a sua ocorrência. Além disso, elas também são consideradas multifatoriais, uma vez que em suas etiologias a influência tanto de fatores genéticos quanto de ambientais, além dos fatores epigenéticos, são claramente reconhecíveis.

Do ponto de vista ambiental, sabe-se que fatores externos possuem um papel significativo no risco de TEA e DI na população. No caso do TEA, a exposição pré-natal ao valproato sódico ou ácido valpróico está relacionado ao aumento de risco à doença ¹³, enquanto que em DI fatores exógenos como abuso de álcool e

complicações no parto, além de má nutrição extrema, são fatores etiológicos para alguns dos casos observados ¹⁰. Entretanto, com exceção desses exemplos, os fatores ambientais ainda carecem de uma devida abordagem que explorem efetivamente o grau de contribuição na fisiopatologia dessas doenças.

No contexto genômico, tanto TEA quanto DI possuem uma arquitetura genética extremamente diversa, podendo estar associados a distintos modelos de herança (monogênicos ou poligênicos ou multifatoriais) e, portanto, contando com a contribuição tanto de variantes raras e comuns presentes na população. Estas variantes podem ser de diferentes tipos, incluindo os grandes arranjos cromossômicos, variantes de número de cópias (*Copy Number Variations*, CNVs), pequenas inserções e deleções (chamadas indels) e variantes de base única ou de ponto (*Single Nucleotide Variants*, SNVs) ^{13,14}.

O conceito de heterogeneidade genética nos transtornos do neurodesenvolvimento pode ser bem compreendido quando observamos um breve panorama histórico do campo das pesquisas genéticas em deficiência intelectual. Os primeiros diagnósticos de DI foram feitos em pacientes com síndrome de Down e em pacientes com a síndrome do X- frágil, cujos rearranjos cromossômicos podiam ser observados apenas com o uso de microscópio óptico. Já a partir do uso de técnicas de bandas citogenéticas, novas anormalidades cromossômicas foram descobertas, o que fez com que esse tipo de mutação fosse reconhecido como a principal causa de DI, correspondendo a cerca de 15% dos casos observados. É interessante notar que a síndrome de Down é a forma genética mais comum de DI, enquanto que o gene *FMR1* (de fragile X mental retardation 1), cuja expansão da repetição de sua sequência nucleotídica CGG em sua região promotora está associada à síndrome do X- frágil, é o gene mais mutado em DI. Em função disso, este gene é rotineiramente testado para o diagnóstico desta doença ¹⁰. O uso de técnicas associadas à análise de bandas citogenéticas também permitiu a identificação de deleções e duplicações recorrentes em autossômos que estão relacionadas a outras síndromes que apresentam DI em seu quadro, como a síndrome de Prader-Willi/ Angelman e Williams, por exemplo.

Com o uso do *microarray* genômico, que permite uma avaliação com maior resolução das sequências gênicas, foi possível identificar deleções e duplicações

menores do que se conseguia visualizar com o cariótipo. Por conta de sua maior resolução e maior porcentagem de identificação do fator etiológico relacionado à manifestação do fenótipo, o teste com *microarray* genômico rapidamente substituiu a tradicional cariotipagem com banda G e ainda hoje é o principal teste sugerido para casos de DI e TEA, especialmente dentre os casos não-sindrômicos. Mais recentemente, o uso das tecnologias de sequenciamento de nova geração possibilitou a identificação de mutações de ponto (*single nucleotide variant*, SNV) com maior facilidade. Com base nos achados obtidos por todas estas tecnologias, tem-se que mais de 700 genes já foram identificados entre os estudos que investigam os diferentes tipos de DI associados aos diferentes modelos de herança, tanto os modelos ligados ao cromossomo X, quanto os modelos autossômicos dominantes e recessivos, além dos modelos oligogênicos ¹⁰. Para o TEA esse número é superior a 1000 genes ¹⁵.

Entretanto, ainda que avanços nas tecnologias de triagem em larga escala como *microarray* e o sequenciamento de nova geração ^{10,16-18}, bem como o uso de grandes coortes de pacientes nos estudos experimentais e compartilhamento generalizado desses dados ¹³ tenham permitido progressos no conhecimento do componente genético dessas doenças, ainda é em uma parcela minoritária dos casos de TEA e DI (especialmente dentre DI moderado) que se consegue identificar alguma alteração no nível genômico que pode ser apontada como o fator etiológico ¹⁶⁻¹⁹. Por outro lado, um grande número de genes implicados na etiologia de DI e TEA fazem parte das mesmas vias moleculares, ^{1,6,10,20} indicando que, embora existam diferentes variantes genéticas associadas a estes transtornos, elas poderiam levar a consequências funcionais semelhantes, as quais poderiam estar refletidas, por exemplo, no nível de transcrição gênica.

1.4 Estudos de expressão gênica global relacionados ao autismo e à deficiência Intelectual

1.4.1 Análises em cérebro post-mortem de pacientes com TEA e DI

Do mesmo modo que as abordagens de triagem em larga escala citadas anteriormente permitiram a descoberta de novos genes associados tanto à TEA quanto à DI, elas também permitiram o estudo do transcriptoma de um tecido de forma global. Esse estudo pode ser feito tanto através da análise de expressão diferencial gene a gene, porém mais recentemente, as análises de transcriptomas também têm sido feitas de forma mais integrativa, com o uso de estratégias oriundas da área de biologia de sistemas, cujo foco é dado na seção 1.5. Uma dessas análises integrativas é a análise de redes de co-expressão (tabela 1). Rapidamente, uma vez que ela também será melhor detalhada na seção 1.5, a análise de redes de co-expressão leva em conta as correlações que cada par possível de genes de um dado conjunto de genes expressos em um tecido ou tipo celular pode apresentar. Com isso, é possível representar graficamente os genes cujas expressões mais se correlacionam em uma rede modular, isto é, que possui sub-redes, denominadas módulos (tabela 1). Em posse desses módulos, é possível testar se os níveis de expressão dos genes que os compõem estão associados a alguma variável biológica de interesse, como a presença de um dado fenótipo ^{21,22}. Algumas das vantagens oferecidas pela análise de redes baseia-se no fato de que como os padrões de expressão de um conjunto de genes são levados em conta concomitantemente, pequenas alterações que não seriam visualizadas em uma análise gene a gene podem ser capturadas, o que particularmente pode ser interessante para doenças cuja etiologia seja heterogênea, como é o caso do TEA e da DI ²².

Quando olhamos para os resultados de trabalhos que utilizaram tecido cerebral *post-mortem*, com efeito podemos observar que essa abordagem tem se mostrado promissora, pois esses trabalhos, conduzidos de forma independente entre si, têm identificado consistentemente nos indivíduos autistas, em sua larga maioria idiopáticos, a desregulação de um módulo de genes de resposta imune associado a células da glia e de um módulo de genes de sinapse e neurotransmissão associado a neurônios ²³⁻²⁵.

Em relação à DI, a maioria dos trabalhos que investigaram a expressão gênica global se concentraram em síndromes específicas como Síndrome de Down e de Rett ²⁶⁻²⁹, sendo que poucos desses trabalhos realizaram análises de módulos de

co-expressão. É o exemplo do estudo de Olmos-Serrano e colegas (2016) ²⁷, um dos estudos sobre expressão gênica global em DI mais recentes encontrados durante nossa revisão sistemática. Neste trabalho, os autores investigaram a causa genética mais comum da deficiência intelectual, a trissomia do cromossomo 21, através do sequenciamento do transcriptoma de diversas regiões cerebrais de pacientes e controles com idades que variavam de períodos mais fetais até a vida adulta. Ao construírem as redes de expressão a partir desses dados, os autores detectaram a presença de 60 módulos, sendo 9 deles enriquecidos para funções e vias biológicas sabidamente afetadas em pacientes com síndrome de Down. O que chamou a atenção dos autores foi a identificação de um módulo que possui genes com a expressão diminuída entre os pacientes com síndrome de Down e que é enriquecido para marcadores de oligodendrócitos e funções como controle de potenciais de ação e mielinização de células nervosas, um processo que ainda não havia sido bem caracterizado em estudos anteriores. Os autores validaram este achado visto no nível transcricional através de estudos anatômicos feitos em modelos animais para esta síndrome, nos quais foram observadas diminuição de mielinização de células nervosas em relação aos controles. Apesar desses achados, ainda não foi explorado se, independente do fator etiológico, há consistência com relação aos módulos desregulados em diferentes casos de DI.

Por fim, é interessante ressaltar que em um estudo feito por Gandal e colaboradores (2018) ³⁰ uma rede de co-expressão gênica robusta foi construída a partir de um conjunto de 700 amostras corticais de cérebros *post-mortem* de diferentes estudos que utilizaram amostras oriundas de controles e pacientes com diferentes transtornos psiquiátricos (TEA, esquizofrenia, transtorno bipolar, depressão e alcoolismo). Aplicando um processo de controle de qualidade e normalização restritivas para levar em conta a sobreposição de amostras entre os estudos e para amenizar os efeitos de covariáveis, os autores ajustaram um modelo linear de efeitos mistos nos dados de expressão gênica que os levaram à detecção de módulos cujos sentidos de desregulação apresentavam padrões similares e/ou distintos entre as doenças analisadas. Um exemplo é o módulo denominado CD4, cujo enriquecimento funcional apontou marcadores de astrócitos, que foi visto com aumento de expressão em TEA, transtorno bipolar e

esquizofrenia. Outro exemplo interessante é o módulo CD2, enriquecido para marcadores de glia, no qual um aumento de expressão foi exclusivamente observado em TEA. Dessa forma, este estudo mostra que tal abordagem pode ser promissora na tentativa de se detectar perfis de expressão gênica que possam ser usados para distinguir DI e TEA.

1.4.2 Análises em células neuronais derivadas de células-tronco pluripotentes induzidas de pacientes com TEA e DI

Apesar dos achados obtidos a partir do tecido cerebral para TEA e DI terem se mostrado relevantes, esse tipo de análise possui uma limitação importante: a disponibilidade do tecido para investigação. O tecido cerebral não é facilmente acessível e sua obtenção depende da doação das famílias dos pacientes para os bancos de tecidos, o que dificulta a obtenção de amostras. Nesse sentido, o uso de modelos celulares derivados de células-tronco pluripotentes induzidas (*induced Pluripotent Stem Cell*, iPSC) ³¹ têm se mostrado de grande importância para as pesquisas sobre os transtornos que afetam o sistema nervoso, uma vez que essa estratégia de investigação permite a geração de células neuronais com o mesmo genoma dos pacientes e, portanto, espera-se que essas células reflitam *in vitro* os mesmos aspectos funcionais que seriam encontrados em modelos *in vivo*. Os estudos que foram feitos até o presente momento e que se valeram dessa estratégia para estudar diversos transtornos neuropsiquiátricos ^{32–35} têm encontrado semelhanças com os dados obtidos de neurônios de tecido *post-mortem* ^{32,36–41}. Em um trabalho conduzido pelo nosso grupo de pesquisa, foi possível observar que, embora os dados de expressão gênica dos neurônios de iPSC se assemelhem aos de cérebro adulto, essa semelhança é ainda maior com os dados de cérebro fetal. Esse resultado é interessante para estudos que investigam doenças que afetam o sistema nervoso logo no início de seu desenvolvimento, como é o caso de DI e TEA. Um outro resultado ainda mais relevante do nosso estudo foi a identificação, a partir de neurônios derivados de iPSC de pacientes com TEA (na sua maioria sem DI), da desregulação do mesmo módulo associado a genes de sinapse (que será

denominado doravante de MNeur1) também encontrado como desregulado em cérebro *post-mortem* de pacientes^{23–25,30,42}.

Já o outro módulo associado a glia e enriquecido de genes relacionados à resposta imune que também foi encontrado desregulado em cérebros *post-mortem* de TEA, não foi detectado em nosso estudo. Isso provavelmente se deve ao fato de que a proporção de células da glia era baixa nas culturas de iPSC. O que é interessante de se observar, é que esses mesmos achados do nosso grupo são semelhantes a outros estudos que também utilizaram o mesmo modelo celular *in vitro*^{43–45}.

Em contrapartida, em um estudo feito com iPSC derivadas de pacientes com DI síndrome (Síndrome do X-Frágil)⁴⁶ o módulo de sinapse encontrado pelos autores não apresentou desregulação nos afetados. Este resultado a princípio se mostra conflitante com os nossos, porém mesmo tendo em vista que os pacientes deste estudo também apresentavam TEA, é importante ressaltar que o mecanismo etiológico avaliado neste estudo é diferente do encontrado no TEA idiopático.

Embora os exemplos de trabalhos em DI que tenham usado a abordagem de módulos de co-expressão sejam poucos, há um número considerável de estudos de expressão gênica global em neurônios derivados de iPSC e em tecido *post-mortem* desses pacientes^{33,47–53}, o que representa uma oportunidade de reanálise desse vasto material que ainda carece de ser explorado com abordagens mais sistêmicas como a análise de redes de co-expressão.

1.5 Biologia de sistemas

Para melhor acompanhamento das ideias e discussões expostas nesta e nas próximas seções, a tabela 1 foi montada com os principais conceitos e termos utilizados para descrever as análises feitas dentro do contexto da biologia de sistemas. Primeiramente, a tabela 1 traz os termos que tratam da composição de uma rede, seguida por metodologias que permitem a quantificação dos dados observados nela e, por fim, as principais métricas e métodos usados na criação dessas redes. Sua elaboração ocorreu com base no trabalho de Parikshak e colaboradores (2015)⁵⁴.

TABELA 1: Principais termos técnicos relacionados às análises de redes.

Conceito/Termo	Definição
Módulo de co-expressão	Subconjunto de uma rede de genes, composto por nós que, nesse caso, são genes cujo padrão de expressão são altamente similares. Em inglês, módulo também pode ser denominado por cluster ou community.
Nós	Componentes moleculares que constituem uma rede, e.g., genes em uma rede de expressão gênica.
Aresta	Componente que representa a relação entre dois nós em uma rede. Quando atribuímos um peso a ela estabelecemos uma relação quantitativa entre os nós de uma rede. Este peso pode ser obtido através de medidas que indicam alguma relação funcional em comum entre os nós, e.g., correlação de expressão de genes.
Hub	Nó que, comparativamente aos outros nós da rede ou do módulo, possui um número alto de arestas, i.e., genes altamente conectados.
<i>RNA-sequencing</i> (RNA-seq)	Tecnologia de sequenciamento de nova geração que mede os níveis de transcritos em uma dada célula e/ou tecido de interesse de maneira global. Nela, o RNA é extraído e a partir dele são geradas bibliotecas de cDNA que passam por um sequenciamento em larga escala.
Hibridização em microarranjos (<i>microarray</i>)	Método de avaliação de expressão gênica global anterior ao RNA-seq. Nele, amostras de DNA ou cDNA marcadas com um fluoróforo são hibridizadas contra sequências gênicas específicas fixas em pequenos poços de uma lâmina de hibridização. Quando

	se hibridizam, o sinal de fluorescência é lido por <i>scanners</i> que capturam esses sinais e os transformam em intensidades que refletem a expressão das sequências analisadas.
Matriz de adjacência	Representação matricial de um grafo. Em redes de co-expressão essa matriz é gerada a partir dos valores de correlação entre todas as permutações possíveis entre os nós em uma rede. Os valores dessa matriz são as arestas da rede e podem representar apenas um dado qualitativo (i.e., se há ou não correlação) ou uma medida ponderada da correlação entre os nós.
Rede não-direcionada (<i>undirect network</i>)	Rede na qual não há direção nas arestas. Em outras palavras, o sentido da associação entre dois nós não é representado.
Rede ponderada	Rede na qual as arestas possuem valores contínuos que refletem a probabilidade de conectividade dos nós.
Método não-supervisionado	Técnica de análise na qual a informação sobre o agrupamento dos dados é obtida a partir da variação intrínseca desses dados, sem o uso de informações <i>a priori</i> sobre algum fator de agrupamento. Este não é o único tipo de método, pois existem métodos supervisionados, onde podemos inserir informações que sabemos <i>a priori</i> sobre os nossos dados.

