Camily Erica de Freitas Rodrigues

**IDENTIFICAÇÃO DE MODELOS CELULARES DE FÁCIL OBTENÇÃO
COM PRESERVAÇÃO DE MÓDULOS DE CO-EXPRESSÃO
DESREGULADOS EM TRANSTORNO DO ESPECTRO AUTISTA E
DEFICIÊNCIA INTELECTUAL**

Dissertação apresentada à Faculdade Israelita de Ciências da Saúde Albert Einstein para obtenção do Título de Mestre em Ciências da Saúde.

São Paulo

2021

Camily Erica de Freitas Rodrigues

**IDENTIFICAÇÃO DE MODELOS CELULARES DE FÁCIL OBTENÇÃO
COM PRESERVAÇÃO DE MÓDULOS DE CO-EXPRESSÃO
DESREGULADOS EM TRANSTORNO DO ESPECTRO AUTISTA E
DEFICIÊNCIA INTELECTUAL**

Dissertação apresentada à Faculdade Israelita de Ciências da Saúde Albert Einstein para obtenção do Título de Mestre em Ciências da Saúde.

Orientadora: Profa. Dra. Karina Griesi Oliveira

São Paulo
2021

Rodrigues, Camilly Erica de Freitas

Identificação de modelos celulares de fácil obtenção com preservação de módulos de co-expressão desregulados em transtorno do espectro autista e deficiência intelectual / Camilly Erica de Freitas Rodrigues. -- São Paulo, 2021.
ix, 49 f.

Dissertação (Mestrado) – Faculdade Israelita de Ciências da Saúde Albert Einstein. Instituto Israelita de Ensino e Pesquisa Albert Einstein. Programa de Pós-Graduação em Ciências da Saúde.

Título em inglês: *Identification of more accessible cell models with preservation of co-expression modules dysregulated in autism spectrum disorder and intellectual disability.*

1. Transtorno do espectro autista. 2. Deficiência intelectual. 3. Transcriptoma. 4. Redes reguladoras de genes.

Elaborada pelo Sistema Einstein Integrado de Bibliotecas

FACULDADE ISRAELITA DE CIÊNCIAS DA SAÚDE ALBERT EINSTEIN

Coordenador do Curso de Pós-Graduação: Prof. Dr. Luiz Vicente Rizzo

Camilly Erica de Freitas Rodrigues

**IDENTIFICAÇÃO DE MODELOS CELULARES DE FÁCIL OBTENÇÃO
COM PRESERVAÇÃO DE MÓDULOS DE CO-EXPRESSÃO
DESREGULADOS EM TRANSTORNO DO ESPECTRO AUTISTA E
DEFICIÊNCIA INTELECTUAL**

Presidente da banca: Profa. Dra. Karina Griesi Oliveira

BANCA EXAMINADORA

Membros titulares:

Profa. Dra. Patricia Maria de Carvalho Aguiar

Prof. Dr. Pedro de Sá Tavares Russo

Prof. Dr. Eduardo Moraes Rego Reis

Membros suplentes:

Profa. Dra. Patricia Severino

Prof. Dr. Omar Julio Sosa

Profa. Dra. Ana Carolina Tahira

Aprovada em: 17/08/21.

Agradecimentos

Aos meus avós Anésio e Raimunda que não mediram esforços em me apoiar durante toda a minha vida em todas as situações.

À minha irmã, Fabienne, que sem dúvida é uma das pessoas mais importantes para mim e foi essencial para que eu chegasse até aqui.

Aos meus pais, Fabio e Edilene, por todo o apoio e incentivo durante o meu aprendizado e em toda essa jornada.

Aos meus tios, Edna e Juracy, que são como meus segundos pais e estiveram presentes em todos os momentos da minha vida.

À minha tia Helem e minhas primas, Jackline e Karol, que sempre estiveram ao meu lado ainda que distantes nesse momento.

A todos os meus colegas e amigos que conquistei durante a realização deste trabalho, em especial Bruna Garbes, que foi essencial durante todo o desenvolvimento deste projeto e ao Paulo Visintin por todo o apoio diário, risadas e companheirismo. Sem vocês tudo teria sido mais difícil, levarei nossa amizade para toda a vida.

A minha orientadora, Dra. Karina Griesi Oliveira, pela oportunidade, confiança, paciência e por todos os ensinamentos, que foram fundamentais para meu crescimento profissional e pessoal. Sem dúvida foi uma das pessoas mais importantes e incríveis em toda essa jornada. Sem a sua dedicação, humanidade e orientação eu não teria chegado até aqui. Serei eternamente grata.

Ao Instituto Israelita de Ensino e Pesquisa Albert Einstein pela disponibilidade de estrutura física e apoio e a todos que de alguma maneira contribuíram para a realização deste trabalho.

Sumário

Agradecimentos	v
Lista de figuras.....	vii
Lista de tabelas	viii
Resumo	ix
1 INTRODUÇÃO.....	1
1.1 Objetivos	3
2 REVISÃO DA LITERATURA.....	4
2.1 Conhecimento atual sobre marcadores genéticos sobre deficiência intelectual e transtorno do espectro autista.....	4
2.2 Estudos de expressão gênica global em células neuronais de deficiência intelectual e transtorno do espectro autista.....	5
2.3 Análise de preservação de módulos em modelos celulares acessíveis	7
3 MÉTODOS.....	9
3.1 Aspectos éticos	9
3.2 Seleção de bancos de dados de transcriptoma de modelos celulares de interesse.....	9
3.3 Normalização dos dados de transcriptoma de modelos celulares de interesse	11
3.4 Análise de preservação de módulos	14
3.5 Definição das redes referências.....	20
3.6 Definição dos módulos de co-expressão desregulados em transtorno do espectro autista e deficiência intelectual	20
4 RESULTADOS.....	24
4.1 Dados selecionados de acordo com os critérios de elegibilidade	24
4.2 Análise de preservação de módulos	28
4.2.1 Papila dérmica	28
4.2.2 Urina	30
4.2.3 Sangue.....	32
4.2.4 Cordão umbilical	36
4.2.5 Sumário dos resultados de preservação por módulo de interesse	39
5 DISCUSSÃO.....	40
6 CONCLUSÕES.....	44
7 REFERÊNCIAS	45
Abstract	
Apêndices	

Lista de figuras

Figura 1. Esquema representando as etapas do processo de normalização quantílica aplicada aos dados de microarray	12
Figura 2. Boxplot da visualização da distribuição dos níveis de expressão gênica de amostras do estudo GSE51546 de cordão umbilical	13
Figura 3. Cluster hierárquico mostrando a relação entre os perfis de expressão de amostras GSE51546 de cordão umbilical	13
Figura 4. Representação das relações de co-expressão entre pares de genes	14
Figura 5. Representação de rede de co-expressão gênica	16
Figura 6. Representação da densidade e conectividade da rede de co-expressão.....	17
Figura 7. Plotagem do resultado da análise de preservação de módulos de acordo com os graus de preservação	19
Figura 8. Comparação das redes de co-expresssão entre um tecido referência e um tecido teste	23
Figura 9. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células de papila dérmica	29
Figura 10. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células epiteliais de urina.....	31
Figura 11. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células do sangue.....	35
Figura 12. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células do cordão umbilical.....	38

Lista de tabelas

Tabela 1. Critérios de inclusão e exclusão na busca de dados de transcriptoma de modelos celulares acessíveis	10
Tabela 2. Pacotes para obtenção e normalização dos dados de transcriptoma de modelos acessíveis	11
Tabela 3. Estatísticas para avaliação de preservação de módulos.....	17
Tabela 4. Descrição dos estudos utilizados na investigação de módulos desregulados em deficiência intelectual.....	22
Tabela 5. Módulos de investigação desregulados em transtorno do espectro autista ..	23
Tabela 6. Dados selecionados de transcriptoma para papila dérmica	24
Tabela 7. Dados selecionados de transcriptoma para urina	25
Tabela 8. Dados selecionados de transcriptoma para sangue.....	25
Tabela 9. Dados selecionados de transcriptoma para cordão umbilical	25
Tabela 10. Dados selecionados seleção de transcriptoma de diferenciação direta	26
Tabela 11. Descrição dos estudos incluídos	27
Tabela 12. Resultado da preservação por módulos de interesse	39

Resumo

Introdução: A deficiência intelectual e os transtornos do espectro autista afetam cerca de 3% da população e podem ser encontrados de forma isolada ou combinados. A desregulação de alguns módulos de genes co-expressos, por exemplo, um módulo de genes ligados a sinapse, tem sido observada em células neuronais de pacientes com transtorno do espectro autista por diversos grupos, incluindo o nosso, sugerindo que o perfil de expressão desses genes poderia ser utilizado como biomarcador para o transtorno. **Objetivos:** Identificar modelos celulares mais acessíveis que possam refletir os mesmos módulos de co-expressão desregulados em células neuronais de pacientes com deficiência intelectual e/ou transtorno do espectro autista. **Métodos:** Por meio de análises de preservação de módulos, nos propusemos a comparar dados de transcriptoma de células neuronais (derivadas de células-tronco pluripotentes induzidas - iPSC ou de cérebro *post-mortem*) com dados públicos de transcriptoma de sangue, células-tronco de papila dérmica e células epiteliais derivadas da urina ou ainda, células neuronais derivadas desses tipos celulares por meio de protocolos de diferenciação direta, buscando identificar se, em algum desses modelos celulares, módulos de co-expressão de interesse se encontram preservados. **Resultados:** Após a busca de dados de transcriptoma dos tecidos de interesse, foi obtido um total de 13 estudos para as análises (2 de papila dérmica, 2 de células epiteliais da urina, 5 de sangue e 4 de cordão umbilical). O módulo relacionado a tradução desregulado em progenitoras neurais derivadas de iPSC e cérebro *post-mortem* foi observado moderadamente preservado em quatro estudos de sangue e em dois estudos de urina. Todos os módulos de resposta imune se mostraram moderadamente preservados em dois estudos de urina. Já em sangue, dois destes módulos obtiveram preservação moderada em três estudos e um módulo também com preservação moderada em cinco estudos. **Conclusão:** Desta forma, nossos resultados sugerem que a análise de expressão destes genes em células epiteliais de urina ou de sangue poderiam refletir a mesma desregulação observada nas células neuronais e de cérebro *post-mortem* de pacientes com transtorno do espectro autista/deficiência intelectual, tendo assim um potencial de aplicação para diagnóstico desses transtornos.