A biologia de sistemas é um campo da ciência que procura estudar os componentes biológicos e suas interações de uma maneira holística, integrando as diferentes camadas de informação que compõem os sistemas biológicos. Através de

uma abordagem orientada a dados, ela constitui-se de um rigoroso método que pode ser tanto supervisionado quanto não- supervisionado (dependendo do método de construção das análises) que permite explorar os mecanismos de regulação molecular e as vias biológicas subjacentes a diversos processos biológicos ⁵⁴. Particularmente no campo da neurociência, o emprego de ferramentas e estratégias de análises oriundas da biologia de sistemas têm propiciado a investigação da arquitetura genética de doenças que acometem o sistema nervoso. Nesse contexto, a análise de redes biológicas revela-se uma maneira interessante de representar grandes conjuntos de dados obtidos principalmente a partir de tecnologias de larga escala, como o transcriptoma, a proteômica e também o metaboloma. A natureza desses dados define a classificação das redes geradas ⁵⁴. Por exemplo, dados de expressão gênica dão origem às redes de co-expressão, enquanto interações físicas entre proteínas originam as chamadas redes PPI (protein- protein interaction networks) e a interação de moléculas oriundas do metabolismo dão origem às redes metabólicas. Ainda, podemos formar redes que levam em conta a combinação de um ou mais tipos desses dados, constituindo uma rede ponderada por diferentes tipos de relações entre os nós ⁵⁵.

Graficamente, as redes são representadas como um conjunto de nós ou vértices e as interações entre esses vértices são representadas por arestas que conectam pares de nós. No caso de redes de co-expressão gênica, foco de nosso trabalho, os nós correspondem aos genes e, as arestas que os conectam, a correlação que a expressão desses genes possuem. Nesse caso os métodos de avaliação da expressão gênica devem ser globais, como o *microarray* e o *RNA-seq* ^{54,56}. Com isso, temos uma estrutura organizacional importante que simultaneamente identifica, agrupa e explora centenas de genes, cujo padrão de expressão é similar através de diferentes condições de interesse ⁵⁵.

De modo geral, o procedimento para a geração de redes de co-expressão envolve os seguintes passos: a partir dos valores de expressão que os genes assumem em uma dada amostra de interesse, uma matriz de adjacência (baseada em um coeficiente de correlação) é construída para cada par possível de genes dos dados de expressão. Acima de um certo limiar aplicado a essa matriz de adjacência, uma lista de nós e de arestas correspondentes é montada para servir de base para a

construção de uma rede que pode levar em conta o sinal da correlação entre os nós ou não. Muitas redes biológicas possuem uma estrutura hierárquica na qual é possível organizar seus nós em sub-redes interconectadas, denominadas de módulos ⁵⁴. Do ponto de vista intramodular, a conectividade dos nós pode identificar quais genes são hubs em um módulo. Ser um hub é um indicativo da importância do gene em um processo de interesse. Genes que codificam reguladores transcricionais ou que servem como marcadores celulares comumente são identificados como hubs ⁵⁴. Mais interessante, a anotação funcional de um hub pode fornecer um indicativo da principal função biológica exercida pelos genes que compõem o módulo e cuja função biológica ainda se é desconhecida.

Adicionalmente, tanto a modularidade quanto a conectividade de uma rede podem ser ordenadas. Isso permite comparar redes geradas entre um estudo de referência e um de interesse para avaliar se elas são preservadas ou não ⁵⁴. Outra possível comparação que também pode ser feita é a verificação da presença de comunidades (ou meta-módulos) que representem genes consistentemente co-regulados em um mesmo módulo ao longo de diferentes estudos. Particularmente, este tipo de análise é empregado nas análises de nosso laboratório e tem gerado resultados relevantes no campo das doenças inflamatórias crônicas como a artrite idiopática juvenil ⁵⁷ e no campo da vacinologia ⁵⁸. Por fim, é possível avaliar se um módulo ou a posição de um dado gene dentro de um módulo muda entre uma condição de controle versus alguma condição de interesse. O que pode indicar desregulações nas expressões de um ou mais módulos que possam estar associados a alguma doença, como TEA ou DI. O mais interessante é que essas análises fornecem um método de priorização de genes com maior alteração para análises posteriores que eventualmente investiguem com mais profundidade o papel desses genes na fisiopatologia da doença ⁵⁴.

JUSTIFICATIVA

Exceto quando há associação com síndromes bem descritas, o diagnóstico de TEA e DI é realizado somente com base na avaliação comportamental dos pacientes. Isso traz incertezas e atraso no diagnóstico, além da impossibilidade de fazer previsões a respeito do prognóstico do indivíduo afetado. Desse modo, a identificação de biomarcadores para essas duas condições é de grande importância para o estabelecimento de uma intervenção terapêutica que seja precoce e direcionada. Porém sabe-se que dada a heterogeneidade genética dessas doenças, ainda é difícil estabelecer o diagnóstico de DI e TEA com base apenas em informações genômicas, mesmo com o avanço das tecnologias de sequenciamento em larga escala. Entretanto, do ponto de vista do fluxo da informação nos sistemas biológicos, quando olhamos para níveis funcionais como o nível transcricional, as alterações encontradas nesse nível em células neuronais de pacientes autistas têm se mostrado homogêneas e consistentes entre diferentes estudos. Todavia, ainda não se sabe se essas alterações seriam exclusivas de TEA ou compartilhadas com DI, uma vez que essas duas condições possuem uma sobreposição clínica relevante. Ainda, por outro lado, também não está claro se haveria desregulação de grupos de genes exclusivamente relacionados a DI, representando assim, uma nova oportunidade para a reanálise de dados de transcriptômica tanto de TEA quanto de DI que estão depositados em banco de dados de acesso livre.

OBJETIVOS

Objetivo geral

De acordo com a explanação da sessão anterior, nosso objetivo geral neste projeto é:

- Identificar se padrões de regulação de módulos de co-expressão podem ser usados para distinguir DI e TEA;

Objetivos específicos

Tendo em mente os objetivos gerais, nossos objetivos específicos neste projeto são:

- Realizar uma busca sistemática por trabalhos de expressão gênica global feitos a partir de tecido cerebral *post-mortem* de indivíduos com TEA isolado, TEA + DI e DI isolado.
- Gerar de forma independente para cada estudo de cada transtorno redes de co-expressão gênica a partir dos dados de transcriptoma depositados em bancos públicos, verificando-se se há módulos desregulados nas condições patológicas;
- Verificar se há consistência entre os achados de diferentes estudos para um mesmo transtorno, ou seja, se o mesmo módulo detectado como desregulado nos pacientes de um estudo encontra-se desregulado em outro estudo da mesma condição patológica;
- Comparar os padrões de desregulação de expressão de módulos entre as três condições em foco, TEA isolado, DI isolado e TEA+DI, buscando identificar se divergem ou se são iguais entre si.

MATERIAIS E MÉTODOS

Revisão sistemática dos estudos em transcriptômica dos transtornos do espectro autista e das deficiências intelectuais

Um processo de revisão sistemática dos trabalhos publicados entre os anos 2000 a 2019 sobre transcriptômica em cérebro *post-mortem* tanto de TEA quanto de DI foi realizado de acordo com as orientações do manual publicado pelo grupo PRISMA ⁵⁹. Uma vez que tanto para TEA quanto para DI podemos observar casos idiopáticos e sindrômicos, incluímos em nossa busca as principais síndromes relacionadas a essas duas condições, dentre elas, síndrome de Down, síndrome do X-frágil, síndrome de Prader-Willi/Angelman, síndrome de Rett e síndrome de Williams, além das formas idiopáticas denominadas simplesmente de autismo idiopático e deficiência intelectual idiopática, totalizando 7 buscas sistemáticas. A tabela 1 da seção apêndice 1 lista todas as doenças incluídas em nosso estudo juntamente com a prevalência estimada da presença de TEA em cada uma das síndromes genéticas selecionadas ^{2,10}. Segundo o guia proposto pelo grupo PRISMA, o processo de revisão sistemática deve possuir 4 etapas essenciais: fase de identificação, fase de triagem, de elegibilidade, e de inclusão (identification, screening, eligibility e inclusion phases, respectivamente). Para cada uma dessas fases os critérios de inclusão e exclusão foram estabelecidos conforme é possível visualizar na figura 1.

FIGURA 1: Critérios de inclusão e exclusão estabelecidos em cada uma das etapas de revisão sistemática feita para cada uma das doenças relacionadas ao TEA e a DI.

Na fase de identificação foram utilizados uma combinação de termos que retornaram trabalhos de transcriptômica realizados a partir tanto de *microarray* quanto de *RNA-seq* conforme cada uma das doenças relacionadas à DI ou TEA. Na fase de avaliação, os resumos dos trabalhos encontrados na fase de identificação foram baixados do banco de dados do PubMed e lidos. Trabalhos que não utilizaram amostras humanas ou não oriundas de córtex foram removidos. Na etapa de elegibilidade os trabalhos selecionados foram lidos integralmente. Nessa etapa também não foi exigido um número mínimo amostral e os trabalhos de revisão foram mantidos para eventual inclusão de trabalhos que tenham sido usados nas revisões encontradas mas que por ventura não tenham sido encontradas pela nossa busca. Na última etapa de inclusão somente trabalhos de expressão gênica global com amostras de córtex humano de pacientes e controles foram utilizados. Esses trabalhos precisaram ter no mínimo 12 amostras entre pacientes e controles que também precisaram ser pareados para covariáveis como idade, sexo e PMI (*post-mortem interval*). Quando disponíveis, os dados clínicos sobre a presença de DI nos trabalhos de TEA também foram coletados.

A implementação das buscas bibliográficas relacionadas às revisões sistemáticas foi feita com o pacote *rentrez*⁶⁰, o qual provê uma interface da linguagem de programação R com a API 'eUtils' do NCBI (National Center for Biotechnology Information). Os termos usados nas estruturas de busca (chamadas

de *queries*) inseridas no rentrez podem ser consultados na tabela 2 da seção de anexos. Com o pacote rentrez realizamos buscas nas bases de dados do *PubMed* (<https://ncbi.nlm.nih.gov/pubmed/>) e do *Gene Expressed Omnibus* (GEO) (<https://ncbi.nlm.nih.gov/geo/>) - dois dos principais bancos para depósitos de artigos científicos na grande área das ciências da saúde e dados de expressão gênica, respectivamente.

Preparação dos dados para análise e geração dos módulos pela ferramenta Cemitool

Os dados dos estudos selecionados para nossas análises foram em sua maioria obtidos da plataforma GEO do NCBI, com exceção dos trabalhos de Gupta e colegas (2014) ²⁴, cujos dados estavam depositados no site do laboratório do grupo, e do trabalho de Garbett e colegas (2008) ⁶¹, cujos dados foram obtidos a partir do material suplementar do trabalho de Gandal e colegas (2018) ³⁰.

Individualmente, para todos estes trabalhos aplicamos a normalização quantílica para compararmos a expressão dos genes entre as amostras de um mesmo estudo, usando assim um pipeline padronizado. Nessa etapa os pacotes utilizados dependeram do tipo de plataforma utilizada por estudo, por exemplo, para dados cuja plataforma de *microarray* era da Illumina, a função *normalizeBetweenArrays* do pacote *limma* foi utilizada ⁶², enquanto que, para estudos cuja plataforma era da *affymetrix* as funções *rma* e/ou *justRMA* do pacote *affy* foram as escolhidas ⁶³.

Mesmo para os dados já normalizados pelo método quantílico pelos seus autores originais, a normalização aplicada por nós não alterou os dados de expressão gênica, conforme é possível compreender com a figura 2 que esquematiza o algoritmo que constitui esse método de normalização.

FIGURA 2: Esquema representando as etapas do processo de normalização quantílica aplicada aos dados de *microarray*. Figura adaptada do curso de Harvard PH525X - Data Analysis for Genomics ministrado pelo professor Rafael Irizarry

Na etapa seguinte, a anotação das *probes* para identificar os genes para as quais essas *probes* mapeiam foi feita. O pacote *biomaRt* ⁶⁴ foi utilizado e informações como nome do gene, código *ensembl* e posição cromossômica no genoma humano foram retidas. Por fim, uma vez que em uma plataforma de *microarray* diferentes *probes* mapeiam um mesmo gene, as *probes* que mapearam um mesmo gene foram sumarizadas pela sua média.

Para os trabalhos cujos dados de expressão foram obtidos através de plataformas de *RNA-seq*, todos os dados utilizados em nosso estudo foram disponibilizados pelos seus autores originais já com os dados de contagem normalizados tanto por tamanho de genes como por tamanho de biblioteca através das funções *withinLaneNormalization* e *betweenLaneNormalization* do pacote *EDA-seq* com uma transformação logarítmica aplicada ⁶⁵.

Neste projeto a análise de redes de co-expressão foi feita com o pacote CEMiTool ⁶⁶ separadamente para cada um dos estudos selecionados. As redes construídas foram redes ponderadas do tipo assinadas (*weighted signed networks*), ou seja, redes que consideram apenas as correlações positivas entre seus nós, e elas foram construídas com o número mínimo de genes por módulos variando de 30 a 40 genes entre os estudos. Dentre as possíveis análises que podem ser

realizadas, o pacote CEMiTool é capaz de identificar módulos de gene co-expressos de maneira automatizada, além de filtrar os genes que não apresentam uma expressão similar aos demais. Os metadados das amostras dos estudos de TEA idiopático contendo as informações clínicas dos pacientes a respeito da presença ou não de DI foram compiladas manualmente usando as informações disponíveis de todos os estudos selecionados para poder atribuir o status de DI para o maior número de amostras possíveis ^{23,24,61,67}. Com isso, conseguimos recuperar, por exemplo, o status da presença de DI para todas as amostras do estudo de Voineagu e colegas (2011) ²³.

A fim de tornar válido nosso pipeline de geração das redes de co-expressão, para os estudos que originalmente construíram estas redes, realizamos uma análise de sobreposição entre os módulos que obtivemos a partir desses estudos e os módulos originalmente descritos pelos autores, obtendo uma boa consistência (dados não reportados).

Análises de associação com as doenças originalmente estudadas e com as informações clínicas de deficiência intelectual

Uma vez gerados os módulos de co-expressão para cada estudo, realizamos a análise de associação da expressão desses módulos com o status da doença investigada em cada um dos estudos originais bem como com as informações clínicas a respeito da presença de DI nos estudos feitos com TEA idiopático. A análise foi feita utilizando-se as funções *cor*, *corPvalueStudent* e *labeledHeatmap* do pacote WGCNA ²² que permitem verificar se o *module eigengene*, que pode ser definido como o primeiro componente principal de um dado módulo, correlaciona-se significativamente com alguma variável biológica ou técnica das amostras utilizadas na construção desses módulos, como por exemplo, o *status* da doença, já que o valor do *module eigengene* pode ser considerado um representante dos perfis de expressão gênica em um módulo ²². Para os estudos de TEA idiopático, como foi possível obter as informações clínicas a respeito da presença ou ausência de DI entre os pacientes com TEA para a maioria das amostras dos estudos, conforme é apresentado na seção de resultados, além de avaliarmos a presença de

desregulações de expressão entre os módulos com base no *design* experimental originalmente feito pelos autores dos trabalhos, isto é pacientes versus controles, também analisamos a presença de desregulações em outros dois cenários: um cenário no qual temos a presença exclusiva de TEA entre as amostras dos pacientes versus os controles mais o restante dos pacientes que possuem DI associado à TEA (cenário somente TEA); e um segundo cenário no qual temos as amostras dos pacientes que possuem TEA acompanhado de DI versus os controles mais os pacientes que não possuem DI associado ao TEA (cenário TEA + DI), conforme é possível visualizar com a tabela 1. Com isso, dentre os estudos de TEA, podemos avaliar se há especificidade de desregulações de expressão a depender da comorbidade com DI ou não.

TABELA 1: Representação das condições experimentais dos estudos de TEA idiopático utilizadas para avaliar desregulação de expressão gênica em TEA e DI considerando as informações clínicas a respeito de DI entre as amostras dos pacientes com TEA.

Amostras	Design experimental original	TEA	TEA + DI
controle 1	0	0	0
paciente 1-TEA	1	1	0
paciente 2 - TEA + DI	1	0	1

Análises de sobreposição de módulos entre os estudos

Muitas estatísticas de validação de agrupamentos (*clusters*) propostas na literatura podem ser transformadas em estatísticas de preservação de módulos ⁶⁸. Estas técnicas estatísticas podem ser divididas em quatro grandes categorias: estatísticas de tabulação cruzada, densidade, separabilidade e estabilidade. A estatística de tabulação cruzada compara as atribuições de módulos (ou *clusters*) na rede de co-expressão de referência (da qual os módulos foram originalmente obtidos) e na rede na qual estamos interessados em saber se o módulo está

presente (rede teste). Com isso, ela requer que o procedimento de construção de redes seja aplicado tanto nos dados de referência quanto nos dados de teste. Já as categorias densidade, separabilidade e estabilidade são estatísticas propriamente relacionadas ao estudo de grafos. As métricas de densidade e separabilidade oferecem a vantagem de não se fazer necessário a construção de redes no conjunto de dados de teste. Geralmente, elas avaliam os módulos pela forma com a qual os nós e as arestas são semelhantes e/ou dissimilares em cada módulo. Por último, a estatística de preservação baseada em estabilidade de redes é mais relevante no contexto da comparação de procedimentos de agrupamento (i.e., de construção de redes), em vez de quantificar a preservação de *clusters* ⁶⁸.