Descritores: Transtorno do espectro autista; Deficiência intelectual; Transcriptoma; Redes reguladoras de genes

1 INTRODUÇÃO

A deficiência intelectual (DI) e o transtorno do espectro autista (TEA) são transtornos do neurodesenvolvimento que, em conjunto, afetam cerca de 3% da população.^(1,2) Estes transtornos podem ser encontrados de forma isolada nos indivíduos ou combinados: estima-se que 10% dos indivíduos com DI apresentam TEA associado,⁽²⁾ e 45% dos indivíduos com TEA apresentam DI.⁽³⁾

A DI caracteriza-se por comprometimento no funcionamento intelectual e no comportamento adaptativo com manifestação antes dos 18 anos de idade e por um $QI < 70$.⁽²⁾ As manifestações podem ocorrer em diferentes graus de intensidade que de acordo com o manual diagnóstico e estatístico de transtornos mentais (DSM-V), podem ser classificados em: leve, moderado, grave e profundo.⁽⁴⁾ Os critérios diagnósticos clássicos para esse transtorno são: limitação significativa das habilidades gerais e mentais e limitações no comportamento adaptativo que se evidenciam durante a infância ou adolescência.⁽⁵⁾ Tais *déficits* comprometem a adaptação do indivíduo a uma vida independente bem como a obter responsabilidade social no cotidiano, incluindo comunicação, participação social e desempenho acadêmico ou profissional.⁽⁵⁾

Na etiologia de DI podem estar envolvidos fatores genéticos, ambientais e sociais e até mesmo a presença de dois ou mais destes.⁽⁶⁾ Estima-se que os casos idiopáticos constituem cerca de 30 a 50% dos casos.⁽⁵⁾ Em um estudo no qual foi realizado sequenciamento do exoma em pacientes com DI, a porcentagem de casos em que foi possível se detectar variantes possivelmente causais esteve entre 16 a 68%, indicando que parte da etiologia genética deste transtorno ainda é desconhecida.⁽⁷⁾ Dentre as causas genéticas mais comuns, estão mutações pontuais e alterações cromossômicas,⁽⁸⁾ incluindo particularmente as aneuploidias já conhecidas como as trissomias dos cromossomos 21, 13 ou 18, duplicações e deleções microscópicas e submicroscópicas (CNVs – *copy number variations*), as quais constituem importantes causas sindrômicas de DI.⁽⁵⁾

Já o diagnóstico de TEA, segundo os critérios estabelecidos pelo DSM-V, consiste em (a) *déficits* persistentes na comunicação e na interação social e (b) padrões restritos e/ou repetitivos de comportamento, interesses ou atividades, classificados em diferentes níveis de gravidade.⁽⁴⁾ Além destas características centrais,

o TEA pode vir acompanhado de outras comorbidades psiquiátricas e cognitivas comuns tais como: transtorno de ansiedade social, transtorno desafiador de oposição, transtorno de *déficit* de atenção/hiperatividade e DI.⁽⁹⁾ Além disso, frequentemente são relatadas anormalidades do sistema imunológico, distúrbio gastrointestinal, distúrbios do sono e epilepsia.⁽⁹⁾ O diagnóstico pode ser estabelecido por meio da aplicação dos critérios de diagnóstico para o transtorno de acordo com o DSM-V⁽¹⁰⁾ ou por meio de escalas diagnósticas de comportamento,⁽¹¹⁾ nos quais o grau de gravidade de DI e de linguagem e a presença de outras comorbidades podem ser especificados.⁽¹⁰⁾

O TEA pode ser resultado de disfunção cerebral decorrente de uma complexa interação entre fatores genéticos, epigenéticos e ambientais⁽¹⁰⁾ com risco aumentado em 3 a 4 vezes no sexo masculino.⁽¹²⁾ Estima-se que fatores de risco genéticos raros como síndromes associadas ao TEA, anormalidades cromossômicas e mutações penetrantes representam 20% dos casos de TEA. De fato, cerca de 40 a 80% dos pacientes com determinadas síndromes genéticas, tais como síndrome de X-frágil, síndrome de Rett e esclerose tuberosa possuem TEA associado.⁽¹³⁾ Entre os casos de TEA não sindrômicos e idiopáticos, 5% são causados por variantes de novo.⁽⁹⁾ Apesar de vários genes já terem sido relacionados ao transtorno, apenas 10 a 38% dos casos de TEA relatados estão associados a fatores genéticos conhecidos.⁽¹¹⁾

Portanto, exceto quando associados a síndromes bem caracterizadas que podem ser diagnosticadas por sinais clínicos particulares e/ou alterações genéticas já conhecidas (ex.: síndrome de Down ou do X-frágil), DI e TEA não apresentam características fisiológicas ou anatômicas específicas, ou seja, não há um biomarcador para estes transtornos. A identificação de biomarcadores é dificultada pelo fato de que a etiologia desses transtornos é bastante variada, como já citado, podendo estar envolvidos diferentes fatores genéticos, epigenéticos e ambientais.⁽¹⁴⁾ Além disso, genes comuns podem estar envolvidos nas duas patologias como evidenciado por achados mutacionais em genes relacionados à regulação sináptica, tais como *SYNGAP1*, *FMRP*, *MECP2*, *NEUROLIGINS/NEURONEXINS* e *SHANK2/3* tanto em pacientes com TEA quanto em pacientes com DI.⁽¹⁵⁾

Dada a sobreposição clínica ou mesmo o curso concorrente das duas patologias, o diagnóstico diferencial de DI e TEA se torna bastante dificultoso baseado apenas em características clínicas. Além disso, é possível que indivíduos com TEA ou DI desenvolvam outros distúrbios do neurodesenvolvimento, como já citado anteriormente, o que impõe ainda mais dificuldades para o estabelecimento do

diagnóstico.⁽¹⁴⁾ Mesmo sabendo que não há medicação para curar essas condições, os pais de uma criança que tem sua qualidade de vida prejudicada por comprometimento de suas habilidades sócio-comunicativas, adaptativas e cognitivas têm a necessidade de conferir uma identidade diagnóstica para seu filho, pois assim podem se concentrar em informações mais específicas sobre a doença, como lidar com ela e quais são as opções de tratamento disponíveis. Particularmente, os pais têm grande anseio em saber quais são as perspectivas prognósticas, uma vez que os quadros de DI e TEA são muito variados, por exemplo, podendo haver DI ou não associada ao quadro de TEA, ou grande variabilidade do nível de déficit intelectual. Em alguns casos, crianças com TEA e DI apresentam comprometimento social e de comunicação mais graves do que aquelas com apenas um dos transtornos.⁽¹⁶⁾ Além disso, é sabido que a intervenção precoce possibilita um melhor prognóstico para esses indivíduos.^(4,17-20)

Assim, delinear estratégias diagnósticas moleculares baseadas em biomarcadores de DI e TEA é importante para que se possa realizar um diagnóstico mais preciso, rápido e eventualmente precoce dessas condições, fornecendo às famílias informações importantes acerca de risco de recorrência, prognóstico e possibilitando que intervenções terapêuticas sejam realizadas mais cedo, além de poder auxiliar no entendimento dos mecanismos comuns e distintos entre as duas patologias.

1.1 Objetivos

1. Realizar levantamento de dados de expressão gênica de tecidos de interesse tais como células do sangue, células-tronco de papila dérmica, células sanguíneas de cordão umbilical, células epiteliais derivadas da urina (ou células neuronais derivadas destes tipos celulares através de protocolos de diferenciação direta);

2. Realizar análise de preservação de módulos de expressão dos tecidos de interesse com dados de células neuronais e cérebro *post-mortem* para identificar se em tecidos de interesse há preservação de módulos de co-expressão de interesse, ou seja, os módulos de co-expressão desregulados em células neuronais (neurônios, astrócitos ou progenitoras neurais) de indivíduos com deficiência intelectual e/ou transtorno do espectro autista.

2 REVISÃO DA LITERATURA

2.1 Conhecimento atual sobre marcadores genéticos sobre deficiência intelectual e transtorno do espectro autista

Muito se avançou no conhecimento dos fatores genéticos relacionados a estas condições após o desenvolvimento de tecnologias de triagem em larga escala como *microarray* e o sequenciamento de nova geração.^(21–23) No entanto, ainda é em uma parcela minoritária dos casos de TEA e DI (especialmente entre DI moderada) que se consegue identificar alguma alteração no nível genômico que pode ser apontada como o fator etiológico.^(24,25)

Esta dificuldade está relacionada a diversos aspectos. Primeiramente, DI e TEA são extremamente heterogêneos do ponto de vista genético, ou seja, diferentes genes podem estar alterados em diferentes pacientes, e ainda não há total conhecimento de quais seriam exatamente estes genes. Além disso, particularmente em TEA e DI moderada, muitos casos podem estar relacionados a um modelo de herança oligo/poligênico, ou seja, seriam causados por um conjunto de alterações genéticas de impacto baixo ou moderado, que individualmente não causam a doença.^(2,26,27) Em função disso, essas variantes podem estar presentes em indivíduos normais da população, e assim, é difícil aferir associação entre estas alterações e o fenótipo.⁽¹²⁾

Finalmente, como já citado, em alguns casos, a etiologia do transtorno pode estar relacionada a fatores epigenéticos ou ambientais, estes últimos agindo isoladamente ou em combinação com fatores genéticos de risco para resultar no fenótipo. Dessa forma, ainda não é possível estabelecer um critério diagnóstico para DI e TEA baseado apenas em marcadores genômicos. Por outro lado, um grande número de genes implicados na etiologia de DI e TEA fazem parte das mesmas vias moleculares,^(14,26–29) indicando que, embora existam diferentes variantes genéticas associadas a estes transtornos, elas poderiam levar a consequências funcionais semelhantes, que poderiam estar refletidas, por exemplo, no nível de transcrição gênica.