Neste projeto o tipo de método de similaridade e/ou consenso de redes utilizado foi o baseado em estatísticas de tabulação cruzada implementado na função *mod_stat_overlaps* do pacote CEMiTool na versão *filter* ⁶⁹. 25 Este método oferece a vantagem de ser mais simples, já que trata-se da comparação das listas de genes que compõem os módulos dos diferentes estudos.

Na sessão de resultados, são apresentados os dados de consenso de redes com significância estatística corrigida a partir de 0.05 para todos os módulos de interesse, ou seja, aqueles encontrados desregulados em TEA, DI ou TEA+DI. No máximo, somente as duas sobreposições de maior significância entre cada estudo são reportadas nos resultados.

Análises de enriquecimento funcional

A análise de enriquecimento funcional é uma abordagem comumente utilizada para determinar se funções ou processos biológicos conhecidos estão sobre-representados, i.e. enriquecidos, em uma lista de genes obtidos experimentalmente, como por exemplo, uma lista de genes pertencentes a um módulo co-expressão. Segundo Wang e colegas (2012) ⁷⁰, autores do pacote *clusterProfile* utilizado em nossas análises, o p-valor associado a essa análise pode ser calculado através de uma distribuição hipergeométrica descrita pela seguinte fórmula:

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

onde N é o número total de genes expresso do estudo (universo), M é o número de genes dentro desse universo que são anotados para o conjunto de genes de interesse (como uma via ou processo biológico), n é o tamanho da lista de genes de interesse (e.g., um módulo de co-expressão) e k é o número de genes nessa lista que são anotados no conjunto de genes ⁷⁰ (via de sinalização de mTOR, por exemplo). Esse cálculo corresponde a uma versão unilateral do teste exato de Fisher e os p-valores gerados devem ser ajustados para comparações múltiplas ⁷⁰. Em nosso estudo essa correção é feita pelo método de *Bonferroni* e somente resultados com significância menores ou iguais a 0.05 são considerados em nosso estudo. Com relação aos bancos consultados na nossa análise, estes incluem o Gene Ontology ^{71,72}, que permite obter informações biológicas com respeito às funções moleculares, aos componentes celulares e aos processos biológicos com os quais os genes de nossos módulos estão envolvidos e Cell Marker, um extenso banco de dados focado principalmente em marcadores celulares para vários tipos de células em tecidos humanos ⁷³.

Sobreposição com o trabalho de Gandal e colegas (2018)

Um trabalho com o qual estamos interessados em saber se há sobreposições com os nossos módulos é o trabalho de Gandal e colegas (2018) ³⁰, descrito na seção de introdução. Este trabalho realizou uma análise com uma abordagem similar a nossa, porém com outros transtornos psiquiátricos que acometem os pacientes em fases mais adultas do desenvolvimento, sem levar em conta os transtornos relacionados à DI. A função *userListEnrichment* do pacote WGCNA ²² foi utilizada para verificar se há sobreposições entre os módulos desregulados encontrados em nosso estudo com os módulos de Gandal e colegas(2018).

Somente as duas maiores sobreposições com cada lista utilizada são reportadas nos resultados.

RESULTADOS

Estudos selecionados para a análise de redes de co-expressão

Com base no guideline proposto pelo grupo PRISMA e utilizando o pacote *rentrez*, foram feitas buscas nas bases de dados do PubMed e do GenBank entre os meses de julho a outubro de 2019. A figura 3 resume o número de trabalhos de transcriptômica feitos em cérebro *post-mortem* de TEA e DI que foram encontrados em cada uma das etapas da busca sistemática.

FIGURA 3: 741 trabalhos foram encontrados na primeira fase da revisão sistemática chamada de fase de identificação. A figura mostra o número de trabalhos de acordo com a inclusão ou não nas fases seguintes do processo de revisão sistemática, somando-se todos os trabalhos obtidos para todas as condições patológicas consideradas em nossa revisão. O número de trabalhos selecionados na fase final de inclusão considera os estudos encontrados em duplicata, uma vez que um mesmo estudo foi encontrado nas diferentes buscas que fizemos por doença. Essa situação ocorreu apenas para 1 estudo da fase de inclusão. Como ao longo do ano de 2020 um novo estudo de transcriptômica envolvendo amostras de córtex de pacientes com síndrome de Rett e controles foi publicado ⁷⁴ este foi o único trabalho incluído em nosso estudo cuja inclusão não se deu via a revisão sistemática realizada.

Na última fase da busca sistemática, após removermos os resultados duplicados, 8 estudos de transcriptômica feitos a partir de amostras de cérebro *post-mortem* foram selecionados para a análise de construção de redes de co-expressão. Dentre esses 8 estudos, 4 são de autismo idiopático, 2 de síndrome de Rett e 2 de síndrome de Down conforme a descrição dos estudos a seguir.

Voineagu et. al. (2011) - GSE28521

No estudo de Voineagu e colegas (2011) ⁴² envolvendo 19 casos de autistas e 17 controles, 58 amostras de córtex pré-frontal e do giro superior temporal (29 autistas e 29 controles) além de 21 amostras cerebelares (11 autistas e 10 controles) foram utilizadas em análises de expressão diferencial, as quais revelaram a presença de 444 genes diferencialmente expressos entre os córtices dos pacientes versus os dos controles e apenas 2 genes diferencialmente expressos entre os cerebelos dos autistas versus os dos controles. Diante desses resultados os autores refinaram as análises para os dados de córtex cerebral e verificaram que entre os genes cuja expressão estava diminuída nos pacientes em relação aos controles as principais ontologias biológicas enriquecidas eram relacionadas às funções sinápticas, enquanto que, entre os genes cuja expressão encontrava-se aumentada estes relacionavam-se às respostas imune e inflamatória. Ao construir redes de co-expressão gênica usando somente dados de autistas ou dos controles, os autores verificaram que há uma diminuição da diferença de expressão gênica entre as regiões corticais dos cérebros dos autistas em comparação aos dos controles, sugerindo uma atenuação dos processos de diferenciação cortical, típica do desenvolvimento normal, nos cérebros dos autistas. Segundo Voineagu e colegas (2011), tal evidência é suportada por achados anatômicos independentes, indicando que ela poderia ser um dos motores da patofisiologia da doença. Ao construir redes utilizando dados tanto dos pacientes quanto dos controles, os autores encontraram a presença de 2 módulos associados ao TEA: o módulo M12 com a expressão gênica diminuída em relação aos controles e enriquecido para marcadores neuronais cujas funções são associadas às funções sinápticas e o

módulo M16 com expressão gênica aumentada com relação aos controles e enriquecidos para marcadores de astrócitos e de glia ativada, além de funções biológicas relacionadas aos processos imunológicos e inflamatórios.

Gupta et. al (2014)

Usando 104 amostras de áreas como córtex frontal, córtex occipital e córtex pré-frontal, Gupta e colegas (2014) ²⁴ sequenciaram o transcriptoma de 57 amostras de controles (40 indivíduos únicos) e 47 amostras de autistas idiopáticos (32 indivíduos únicos) no estudo de maior número amostral entre todos os 8 estudos analisados neste trabalho.

Ao construir redes de co-expressão usando o pacote WGCNA os autores encontraram 12 módulos, sendo 3 deles diferencialmente expressos entre controles e pacientes: mod1, mod5 e mod6. O módulo 5 apresentou a maior significância de associação ao TEA e seus genes tinham a expressão aumentada nesta condição com relação aos controles. Seu enriquecimento funcional apontou aumento da presença de genes marcadores gliais, em particular, de células microgliais em estado M2 e da via de interferon do tipo 1. Além disso, juntamente com o módulo 7, o módulo 5 apresentou sobreposição significativa com o módulo M16 do estudo de Voineagu e colegas (2011) que também foi encontrado como associado ao TEA e enriquecido para marcadores de células gliais. Os módulos 1 e 6 são módulos enriquecidos para marcadores neuronais e funções como transmissão sináptica, sendo que o mod1 mostrou expressão diminuída e o mod6 expressão aumentada entre os autistas. Juntamente com o módulo 2 (não associado ao TEA) os módulos 1 e 6 possuem sobreposição com o módulo M12 do estudo de Voineagu et. al (2011) que é associado ao TEA e cuja expressão está diminuída nos autistas além de ser também enriquecido para marcadores neuronais e funções sinápticas.

Chow et. al (2011) - GSE28475

Publicado em 2011, o estudo de Chow e colegas ⁶⁷ comparou a performance de dois métodos de captura de RNA mensageiro feitos por plataformas de detecção de expressão de RNA a partir de amostras artificialmente degradadas e a partir de

amostras de cérebro *post-mortem* congeladas versus fixadas em formalina de pacientes autistas e controles, respectivamente.

A plataforma de avaliação de RNA baseada no priming aleatório na fase de síntese de cDNA durante o preparo de biblioteca para o sequenciamento (DASL) é apresentada pelo autores como uma alternativa à metodologia tradicional, baseada em primers que capturam RNA mensageiros pela cauda polyA, para o sequenciamento de RNA e, particularmente, de RNA parcialmente degradado como é o caso de amostras fixadas em formalina. Segundo os resultados dos autores, o método DASL mostrou-se mais confiável em relação ao método tradicional na detecção de expressão gênica feita a partir de amostras de cérebro *post-mortem* com RNA artificialmente degradadas em comparação com suas respectivas amostras intactas. Ainda com relação às vantagens do método DASL, este mostrou-se superior na detecção da expressão gênica de amostras oriundas de tecido cerebral congelado em comparação ao método de captura de cauda polyA. Por outro lado, o mesmo resultado não foi observado entre os dados obtidos a partir de amostras fixadas em formalina. Em nosso estudo, somente amostras congeladas de cérebro *post-mortem* de pacientes e controles avaliadas pela técnica de DASL foram utilizadas.

Garbett et. al (2008)

No estudo de Garbett e colegas (2008) ⁶¹, o mais antigo dos trabalhos envolvendo autismo idiopático, 12 amostras oriundas do giro superior temporal de 6 pacientes e 6 controles tiveram o seu transcriptoma sequenciados pela plataforma de *microarray* da Affymetrix Human Genome 133 plus 2. Com a limitação das ferramentas e das análises presentes na época de realização do estudo, os autores realizaram uma análise de expressão diferencial baseando-se na diferença das médias dos logaritmos das intensidades dos sinais de expressão entre as amostras dos pacientes versus os controles (GC-RMA normalized expression levels) validando seus achados com diferentes estratégias de testes estatísticos. Com isso, dos 221 genes reportados como diferencialmente expressos pelos autores, 186 tinham expressão aumentada entre os autistas em comparação aos controles, enquanto

que 35 tinham expressão diminuída. Ao realizar análises de enriquecimento funcional baseadas em buscas manuais na literatura disponível no NCBI, os autores encontraram entre os genes com expressão aumentada um enriquecimento para vias relacionadas ao sistema imune e transcritos responsivos às citocinas ⁶¹ enquanto que, entre os genes com expressão diminuída, os autores encontraram enriquecimento de funções relacionadas à diferenciação neuronal, um resultado que, segundo os autores, é sugestivo para a ideia de há alterações nos processos de maturação e conectividade dos neurônios nos pacientes autistas.

Lin et al (2016) - GSE75303

A síndrome de Rett é uma doença do neurodesenvolvimento causada por mutações no gene *MECP2* localizado no cromossomo X. A maioria dessas mutações são do tipo *de novo* e são originadas durante a gametogênese paterna. A presença de mutações relacionada à Rett em hemizigose em fetos masculinos, em geral, não são toleradas, o que faz com que a doença ocorra quase que exclusivamente nos indivíduos do sexo feminino. Estima-se que a doença afeta cerca de 1 em cada 10000 meninas e seus primeiros sintomas se iniciam a partir dos 6 meses de vida. A síndrome é caracterizada pela perda de interesse pelo meio, regressão da habilidade de comunicação, presença de deficiência intelectual, perda de habilidades manuais intencionais e presença de movimentos estereotipados das mãos, além de epilepsia ^{28,75}.

No artigo de Lin e colegas (2016) ²⁸ o transcriptoma do córtex temporal e frontal de 3 pacientes e 4 controles foram sequenciados pelas técnicas de *microarray* e *RNA-seq* para investigar as mudanças globais de expressão gênica em pacientes com a síndrome de Rett. Como no sequenciamento com *RNA-seq* os autores tiveram de juntar as amostras de RNA dos córtices frontal e temporal de seus respectivos doadores devido à baixa quantidade de RNA disponível, somente os dados gerados a partir da técnica de *microarray*, feitos separadamente para cada amostra de cada tipo de córtex, foram usados em nosso estudo. A partir dos dados de *RNA-seq*, os autores encontraram 244 genes diferencialmente expressos entre os pacientes versus os controles. Desses, 151 possuíam expressão diminuída

enquanto que os outros 93 genes possuíam expressão aumentada. A partir desses resultados os autores fizeram duas comparações: a primeira foi uma validação desses resultados com os achados da técnica de *microarray* e a segunda foi uma sobreposição com a lista de genes diferencialmente expressos (DEGs) obtidas a partir de modelos animais para a doença.

Na primeira comparação os autores observaram que, 174 dos 244 genes tidos como DEGs pela técnica de *RNA-seq* foram detectados pelo *microarray* sendo que 73% possuíam a mesma direção de alteração detectada pelo *RNA-seq*. Já na segunda comparação, os autores viram que 13 genes diferencialmente expressos entre os modelos animais também foram detectados com o mesmo padrão de alteração nos dados obtidos a partir de humanos. Enquanto a análise de enriquecimento funcional entre os genes com expressão aumentada não revelou nenhuma via ou processo significativamente enriquecidos, entre os genes com expressão diminuída, revelou-se um enriquecimento de genes relacionados à resposta imune, em particular à via do sistema de complemento, o que segundo os autores pode ser uma alteração que contribui para a patofisiologia da síndrome de Rett.

Aldinger et al (2020) - GSE128380

Publicado como um artigo descritor de dados, o trabalho de Aldinger e colegas (2020) ⁷⁴ foi o único trabalho a ser selecionado pelo nosso estudo fora da análise de revisão sistemática que realizamos, uma vez que essa análise se estendeu até o ano de 2019 e o trabalho foi publicado em 2020. Nele, os autores sequenciaram o transcriptoma de regiões como córtex temporal e cingular de cérebro *post-mortem* de 4 pacientes do sexo feminino com síndrome de Rett e 4 controles com idades pareadas. Os autores não realizaram análises de rede de co-expressão, mas sim uma meta-análise de expressão gênica diferencial, no qual os dados originais foram analisados conjuntamente com os dados de Lin e colegas (2016) ²⁸ e Gogliotti e colegas (2018) ⁷⁶ que não disponibilizou seus dados em um formato de tabela de contagem e, por isso, não foi incluído em nosso estudo. Modelos estatísticos de variância foram ajustados separadamente para cada um dos estudos e, em seguida,

combinados em um modelo de efeitos aleatórios. A meta-análise dos três trabalhos revelou a presença de 1455 genes diferencialmente expressos entre pacientes e controles. Não foram reportadas informações sobre a quantidade de genes com aumento ou diminuição de expressão e nem resultados de análises de enriquecimento funcional. Para comparar os resultados com os dados originais, os autores calcularam a correlação de Spearman entre o Z-score obtido na meta-análise com o log da variação da expressão gênica (*log fold change*) dos genes encontrados em cada um dos estudos. Com essa análise, foi detectado uma correlação de 0.78 entre os dados gerados por Aldinger e colegas (2020) e os dados gerados por Lin e colegas (2016). Os dados de cerebelo do estudo de Gogliotti e colegas (2018) foram os que menos se correlacionaram com o restante das amostras dos outros estudos, todas oriundas de córtex cerebral.

Lockstone et al (2017) - GSE5390

Causada pela trissomia total ou parcial do cromossomo 21, a síndrome de Down é uma doença genética comum que afeta cerca de 1 a cada 800 nascimentos na população humana ²⁹. É caracterizada por déficits cognitivos e de linguagem juntamente com problemas neuromotores. Além disso, é comum os pacientes apresentarem face achatada e fissuras palpebrais oblíquas, além de maior propensão a problemas cardíacos congênitos, envelhecimento precoce e desenvolvimento de mal de Alzheimer ^{29,75}.

No trabalho de feito por Lockstone e colegas (2007) ²⁹, amostras de córtex frontal da porção dorsolateral de 7 pacientes e 8 controles foram sequenciadas para avaliar os efeitos da presença de genes trissômicos na expressão gênica global no cérebro desses pacientes. Ao realizar a análise de expressão diferencial, os autores identificaram 400 genes com expressão aumentada e 163 com expressão diminuída entre os pacientes versus os controles. Mais de 25% dos genes do cromossomo 21 eram genes diferencialmente expressos, ao passo que somente 4.4% dos genes de outros cromossomos também apresentaram alterações de expressão gênica. Devido à comorbidade com a doença de Alzheimer, genes relacionados a essa doença foram analisados com maior atenção. Foi o caso do gene precursor da proteína

amilóide (APP) responsável pelas placas beta amiloides típicas do processo de neurodegeneração da doença, mas que no estudo de Lockstone e colegas (2017) não apresentou expressão diferencial. Entretanto, é interessante ressaltar que outros genes associados ao processamento da proteína APP eram diferencialmente expressos. Com relação às vias e os processos biológicos enriquecidos pelos genes diferencialmente expressos, os autores identificaram processos como sistema imune, proliferação e ciclo celular, funcionamento de citoesqueleto e tráfico de vesículas, transporte de lipídeos e resposta ao estresse oxidativo, todos com aumento de expressão e processos de transmissão sináptica dentre os com expressão diminuída. Segundo os autores, os achados corroboram a teoria de que mudanças transcricionais secundárias ocorrem como consequência da trissomia do cromossomo 21.

Olmos-Serrano et al. (2016) - GSE59630

O estudo de maior número amostral selecionado em nosso trabalho foi o estudo de Olmos-Serrano e colegas (2016) ²⁷ que utilizou 116 amostras de cérebro *post-mortem* de homens e mulheres com síndrome de Down e controles. Essas amostras eram das áreas do córtex pré-frontal e cerebelar e refletiam diferentes períodos do desenvolvimento humano, desde períodos fetais até a idade adulta. Os objetivos dos autores eram avaliar as mudanças na expressão gênica global no cérebro de pacientes com essa síndrome e compará-las tanto entre as diferentes regiões cerebrais analisadas quanto entre as diferentes fases do desenvolvimento humano. Ao realizarem as análises de redes, os autores identificaram 60 módulos dos quais 9 estavam enriquecidos para funções e vias biológicas sabidamente afetadas nessa síndrome. Além disso, os autores identificaram um módulo M43 enriquecido para termos como regulação do potencial de ação e mielinização de axônios que ainda não haviam sido bem caracterizados em SD. Esse módulo apresentou expressão diminuída nos pacientes. Utilizando modelos animais, os autores validaram os achados *in silico* e observaram que as diferenças nos processos de mielinização e da atividade elétrica entre os animais controles e mutados se intensificaram ao longo do desenvolvimento refletindo as alterações no

padrão de expressão do módulo M43 entre pacientes e controles para as diferentes idades observadas.

Caracterização das amostras

Na tabela 2 é possível observar a descrição de algumas das principais características dos estudos incluídos na análise de construção de redes, incluindo-se o número de amostras que efetivamente usamos de cada um deles, plataforma de *microarray* ou sequenciamento e origem do tecido. O número de indivíduos controles e pacientes de cada estudo pode ser consultado na tabela 3 do apêndice I.

TABELA 2: Descrição dos 8 estudos selecionados para a análise de redes de coexpressão.

Estudo	Doença	Amostras		Plataforma de sequenciamento	Número GSE	Tecido/Tipo celular
		Controles	Casos			
Voineagu <i>et al.</i> (2011)	Autismo idiopático	29	23	Illumina HumanRef-8 v3.0	GSE28521	BA9, BA41
Gupta <i>et al.</i> (2014) *	Autismo idiopático	57	47	Illumina HiSeq 2000	-	BA10/BA19/BA44
Chow <i>et al.</i> (2011)	Autismo idiopático	17	14	Illumina HumanRef-8 v3.0	GSE28475	BA9/BA46
Garbett <i>et al.</i> (2008)**	Autismo idiopático	6	6	Affymetrix HG-U133 plus2	-	BA41/BA42
Lin <i>et al.</i> (2016)	Síndrome de Rett	6	6	Illumina HumanHT-12 V3.0	GSE75303	BA9/ BA21/ BA22/ or BA41/42
Aldinger <i>et al.</i> (2019)	Síndrome de Rett	8	8	Illumina HiSeq 2000	GSE128380	temporal and cingulate cortex
Lockstone <i>et al.</i> (2007)	Síndrome de Down	8	7	Affymetrix HG-U133A	GSE5390	BA9/B46

Olmos-Se rrano <i>et</i> <i>al.</i> (2016)	Síndrome de Down	15	14	Affymetrix HuEx-1_0-st	GSE59630	BA9/BA46 BA22 BA44/BA45 BA40 BA20
--	---------------------	----	----	---------------------------	----------	---

** Disponível em: <http://www.arkinglab.org/resources/> .

* Inicialmente disponibilizado em <http://mirnicslab.vanderbilt.edu/mirnicslab/> porém obtido a partir de https://github.com/mgandal/Shared-molecular-neuropathology-across-major-psychiatric-disorders-parallel-polygenic-overlap/tree/master/raw_data/Microarray/Garbett

Quando consideramos todos os 8 estudos, 271 amostras oriundas do córtex frontal e córtex temporal foram selecionadas para nosso trabalho (figura 4 A e figura 5 A). Em todos os estudos houve paridade entre as variáveis sexo, idade e intervalo *post-mortem*.

Observamos que há uma sobreposição entre as amostras utilizadas nos 4 estudos de autismo idiopático^{23,24,61,67} : das 199 amostras presentes nos 4 estudos, 100 são de pacientes e controles únicos e 33 estão presentes em pelo menos 2 estudos (figura 4 A). Apesar dessa sobreposição, a presença das informações clínicas dessas amostras (em particular, a associação do quadro com DI) variou bastante entre os estudos bem como o formato no qual elas foram disponibilizadas pelos autores originais. Desse modo, compilamos manualmente essas informações e, assim, conseguimos maximizar o número de amostras para as quais pudemos atribuir o status de DI comórbida em todos os estudos (tabela 3). Assim, com exceção do estudo de Garbett e colaboradores (2008), no qual não houve amostras com TEA isolado, foi possível conseguir um número suficiente de amostras para os dois grupos (TEA isolado e TEA+DI) para permitir a análise desses grupos separadamente nas análises de associação entre expressão das redes e status clínico (figura 4 B).

TABELA 3: Informações clínicas a respeito do número de amostras de pacientes com TEA que também possuem DI.

Estudo (nº GSE)	Controles	Casos (nº total)	Casos conforme presença de DI		
			TEA	TEA + DI	NA*
Voineagu <i>et</i>	29	23	8	15	0

<i>al.</i> (2011) (GSE28521)					
Gupta <i>et al.</i> (2014)	57	47	10	22	15
Chow <i>et al.</i> (2011) (GSE28475)	17	14	5	7	2
Garbett <i>et al.</i> (2008)	6	6	0	4	2

*Os casos em que a informação a respeito da presença de DI entre os autistas não estavam disponíveis foram incluídos na coluna 'NA'.

FIGURA 4: A) Diagrama de Venn representando a sobreposição das amostras únicas entre os estudos de autismo idiopático. B) Número de amostras cuja informação a respeito da presença de DI entre os pacientes com autismo idiopático era ausente, presente ou indisponível por estudo.

Geração das redes de co-expressão

Por uma questão computacional, antes de cada um dos estudos passarem para a etapa de construção de módulos, as tabelas de expressão gênica foram filtradas de modo que somente aproximadamente 8 mil genes com maior variância de expressão fossem mantidos em cada estudo (figura 5 B). Redes ponderadas do tipo assinadas (*weighted signed networks*) foram construídas com o número mínimo de genes por módulos variando de 30 a 40 entre os estudos. A maioria dos

parâmetros utilizados na construção das redes foram mantidos conforme o padrão definido pelo pacote CEMiTool, exceto aqueles disponíveis na tabela 4. O número de módulos gerados por estudo pode ser observado na figura 5 B.

TABELA 4: Parâmetros específicos utilizados na construção das redes de coexpressão por estudo.

Estudo	Parâmetros		
	beta	r2	min_ngen*
Voineagu <i>et al.</i> (2011) (TEA - GSE28521)	8	0.814	40
Gupta <i>et al.</i> (2014) (TEA)	10	0.803	40
Chow <i>et al.</i> (2011) (TEA - GSE28475)	12	0.818	40
Garbett <i>et al.</i> (2008) (TEA)	14	0.819	100
Lin <i>et al.</i> (2016) (Rett - GSE75303)	16	0.805	100
Aldinger <i>et al.</i> (2020) (Rett - GSE128380)	16	0.887	100
Lockstone <i>et al.</i> (2017) (Down - GSE5390)	26	0.833	30
Olmos-Serrano <i>et al.</i> (2016) (Down - GSE59630)	10	0.807	30

*min_ngen = número mínimo de genes por módulo. ** n_mods = número de módulos

FIGURA 5: A) Esquema representando as regiões cerebrais amostradas em cada um dos 8 estudos selecionados para a análise de construção de redes de co-expressão. B) Número de genes por estudo usados na construção das redes. No alto de cada barra há o respectivo número de módulos obtidos com a ferramenta CEMiTool. No total foram gerados 133 módulos entre todos os 8 estudos.

Análises de associação entre expressão do módulo e presença do transtorno

Para identificar quais módulos das redes geradas encontram-se desregulados nos pacientes, realizamos a análise de associação entre a expressão de cada um dos módulos dos estudos, sumarizada pelo valor *eigengene*, com o status das doenças originalmente estudadas nos trabalhos de TEA e DI. Para os trabalhos de autismo idiopático, também realizamos a análise de associação da expressão dos módulos com o status da presença de DI entre os pacientes (associação com TEA isolado ou com TEA+DI). Os resultados são apresentados na tabela 5 e dentre os 133 módulos gerados em nosso estudo a partir dos 8 trabalhos selecionados, 16 módulos possuem associação entre sua expressão e a presença do transtorno analisado originalmente nos 8 estudos selecionados. Além disso, dos 71 módulos obtidos entre estudos de TEA idiopático, 4 módulos apresentaram associação com o quadro clínico de TEA isolado (e não com o quadro clínico de TEA + DI), enquanto 7 módulos foram vistos significativamente associados ao grupo TEA+DI (conforme descrito na tabela 1 da seção de materiais e métodos).

TABELA 5: Associação com o status do transtorno e com a presença ou não de DI nos estudos de TEA idiopático.

Estudo (Transtorno)	Módulo	Associação transtorno (p-valor)*	Módulo	Associação somente TEA	Módulo	Associação TEA + quadro clínico DI (p-valor)*
GSE28521 (TEA)	M7 M10 M8 M4	-0.506 (0.001) 0.502 (0.002) 0.446 (0.010) -0.424 (0.019)	M8	0.424(0.019)	0	-
Gupta (TEA)	M3 M13 M6	0.471 (1E ⁻⁵) -0.347 (0.010) -0.319 (0.033)	M11	-0.397(1E ⁻³)	M3 M13 M6 M10 M18	0.445 (1E ⁻⁴) -0.394 (0.001) -0.394 (0.001) 0.391 (0.001) 0.311 (0.045)
GSE28475 (TEA)	0	-	M7 M11	0.551 (0.018) 0.509 (0.049)	M1	0.529 (0.031)
Garbett (TEA)	0	-			M7	0.826 (0.011)
GSE75303 (Rett)	M7	-0.830 (0.010)	-	-	-	-
GSE128380 (Rett)	0	-	-	-	-	-
GSE5390 (Down)	M1 M2 M6	0.868 (7E ⁻⁴) -0.844 (0.002) -0.823 (0.004)	-	-	-	-
GSE59630 (Down)	M1 M10 M7 M3 M8	0.602 (0.006) -0.548 (0.023) -0.749 (3E ⁻⁵) -0.641 (0.002) 0.694 (3E ⁻⁴)	-	-	-	-

*Corrigidos pelo método de Bonferroni para múltiplas comparações. As doenças originalmente investigadas foram colocadas junto aos nomes de seus respectivos estudos para facilitar a leitura dos resultados.

Observando os resultados de associação no estudo de Voineagu e colegas (2011), denominado como GSE28521 na tabela 5, podemos observar que os módulos M7 e M4 apresentaram expressão diminuída e o módulo M10 e M8

apresentaram expressão aumentada no grupo total de pacientes com TEA. O módulo M8 também se mostrou com expressão aumentada quando considerada a amostra de pacientes com TEA isolado, porém nenhum módulo foi encontrado associado ao grupo TEA+DI (tabela 5). Com relação aos módulos construídos a partir dos dados de Gupta e colegas (2014), dois módulos se mostraram com diminuição de expressão (M6 e M13) e um módulo com aumento de expressão (M3) no conjunto total de pacientes com TEA. Esses mesmos módulos também se mostraram associados ao grupo TEA+DI, inclusive mantendo a mesma direção de desregulação de expressão nos dois grupos. Além destes, também observamos a presença de módulos exclusivamente correlacionados à situação em que TEA está em associação com DI, os módulos M10 e M18, que apresentaram aumento de expressão. Entre os módulos associados a TEA isolado, o M11, cuja expressão apresentou diminuição, foi o único módulo encontrado.

Já nos estudos de Chow e colegas (2011), denominado GSE28475 na tabela 5, e Garbett e colegas (2008), não identificamos módulos significativamente associados ao grupo total de pacientes com TEA. Entretanto, no estudo de Chow e colegas (2011) identificamos o aumento de expressão dos módulos M7 e M11 nas amostras de TEA isolado e o aumento de expressão do módulo M1 nas amostras de TEA+DI. Já no estudo de Garbett e colegas (2008), apenas o M7 mostrou expressão diminuída entre as amostras de TEA + DI. Não foi possível realizar a análise de associação de módulos desse estudo com TEA isolado pela ausência de amostras com esse quadro clínico (vide tabela 3).

Entre os estudos nos quais o transtorno avaliado originalmente foi a síndrome de Rett, podemos observar que apenas o trabalho de Lin e colegas (2016) encontrou um módulo associado à síndrome, o M7, o qual possui forte diminuição de expressão entre os pacientes. Nos estudos de síndrome de Down, em ambos os trabalhos foi possível identificar módulos desregulados entre os pacientes. No trabalho de Lockstone e colegas (2017) foram detectados 3 módulos, um com aumento de expressão (M1) e dois com diminuição de expressão significativa nos pacientes (M2 e M6), enquanto que no trabalho de Olmos-Serrano e colegas (2016) foi possível observar 5 módulos desregulados. Desses, os módulos M10, M7 e M3

apresentaram diminuição de expressão ao passo que os módulos M1 e M8 tiveram expressão aumentada.

Análises de sobreposição de módulos dos diferentes estudos

A fim de entender se os módulos detectados como associados às patologias estudadas são semelhantes entre os diferentes trabalhos, mostrando consistência de associação para uma mesma doença, e se são iguais ou diferentes entre TEA e DI, realizamos uma análise de sobreposição de módulos. Para isso, utilizamos a função *mod_stat_overlap*, disponível na versão 'filter' do pacote CEMiTool ⁶⁹, o qual é capaz de realizar a análise de sobreposição de todos os pares possíveis de módulos que obtivemos juntamente com o resultado do teste exato de Fisher corrigido. A partir dessa análise, encontramos 513 sobreposições entre 115 módulos com significância estatística a partir de 0.05.

Quando filtramos as sobreposições que continham pelo menos 1 dos módulos associados com o status da doença (módulos apresentados na tabela 5), obtivemos 251 sobreposições significativas. A figura 6 ilustra as 100 sobreposições mais significativas dentre essas 251 em uma rede na qual as arestas indicam as sobreposições que cada um desses módulos possuem entre si e os nós indicam o aumento (vermelho) ou diminuição (verde) da expressão dos módulos desregulados.

FIGURA 6: Rede representando as 100 sobreposições mais significativas que os módulos associados com o status da doença (TEA total, TEA isolado, TEA+DI ou DI síndrômica) fazem com módulos dos 8 trabalhos analisados neste projeto.

A figura 7 mostra os resultados de sobreposição entre os módulos identificados como desregulados em um dado estudo com módulos detectados nos demais estudos. Para apresentação mais compreensível, quando o módulo desregulado de um estudo apresentou sobreposição apenas com um módulo desregulado de outro estudo e este foi o de maior significância, somente este é mostrado. No caso da sobreposição ter sido com mais de um módulo do outro estudo, apresenta-se na figura a sobreposição de maior significância e também todas as sobreposições com outros módulos desregulados. Por fim, quando houve sobreposições somente com módulos não desregulados, apenas o de maior significância é apresentado. A seguir, descrevemos estes resultados primeiramente para as comparações entre os estudos de TEA, em seguida, as comparações entre os estudos de DI e, finalmente, as comparações entre todos os estudos.