2.2 Estudos de expressão gênica global em células neuronais de deficiência intelectual e transtorno do espectro autista

As mesmas abordagens de triagem em larga escala citadas anteriormente também possibilitaram a investigação do transcriptoma de um tecido de forma global. Para análise desse tipo de dado pode ser utilizado como método a construção de redes de co-expressão, frequentemente utilizada em estudos genômicos e de biologia de sistemas para descrever a relação entre um grande número de variáveis.⁽³⁰⁾ Redes de co-expressão gênicas são construídas com base nas relações de expressão entre pares de genes, gerando conjuntos de genes com relações de co-expressão muitos semelhantes, chamados de módulos.⁽³¹⁾ A análise de redes de co-expressão pode ser particularmente mais interessante do que a avaliação de genes individuais para transtornos de etiologia heterogênea como TEA e DI porque, como os padrões de expressão de um conjunto de genes são levados em conta concomitantemente (por meio de um valor que sumariza a expressão de todos os genes de um dado módulo de co-expressão), isso pode capturar pequenas mudanças que seriam perdidas em uma análise gene a gene.⁽³¹⁾ Dessa forma mais sistêmica, pode-se então identificar módulos de genes co-expressos em um dado tecido ou tipo celular cujos níveis de expressão estão desregulados em alguma condição biológica, como por exemplo, em um dado fenótipo.^(14,31,32) Este tipo de análise permite identificar ainda múltiplos níveis de organização molecular como região do cérebro, tipo de célula, organela e vias moleculares usando apenas dados transcricionais e pode, assim, permitir a integração com outras informações, tais como anotações de via, interações de proteína e dados genômicos.⁽³³⁾ Finalmente, a abordagem sistêmica dos dados de expressão pode ser mais interessante para se confrontar resultados de diferentes estudos: enquanto a comparação de listas de genes diferencialmente expressos em uma determinada condição patológica pode mostrar pouca consistência entre distintos trabalhos, a redução dos dados de expressão de milhares de genes a alguns poucos módulos de real significado biológico torna os dados de diferentes estudos mais facilmente comparáveis.

Apesar de existir um número considerável de trabalhos que realizaram estudos de expressão gênica global de cérebro de indivíduos com DI syndrômica, como síndrome de Down e de Rett,^(34–38) apenas poucos deles utilizaram a

abordagem de análise de módulos de co-expressão gênica.⁽³⁹⁾ Assim, ainda não foi explorado se, independentemente do fator genético etiológico, há consistência com relação aos módulos desregulados em diferentes casos de DI.

Por outro lado, essa abordagem tem se mostrado particularmente promissora nos estudos de TEA, uma vez que trabalhos realizados com tecidos cerebrais *post-mortem* têm identificado de forma consistente nos indivíduos autistas 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.^(40–42)

Apesar dos resultados promissores desses estudos, o tecido cerebral não é acessível, dificultando a obtenção de amostras. Nesse sentido, a pesquisa sobre transtornos que afetam o sistema nervoso tem se beneficiado de modelos celulares derivados de células-tronco pluripotentes induzidas (iPSC),⁽⁴³⁾ uma vez que esta abordagem permite a geração de células neuronais com o mesmo genoma de pacientes e, portanto, espera-se que essas células reflitam *in vitro* os mesmos aspectos funcionais que seriam encontrados *in vivo*. De fato, muitos estudos utilizando esse modelo foram conduzidos até o momento investigando diversos transtornos neuropsiquiátricos.^(44–48)

Em TEA, em particular, alguns estudos, incluindo um trabalho de nosso grupo, analisaram transcriptoma de células neuronais derivadas de iPSC, também detectando a desregulação do módulo de genes relacionados a sinapse e neurotransmissão, como visto em cérebro *post-mortem*.^(49–52) Além disso, no estudo,⁽⁵²⁾ foi identificado ainda um módulo em células progenitoras neurais (NPC), composto por genes relacionados ao controle da tradução, também com expressão aumentada nos pacientes com TEA. Curiosamente, este módulo também foi detectado nos neurônios derivados de iPSCs, mas não se apresentou desregulado na comparação entre autistas e controles nessas células. Por outro lado, em estudos com cérebro *post-mortem*, este módulo se mostrou com diferença de expressão nos pacientes com TEA.^(53–55)

De forma conjunta, tais resultados sinalizam que talvez a análise dos níveis de expressão desses módulos isoladamente ou em combinação possa ser usada como biomarcador do transtorno. De fato, um estudo que analisou vários transtornos psiquiátricos (TEA, transtorno bipolar, esquizofrenia, depressão maior e alcoolismo) mostrou que cada um deles possui perfis de desregulação particulares para diferentes

módulos de co-expressão vistos em cérebro, indicando que de fato essa abordagem tem potencial para ser utilizada para diagnóstico.^(53,56)

No entanto, estes trabalhos não analisaram dados de transcriptoma de cérebro de pacientes com DI para entender se há padrões de desregulação de módulos específicos que possibilitem a distinção entre esta condição e TEA. Esta questão foi abordada recentemente em outro trabalho conduzido em paralelo pelo grupo, que resultou na tese de mestrado “Transcriptômica de autismo e deficiências intelectuais: uma abordagem de biologia de sistemas para identificar especificidades e convergências”.⁽⁵⁷⁾ Os resultados desse estudo mostraram que os módulos desregulados em indivíduos com DI (com síndrome de Down, especificamente) são os mesmos módulos que têm sido consistentemente encontrados como desregulados em TEA, ou seja, o módulo associado a sinapse/neurônio e o módulo associado ao sistema imune/glia. Nesse sentido, até o momento, não há consistência de módulos desregulados associados somente a DI ou TEA.

2.3 Análise de preservação de módulos em modelos celulares acessíveis

Como visto acima, perfis de expressão de módulos de co-expressão gênica de células neuronais podem representar um biomarcador para TEA e DI e, assim, esta análise pode ser útil como estratégia diagnóstica. No entanto, o tecido neuronal não é um tecido do qual se possa obter amostras por biópsia. Ainda, embora os modelos neuronais derivados de iPSC se mostrem bastante promissores para a investigação de alterações de expressão gênica em TEA e DI, uma vez que refletem o perfil de expressão visto *in vivo*, sua aplicação para diagnóstico seria difícil, já que o processo de reprogramação e diferenciação é custoso e demanda muito tempo. Por outro lado, um mesmo módulo de co-expressão gênica pode estar presente em diferentes tecidos ou tipos celulares (como exemplificado acima com relação ao módulo de genes de tradução, identificado tanto em progenitores neurais quanto em neurônios). Ou seja, as relações de co-expressão de um grupo de genes podem estar mantidas entre dois tipos celulares distintos, o que pode ser investigado por análises de preservação de módulo.^(30,58)

Assim, identificar outros modelos celulares de coleta acessível ou de fácil geração *in vitro* que reflitam os módulos de co-expressão desregulados em células

neuronaes de pacientes com TEA e DI poderia possibilitar a transposiço desses testes para a prtica clnica. Um modelo interessante a ser explorado, por exemplo, so clulas-tronco mesenquimais de papila drmica, que podem ser obtidas pelo pinamento de pelos ou de fios de cabelo. Estas clulas expressam alguns marcadores neurais, possuindo potencial de se diferenciarem para neurnio.^(59–61) Alm disso, clulas-tronco mesenquimais j se mostraram teis na deteco de diferenas transcrionais e funcionais entre autistas e controles,^(45,62,63) sugerindo que este modelo poderia ser propenso a expressar mdulos de co-expresso presentes em clulas neuronaes. J com relao ao sangue, outro tecido de fcil coleta, h evidncias de que possvel se detectar diferenas de expresso gnica entre autistas e controles nesse tecido, como demonstrado no trabalho,⁽⁶⁴⁾ que identificou aumento de expresso de um mdulo de genes relacionados a traduo em indivduos com TEA. Considerando as evidncias acima expostas e os benefcios de um diagnstico precoce de TEA, o sangue de cordo umbilical se mostra um outro importante modelo para investigao, uma vez que, havendo preservao dos mdulos de interesse nesse tecido, seus nveis de expresso poderiam ser imediatamente testados ao nascimento. Finalmente, um outro modelo interessante a se testar so clulas epiteliaes de urina: ainda que at o momento no haja dados de transcriptoma dessas clulas de autistas ou DI, este modelo celular de fcil coleta e pode ser utilizado como material inicial para diferenciao direta em NPC como j reportado na literatura.⁽⁶⁵⁾

3 MÉTODOS

3.1 Aspectos éticos

Este projeto se utilizou apenas de dados disponíveis em bancos públicos e, portanto, não envolve experimentação com amostras de seres humanos a serem coletadas. Dessa forma, ainda que o mesmo se encontra inserido no contexto do projeto “Identificação de biomarcadores para deficiência intelectual e autismo” (SGPP nº 3529-18), o qual foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos local, por se tratar de um subprojeto que não requiriu aprovação por comitê de ética.