Especificidades e convergências entre os estudos de TEA idiopático

Entre os 4 trabalhos de autismo idiopático, apenas em dois (Voineagu e Gupta) detectamos módulos cuja expressão está associada ao TEA, considerando o conjunto dos pacientes avaliados originalmente pelos trabalhos selecionados. Estes módulos apresentam sobreposição entre si e apresentam-se desregulados na mesma direção: o módulo com diminuição de expressão M4 do estudo de Voineagu, apresenta sobreposição com M6 de Gupta; e os módulos com aumento de expressão M8 e M10 de Voineagu, apresentam sobreposição com o módulo M3 de Gupta. Curiosamente, os módulos citados com diminuição de expressão não apresentam sobreposição com módulos dos estudos de Garbett e Chow ; já os que apresentaram aumento de expressão, apresentam módulos correspondentes nos estudos de Garbett e Chow (ou seja, com sobreposição significativa), porém estes não apresentaram desregulação nestes dois estudos. Também não houve consistência nos resultados de diminuição de expressão do módulo M7 de Voineagu: embora tenham sido detectados módulos com sobreposição a esse nos outros estudos, estes últimos não se apresentaram desregulados. Finalmente, o módulo com diminuição de expressão em Gupta M13 não apresentou sobreposição significativa com nenhum outro módulo dos outros estudos (Figura 7).

Já dentre os módulos associados a TEA isolado (sem DI), detectamos sobreposição entre os módulos M8 de Voineagu e M7 e M11 de Chow, todos apresentando aumento de expressão neste grupo de pacientes. No entanto, os módulos de Gupta e Garbett que apresentam sobreposição com estes últimos, não se mostraram desregulados nos pacientes com TEA isolado. O módulo 11 de Gupta, com diminuição de expressão nos pacientes com TEA isolado, mostra sobreposição com outros módulos de todos os demais estudos, os quais por sua vez, não se mostram desregulados para esta condição. Dessa forma, consideramos que não podemos apontar um módulo que apresenta-se consistentemente desregulado em pacientes com TEA isolado.

Finalmente, na análise de associação com os pacientes TEA+DI, a única consistência de resultados se deu entre os módulos M3 de Gupta e M7 de Garbett, os quais apresentaram sobreposição e aumento de expressão em ambos estudos. No entanto, esses módulos se sobrepõem com outros módulos dos estudos de Voineagu e Chow, os quais por sua vez, não se mostraram desregulados. Com

relação aos demais módulos detectados como desregulados nesse grupo de pacientes (M6, M10, M13 e M18 de Gupta e M1 de Chow), mesmo para aqueles em que se detectou módulos correspondentes nos demais estudos, estes últimos não se mostraram desregulados (figura 7). Estes resultados apontam para uma consistência mais robusta de identificação de desregulação quando consideramos a amostra total de autistas.

Especificidades e convergências entre os estudos de DI síndrômica

Analisando os estudos feitos com pacientes com deficiência intelectual na sua forma síndrômica, dois conjuntos de módulos chamam a atenção. Primeiramente, os módulos M3 e M7 de Olmos-Serrano, que se sobrepõem ao módulo M6 de Lockstone, todos com diminuição de expressão em pacientes com Down. Estes módulos apresentaram sobreposição com o M1 de Aldinger e com dois módulos de Lin, M1 e M6, que em Rett não se mostraram desregulados. Dessa forma, estes dados parecem apontar para uma especificidade de desregulação desses módulos em pacientes com Down, em contraponto a outra forma de DI síndrômica, a síndrome de Rett. O segundo conjunto é representado pelos módulos M8 e M1 de Olmos-Serrano, com aumento de expressão em Síndrome de Down, que se sobrepõem respectivamente aos módulos M1 e M2 de Lockstone, que encontram-se com expressão aumentada e diminuída, respectivamente, para a mesma síndrome. Estes módulos também apresentam sobreposição com módulos dos estudos de Síndrome de Rett: M8 e M2 de Aldinger, não desregulados, e M7 de Lin, que encontra-se com expressão diminuída nos pacientes de Rett. Estes dados parecem sugerir que estes módulos apresentam uma consistência de desregulação em DI, com aumento de expressão em SD, embora a direção desta desregulação tenha sido diferente para o módulo M2 de Lockstone, e diminuição em Rett.

Especificidades e convergências entre os estudos de TEA idiopático e DI síndrômica

Quando comparamos os módulos detectados nos estudos de TEA com os módulos detectados nos estudos com formas síndrômicas de DI, com foco nos

módulos desregulados em qualquer um dos estudos, notamos dois blocos principais de sobreposição entre eles.

Um bloco abarca os módulos M4 e M7 de Voineagu, M6 de Gupta, M6 de Lockstone, M3, M7 e M10 de Olmos-Serrano, M1 de Aldinger e M4 e M6 de Lin. Interessante notar que estes módulos se apresentaram com expressão diminuída nos pacientes com TEA e pacientes com Síndrome de Down, mas não apresentaram desregulação nos pacientes com Síndrome de Rett conforme é possível observar na figura 8 na qual a as cores dos nós indicam a diminuição (verde) ou não alteração (branco) da expressão dos módulos.

FIGURA 8: Grafo representando as sobreposições dos módulos do bloco 1, majoritariamente com diminuição de expressão.

O outro bloco integra os módulos M8 e M10 de Voineagu, M3 de Gupta, M1, M7 e M11 de Chow, M7 de Garbett, M1 e M2 de Lockstone, M1 e M8 de Olmos-Serrano, M8 e M2 de Aldinger e M7 de Lin. Estes módulos, nos seus respectivos estudos, se encontram com aumento de expressão nos pacientes com TEA (note que para Garbett e Chow isso é verdadeiro apenas quando consideradas as condições TEA+DI ou TEA isolado; vide seções anteriores), mas com diminuição

de expressão em pacientes com DI do estudo de Lin. Curiosamente, no estudo com pacientes com Síndrome de Down, a desregulação dos módulos correspondentes M1 e M8 de Olmos-Serrano e M1 de Lockstone, se apresenta no sentido de super-expressão como o visto em autistas e contrário ao visto em pacientes com Rett do estudo de Lin. Por fim, os módulos do estudo de Aldinger com sobreposição a este grupamento, M2 e M8, não apresentaram desregulação, conforme é possível observar na figura 9 na qual os nós coloridos indicam o aumento (vermelho) ou diminuição (verde) da expressão dos módulos desregulados.

FIGURA 9: Grafo representando as sobreposições dos módulos do bloco 2 majoritariamente com aumento de expressão.

Estes resultados sugerem que o primeiro grupo de módulos pode estar com expressão diminuída especificamente em TEA e SD, mas não em RTT, enquanto o segundo grupo pode apresentar padrão de desregulação específico entre TEA e DI, onde em TEA e SD esse padrão é de aumento de expressão, apesar do sentido contrário de desregulação do módulo M2 de Lockstone, enquanto que em Rett é de diminuição de expressão.

Sobreposição de Módulos

FIGURA 7: Sobreposições entre cada um dos módulos associados aos transtornos avaliados em nosso trabalho (linhas à esquerda da tabela coloridos de acordo com o sentido da desregulação apresentada) com os módulos de outros estudos (células da tabela coloridas de acordo com o valor negativo do logaritmo do p-valor ajustado de sobreposição). Somente são mostradas as sobreposições entre módulos associados a algum transtorno e as sobreposições de maior significância por estudo (essas duas situações podem ou não

coincidirem, *i.e.*, a sobreposição de maior significância pode ser entre dois módulos associados aos transtornos avaliados em nosso trabalho). Sem asterisco: módulo associado somente ao transtorno avaliado pelo trabalho original. * associado exclusivamente à TEA. ** associado exclusivamente à TEA+DI. *** associado ao transtorno avaliado pelo trabalho original e à TEA. **** associado ao transtorno avaliado pelo trabalho original e à TEA+DI.

Análises de enriquecimento funcional

A fim de entender a constituição em termos de função dos módulos detectados como desregulados nos dois conjuntos de módulos que apresentaram padrões de sobreposição distintos entre si nesse estudo, realizamos uma análise de enriquecimento funcional para estes módulos, conforme resultados apresentados na tabela 6 e figura 10 onde os 5 termos com maior significância de enriquecimento para cada módulo são apresentados.

Dentre os módulos do bloco 1, módulos M4 e M7 de Voineagu, M6 de Gupta, M6 de Lockstone, M3, M7 e M10 de Olmos-Serrano, M1 de Aldinger e M4 e M6 de Lin, amplamente detectados com diminuição de expressão, com exceção dos trabalhos relacionados a Rett que não apresentaram alterações de expressão, os termos mais observados como enriquecidos neste conjunto de módulos foram termos relacionados a organização e modulação da transmissão sináptica. Porém, a partir do banco de dados Cell Marker ⁷³, somente o módulo M11 de Gupta apresentou enriquecimento significativo para marcadores celulares neuronais. Nesse caso, o termo encontrado era relacionado às células piramidais, que são os principais neurônios excitatórios do córtex humano e o módulo M3 de Olmos-Serrano que também apresentou enriquecimento para marcadores de tipos celulares relacionados à neurônios.

Já entre os módulos do bloco 2, majoritariamente com aumento de expressão em TEA e SD, mas com diminuição de expressão em um módulo de Rett, os termos mais frequentemente identificados como enriquecidos para genes deste conjunto de módulos relacionam-se à transcrição, regulação do sistema imune, ativação e proliferação de neutrófilos e leucócitos. Com relação ao enriquecimento de

marcadores celulares, foi possível observar o termo microglia na grande maioria desses módulos, além do termo astrócitos para alguns deles, como nos módulos de Olmos-Serrano. As exceções foram os módulos M10 e M18 de Gupta, o módulo M1 de Chow e o M1 de Olmos-Serrano. O primeiro não apresentou enriquecimento significativo para termos funcionais e nem de marcadores celulares significativos, enquanto que o restante dos módulos apresentaram enriquecimento para termos relacionado a sinapse, porém sem enriquecimento de marcadores celulares deste tipo de célula.

TABELA 6: Termos relacionados aos tipos celulares do banco Cell Marker ⁷³ enriquecidos para os genes dos módulos desregulados em TEA e DI sintromico.

Módulo	Termo	P-valor ajustado
GSE28521.M4 Voineagu <i>et al.</i> (2011)	-	-
GSE28521.M7 Voineagu <i>et al.</i> (2011)	Fetal gonad, Normal, Retinoid acid signaling-responsive fetal germ cell	0.03
Gupta.M6	-	-
Gupta.M11	Undefined, Normal, Pyramidal cell	0.03
Gupta.13	-	-
GSE5390.M6 Lockstone <i>et al.</i> (2017)	-	-
GSE59630.M3 Olmos-Serrano <i>et al.</i> (2016)	Brain, Normal, Neuron Embryonic prefrontal cortex, Normal, Oligodendrocyte progenitor cell	1.49e-04 1.94e-2
GSE59630.M7 Olmos-Serrano <i>et al.</i> (2016)	-	-
GSE59630.M10 Olmos-Serrano <i>et al.</i> (2016)	-	-
GSE28521.M8 Voineagu <i>et</i>	Embryonic prefrontal cortex, Normal, Microglial cell	8.4e-42

<i>al.</i> (2011)		
GSE28521.M10 Voineagu <i>et al.</i> (2011)	Liver, Hepatocellular Cancer, Exhausted CD8+ T cell Embryonic prefrontal cortex, Normal, Microglial cell	0.002 0.02
Gupta.M3	Embryonic prefrontal cortex, Normal, Astrocyte	4e-11
Gupta.M10	-	-
Gupta.M18	-	-
GSE28475.M1 Chow <i>et al.</i> (2011)	Undefined, Normal, B cell Pancreatic islet, Normal, Acinar cell Lung, Normal, Basal cell	2.76E-10 5.48E-10 5.81E-09
GSE28475.M7 Chow <i>et al.</i> (2011)	Large intestine, Normal, Paneth cell Embryonic prefrontal cortex, Normal, Microglial cell Blood, Normal, CD1C-CD141- dendritic cell	1.66E-22 2.79E-17 1.16E-09
GSE28475.M11 Chow <i>et al.</i> (2011)	Embryonic prefrontal cortex, Normal, Microglial cell Large intestine, Normal, Paneth cell Blood, Normal, CD1C-CD141- dendritic cell	4.47E-43 5.18E-32 2.14E-14
Garbett.M7	Brain, Normal, Astrocyte Fetal gonad, Normal, Granulosa cell Embryonic prefrontal cortex, Normal, Astrocyte	5.35E-05 2.49e-3 2.48e-3
GSE5390.M1 Lockstone <i>et al.</i> (2017)	Embryonic prefrontal cortex, Normal, Astrocyte Brain, Normal, Astrocyte Embryonic prefrontal cortex, Normal, Microglial cell	7.03e-34 1.73e-21 1.00e-09
GSE5390.M2 Lockstone <i>et al.</i> (2017)	Heart, Normal, Cardiomyocyte Undefined, Normal, T cell Large intestine, Normal, Enterocyte	5.83e-3 5.83e-3 5.83e-3
GSE59630.M1 Olmos-Serrano <i>et al.</i> (2016)	Pancreatic islet, Normal, Acinar cell Kidney, Renal Cell Carcinoma, Neutrophil Lung, Normal, Basal cell	5.56e-13 1.39e-08 6.42e-07
GSE59630.M8 Olmos-Serrano <i>et al.</i> (2016)	Embryonic prefrontal cortex, Normal, Microglial cell Large intestine, Normal, Paneth cell Umbilical cord blood, Normal, Granulocyte-monocyte progenitor	1.08e-24 5.87e-15 5.62e-06
GSE75303.M7 Lin <i>et al.</i> (2016)	Embryonic prefrontal cortex, Normal, Microglial cell Large intestine, Normal, Paneth cell Large intestine, Normal, MKI67+ progenitor cell	2.50e-81 8.88e-42 8.26e-17

FIGURA 10: Funções moleculares, componentes celulares e processos biológicos enriquecidos entre os módulos gerados em nosso estudo e encontrados como desregulados em casos com TEA, TEA isolado, TEA+DI e DI e DI síndrômica. Os retângulos tracejados em cinza que circundam os gráficos indicam blocos de módulos vistos sobrepostos em nossas análises. Os termos são oriundos do banco Gene Ontology⁷¹.

Validação com estudos robustos de transcriptômica

Como mencionado anteriormente, Gandal e colaboradores (2018)³⁰ utilizaram estratégia semelhante a aplicada neste trabalho para comparar a expressão gênica em tecido cerebral de diferentes transtornos psiquiátricos, incluindo TEA, esquizofrenia, transtorno bipolar, depressão e alcoolismo, tendo detectado perfis de desregulações de expressão de módulos associados a tipos celulares bem definidos. Assim, a fim de validarmos nossos resultados, realizamos uma análise de sobreposição entre os módulos do estudo de Gandal e os módulos desregulados de nosso estudo (figura 11).

Dentre os módulos do bloco 1 majoritariamente enriquecidos para termos relacionados à sinapse e a marcadores neuronais e amplamente detectados com diminuição de expressão, é possível observar que os módulos M7 e M4 de Voineagu, M6 e M11 de Gupta e M10 e M7 de Olmos-Serrano se sobrepuseram todos ao módulo CD1 de Gandal e colegas (2018), enquanto que os módulos M6 de Lockstone e M3 de Olmos-Serrano se sobrepuseram ao módulo CD5 de Gandal. Além disso, o M3 de Olmos-Serrano também se sobrepôs ao módulo CD13 de Gandal. Os módulos CD1, CD5 e CD13 do estudo de Gandal e colegas (2018) são todos módulos associados a neurônios e foram detectados com expressão reduzida entre os pacientes com TEA³⁰.

Já dentre os módulos do bloco 2, largamente para termos relacionados às células da glia e com expressão aumentada, com exceção do módulo M7 de Lin associado à Rett, os módulos M8 e M10 de Voineagu, M3 de Gupta, M7 e M11 de Chow, M7 de Garbett, M2 de Lockstone e M8 de Olmos-Serrano se sobrepuseram aos módulos CD4 e CD11 de Gandal e colegas (2018), enquanto que os módulos M1 de Lockstone e M1 de Olmos-Serrano se sobrepuseram exatamente aos mesmos módulos, CD2 e CD6 de Gandal. Além disso, o único módulo de SD do bloco 2 com diminuição de expressão, M2 de Lockstone, se sobrepôs ao módulo CD3 de Gandal, um módulo enriquecido de marcadores de oligodendrócitos. Por fim, o módulo M7 de Lin e colegas (2016) com expressão diminuída em Rett, se sobrepôs com uma grande significância com o módulo CD11 de Gandal e colegas

(2018) visto consistentemente sobreposto aos módulos dos estudos de TEA com aumento de expressão. Dentre as exceções, os módulos M10 de Gupta não se sobrepõem a nenhum módulo de Gandal, o M18, também de Gupta, se sobrepõe ao módulo CD1 de Gandal, visto com diminuição de expressão em TEA e o M1 de Chow que se sobrepõe somente a um módulo não associado a nenhum tipo celular específico, CD6. Os módulos CD2 e CD11 de Gandal são enriquecidos para marcadores de células da microglia, enquanto que o módulo CD4 é associado a astrócitos.