3.2 Seleção de bancos de dados de transcriptoma de modelos celulares de interesse

A busca por dados de transcriptoma de modelos celulares de fácil acesso (células-tronco de papila dérmica, células de sangue, células epiteliais de urina, células sanguíneas de cordão umbilical, ou modelos neuronais derivados destes tipos celulares por diferenciação direta) foi executada no banco Gene Expression Omnibus (GEO)¹. O processo de busca foi realizado de acordo com as orientações do manual publicado pelo grupo PRISMA⁽⁶⁶⁾ seguindo as seguintes etapas: fase de identificação, fase de triagem, de elegibilidade e de inclusão. Na fase de identificação foram incluídos termos específicos para cada tipo celular para a busca de dados tanto de *microarray* e *RNAseq*. Na fase de avaliação, os resumos foram lidos no próprio *site* do GEO e foram excluídos os estudos: (a) realizados com amostras não humanas ou amostras de tecidos de não interesse; (b) realizados com tipo celular inadequado (ex: *single-cell* e amostras de células cancerígenas) e (c) cujos dados não eram de transcriptoma (ex.: ATAC-seq). Na etapa de elegibilidade o desenho do estudo foi avaliado, tendo sido excluídos estudos: (a) com amostras em cultura que sofreram algum tipo de tratamento; ou (b) trabalhos sem informações suficientes sobre as amostras ou sobre a plataforma utilizada. Dados clínicos foram coletados quando disponíveis. Cabe

¹ <https://www.ncbi.nlm.nih.gov/geo/>

ressaltar que a presença de condições patológicas não foi utilizada como critério de exclusão do estudo ou das amostras dos indivíduos afetados, pois não espera-se que tais condições levem à interrupção completa das relações de co-expressão dos genes expressos em determinado tecido. Essas relações estão mais ligadas ao caráter do tecido, e condições patológicas mais provavelmente afetam níveis de expressão de alguns módulos (exceção se faz a tecidos tumorais, por exemplo, em que, nesse caso sim, é possível que a rede esteja consideravelmente desconfigurada, uma vez que o caráter daquele tipo celular já não é o mesmo do tecido original).

Na fase de inclusão foram utilizados somente trabalhos com pelo menos 12 amostras. Este número foi escolhido baseado na experiência prévia do grupo, que vem observando que números de amostras inferiores a 11 em geral não possibilitam a construção de redes de co-expressão confiáveis.

Na tabela 1 é possível visualizar os critérios de inclusão e exclusão utilizados em cada uma das etapas.

Tabela 1. Critérios de inclusão e exclusão na busca de dados de transcriptoma de modelos celulares acessíveis

Tecidos	Etapas			
	Identificação	Avaliação	Elegibilidade	Inclusão
Papila dérmica	<i>Dermal papilla, hair follicle, papilla tissue</i>	Amostras não humanas ou de não interesse, tipo celular inadequado e dados não transcriptoma	Tratamento em cultura e informações insuficientes	Número amostral com pelo menos 12 amostras
Urina	<i>Urine cells, urinary mesenchymal cells,</i>			
Sangue	<i>Peripheral blood, blood cells</i>			
Cordão umbilical	<i>Cord blood, umbilical cord, umbilical vein</i>			
Diferenciação direta	<i>Transdifferentiation, direct conversion, direct induction, neuron derived</i>			

Os dados de expressão gênica dos estudos elegíveis foram baixados com o uso da ferramenta GEOquery do *software* R, o qual permite a importação dos dados do banco GEO de cada estudo para o *software*.

Vale ressaltar que, como nosso objetivo é verificar se as relações de co-expressão entre os genes de um módulo desregulado em células neuronais de pacientes com autismo ou DI encontram-se preservadas em um tecido acessível, os dados de transcriptoma desses tecidos teste não precisam ser de pacientes com os transtornos em questão.

3.3 Normalização dos dados de transcriptoma de modelos celulares de interesse

Os dados de cada estudo elegível foram normalizados antes de seguirmos para as análises posteriores. A normalização é uma etapa essencial no pré-processamento dos dados, particularmente em estudos de expressão gênica.⁽⁶⁷⁾ Particularmente em estudos utilizando *microarray*, alguns vieses podem ocorrer devido a diferenças no nível de expressão de genes específicos entre tecidos diferentes ou em razão de diferenças no processo de realização do experimento. Para a redução destes vieses, foi aplicada em cada estudo a normalização quantílica que permite que os dados de intensidade de hibridização das *probes* sejam normalizados para que estas possam ser comparáveis entre as amostras.⁽⁶⁸⁾

O pacote de normalização variou de acordo com a plataforma de sequenciamento utilizada em cada estudo, e, portanto, um *script* de programação diferente foi utilizado para cada situação. Todos os pacotes utilizados estão depositados no Bioconductor², um projeto de *software* livre, de código e desenvolvimento abertos para a análise e compreensão de dados genômicos. Os pacotes utilizados, com suas respectivas funções e plataformas de uso estão descritos abaixo na tabela 2.

Tabela 2. Pacotes para obtenção e normalização dos dados de transcriptoma de modelos acessíveis

Pacote	Função	Plataforma
GEOquery	Obtenção de dados do banco GEO	
Affy	Análise exploratória de oligonucleotídeos	Affymetrix Human Genome U133 Plus 2.0 Array
Oligo	Análise de oligonucleotídeos (expressão/SNP/éxon)	Affymetrix Human Gene 1.0 ST Array
Limma	Análise de dados, modelos lineares e expressão diferencial para dados de <i>microarray</i>	Illumina HumanHT-12 V4.0 expression beadchip

* Single Nucleotide Polymorphism.

Embora desenhados para dados de diferentes plataformas, todos estes pacotes utilizam a normalização quantílica, a qual está descrita na figura 1. A fim de facilitar a manipulação dos dados, executando um mesmo *pipeline* para todos os estudos, todas as tabelas de expressão utilizadas em nosso estudo foram submetidas

² <https://www.bioconductor.org/>

à normalização quantílica, ainda que os dados já houvessem sido normalizados pelos autores originais dos estudos. Importante notar, no entanto, que a ressubmissão de um dado à nova normalização não altera os dados iniciais de expressão gênica (Figura 1).

Dado bruto	Ordenar crescentemente os valores por amostra (coluna)	Tirar média por linhas e substituir os valores pela média	Reordenar os valores médios na ordem original																																																																																
<table border="1"> <tr><td>2</td><td>4</td><td>4</td><td>5</td></tr> <tr><td>5</td><td>14</td><td>4</td><td>7</td></tr> <tr><td>4</td><td>8</td><td>6</td><td>9</td></tr> <tr><td>3</td><td>8</td><td>5</td><td>8</td></tr> <tr><td>3</td><td>9</td><td>3</td><td>5</td></tr> </table>	2	4	4	5	5	14	4	7	4	8	6	9	3	8	5	8	3	9	3	5	<table border="1"> <tr><td>2</td><td>4</td><td>3</td><td>5</td></tr> <tr><td>3</td><td>8</td><td>4</td><td>5</td></tr> <tr><td>3</td><td>8</td><td>4</td><td>7</td></tr> <tr><td>4</td><td>9</td><td>5</td><td>8</td></tr> <tr><td>5</td><td>14</td><td>6</td><td>9</td></tr> </table>	2	4	3	5	3	8	4	5	3	8	4	7	4	9	5	8	5	14	6	9	<table border="1"> <tr><td>3.5</td><td>3.5</td><td>3.5</td><td>3.5</td></tr> <tr><td>5.0</td><td>5.0</td><td>5.0</td><td>5.0</td></tr> <tr><td>5.5</td><td>5.5</td><td>5.5</td><td>5.5</td></tr> <tr><td>6.5</td><td>6.5</td><td>6.5</td><td>6.5</td></tr> <tr><td>8.5</td><td>8.5</td><td>8.5</td><td>8.5</td></tr> </table>	3.5	3.5	3.5	3.5	5.0	5.0	5.0	5.0	5.5	5.5	5.5	5.5	6.5	6.5	6.5	6.5	8.5	8.5	8.5	8.5	<table border="1"> <tr><td>3.5</td><td>3.5</td><td>5.0</td><td>5.0</td></tr> <tr><td>8.5</td><td>8.5</td><td>5.5</td><td>5.5</td></tr> <tr><td>6.5</td><td>5.0</td><td>8.5</td><td>8.5</td></tr> <tr><td>5.0</td><td>5.5</td><td>6.5</td><td>6.5</td></tr> <tr><td>5.5</td><td>6.5</td><td>3.5</td><td>3.5</td></tr> </table>	3.5	3.5	5.0	5.0	8.5	8.5	5.5	5.5	6.5	5.0	8.5	8.5	5.0	5.5	6.5	6.5	5.5	6.5	3.5	3.5
2	4	4	5																																																																																
5	14	4	7																																																																																
4	8	6	9																																																																																
3	8	5	8																																																																																
3	9	3	5																																																																																
2	4	3	5																																																																																
3	8	4	5																																																																																
3	8	4	7																																																																																
4	9	5	8																																																																																
5	14	6	9																																																																																
3.5	3.5	3.5	3.5																																																																																
5.0	5.0	5.0	5.0																																																																																
5.5	5.5	5.5	5.5																																																																																
6.5	6.5	6.5	6.5																																																																																
8.5	8.5	8.5	8.5																																																																																
3.5	3.5	5.0	5.0																																																																																
8.5	8.5	5.5	5.5																																																																																
6.5	5.0	8.5	8.5																																																																																
5.0	5.5	6.5	6.5																																																																																
5.5	6.5	3.5	3.5																																																																																

Representação da normalização dos valores de expressão gênica entre as amostras. Os valores iniciais brutos são ordenados crescentemente, a média de valores de cada linha é calculada e então os valores são reordenados na ordem em que estavam inicialmente.