FIGURA 11: Sobreposição dos módulos desregulados gerados a partir dos trabalhos feitos com TEA idiopáticos e DI síndrômica com os módulos obtidos por Gandal e colegas (2018)

DISCUSSÃO E CONCLUSÃO

Por conta de sua base genética heterogênea, os transtornos do espectro do autismo e a deficiência intelectual, quando não associados à síndromes, ainda possuem um diagnóstico difícil de ser estabelecido com base apenas em informações oriundas do genoma. Todavia, quando analisamos níveis mais funcionais do fluxo da informação nos sistemas biológicos, como o nível transcricional, as alterações encontradas em células neuronais de pacientes autistas têm se mostrado mais homogêneas e consistentes entre diferentes estudos ^{42,43,45}. Embora outros trabalhos tenham explorado diferenças e semelhanças em termos de desregulação de módulos de co-expressão entre diferentes transtornos psiquiátricos, incluindo TEA ³⁰, nenhum estudo até o momento realizou essa comparação entre TEA e DI. Essa investigação é bastante importante dado que essas duas condições podem ocorrer concomitantemente em alguns pacientes. Assim, investigar se as alterações vistas em TEA são compartilhadas com DI, ou se há alterações da expressão gênica de grupos de genes exclusivamente relacionados a uma condição ou a outra é importante porque elas podem nos ajudar a identificar possíveis biomarcadores para essas condições que eventualmente permitiriam um diagnóstico precoce e o estabelecimento de intervenções terapêuticas mais adiantadas, melhorando assim, o prognóstico dos indivíduos afetados. Nesse sentido, o presente projeto teve por objetivo identificar se padrões de regulação de módulos de co-expressão podem ser usados para distinguir DI e TEA. Através da realização de uma busca sistemática para identificar trabalhos de expressão gênica global feitos a partir de tecido cerebral *post-mortem* de indivíduos com TEA isolado, TEA + DI e DI isolado, geramos redes de co-expressão gênica a partir dos dados de transcriptoma depositados em bancos públicos para cada estudo de cada transtorno selecionado em nossa revisão. Para cada módulo gerado em nossas análises, verificamos quais eram desregulados nas condições patológicas consideradas em nosso trabalho e buscamos encontrar consistências entre os achados de diferentes estudos para um mesmo transtorno. Além disso, também comparamos os padrões de desregulação de expressão de módulos entre as condições foco de nosso estudo, isto é, TEA

isolado, DI isolado e TEA+DI, buscando identificar se essas desregulações divergem ou se são iguais entre si.

Apesar dos transtornos do espectro autista e das formas sindrômicas de deficiência intelectual representarem formas bastante graves de acometimento do neurodesenvolvimento, é interessante notar que há poucos estudos de transcriptoma em amostras humanas envolvendo esses transtornos e síndromes. O maior número de estudos de expressão gênica global envolvendo cérebro *post-mortem* com número amostral suficiente para uma análise sistêmica, como a realizada em nosso trabalho, foi observado entre os estudos envolvendo TEA idiopático, enquanto que entre as síndromes envolvendo DI frequentemente os dados de transcriptoma eram de modelos animais, especialmente camundongo, ou quando de humanos, com um número amostral bastante reduzido. Somente trabalhos envolvendo síndrome de Down e Rett passaram em nossos filtros de seleção de estudos, muito provavelmente em razão do fato de que, embora estudos com expressão gênica global tenham se iniciado nos anos 2000 na era pós-genômica, a obtenção de amostras cerebrais para esses tipos de transtornos é ainda muito limitada e restrita a poucos bancos de tecidos no mundo como pudemos observar durante o levantamento de nossos estudos.

Nossa abordagem de biologia de sistemas identificou desregulação da expressão gênica de módulos construídos a partir de estudos de TEA idiopático e de formas sindrômicas de DI, que incluíram estudos feitos a partir de síndrome de Down e de Rett. Uma vez que no caso dos estudos de TEA conseguimos obter a informação de quais amostras possuíam somente TEA e TEA associado à DI, investigamos quais módulos possuíam desregulação associada ao desenho experimental original dos estudos, mas também aos casos em que TEA se apresenta isoladamente ou em conjunto com DI. Nesse sentido, quando avaliamos individualmente os 4 estudos de TEA foi possível encontrar módulos que se associaram de modo coerente às 3 condições avaliadas em nosso trabalho, isto é, módulos vistos como associados somente ao TEA não foram vistos como associados à presença conjunta de TEA + DI. Entretanto, ao fazermos a sobreposição dos módulos dos diferentes estudos, um dado módulo desregulado em TEA isolado, se sobrepôs significativamente com outro módulo que foi detectado

como desregulado simultaneamente à condição experimental de seu estudo original como também à TEA+DI. É caso do módulo M7 de Chown e colegas (2011) com o módulo M3 de Gupta e colegas (2014).

A maior consistência de desregulação de módulos é observada quando consideramos o conjunto amostral total de autistas, particularmente, dos estudos de Voineagu e colegas (2011) e Gupta e colegas (2014), nos quais os módulos M4 de Voineagu e M6 de Gupta possuem expressão diminuída e os módulos M8 e M10 de Voineagu e M3 de Gupta possuem expressão aumentada. Desse modo, mesmo com a recuperação das informações clínicas a respeito do *status* de DI para a maioria das amostras dos estudos de TEA, não foi possível detectar módulos consistentemente desregulados quando estratificamos os pacientes de TEA entre aqueles que possuem somente TEA daqueles que possuem TEA conjuntamente com DI.

Quando consideramos conjuntamente os módulos desregulados dos estudos de TEA idiopático com os módulos dos estudos de DI sindrômica, observamos a presença de dois cenários distintos. O primeiro é formado por um conjunto de módulos sobrepostos entre si, M4 e M7 de Voineagu, M6 de Gupta, M6 de Lockstone, M3 e M7 de Olmos-Serrano, o M1 de Aldinger e M4 e M6 de Lin, que apresentam um padrão de regulação onde em TEA e Down é diminuída, mas em Rett não é alterada. Neste bloco de módulos, os termos mais enriquecidos foram relacionados a organização e modulação da transmissão sináptica, além de marcadores de células neuronais. Quando sobrepostos com os módulos do trabalho de Gandal, esses módulos majoritariamente se sobrepuseram aos módulos CD1, CD5 e CD13, todos módulos associados a neurônios e detectados com expressão reduzida entre os pacientes com TEA ³⁰. É interessante notar que o envolvimento do módulo CD1 na patologia de TEA é ainda suportado por achados genéticos, uma vez que ele possui enriquecimento para variantes gênicas raras com mutações do tipo *de novo* não sinônimas de TEA e de esquizofrenia, além de variantes de número de cópias (CNV) tanto de TEA quanto de esquizofrenia ³⁰. O segundo cenário, composto pelo bloco de módulos sobrepostos M8 e M10 de Voineagu, M3 de Gupta, M1, M7 e M11 de Chow, M7 de Garbett, M1 e M2 de Lockstone, M1 e M8 de Olmos-Serrano, M8 e M2 de Aldinger e M7 de Lin indica um perfil de regulação que

em TEA e DI é aumentada, mas que em Rett é diminuída. Nesse cenário, os termos mais frequentemente identificados como enriquecidos para genes deste conjunto de módulos relacionam-se às diversas funções de regulação do sistema imune e às diversas células que compõem o grupo das células da glia, em especial, as microglias. As desregulações de expressão gênica detectadas em nosso estudo como associadas às microglias e presentes tanto em TEA quanto em DI reforçam o papel importante que essas células possuem na regulação da atividade sináptica, principalmente ao longo do desenvolvimento do sistema nervoso.

Não observamos nenhuma desregulação dos módulos obtidos a partir do estudo Aldinger e colegas (2020). Esse foi o único estudo a utilizar amostras do córtex cingular. Como perderíamos este estudo caso desconsiderássemos as amostras dessa região cerebral, em razão do baixo número amostral restante que impossibilitaria a construção de redes de co-expressão gênica, decidimos manter o estudo. Além disso, durante as análises de identificação de módulos desregulados, também verificamos se os dados desse e dos outros estudos possuíam desregulações de módulos associadas às regiões cerebrais amostradas, bem como com outras covariáveis, quando disponíveis, que pudessem afetar nossos resultados. Nessa etapa, a expressão dos módulos do estudo de Aldinger e colegas (2020) não se mostrou correlacionada a nenhuma dessas covariáveis, o que interpretamos como um sinal positivo de que os dados deste estudo não eram afetados significativamente por essas covariáveis. Entretanto, dado que a expressão gênica global de células do sistema nervoso é influenciada pela região cerebral considerada, não podemos descartar totalmente a hipótese de que esse fator possa ter contribuído para os resultados obtidos para este estudo.

Dada a dificuldade e escassez de se obter amostras de tecido cerebral, o uso de modelos alternativos que reproduzam os mesmos aspectos morfofuncionais de células neuronais possuem grande importância para as pesquisas sobre os transtornos relacionados ao sistema nervoso. Nesse sentido, um modelo que revelou ser efetivo na modelagem de células neurais em um período do desenvolvimento crítico para o autismo é o desenvolvido a partir de células progenitoras neurais e neurônios derivados de iPSC e explorado pelo nosso grupo recentemente em um trabalho publicado na revista *Molecular Psychiatry* ⁴². Nele,

identificamos nos pacientes autistas a desregulação do módulo MNeur1, o mesmo módulo de genes sinápticos identificado como desregulado nos trabalhos de transcriptoma com cérebros post-mortem de autistas. Quando analisamos as sobreposições dos módulos desregulados obtidos a partir dos estudos de DI com os módulos deste nosso estudo, verificamos que nosso módulo MNeur1 se sobrepõem à módulos de SD com diminuição de expressão, como o M3 de Olmos-Serrano, por exemplo, e à módulos de Rett não detectados com alteração de expressão, como o M4 de Lin, apresentando assim, o mesmo padrão de desregulação observado entre os módulos do bloco 1, onde em TEA e SD possuem diminuição de expressão, mas que em Rett não há alterações (dados não reportados).

É sabido pelo grupo que as diferenças observadas podem ser fruto de uma questão matemática com relação à definição dos parâmetros de construção de uma rede de co-expressão, afinal são pelo menos 30 parâmetros que podem ser ajustados durante a construção de uma rede e ao construí-las separadamente por estudo podemos dividir módulos de um estudo que eventualmente estejam concatenados em outro, e isso pode fazer com o que o módulo seja detectado como desregulado em um estudo, mas em outro não. Além disso, as variações oriundas das diferenças entre as plataformas de sequenciamento usadas pelos estudos bem como a presença de covariáveis biológicas, que apesar de serem amenizadas através da escolha de estudos que utilizaram amostras pareadas para variáveis como sexo, idade e intervalo *post-mortem*, ainda assim podem impactar nos resultados observados. Para contornar esses problemas, é possível analisar os dados de todos os estudos juntos como um único grande estudo em uma metanálise que aplique um rigoroso *pipeline* de controle dos efeitos de covariáveis, além do controle dos efeitos da sobreposição amostral presente entre os estudos e também de controle dos chamados efeitos de lote (*batch effects*) que surgem ao analisarmos dados de transcriptoma de diferentes estudos conjuntamente, empregando para isso, modelos estatísticos lineares de efeitos mistos como Gandal e colegas (2018) fizeram em seu estudo publicado na revista Science ³⁰.

Se outros padrões de desregulação transcricional podem ou não distinguir pacientes com TEA dos pacientes com TEA em comorbidade com DI é uma questão que continua em aberto. Nesse sentido, futuros trabalhos deverão se atentar ainda

mais para a questão do uso de métodos e análises mais reprodutíveis, incluindo a caracterização genética e clínica detalhada de suas amostras, e de tecnologias de avaliação do perfil de expressão gênica global que permitam lidar melhor com a presença de covariáveis biológicas que sabidamente afetam dados de expressão de um estudo de transcriptômica realizado tanto com *microarray* quanto com *RNA-seq*. Nesse sentido, o trabalho de Velmeshev e colegas (2019) ⁷⁷ apresenta uma abordagem promissora, pois é o primeiro trabalho a utilizar a técnica de *single cell RNA-sequencing* em amostras corticais de cérebro *post-mortem* de pacientes autistas sem deficiência intelectual para identificar alterações de expressão gênica associadas a tipos celulares específicos. Por exemplo, as desregulações transcricionais detectadas pelos autores afetam, especialmente, a expressão de genes relacionados à sinapse e ao neurodesenvolvimento de neurônios corticais das camadas 2 e 3 que participam de projeções neuronais córtico-corticais, mas não de projeções corticofugais ⁷⁷. O mais interessante neste trabalho é que, dado que metade dos pacientes autistas possuíam epilepsia esporádica, os autores foram capazes de separar as alterações de expressão gênica associadas às convulsões epiléticas daquelas associadas somente ao TEA, representando assim um grande potencial para o entendimento de como outras comorbidades, como a deficiência intelectual, interagem e afetam o quadro clínico de pacientes com TEA.

Apêndice I - Tabelas

TABELA 1: Prevalência de autismo nas síndromes genéticas abordadas em nosso projeto obtidas a partir dos trabalhos de Garg e colegas (2018) ² e Visser e colegas (2016) ¹⁰

Transtorno	Prevalência de TEA nas síndromes genéticas
Síndrome de Down	5.1%
Síndrome do X-Frágil	21 - 47%
Síndrome de Prader- Willi e Angelman	18- 38% Prader- Willi 42 - 100% Algelman
Síndrome de Rett	~61% (indivíduos do sexo feminino)
Síndrome de Williams	~50%
Autismo idiopático	-
Deficiência intelectual idiopática	-

TABELA 2: Termos de interesse utilizados para a construção das estruturas de busca (*queries*) implementadas no pacote *rentrez* para a realização das buscas sistemáticas dos trabalhos de transcriptômica de cérebro *post-mortem* para cada um dos transtornos.

Transtorno	Termos de interesse
Síndrome de Down	("down syndrome" ou "trisomy 21") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")
Síndrome do X-Frágil	("fragile X syndrome" ou "FXS" ou "FMRP syndrome") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")
Síndrome de Prader- Willi e Angelman	("prader-willi syndrome" ou "PWS" ou syndrome" ou "15q11.2-q13") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")

Síndrome de Rett	("rett syndrome" ou "rett" ou "Rett" ou "MECP2 syndrome") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")
Síndrome de Williams	("williams-beuren syndrome" ou "WBS ou "7q11.23") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")
Autismo idiopático	("autism spectrum disorders" ou "idiopathic autism" ou "autism") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")
Deficiência intelectual idiopática	("intellectual disability" ou "idiopathic intellectual disability") E ("post mortem brain" ou "post-mortem brain" ou "brain") E ("microarray" ou "RNAseq" ou "RNA-seq" ou "global gene expression" ou transcriptome) E ("publication date between 2000- 2019")

TABELA 3: Descrição do número de indivíduos controles e pacientes utilizados em cada um dos 8 estudos selecionados para a análise de redes de coexpressão.

Estudo	Doença	Indivíduos		Número GSE
		Controles	Pacientes	
Voineagu <i>et al.</i> (2011)	Autismo idiopático	17	19	GSE28521
Gupta <i>et al.</i> (2014) *	Autismo idiopático	40	32	-
Chow <i>et al.</i> (2011)	Autismo idiopático	29	28	GSE28475
Garbett <i>et al.</i> (2008)**	Autismo idiopático	6	6	-

Lin <i>et al.</i> (2016)	Síndrome de Rett	4	4	GSE75303
Aldinger <i>et al.</i> (2019)	Síndrome de Rett	4	4	GSE128380
Lockstone <i>et al.</i> (2007)	Síndrome de Down	9	10	GSE5390
Olmos-Se rrano <i>et</i> <i>al.</i> (2016)	Síndrome de Down	15	14	GSE59630

Apêndice II - Trabalhos Publicados

Molecular Psychiatry
<https://doi.org/10.1038/s41380-020-0669-9>

ARTICLE

Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder

K. Griesi-Oliveira^{1,2} · M. S. Fogo^{1,2} · B. G. G. Pinto¹ · A. Y. Alves¹ · A. M. Suzuki² · A. G. Morales² · S. Ezquina^{1,2} · O. J. Sosa³ · G. J. Sutton⁴ · D. Y. Sunaga-Franze⁵ · A. P. Bueno¹ · G. Seabra⁶ · L. Sardinha¹ · S. S. Costa² · C. Rosenberg² · E. C. Zach⁷ · A. L. Sertie¹ · D. Martins-de-Souza^{1,6,8,9} · E. M. Reis¹⁰ · I. Voineagu⁴ · M. R. Passos-Bueno²

Received: 27 August 2019 / Revised: 11 December 2019 / Accepted: 28 January 2020
© The Author(s) 2020. This article is published with open access

Abstract

Evaluation of expression profile in autism spectrum disorder (ASD) patients is an important approach to understand possible similar functional consequences that may underlie disease pathophysiology regardless of its genetic heterogeneity. Induced pluripotent stem cell (iPSC)-derived neuronal models have been useful to explore this question, but larger cohorts and different ASD endophenotypes still need to be investigated. Moreover, whether changes seen in this *in vitro* model reflect previous findings in ASD postmortem brains and how consistent they are across the studies remain underexplored questions. We examined the transcriptome of iPSC-derived neuronal cells from a normocephalic ASD cohort composed mostly of high-functioning individuals and from non-ASD individuals. ASD patients presented expression dysregulation of a module of co-expressed genes involved in protein synthesis in neuronal progenitor cells (NPC), and a module of genes related to synapse/neurotransmission and a module related to translation in neurons. Proteomic analysis in NPC revealed potential molecular links between the modules dysregulated in NPC and in neurons. Remarkably, the comparison of our results to a series of transcriptome studies revealed that the module related to synapse has been consistently found as upregulated in iPSC-derived neurons—which has an expression profile more closely related to fetal brain—while downregulated in postmortem brain tissue, indicating a reliable association of this network to the disease and suggesting that its dysregulation might occur in different directions across development in ASD individuals. Therefore, the expression pattern of this network might be used as biomarker for ASD and should be experimentally explored as a therapeutic target.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41380-020-0669-9>) contains supplementary material, which is available to authorized users.

✉ K. Griesi-Oliveira
karina.griesi@einstein.br

✉ M. R. Passos-Bueno
passos@ib.usp.br

¹ Hospital Israelita Albert Einstein, São Paulo, Brazil

² Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

³ Programa Interunidades de Pós-Graduação em Bioinformática, Universidade de São Paulo, São Paulo, Brazil

⁴ School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia

⁵ Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

⁶ Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), São Paulo, Brazil

⁷ Núcleo de Neurociências e Comportamento, Departamento de Psicologia Experimental, Instituto de Psicologia, Universidade de São Paulo, São Paulo, Brazil

⁸ Instituto Nacional de Biomarcadores em Neuropsiquiatria (INBION), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), São Paulo, Brazil

⁹ Experimental Medicine Research Cluster (EMC), University of Campinas, Campinas, Brazil

¹⁰ Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Published online: 14 February 2020

SPRINGER NATURE

ACE2 Expression Is Increased in the Lungs of Patients With Comorbidities Associated With Severe COVID-19

Bruna G. G. Pinto,^{1,2} Antonio E. R. Oliveira,^{1,2} Youvika Singh,¹ Leandro Jimenez,¹ Andre N. A. Gonçalves,¹ Rodrigo L. T. Ogava,¹ Rachel Creighton,² Jean Pierre Schatzmann Peron,^{3,4} and Helder I. Nakaya^{1,2,5}

¹Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, ²Department of Bioengineering, University of Washington, Seattle, Washington, USA, ³Department of Immunology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil, and ⁴Scientific Platform Pasteur, University of São Paulo, São Paulo, Brazil

Patients who died from COVID-19 often had comorbidities, such as hypertension, diabetes, and chronic obstructive lung disease. Although angiotensin-converting enzyme 2 (ACE2) is crucial for SARS-CoV-2 to bind and enter host cells, no study has systematically assessed the ACE2 expression in the lungs of patients with these diseases. Here, we analyzed over 700 lung transcriptome samples from patients with comorbidities associated with severe COVID-19 and found that ACE2 was highly expressed in these patients compared to control individuals. This finding suggests that patients with such comorbidities may have higher chances of developing severe COVID-19. Correlation and network analyses revealed many potential regulators of ACE2 in the human lung, including genes related to histone modifications, such as *HAT1*, *HDAC2*, and *KDM5B*. Our systems biology approach offers a possible explanation for increased COVID-19 severity in patients with certain comorbidities.

Keywords. angiotensin converting enzyme 2; COVID-19; SARS-CoV-2; KDM5B.

Recent studies of the epidemiological characteristics of coronavirus disease 2019 (COVID-19) have revealed that severe infection is more likely in people with an existing chronic medical condition. Two independent studies of infected populations in Wuhan, China found that approximately half the subjects with COVID-19 had an existing comorbidity [1, 2]. In a study of 1099 patients across mainland China, 38.7% of patients with comorbidities progressed to severe infection [3] and in a study of 52 inpatients in Wuhan, 67% of patients with comorbidities died [2]. The most common comorbidities reported in these studies were hypertension, diabetes, cerebrovascular disease, chronic obstructive lung disease, and coronary heart disease [1–3]. Other comorbidities such as carcinoma, chronic kidney disease, chronic liver disease, digestive system disease, and nervous system disease have also been reported in patients with COVID-19 [1, 2, 4]. A better understanding of the link between these conditions and COVID-19 infection is required to inform improved treatment and prevention interventions.

The molecular mechanism responsible for the increased disease severity in patients with these comorbidities is not fully understood, but previous studies suggest a role for angiotensin-converting enzyme 2 (ACE2) [5]. ACE2 is a membrane protein

required for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to bind and enter cells [6–8]. After binding, viral entry is facilitated by activation of the viral spike glycoprotein and cleavage of the C-terminal segment of ACE2 by proteases like TMPRSS2 and FURIN that are readily expressed in lung tissue [9–11]. ACE2 is only moderately expressed in healthy lung tissue compared to the heart, kidneys, and testes [12], but staining of lung tissue sections from adults with pulmonary hypertension has revealed increased ACE2 protein in the endothelium of pulmonary arteries compared to healthy controls [13]. A comprehensive analysis of single-cell RNA sequence (RNA-seq) datasets revealed that ACE2 was coexpressed with *TMPPSS2* within ileal absorptive enterocytes, nasal goblet secretory cells, and lung type 2 pneumocytes [14]. ACE2 upregulation has also been observed in animal models of liver fibrosis [15]. However, the reason for this upregulation remains unclear, and a link to other COVID-19 comorbidities has not been determined.

Here, we showed that expression of the gene encoding the ACE2 receptor in lung tissue is upregulated in diseases reported to be comorbidities associated with severe COVID-19. We also used systems biology approaches, including coexpression analysis, meta-analysis, and network analysis, to determine a potential cause of the ACE2 upregulation. From this analysis, we found that ACE2 expression could be regulated by enzymes that modify histones, including KDM5B. This identification of a common molecular mechanism of increased COVID-19 severity in patients with diverse comorbidities could direct the development of interventions to reduce the infection risk and disease severity in this population.

Received 30 March 2020; editorial decision 3 June 2020; accepted 5 June 2020; published online June 11, 2020.

*B. G. G. P. and A. E. R. O. contributed equally.

Correspondence: Helder I. Nakaya, PhD, Av. Prof. Lúcio Martins Rodrigues, 370, São Paulo, SP, Brazil, CEP: 05508-020 (hlnakaya@usp.br).

The Journal of Infectious Diseases® 2020;222:556–63

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com. DOI: 10.1093/infdis/jiaa332

Apêndice III - Projetos Adicionais

As seguintes atividades foram realizadas concomitantemente à realização deste projeto:

- Participação como co-organizadora do grupo R-ladies - capítulo São Paulo. O R-Ladies é uma organização mundial que promove a diversidade de gênero na comunidade da linguagem R. O R-Ladies São Paulo integra a organização R-Ladies Global, em São Paulo e realiza periodicamente meetups e mentorias para promover a linguagem computacional estatística R em um ambiente seguro e amigável para pessoas que se identificam com o gênero feminino, portanto, mulheres cis, mulheres trans, bem como pessoas não-binárias e queer ⁷⁸.
- Apresentação do 5o MeetUp do R-ladies São Paulo intitulado “Minicurso de Introdução ao R”.
- Apresentação do 6o MeetUp do R-ladies São Paulo intitulado “Oficina de R - Intermediário”.
- Apresentação do 10o MeetUp do R-ladies São Paulo intitulado “Cases de DataViz com R”.
- Participação como co-organizadora do Núcleo de Estudos em Bioinformática (NEB) no Hospital Israelita Albert Einstein no qual apresentou os seguintes encontros: “Introdução à Computação Para Bioinformática - Parte I” e “Introdução à Computação Para Bioinformática - Parte II”.
- Participação como co-organizadora do “Workshop de Transcriptômica” realizado no Hospital Israelita Albert Einstein em setembro de 2018 com a participação da colaboradora internacional Dr.a Irina Voineagu da University of New South Wales - Sidney, Austrália.

- Elaboração de pôster para o evento IV SER – Seminário Internacional de Estatística com R intitulado “Ensino de R na comunidade R-ladies - capítulo São Paulo”.
- Orientação do aluno de pré-iniciação científica Edson Kenso Takei no programa de pré-iniciação científica “Cientista Aprendiz” do colégio Dante Alighieri que atualmente encontra-se no Canadá onde conseguiu uma vaga no programa “The Foundation for Student Science and Technology”.
- Participação como monitora no Workshop de Visualização de Dados com o R na 14ª aMostra Estatística do IME-USP.
- Apresentação do 16º MeetUp da R-ladies - Curso Básico de R.
- Paralelamente à execução do projeto de mestrado desde março de 2020, faz exercício do cargo de analista de dados na agência de publicidade AlmapBBDO.
- Apresentação do 19º MeetUp das R-ladies - Contribuindo via GitHub no RStudio - RLadies SP no evento “Hacktoberfest”.
- Professora no curso de especialização em bioinformática da faculdade FACIS no primeiro semestre de 2021

Apêndice IV - Pacotes utilizados

Todos os pacotes e configurações de máquina usados para realizar as análises deste projeto encontram-se presentes nesta sessão. Os códigos criados para a realização de todas as análises deste projeto serão disponibilizados via *GitHub* tão logo seja submetido para publicação em revista especializada o manuscrito relativo às análises e resultados deste projeto.

— Session info —

```
setting value
version R version 3.6.2 (2019-12-12)
os      Ubuntu 18.04.3 LTS
system  x86_64, linux-gnu
ui      RStudio
language (EN)
collate en_US.UTF-8
ctype   pt_BR.UTF-8
tz      America/Sao_Paulo
date    2020-01-30
```

— Packages —

package	* version	date	lib	source
acepack	1.4.1	2016-10-29	[1]	CRAN (R 3.6.1)
AnnotationDbi	1.47.0	2019-05-02	[1]	Bioconductor
assertthat	0.2.1	2019-03-21	[1]	CRAN (R 3.6.1)
backports	1.1.4	2019-04-10	[1]	CRAN (R 3.6.1)
base64enc	0.1-3	2015-07-28	[1]	CRAN (R 3.6.1)
Biobase	2.45.1	2019-09-04	[1]	Bioconductor
BiocGenerics	0.31.6	2019-09-19	[1]	Bioconductor
BiocParallel	1.19.3	2019-09-26	[1]	Bioconductor
bit	1.1-14	2018-05-29	[1]	CRAN (R 3.6.1)
bit64	0.9-7	2017-05-08	[1]	CRAN (R 3.6.1)
bitops	1.0-6	2013-08-17	[1]	CRAN (R 3.6.1)
blob	1.2.0	2019-07-09	[1]	CRAN (R 3.6.1)
caTools	1.17.1.2	2019-03-06	[1]	CRAN (R 3.6.1)
CEMiTool	1.9.2	2019-05-23	[1]	Bioconductor
checkmate	1.9.4	2019-07-04	[1]	CRAN (R 3.6.1)
cli	1.1.0	2019-03-19	[1]	CRAN (R 3.6.1)
cluster	2.1.0	2019-06-19	[4]	CRAN (R 3.6.1)
clusterProfiler	3.13.0	2019-05-02	[1]	Bioconductor
coda	0.19-3	2019-07-05	[1]	CRAN (R 3.6.1)
codetools	0.2-16	2018-12-24	[4]	CRAN (R 3.6.1)

colorspace	1.4-1	2019-03-18	[1]	CRAN (R 3.6.1)
cowplot	1.0.0	2019-07-11	[1]	CRAN (R 3.6.1)
crayon	1.3.4	2017-09-16	[1]	CRAN (R 3.6.1)
data.table	1.12.2	2019-04-07	[1]	CRAN (R 3.6.1)
DBI	1.0.0	2018-05-02	[1]	CRAN (R 3.6.1)
DEoptimR	1.0-8	2016-11-19	[1]	CRAN (R 3.6.1)
digest	0.6.21	2019-09-20	[1]	CRAN (R 3.6.1)
DO.db	2.9	2019-08-08	[1]	Bioconductor
doParallel	1.0.15	2019-08-02	[1]	CRAN (R 3.6.1)
DOSE	3.11.2	2019-07-31	[1]	Bioconductor
dplyr	0.8.3	2019-07-04	[1]	CRAN (R 3.6.1)
DT	0.8	2019-08-07	[1]	CRAN (R 3.6.1)
dynamicTreeCut	1.63-1	2016-03-11	[1]	CRAN (R 3.6.1)
enrichplot	1.5.0	2019-05-02	[1]	Bioconductor
europepmc	0.3	2018-04-20	[1]	CRAN (R 3.6.1)
evaluate	0.14	2019-05-28	[1]	CRAN (R 3.6.1)
farver	1.1.0	2018-11-20	[1]	CRAN (R 3.6.1)
fastcluster	1.1.25	2018-06-07	[1]	CRAN (R 3.6.1)
fastmatch	1.1-0	2017-01-28	[1]	CRAN (R 3.6.1)
ff	2.2-14	2018-05-15	[1]	CRAN (R 3.6.1)
ffbase	0.12.7	2018-11-06	[1]	CRAN (R 3.6.1)
fgsea	1.11.0	2019-05-02	[1]	Bioconductor
fit.models	0.5-14	2017-04-06	[1]	CRAN (R 3.6.1)
foreach	1.4.7	2019-07-27	[1]	CRAN (R 3.6.1)
foreign	0.8-74	2019-12-26	[4]	CRAN (R 3.6.2)
Formula	1.2-3	2018-05-03	[1]	CRAN (R 3.6.1)
gdata	2.18.0	2017-06-06	[1]	CRAN (R 3.6.1)
GeneOverlap	1.21.0	2019-05-02	[1]	Bioconductor
ggdendro	0.1-20	2016-04-27	[1]	CRAN (R 3.6.1)
ggforce	0.3.0	2019-08-09	[1]	CRAN (R 3.6.1)
ggplot2	3.2.1	2019-08-10	[1]	CRAN (R 3.6.1)
ggplotify	0.0.4	2019-08-06	[1]	CRAN (R 3.6.1)
ggpmisc	0.3.1	2019-04-02	[1]	CRAN (R 3.6.1)
ggraph	1.0.2	2018-07-07	[1]	CRAN (R 3.6.1)
ggrepel	0.8.1	2019-05-07	[1]	CRAN (R 3.6.1)
ggridges	0.5.1	2018-09-27	[1]	CRAN (R 3.6.1)
ggthemes	4.2.0	2019-05-13	[1]	CRAN (R 3.6.1)
ggthemr	1.1.0	2019-10-07	[1]	Github (cttobin/ggthemr@10eb50a)
glue	1.3.1	2019-03-12	[1]	CRAN (R 3.6.1)
GO.db	3.8.2	2019-08-08	[1]	Bioconductor
GOSemSim	2.11.0	2019-05-02	[1]	Bioconductor
gplots	3.0.1.1	2019-01-27	[1]	CRAN (R 3.6.1)
graph	1.63.0	2019-05-02	[1]	Bioconductor
gRbase	1.8-4.5	2019-09-28	[1]	CRAN (R 3.6.1)
gridExtra	2.3	2017-09-09	[1]	CRAN (R 3.6.1)
gridGraphics	0.4-1	2019-05-20	[1]	CRAN (R 3.6.1)