Fonte: Adaptado de Harvard University. Data analysis for genomics. Cambridge (MA): edX; C2021 [cited 2021 May 19]. Available from:

<https://courses.edx.org/courses/HarvardX/PH525x/1T2014/course/#>⁽⁶⁹⁾

Figura 1. Esquema representando as etapas do processo de normalização quantílica aplicada aos dados de *microarray*

Após esta etapa, foi realizada a anotação das *probes* para identificação dos genes aos quais estas estão mapeadas. Para isso, foi utilizado o pacote “biomaRt” para retornar o nome do gene, código *ensembl* e posição cromossômica correspondente de cada *probe*. Além disso, a expressão de genes que estavam representados por mais de uma *probe* na plataforma foi sumarizada pela média das diferentes *probes*. Após isso, os dados de cada estudo passaram por um processo de remoção de amostras *outliers*, ou seja, amostras com perfil de expressão discrepante, por meio da análise desses perfis e suas relações por meio de análises gráficas como *boxplot* e *cluster* hierárquico (Figuras 2 e 3, respectivamente). Isso foi feito para controle de qualidade, de modo a trabalharmos com dados livres de estimativas enviesadas para seguir com as próximas análises.

A inspeção visual da figura permite identificar duas amostras (indicadas pelas setas em vermelho) cuja distribuição dos níveis de expressão dos diferentes genes mensurados se mostram discrepantes com relação às demais (amostras 11 e 41).

Figura 2. *Boxplot* da visualização da distribuição dos níveis de expressão gênica de amostras do estudo GSE51546 de cordão umbilical

Árvore de clusterização e visualização de *outliers* indicados pelas amostras destacadas em vermelho.

Figura 3. *Cluster* hierárquico mostrando a relação entre os perfis de expressão de amostras GSE51546 de cordão umbilical

3.4 Análise de preservação de módulos

Para um melhor entendimento da análise de preservação de módulo, detalharemos algumas conceitualizações importantes a seguir.

Como já mencionado, o método de construção de uma rede de co-expressão gênica é baseado nas correlações de expressão entre todos os pares possíveis de genes de um dado de transcriptoma (Figura 4).⁽³¹⁾ Em uma representação gráfica, os nós dessa rede são os genes, e a espessura das arestas, os índices de correlação de expressão entre eles (Figura 5). Conjuntos de genes altamente correlacionados são reunidos em módulos específicos nessa rede.

Representação esquemática das relações de expressão entre pares de genes. A e B representam ausência de relação de co-expressão entre os genes 15 e 4 e entre os genes 13 e 9; C representa uma fraca relação de co-expressão positiva entre os genes 13 e 3; e D representa uma forte relação positiva de co-expressão entre os genes 3 e 6.

Fonte: Adaptado de Dam S, Vösa U, Van der Graaf A, Franke L, Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* 2018;19(4):575–92. Figure 1. Example of a co-expression network analysis; p. 576.⁽⁷⁰⁾

Figura 4. Representação das relações de co-expressão entre pares de genes

Assim, nota-se que a construção dos módulos de co-expressão a partir de dados de transcriptoma é feita estritamente de forma matemática, sem nenhum tipo de *input* de informações biológicas. Uma das ferramentas mais utilizadas para este tipo de análise é o pacote desenvolvido em linguagem R chamado Weighted Gene Co-Expression Network Analysis (WGCNA). Este pacote contém um conjunto abrangente de funções que podem ser divididas nas seguintes categorias: construção de redes, detecção de módulos, seleção de genes, cálculos de propriedades topológicas, simulação de dados, visualização e interface com *software* externo.⁽³¹⁾

A detecção dos módulos é baseada na criação do padrão topológico *Scale-free* (livre de escala) baseada na correlação de Pearson como medida de similaridade dos valores de expressão utilizados ⁽³²⁾. Essa matriz de similaridade é transformada em uma matriz de adjacência elevada a uma potência β (*soft-threshold*) a qual afeta diretamente em como aderente será ao modelo livre de escala de forma que, quanto maior essa potência, menor é a conectividade média da rede. A adesão a este modelo é medida por um ajuste de regressão linear ($R^2=0.8$) que quantifica até que ponto o número de vértices adjacentes a ele segue uma lei do poder.⁽³²⁾

Para a construção dos módulos, os pares de genes são clusterizados baseado na distância entre os valores de correlação de Pearson. A matriz do *soft-threshold* é transformada em um elemento de matriz de sobreposição topológica-TOM. Nesse processo, é estipulado um número mínimo de genes em cada módulo de forma que os valores TOM são baseados entre um (mais distante) e zero (mais próximo).⁽³²⁾

Após a construção desses módulos, a eles são atribuídos aleatoriamente nomes de identificação que podem ser números ou cores. Essa abordagem de construção de redes tem sido utilizada com o objetivo de investigar as propriedades dos módulos de co-expressão em diferentes condições, tecidos ou doenças.⁽³⁰⁾

Representação dos componentes de uma rede de co-expressão gênica. Cada círculo, ou também chamado nó, representa um gene. As arestas representam as correlações entre os pares de genes, onde a espessura da aresta indica o índice de correlação entre os genes. O grupo mais altamente conectado representa um módulo de co-expressão dentro dessa rede.

Fonte: Adaptado de Dam S, Vösa U, Van der Graaf A, Franke L, Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* 2018;19(4):575–92. Figure 1. Example of a co-expression network analysis; p. 576.⁽⁷⁰⁾

Figura 5. Representação de rede de co-expressão gênica

Comparando-se dois módulos obtidos em estudos distintos, ainda que estes tenham uma sobreposição significativa com relação às listas de genes que os compõem, não necessariamente as relações de co-expressão entre estes genes é a mesma nos dois módulos. Para este tipo de avaliação é necessária uma análise de preservação de módulos, a qual também pode ser realizada com o pacote WGCNA. Como descrito acima, rede define-se como o conjunto total de módulos e seus genes componentes identificados em um dado de expressão gênica. A análise de preservação de módulo é baseada em uma análise estatística que determina as propriedades entre duas redes diferentes. Para isso, uma rede é utilizada como referência e o que se investiga é se as relações de co-expressão permanecem semelhantes em uma segunda rede teste, ou seja, se as propriedades de conexão entre os genes se mantêm preservadas entre elas.⁽³⁰⁾ Assim, pode-se investigar se um mesmo módulo de co-expressão se encontra preservado entre duas condições, ou seja, se as relações de co-expressão de um mesmo grupo de genes se mantêm, por exemplo, entre pacientes e controles, ou entre dois tipos celulares.^(30,58)

São aplicadas várias estatísticas de preservação, as quais inicialmente são sumarizadas em duas estatísticas essenciais para avaliação da

preservação. São elas: densidade e conectividade. A densidade é utilizada para determinar se os genes de um módulo permanecem fortemente conectados na rede teste e conectividade determina se o padrão de conexão entre os genes na rede referência é semelhante ao da rede teste⁽³⁰⁾ (Figura 6).

Conectividade e densidade como parâmetros estatísticos de preservação aplicados na análise de preservação de módulos. Cada círculo em preto representa um gene, sendo o tamanho do círculo indicativo da conectividade do gene com o módulo, ou seja, quanto maior o círculo, maior número de conexões do gene com o módulo. As arestas em vermelho e em verde representam a conectividade entre os genes, sendo a aresta vermelha representação de correlação positiva e a verde correlação negativa. Quanto maior a espessura da aresta, maior a conectividade entre os genes. Nota-se que o padrão de conectividade e densidade é diferente entre as redes A e B. Os genes altamente conectados na rede A são diferentes dos genes altamente conectados na rede B. Além disso, o padrão de conexão entre os genes também é diferente, representado pelas diferentes espessuras das arestas entre os genes da rede A e B.

Fonte: Adaptado de Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7(1):e1001057. Figure 7. Detailed preservation analysis of KEGG pathways between human and chimp brain networks; p. 16.⁽³⁰⁾

Figura 6. Representação da densidade e conectividade da rede de co-expressão

Assim as estatísticas de preservação são sumarizadas em $Z_{density}$ e $Z_{connectivity}$ correspondentes a densidade e conectividade, respectivamente. Estes parâmetros, por sua vez, são gerados com base em diferentes estatísticas calculadas utilizando-se os dados de expressão da rede teste para cada conjunto de genes que compõe um módulo na rede referência, as quais estão descritas na tabela 3.

Tabela 3. Estatísticas para avaliação de preservação de módulos

Nome	Tipo	Descrição
meanCor	Densidade/conectividade	Média da preservação dos padrões de conectividade dentro de um módulo
meanAdj	Densidade	Adjacência média dos genes que compõem o módulo
propVarExpl	Densidade	Proporção de variância explicada. Quanto maior o valor, mais preservado o módulo
meanKME	Densidade/conectividade	Média de intensidade de conexão do gene com o módulo

continua...

...continuação

cor.kIM	Conectividade	Soma das intensidades de conexão com outros genes dentro do mesmo módulo
cor.kME	Conectividade	Coefficiente de intensidade de conexão do gene com o módulo
cor.cor	Conectividade	Medida da preservação dos padrões de conectividade dentro de um módulo

Fonte: Adaptado de Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7(1):e1001057. Table 3. Overview of the performance of various module preservation statistics in our simulation studies; p. 21.⁽³⁰⁾

A densidade é então calculada da seguinte forma:

$$Z_{\text{density}} = \text{median} (Z_{\text{meanCor}}, Z_{\text{meanAdj}}, Z_{\text{propVarExpl}}, Z_{\text{meanKME}})$$

Similarmente, a conectividade é sumarizada da seguinte forma:

$$Z_{\text{connectivity}} = \text{median}(Z_{\text{cor.kIM}}, Z_{\text{cor.kIM}}, Z_{\text{cor.cor}})$$

Após a densidade e conectividade terem sido calculadas, estas são sumarizadas em um parâmetro estatístico Z_{summary} gerado pela fórmula abaixo:

$$Z_{\text{summary}} = \frac{Z_{\text{density}} + Z_{\text{connectivity}}}{2}$$

A partir do Z_{summary} , é possível analisar, por exemplo, se os parâmetros de conectividade dos genes contidos em cada módulo de um tecido de referência são estatisticamente diferentes ou iguais quando comparados com os padrões de conectividade destes mesmos elementos em um segundo tecido, como proposto nesse trabalho. Quanto maior o valor do Z_{summary} , mais forte é a evidência de que o valor observado da estatística de preservação é significativamente maior do que o esperado por acaso.⁽³⁰⁾ Os resultados gerados para avaliar a preservação são representados pelo seguinte ranqueamento: $Z_{\text{summary}} < 2$ é indicativo de módulo pouco preservado, $2 < Z_{\text{summary}} < 10$, módulo moderadamente preservado e $Z_{\text{summary}} > 10$, módulo altamente preservado. Além do parâmetro Z_{summary} , a preservação do módulo é também avaliada por um valor-p estatístico sumarizado a uma medida descritiva P_{summary} . É calculado um valor log-p que sumariza os valores de Z_{scores} individuais, de forma que quanto menor o valor de P_{summary} , maior a evidência de preservação do módulo indicando assim, relação inversa à medida de Z_{summary} . Tal ranqueamento, já estabelecido na literatura para estatística de preservação de módulos, foi utilizado em nosso estudo.⁽³⁰⁾

Esses valores são corrigidos visto que, ao analisar várias estatísticas de preservação individualmente para qualquer indicação de preservação de módulo, como já mencionado acima, deve haver uma correção para essas comparações múltiplas. Para isso, é utilizada a correção de Bonferroni que determina o limite a ser utilizado em vários testes por meio da função `modulePreservation` também do pacote WGCNA.⁽³⁰⁾

Ao final da análise, os módulos da rede referência são plotados em um gráfico de “Preservation $Z_{summary}$ ”, de acordo com o ranqueamento de critérios de preservação do Z_{score} , como já citado acima. Cada módulo é mostrado dentro do parâmetro referente ao nível de preservação, como exemplificado na figura 7.

Cada módulo da rede referência é plotado de acordo com seu índice $Z_{summary}$ e tamanho do módulo, e linhas horizontais demarcam os três níveis de preservação (não preservado, moderadamente preservado e altamente preservado).

Figura 7. Plotagem do resultado da análise de preservação de módulos de acordo com os graus de preservação

3.5 Definição das redes referências

Foram consideradas como redes referência, as redes obtidas a partir dos estudos de Griesi-Oliveira et al.⁽⁵²⁾ e Gandal et al.⁽⁵³⁾ para análise de transcriptoma de células neuronais com pacientes autistas descritos a seguir.

No estudo de Griesi-Oliveira et al.⁽⁵²⁾ um total de 12 amostras foram utilizadas, sendo 6 de pacientes com TEA e 6 controles, todos do sexo masculino. Destes, 5 eram pacientes com TEA de alto funcionamento e 1 de baixo funcionamento (QI<70). Nenhum dos pacientes apresentava síndrome de X-frágil. Este dado foi selecionado como uma de nossas redes referência pois trata-se de transcriptomas oriundos de NPC e neurônios derivados de iPSC, as quais, segundo análise dos próprios autores, apresentam perfil de expressão mais próximo a cérebros fetais do que cérebros adultos.

Por sua vez, os dados do estudo de Gandal et al.⁽⁵³⁾ foram utilizados como uma segunda rede referência em nosso trabalho, representando um perfil de expressão de cérebro adulto. Embora existam diversos estudos que realizaram transcriptoma de cérebro *post-mortem* de indivíduos com autismo, optamos pelo uso deste estudo em particular por este se tratar de uma metanálise que incluiu dados de cérebro *post-mortem* referentes a oito estudos de doenças neuropsiquiátricas, sendo dois estudos de TEA, gerando um total de 1768 amostras distribuídas entre sexo masculino e feminino. Destas, 51 são amostras de pacientes com TEA, 559 de Esquizofrenia, 222 de transtorno bipolar e 936 controles. Todas as amostras incluídas se restringiram a áreas do córtex frontal e temporal do cérebro do período pós-natal.

3.6 Definição dos módulos de co-expressão desregulados em transtorno do espectro autista e deficiência intelectual

Para um melhor entendimento das descrições a seguir, é importante ressaltar que, como o método de geração de redes de co-expressão gênica pelo WGCNA é estritamente matemático e os dados de expressão de diferentes estudos nunca serão iguais, é extremamente improvável haver um módulo exatamente idêntico em termos de composição gênica entre dois estudos. O que pode ocorrer, sim, é uma sobreposição significativa entre os genes componentes do módulo de um estudo em

relação ao de outro. Pela mesma razão, os nomes atribuídos aos módulos, como já citado anteriormente, são aleatórios e não têm relação com parâmetros de composição biológica dos mesmos. Assim, um módulo nomeado como “módulo 1” em um estudo, não é necessariamente o mesmo módulo identificado pelo número 1 em outro estudo (ou seja, não têm a mesma composição de genes). Dessa forma, para indicar módulos que são semelhantes (possuem semelhante composição de genes) utilizamos o termo “módulos correspondentes”.

A definição dos módulos desregulados em TEA e DI, cujas preservações foram analisadas em outros tecidos no presente estudo foi embasada em dados da literatura, em estudos prévios do grupo e também em estudos que se desenvolveram em paralelo a este.

Como evidenciado no estudo de Griesi-Oliveira et al.⁽⁵²⁾ um módulo de co-expressão em neurônios, composto por genes relacionados particularmente a sinapse e canais controlados por voltagem, tem sido consistentemente identificado como desregulado em pacientes com TEA em diversos estudos de transcriptômica, quer sejam de cérebro *post-mortem*, particularmente do córtex frontal e temporal, quer sejam com neurônios derivados de iPSC. Este módulo, identificado como módulo M_{Neur}1-turquoise no referido estudo, é correspondente aos módulos M12 de Voineagu et al.,⁽⁴²⁾ m1 de Gupta et al.,⁽⁴⁰⁾ CTX.M16 de Parikshak et al.⁽⁴¹⁾ e geneM31, geneM1 e geneM23 de Gandal et al.⁽⁵³⁾ Outro módulo que tem sido consistentemente encontrado como desregulado em tecido neuronal de indivíduos autistas é um módulo de genes relacionados a resposta imune e células gliais (identificado como módulo M16 no estudo de Voineagu et al.,⁽⁴²⁾ m5 em Gupta et al.,⁽⁴⁰⁾ CTX.M19 em Parikshak et al.,⁽⁴¹⁾ geneM3, geneM6 e geneM3 em Gandal et al.⁽⁵³⁾ Justamente por ser expresso em células da glia, este módulo não emergiu nas análises de neurônios derivados de iPSC de Griesi-Oliveira et al.⁽⁵²⁾

Finalmente, um módulo de genes co-expressos relacionados ao controle de tradução (identificado como módulo M_{NPC}10-blue) mostrou aumento de expressão nas NPCs dos indivíduos com autismo avaliados no estudo de Griesi-Oliveira et al.⁽⁵²⁾ Curiosamente, este módulo encontra-se preservado nos neurônios derivados dessas mesmas NPCs (correspondendo ao módulo M_{Neur}19-greenyellow), mas não apresenta desregulação nessas células. Por outro lado, em estudos com cérebro *post-mortem*, este módulo se mostrou com diferença de expressão nos pacientes com TEA, correspondendo aos módulos M1 e M25 de Lombardo et al.^(54,55) e

geneM15 de Gandal et al.⁽⁵³⁾ uma discussão sobre estas correspondências também podem ser vista em Lombardo et al.⁽⁵⁴⁾

Já com relação a DI, poucos foram os trabalhos na literatura que avaliaram expressão gênica em células neuronais desses indivíduos utilizando a estratégia de análise de redes de co-expressão. Assim, os módulos desregulados em DI, portanto de interesse para este estudo, incluídos aqui, foram obtidos de um projeto conduzido em paralelo por uma outra autora do nosso grupo de pesquisa já mencionado acima.⁽⁵⁷⁾ Os dados do estudo em questão foram oriundos de amostras de cérebro *post-mortem* de 3 estudos de pacientes com DI sindrômica. A descrição destes estudos com suas respectivas identificações no GEO são mostradas na tabela 4.

Tabela 4. Descrição dos estudos utilizados na investigação de módulos desregulados em deficiência intelectual

GSE	Amostras DI				Síndrome associada
	Masculino	Feminino	Controles	Total	
GSE5390	5	5	5	15	Síndrome de Down
GSE59630	33	25	58	116	Síndrome de Down
GSE75303		6	6	12	Síndrome de Rett

DI: deficiência intelectual.

Após as análises desse trabalho, foi demonstrado que os módulos desregulados em pacientes com DI são os mesmos módulos que têm sido consistentemente encontrados como desregulados em TEA, ou seja, o módulo associado a sinapse e o módulo associado ao sistema imune/glia já mencionados acima.

Desta forma, três conjuntos de módulos foram definidos como alvo de investigação no presente estudo: um conjunto de módulos de co-expressão composto por genes relacionados a sinapse; um conjunto de módulos composto por genes relacionados a resposta imune; e um conjunto de módulos composto por genes relacionados a tradução. A identificação desses módulos dentro das redes referência selecionadas é mostrada na tabela 5.