gtable	0.3.0	2019-03-25	[1]	CRAN (R 3.6.1)
gtools	3.8.1	2018-06-26	[1]	CRAN (R 3.6.1)
Hmisc	4.2-0	2019-01-26	[1]	CRAN (R 3.6.1)
hms	0.5.1	2019-08-23	[1]	CRAN (R 3.6.1)
htmlTable	1.13.1	2019-01-07	[1]	CRAN (R 3.6.1)
htmltools	0.3.6	2017-04-28	[1]	CRAN (R 3.6.1)
htmlwidgets	1.3	2018-09-30	[1]	CRAN (R 3.6.1)
httr	1.4.1	2019-08-05	[1]	CRAN (R 3.6.1)
igraph	1.2.4.1	2019-04-22	[1]	CRAN (R 3.6.1)
impute	1.59.0	2019-05-02	[1]	Bioconductor
intergraph	2.0-2	2016-12-05	[1]	CRAN (R 3.6.1)
IRanges	2.19.16	2019-09-13	[1]	Bioconductor
iterators	1.0.12	2019-07-26	[1]	CRAN (R 3.6.1)
jsonlite	1.6	2018-12-07	[1]	CRAN (R 3.6.1)
KernSmooth	2.23-16	2019-10-15	[4]	CRAN (R 3.6.1)
knitr	1.25	2019-09-18	[1]	CRAN (R 3.6.1)
lattice	0.20-38	2018-11-04	[4]	CRAN (R 3.6.1)
latticeExtra	0.6-28	2016-02-09	[1]	CRAN (R 3.6.1)
lazyeval	0.2.2	2019-03-15	[1]	CRAN (R 3.6.1)
limma	3.41.16	2019-09-09	[1]	Bioconductor
magrittr	1.5	2014-11-22	[1]	CRAN (R 3.6.1)
MASS	7.3-51.5	2019-12-20	[4]	CRAN (R 3.6.2)
Matrix	1.2-18	2019-11-27	[4]	CRAN (R 3.6.1)
matrixStats	0.55.0	2019-09-07	[1]	CRAN (R 3.6.1)
memoise	1.1.0	2017-04-21	[1]	CRAN (R 3.6.1)
munsell	0.5.0	2018-06-12	[1]	CRAN (R 3.6.1)
mvtnorm	1.0-11	2019-06-19	[1]	CRAN (R 3.6.1)
network	1.15	2019-04-02	[1]	CRAN (R 3.6.1)
nnet	7.3-12	2016-02-02	[4]	CRAN (R 3.6.1)
pcaPP	1.9-73	2018-01-14	[1]	CRAN (R 3.6.1)
pillar	1.4.2	2019-06-29	[1]	CRAN (R 3.6.1)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN (R 3.6.1)
plyr	1.8.4	2016-06-08	[1]	CRAN (R 3.6.1)
polyclip	1.10-0	2019-03-14	[1]	CRAN (R 3.6.1)
pracma	2.2.5	2019-04-09	[1]	CRAN (R 3.6.1)
preprocessCore	1.47.1	2019-05-13	[1]	Bioconductor
prettyunits	1.0.2	2015-07-13	[1]	CRAN (R 3.6.1)
progress	1.2.2	2019-05-16	[1]	CRAN (R 3.6.1)
purrr	0.3.2	2019-03-15	[1]	CRAN (R 3.6.1)
qvalue	2.17.0	2019-05-02	[1]	Bioconductor
R6	2.4.0	2019-02-14	[1]	CRAN (R 3.6.1)
RColorBrewer	1.1-2	2014-12-07	[1]	CRAN (R 3.6.1)
Rcpp	1.0.2	2019-07-25	[1]	CRAN (R 3.6.1)
reshape2	1.4.3	2017-12-11	[1]	CRAN (R 3.6.1)
Rgraphviz	2.29.0	2019-05-02	[1]	Bioconductor
rlang	0.4.0	2019-06-25	[1]	CRAN (R 3.6.1)

rmarkdown	1.14	2019-07-12	[1]	CRAN (R 3.6.1)
robust	0.4-18.1	2019-07-02	[1]	CRAN (R 3.6.1)
robustbase	0.93-5	2019-05-12	[1]	CRAN (R 3.6.1)
rpart	4.1-15	2019-04-12	[4]	CRAN (R 3.6.1)
rrcov	1.4-7	2018-11-15	[1]	CRAN (R 3.6.1)
RSQLite	2.1.2	2019-07-24	[1]	CRAN (R 3.6.1)
rstudioapi	0.10	2019-03-19	[1]	CRAN (R 3.6.1)
rvcheck	0.1.3	2018-12-06	[1]	CRAN (R 3.6.1)
S4Vectors	0.23.23	2019-09-13	[1]	Bioconductor
scales	1.0.0	2018-08-09	[1]	CRAN (R 3.6.1)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN (R 3.6.1)
sna	2.4	2016-08-08	[1]	CRAN (R 3.6.1)
statnet.common	4.3.0	2019-06-02	[1]	CRAN (R 3.6.1)
stringi	1.4.3	2019-03-12	[1]	CRAN (R 3.6.1)
stringr	1.4.0	2019-02-10	[1]	CRAN (R 3.6.1)
survival	3.1-8	2019-12-03	[4]	CRAN (R 3.6.2)
tibble	2.1.3	2019-06-06	[1]	CRAN (R 3.6.1)
tidyr	0.8.3	2019-03-01	[1]	CRAN (R 3.6.1)
tidyselect	0.2.5	2018-10-11	[1]	CRAN (R 3.6.1)
triebeard	0.3.0	2016-08-04	[1]	CRAN (R 3.6.1)
tweenr	1.0.1	2018-12-14	[1]	CRAN (R 3.6.1)
UpSetR	1.4.0	2019-05-22	[1]	CRAN (R 3.6.1)
urltools	1.7.3	2019-04-14	[1]	CRAN (R 3.6.1)
vctrs	0.2.0	2019-07-05	[1]	CRAN (R 3.6.1)
viridis	0.5.1	2018-03-29	[1]	CRAN (R 3.6.1)
viridisLite	0.3.0	2018-02-01	[1]	CRAN (R 3.6.1)
WGCNA	1.68	2019-05-23	[1]	CRAN (R 3.6.1)
withr	2.1.2	2018-03-15	[1]	CRAN (R 3.6.1)
xfun	0.9	2019-08-21	[1]	CRAN (R 3.6.1)
xml2	1.2.2	2019-08-09	[1]	CRAN (R 3.6.1)
yaml	2.2.0	2018-07-25	[1]	CRAN (R 3.6.1)
zeallot	0.1.0	2018-01-28	[1]	CRAN (R 3.6.1)

[1] /home/bruna_garbes/R/x86_64-pc-linux-gnu-library/3.6

[2] /usr/local/lib/R/site-library

[3] /usr/lib/R/site-library

[4] /usr/lib/R/library

REFERÊNCIAS

1. Bourgeron, T. *From the genetic architecture to synaptic plasticity in autism spectrum disorder*. *Nat.Rev.Neurosci.* vol. 16 (2015).
2. Garg, S. & Green, J. Studying child development in genetic models of ASD. *Genet. Models Mol. Pathw. Underlying Autism Spectr. Disord.* **241**, 159–192.
3. Gokoolparsadh, A. Searching for convergent pathways in autism spectrum disorders: insights from human brain transcriptome studies. *Cell Mol Life Sci* **73**, 4517–4530 (2016).
4. Association, A. P. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®*. (American Psychiatric Pub, 2013).
5. Baio, J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites. *U. S.* **67**, 1–23 (2014).
6. Srivastava, A. K. & Schwartz, C. E. Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms. *Neurosci Biobehav Rev* **46**, 161–174 (2014).
7. Ronski, M., Sevcik, R. A., Barton-Hulsey, A. & Whitmore, A. S. Early Intervention and AAC: What a Difference 30 Years Makes. *Augment Altern Commun* **31**, 181–202 (2015).
8. Lai, D.-C. Predictors of effectiveness of early intervention on children with intellectual disability: a retrospective cohort study. *BMC Pediatr* **14**, 624 (2014).
9. Tonge, B. J., Bull, K., Brereton, A. & Wilson, R. A review of evidence-based early 39 intervention for behavioural problems in children with autism spectrum disorder: the core components of effective programs, child-focused interventions and comprehensive treatment models. *Curr Opin Psychiatry* **27**, 158–165 (2014).
10. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**, 9–18 (2016).
11. Wilkins, J. & Matson, J. L. A Comparison of Social Skills Profiles in Intellectually

- Disabled Adults With and Without ASD. *Behav Modif* **33**, 143–155 (2009).
12. Galli Carminati, G., Gerber, F., Baud, M. A. & Baud, O. Evaluating the effects of a structured program for adults with autism spectrum disorders and intellectual disabilities. *Res Autism Spectr Disord* **1**, 256–265 (2007).
 13. Ramaswami, G. & Geschwind, D. H. Genetics of autism spectrum disorder. *Neurogenetics Part I* **147**, 321–329.
 14. De Rubeis, S. & Buxbaum, J. D. Genetics and genomics of autism spectrum disorder: embracing complexity. *Hum Mol Genet* **24**, 24–31 (2015).
 15. Huguet, G., Benabou, M. & Bourgeron, T. *The Genetics of Autism Spectrum Disorders. in A Time for Metabolism and Hormones*. vol. 9 (Springer International Publishing, 2016).
 16. Harripaul, R., Noor, A., Ayub, M. & Vincent, J. B. The Use of Next-Generation Sequencing for Research and Diagnostics for Intellectual Disability. *Cold Spring Harb Perspect Med* **7**, 026864 (2017).
 17. Pinto, D. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
 18. Iossifov, I. De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* **74**, 285–299 (2012).
 19. Griesi-Oliveira, K. & Sertié, A. L. Autism spectrum disorders: an updated guide for genetic counseling. *Einstein* **15**, 233–238 (2017).
 20. Poot, M., Smagt, J. J., Brilstra, E. H. & Bourgeron, T. Disentangling the Myriad Genomics of Complex Disorders, Specifically Focusing on Autism, Epilepsy, and Schizophrenia. *Cytogenet Genome Res* **135**, 228–240 (2011).
 21. Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat Appl Genet Mol Biol* **4**, (2005).
 22. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network

- analysis. *BMC Bioinformatics* **9**, 1, (2008).
23. Voineagu, I. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
 24. Gupta, S. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* **5**, 1206 (2014).
 25. Parikshak, N. N. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
 26. Pelleri, M. C. *Integrated Quantitative Transcriptome Maps of Human Trisomy 21 Tissues and Cells*. (2018). doi:10.3389/fgene.2018.00125.
 27. Olmos-Serrano, J. L. Down Syndrome Developmental Brain Transcriptome Reveals Defective Oligodendrocyte Differentiation and Myelination. *Neuron* **89**, 1208–1222 (2016).
 28. Lin, P. Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q Genes in Rett syndrome. *BMC Genomics* **17**, 7951 (2016).
 29. Lockstone, H. E. Gene expression profiling in the adult Down syndrome brain. *Genomics* **90**, 647–660 (2007).
 30. Gandal, M. J. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
 31. Takahashi, K. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).
 32. Brennand, K. J. Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).
 33. Weick, J. P. Deficits in human trisomy 21 iPSCs and neurons. in *Proceedings of the National Academy of Sciences* vol. 110 9962–9967 (2013).
 34. Griesi-Oliveira, K. Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons. *Mol Psychiatry* **20**, 1350–1365 (2015).
 35. Machado, C. O. F. Collybistin binds and inhibits mTORC1 signaling: a potential novel

- mechanism contributing to intellectual disability and autism. *Eur J Hum Genet* **24**, 59–65 (2016).
36. Courchesne, E. Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology* **57**, 245–254 (2001).
 37. Chappleau, C. A. Dendritic spine pathologies in hippocampal pyramidal neurons from Rett syndrome brain and after expression of Rett-associated MECP2 mutations. *Neurobiol Dis* **35**, 219–233 (2009).
 38. Walsh, R. M. & Hochedlinger, K. Modeling Rett Syndrome with Stem Cells. *Cell* **143**, 499–500 (2010).
 39. Marchetto, M. C. N. A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells. *Cell* **143**, 527–539 (2010).
 40. Sweet, R. A., Henteleff, R. A., Zhang, W., Sampson, A. R. & Lewis, D. A. Reduced Dendritic Spine Density in Auditory Cortex of Subjects with Schizophrenia. *Neuropsychopharmacology* **34**, 374–389 (2009).
 41. Marchetto, M. C. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry* **22**, 820–835 (2017).
 42. Griesi-Oliveira, K. *et al.* Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. *Mol. Psychiatry* (2020) doi:10.1038/s41380-020-0669-9.
 43. Flora M. Vaccarino, J. M. FOXP1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).
 44. Schafer, S. T. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat Neurosci* **22**, 243–255 (2019).
 45. DeRosa, B. A. Convergent Pathways in Idiopathic Autism Revealed by Time Course Transcriptomic Analysis of Patient-Derived Neurons. *Sci Rep* **8**, 8423 (2018).
 46. Boland, M. J. Molecular analyses of neurogenic defects in a human pluripotent stem

- cell model of fragile X syndrome. *Brain* **343**, 357 (2017).
47. Gonzales, P. K. *Transcriptome analysis of genetically matched human induced pluripotent stem cells disomic or trisomic for chromosome 21*. (2018).
doi:10.1371/journal.pone.0194581.
 48. Ovchinnikov, D. A., Korn, O., Virshup, I., Wells, C. A. & Wolvetang, E. J. The Impact of APP on Alzheimer-like Pathogenesis and Gene Expression in Down Syndrome iPSC-Derived Neurons. *Stem Cell Rep.* **11**, 32–42 (2018).
 49. Landucci, E. iPSC-derived neurons profiling reveals GABAergic circuit disruption and acetylated α -tubulin defect which improves after iHDAC6 treatment in Rett syndrome. *Exp Cell Res* **368**, 225–235 (2018).
 50. Halevy, T., Czech, C. & Benvenisty, N. Molecular Mechanisms Regulating the Defects in Fragile X Syndrome Neurons Derived from Human Pluripotent Stem Cells. *Stem Cell Rep.* **4**, 37–46 (2015).
 51. Khattak, S. Human induced pluripotent stem cell derived neurons as a model for Williams-Beuren syndrome. *Mol Brain* **8**, 239 (2015).
 52. Adamo, A. 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet* **47**, 132–141 (2015).
 53. Letourneau, A. Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* **508**, 345–350 (2014).
 54. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* **16**, 441–458 (2015).
 55. Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M. & Ligterink, W. Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci* **7**, 4947 (2016).
 56. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based

- approach to human disease. *Nat Rev Genet* **12**, 56–68 (2011).
57. Prada-Medina, C. A. *Estudando a artrite idiopática juvenil mediante uma abordagem de medicina de redes*. (Faculdade de Ciências Farmacêuticas da Universidade de São Paulo, 2019).
 58. Lima, D. S. Long noncoding RNAs are involved in multiple immunological pathways in response to vaccination. *Proc Natl Acad Sci U A* **116**, 17121–17126 (2019).
 59. Liberati, A. *The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration*. (2009). doi:10.1371/journal.pmed.1000100.
 60. Winter, D., David & Winter, J. rentrez. An R package for the NCBI eUtils API. *R J.* **9**, 520 (2017).
 61. Garbett, K. Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol Dis* **30**, 303–311 (2008).
 62. Ritchie, M. E. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, 47–47 (2015).
 63. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix 45 GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
 64. Durinck, S. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
 65. LudwigGeistlinger [Ctb, D. R. [Aut, Cre, Cph], Sandrine Dudoit [Aut]. *EDASeq*. (Bioconductor, 2017). doi:10.18129/B9.BIOC.EDASEQ.
 66. Russo, P. S. T. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* **19**, 17973 (2018).
 67. Chow, M. L. Genome-wide expression assay comparison across frozen and fixed postmortem brain tissue samples. *BMC Genomics* **12**, 434 (2011).
 68. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. *Is My Network Module*

- Preserved and Reproducible?* (2011). doi:10.1371/journal.pcbi.1001057.
69. CSBL, R., Pedro S. T. CEMiTool. <https://github.com/csbl-usp/CEMiTool/tree/filter>.
 70. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
 71. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
 72. The Gene Ontology Consortium *et al.* The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
 73. Zhang, X. *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–D728 (2019).
 74. Aldinger, K. A. *et al.* Transcriptome data of temporal and cingulate cortex in the Rett syndrome brain. *Sci. Data* **7**, 192 (2020).
 75. Otto, P. A., Mingroni Netto, R. C. & Guimarães Otto, P. *Genética médica.* (2013).
 76. Gogliotti, R. G. *et al.* Total RNA Sequencing of Rett Syndrome Autopsy Samples Identifies the M₄ Muscarinic Receptor as a Novel Therapeutic Target. *J. Pharmacol. Exp. Ther.* **365**, 291–300 (2018).
 77. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685–689 (2019).
 78. Global, R. -ladies. *R-ladies Available at.*