Tabela 5. Módulos de investigação desregulados em transtorno do espectro autista

	ID do módulo em NPCs derivadas de iPSC*	ID do módulo em Neurônios derivados de iPSC*	ID do módulo em cérebro <i>post-mortem</i> **
Módulo de genes relacionado a tradução	M _{NPC} 10-blue	M _{Neur} 19-greenyellow	M _{Brain} 15-midnightblue
Módulo de genes relacionado a sinapse		M _{Neur} 1-turquoise	M _{Brain} 1-turquoise, M _{Brain} 23- darkturquoise e M _{Brain} 31- paleturquoise
Módulo de genes relacionado a resposta imune/glia			M _{Brain} 2-blue, M _{Brain} 3-brown e M _{Brain} 6-red

*O nome de identificação dos módulos do estudo de Gandal et al.⁽⁵³⁾ foi ligeiramente alterado a fim de facilitar a clareza nas descrições de nossos resultados (ex.: o módulo identificado como geneM15 no estudo de Gandal et al.⁽⁵³⁾ é identificado aqui como M_{Brain}15-midnightblue).

Fonte: * Griesi-Oliveira K, Fogo MS, Pinto BG, Alves AY, Suzuki AM, Morales AG, et al. Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. Mol Psychiatry. 2020. [Epub ahead of print].⁽⁵²⁾

** Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al.; PsychENCODE Consortium, Peters MA, Gerstein M, Liu C, Iakoucheva LM, Pinto D, Geschwind DH. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science. 2018;362(6420):eaat8127. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder.⁽⁵³⁾

Um esquema das comparações realizadas encontra-se resumido na figura 8.

Fonte: Adaptado de Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7(1):e1001057. Figure 7. Detailed preservation analysis of KEGG pathways between human and chimp brain networks; p. 16.⁽³⁰⁾

Figura 8. Comparação das redes de co-expressão entre um tecido referência e um tecido teste

4 RESULTADOS

4.1 Dados selecionados de acordo com os critérios de elegibilidade

Após a busca de dados de transcriptoma dos tecidos de interesse no banco de dados GEO³, foram excluídos trabalhos realizados com células não humanas ou em tipos celulares que não os de interesse deste estudo (ex.: trabalhos com *single—cell*, com amostras tumorais ou tecidos de não interesse), e trabalhos cujos dados não eram de transcriptoma (ex.: dados de metiloma, ATAC-seq, etc.), restando 47 potenciais estudos a serem incluídos em nossas análises. Em seguida, foi cuidadosamente verificado se cada estudo atendia aos critérios de elegibilidade previamente estabelecidos: células não submetidas a nenhum tipo de tratamento; disponibilidade de dados com identificação adequada da plataforma utilizada e, finalmente foi aplicado o critério de inclusão com número amostral maior que 12. As avaliações dos critérios de avaliação, elegibilidade e inclusão para cada estudo encontram-se descritas nas tabelas abaixo para cada tecido separadamente (Tabelas 6 a 10), destacando-se negrito os itens que determinaram a exclusão do estudo.

Tabela 6. Dados selecionados de transcriptoma para papila dérmica

GSE	Avaliação dos critérios de elegibilidade			Incluído
	Nº amostral	Tratamento	Outras observações	
GSE44765	18	não		sim
GSE72894	6	não		não
GSE66664	140	sim		não
GSE78606	2	não		não
GSE61511	6	não		não
GSE31324	12	não		sim
GSE77825	6	não	RNA não codificante	não
GSE62798	8	não		não
GSE117360	2	não		não
GSE69299	9	não		não
GSE109632	6	não		não
GSE66663	6	não		não

³ <https://www.ncbi.nlm.nih.gov/geo/>

Tabela 7. Dados selecionados de transcriptoma para urina

GSE	Avaliação dos critérios de elegibilidade			Incluído
	N° amostral	Tratamento	Outras observações	
GSE36395	6	não		não
GSE123247	4	não		não
GSE128347	228	não	<i>single-cell</i>	não
GSE128348	207	não	<i>single-cell</i>	não
GSE64306	32	não	RNA não codificante	não
GSE114477	6	não		não
GSE68020	50	não	tipo celular inadequado	não
GSE75545	4	não		não
GSE48318	4	não		não
GSE128359	47	não	RNA não codificante	não
GSE85885	6	não		não
GSE28242	13	não		sim
GSE31189	40	não		sim
GSE140627	10	não		não

Tabela 8. Dados selecionados de transcriptoma para sangue

GSE	Avaliação dos critérios de elegibilidade			Incluído
	N° amostral	Tratamento	Outras observações	
GSE53163	9	não		não
GSE54536	10	não		não
GSE66486	4	não		não
GSE47621	8	não		não
GSE54017	8	não		não
GSE48556	139	não	Informações insuficientes para conversão dos IDs das <i>probes</i>	não
GSE126124	98	não		sim
GSE131793	20	não		sim
GSE129166	117	não		sim
GSE4607	123	não		sim
GSE128557	28	sim		não
GSE46510	154	não		sim

Tabela 9. Dados selecionados de transcriptoma para cordão umbilical

GSE	Avaliação dos critérios de elegibilidade			Incluído
	N° amostral	Tratamento	Outras observações	
GSE60403	16	não		sim
GSE51546	12	não		sim
GSE48354	38	não		sim
GSE123302	224	não		sim

Tabela 10. Dados selecionados seleção de transcriptoma de diferenciação direta

GSE	Avaliação dos critérios de elegibilidade			Incluído
	Nº amostral	Tratamento	Outras observações	
GSE113804	6	não		não
GSE69480	6	não		não
GSE51580	5	não		não
GSE52352	1	não		não
GSE68101	8	não		não

Após a aplicação dos filtros, estabelecemos o total de 13 estudos elegíveis para as etapas posteriores (Tabela 11). Dados clínicos destes estudos também foram coletados quando disponíveis. Para os estudos de papila dérmica as amostras se tratavam de amostras controle sem nenhuma doença associada. Para os estudos de urina amostras controle, síndrome da bexiga ou síndrome da dor pélvica estavam presentes entre as amostras. Para dados de sangue algumas doenças ou condições se encontravam associadas como: doença inflamatória intestinal, hipertensão arterial pulmonar, parto espontâneo e síndrome da resposta inflamatória sistêmica. Para os estudos de cordão umbilical além de amostras controle, amostras de indivíduos com diabetes e TEA estavam presentes.

Após a análise para verificação de amostras com perfil de expressão gênica discrepante (*outliers*), fez-se necessário a remoção de amostras dos estudos de cordão umbilical e de quatro estudos de sangue. Os 13 estudos incluídos em nosso trabalho com suas respectivas informações sobre número de amostras, plataforma e número de acesso de acordo com o tipo celular são mostrados na tabela 11.

Tabela 11. Descrição dos estudos incluídos

Tipo celular	Plataforma	GSE*	Acesso	N° amostras inicial	N° amostras final
Papila dérmica	<i>Microarray</i>	GSE44765	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44765	18	18
Papila dérmica	<i>Microarray</i>	GSE31324	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31324	12	12
Urina	<i>Microarray</i>	GSE28242	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28242	13	13
Urina	<i>Microarray</i>	GSE31189	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31189	40	40
Sangue	<i>Microarray</i>	GSE129166	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129166	117	106
Sangue	<i>Microarray</i>	GSE126124	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126124	98	98
Sangue	<i>Microarray</i>	GSE4607	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66099	123	121
Sangue	<i>Microarray</i>	GSE46510	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46510	154	151
Sangue	<i>Microarray</i>	GSE131793	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131793	20	18
Cordão umbilical	<i>Microarray</i>	GSE48354	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48354	38	34
Cordão umbilical	<i>Microarray</i>	GSE123302	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123302	224	219
Cordão umbilical	<i>Microarray</i>	GSE60403	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60403	12	11
Cordão umbilical	<i>Microarray</i>	GSE51546	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51546	16	15

*Gene Expression Omnibus Accession.

4.2 Análise de preservação de módulos

A análise de preservação foi realizada comparando-se os dados de expressão dos estudos selecionados apresentados na tabela 10 aos dados das redes detectadas em NPC derivadas de iPSC, neurônios derivados de iPSC e em cérebro *post-mortem*. Como citado anteriormente, a definição dos módulos desregulados em TEA e DI, portanto, para os quais se avaliou a preservação nos tecidos acessíveis, foi baseada em estudos já realizados pelo grupo e outros dados da literatura (Tabela 5).

Os resultados das análises de preservação foram descritos separadamente por tecido de interesse nas seções de 4.2.1 a 4.2.4. Um sumário dos resultados por módulo de interesse é apresentado na seção 4.2.5 e, finalmente, um sumário dos resultados por rede referência encontra-se nos apêndices 1, 2, 3 e 4.

4.2.1 Papila dérmica

Não foi encontrada alta preservação de nenhum dos módulos de interesse de TEA/DI nos estudos de papila dérmica. O módulo de tradução M_{Brain}15-midnightblue e o módulo de resposta imune/glia M_{Brain}6-red, foram encontrados com preservação moderada em ambos os estudos (Figuras 9C e 9F) (M_{Brain}15-midnightblue: $Z_{score}=4.8$ para GSE31324 e $Z_{score}=2.1$ para GSE44765; M_{Brain}6-red: $Z_{score}=5.9$ para GSE31324 e $Z_{score}=2.9$ para GSE44765).

O módulo de resposta imune/glia M_{Brain}3-brown, se encontrou moderadamente preservado apenas no GSE31324 (Figura 9C) ($Z_{score}=2.4$), da mesma forma que o módulo de sinapse M_{Brain}1-turquoise, mostrou-se moderadamente preservado apenas no GSE44765 (Figura 9F) ($Z_{score}=2.2$).

Figura 9. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro *post-mortem* em células de papila dérmica

4.2.2 Urina

Em células epiteliais da urina, nenhum dos módulos relacionados à sinapse foram detectados como preservados em nenhum dos estudos analisados (Figura 10). Por outro lado, o módulo de tradução $M_{NPC10-blue}$, se mostrou com preservação moderada em ambos os estudos (Figuras 10A e 10D) ($Z_{score}=4.4$ para GSE28242 e $Z_{score}=6$ para GSE31189). Seu módulo correspondente em neurônios, $M_{Neur19-greenyellow}$ também se mostrou moderadamente preservado nos dois estudos (Figuras 10B e 10E) ($Z_{score}=2.7$ para GSE28242 e $Z_{score}=5.6$ para GSE31189). O mesmo foi encontrado para o módulo correspondente em cérebro, $M_{Brain15-midnightblue}$ (Figuras 10C e 10F) ($Z_{score}=5$ para GSE28242 e $Z_{score}=9.6$ para GSE31189).

Da mesma forma, os módulos de resposta imune/glia $M_{Brain2-blue}$, $M_{Brain3-brown}$ e $M_{Brain6-red}$ foram encontrados moderadamente preservados nos dois estudos (Figuras 10C e 10F) ($M_{Brain2-blue}$: $Z_{score}=3.3$ para GSE28242 e $Z_{score}=4$ para GSE31189; $M_{Brain3-brown}$: $Z_{score}=4.4$ para GSE28242 e $Z_{score}=3.2$ para GSE31189; $M_{Brain6-red}$: $Z_{score}=7.2$ para GSE28242 e $Z_{score}=2.5$ para GSE31189). Esses resultados evidenciam uma consistência da preservação moderada dos módulos de genes relacionados à resposta imune/glia e tradução nesse tipo celular.

Figura 10. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células epiteliais de urina

4.2.3 Sangue

Em células do sangue, o módulo de tradução $M_{NPC10-blue}$ se mostrou moderadamente preservado em quatro dos cinco estudos analisados (Figuras 11A, 11D, 11G e 11J) ($Z_{score}=3.9$ para GSE129166, $Z_{score}=3.6$ para GSE126124, $Z_{score}=4.8$ para GSE4607 e $Z_{score}=5.7$ para GSE46510). Seu módulo correspondente em neurônios, $M_{Neur19-greenyellow}$, se mostrou moderadamente preservado em células do sangue desses mesmos estudos, como mostrado nas figuras 11B, 11E, 11H e 11K ($Z_{score}=3.9$ para GSE129166, $Z_{score}=3.9$ para GSE126124, $Z_{score}=4.3$ para GSE4607 e $Z_{score}=5.6$ para GSE46510). Já o módulo correspondente em cérebro, $M_{Brain15-midnightblue}$, foi identificado como moderadamente preservado em 4 estudos (Figuras 11C, 11F, 11I e 11L) ($Z_{score}=3.5$ para GSE129166, $Z_{score}=4.8$ para GSE126124, $Z_{score}=3.7$ para GSE4607 e $Z_{score}=6.9$ para GSE46510).

Com relação aos módulos de resposta imune/glia, o módulo $M_{Brain2-blue}$ foi encontrado moderadamente preservado em todos os estudos (Figuras 11C, 11F, 11I, 11L e 11O) ($Z_{score}=6.2$ para GSE129166, $Z_{score}=3.2$ para GSE126124, $Z_{score}=3.8$ para GSE4607, $Z_{score}=5$ para GSE46510 e $Z_{score}=2.8$ para GSE131793). Já o módulo $M_{Brain3-brown}$ foi encontrado moderadamente preservado em três estudos (Figuras 11C, 11I e 11L) ($Z_{score}=3.8$ para GSE129166, $Z_{score}=3.8$ para GSE4607 e $Z_{score}=2.2$ para GSE46510). Finalmente, o módulo $M_{Brain6-red}$ foi observado como altamente preservado em um dos três estudos e moderadamente preservado em dois estudos (Figuras 11F, 11I, 11O) ($Z_{score}=2.5$ para GSE126124, $Z_{score}=4.4$ para GSE4607 e $Z_{score}=14$ para GSE131793).

Já para os módulos relacionados a sinapse, foi observada preservação moderada apenas para o $M_{Brain1-turquoise}$ em 3 estudos (Figuras 11C, 11F e 11L) ($Z_{score}=4$ para GSE129166, $Z_{score}=2.3$ para GSE126124 e $Z_{score}=2.2$ para GSE46510).

Esses resultados evidenciam uma consistência da preservação moderada dos módulos relacionados a resposta imune/glia e, particularmente, do módulo relacionado à tradução nesse tipo celular.

...continuação

continua...

...continuação

Figura 11. Plotagem da análise de preservação das redes detectadas em células progenitoras neurais, neurônios derivados de iPSC e cérebro post-mortem em células do sangue

4.2.4 Cordão umbilical

Não foi encontrada preservação para nenhum dos módulos relacionados a sinapse e tradução nos estudos de cordão umbilical. O módulo de resposta imune/glia, $M_{\text{Brain}2\text{-blue}}$, foi encontrado moderadamente preservado em dois dos quatro estudos analisados (Figuras 12F e 12I) ($Z_{\text{score}}=3.1$ para GSE12302 e $Z_{\text{score}}=4.4$ para GSE60403). Os outros módulos associados a resposta imune/glia, $M_{\text{Brain}6\text{-red}}$ e $M_{\text{Brain}3\text{-brown}}$, também foram encontrados moderadamente preservados em um estudo cada (Figuras 12C e 12I) ($M_{\text{Brain}6\text{-red}}$: $Z_{\text{score}}=2.5$ para GSE48354; $M_{\text{Brain}3\text{-brown}}$: $Z_{\text{score}}=2.3$ para GSE60403).

NPC

Neurônio

Cérebro post-mortem

GSE48354

GSE123302

continua...

4.2.5 Sumário dos resultados de preservação por módulo de interesse

Os valores de Z_{score} para os estudos em que foi observada preservação moderada ou alta são mostrados na tabela 12.

Tabela 12. Resultado da preservação por módulos de interesse

Módulo		Papila dérmica		Urina		Sangue				Cordão umbilical				
		GSE31324	GSE44765	GSE28242	GSE31189	GSE129166	GSE126124	GSE131773	GSE4607	GSE46510	GSE60403	GSE51546	GSE48354	GSE123302
Tradução	M _{NPC} 10-blue			4.4	6	3.9	3.6		4.8	5.7				
	M _{Neur} 19-greenyellow			2.7	5.6	3.9	3.9		4.3	5.6				
	M _{brain} 15-midnightblue	4.8	2.1	5	9.6	3.5	4.8		3.7	6.9				
Sinapse	M _{Neur} 1-turquoise													
	M _{brain} 1-turquoise		2.2			4	2.3			2.2				
	M _{brain} 23-darkturquoise													
	M _{brain} 31-paleturquoise													
Resposta Imune/Glia	M _{brain} 2-blue			3.3	4	6.2	3.2	2.8	3.8	5	4.4			3.1
	M _{brain} 3-brown	2.4		4.4	3.2	3.8			3.8	2.2	2.3			
	M _{brain} 6-red	5.9	2.9	7.2	2.5		2.5	14	4.4				2.5	

5 DISCUSSÃO

Os consistentes achados de desregulação de módulos de co-expressão em tecido *post-mortem* e em células neuronais derivadas de iPSCs de pacientes com TEA e a investigação de potenciais módulos desregulados associados a DI motivaram a investigação de se esses mesmos módulos estariam preservados em tecidos acessíveis, o que permitiria o uso de um modelo celular mais simples para mensuração dessas diferenças de expressão gênica, facilitando a transposição dessa abordagem para prática clínica.

A abordagem de análises de redes de co-expressão além de permitir a identificação de módulos em um dado transcriptoma, permite também, por meio de análises matemáticas, comparar duas redes de co-expressão distintas, avaliando suas semelhanças. Assim, essa ferramenta se torna muito interessante para verificar se tecidos diferentes podem refletir os mesmos padrões de co-expressão.

De fato, a análise de preservação de módulos tem sido utilizada em vários estudos como estratégia de validação entre diferentes dados de transcriptoma associados a uma mesma doença.^(71–74) A utilização dessa abordagem para verificar a preservação de módulos de co-expressão entre diferentes tecidos, como objetivado nesse trabalho, também já foi aplicada com sucesso em outros trabalhos.^(75–77)

A busca dos dados de transcriptoma de células do sangue, células-tronco de papila dérmica, células sanguíneas de cordão umbilical, células epiteliais derivadas da urina ou células neuronais derivadas destes tipos celulares por meio de protocolos de diferenciação direta exigiu leitura minuciosa de cada estudo e controle para a seleção dos quais fariam parte das análises posteriores. Dentre os critérios de exclusão, observamos que a maioria dos estudos apresentavam número amostral pequeno, o que limitou o número de estudos elegíveis de forma não prevista. Tal fato aconteceu principalmente para os estudos de conversão direta, dentre os quais nenhum estudo obedeceu ao número amostral adequado para as análises. Além disso, estudos com informações insuficientes para conversão dos IDs das *probes*, células que sofreram tratamento, tipo celular inadequado e dados não-transcriptoma foram observados em alguns estudos levantados, o que levou a exclusão dos mesmos.

Foram investigados se os módulos de interesse (desregulados em TEA e/ou DI) associados à sinapse, tradução e resposta imune/glia se apresentavam

preservados em modelos celulares mais acessíveis. Com relação ao módulo associado à sinapse e canais controlados por voltagem, por sua alta especificidade funcional, era pouco esperado que se apresentasse preservado nos modelos selecionados, embora tenhamos observado preservação moderada do módulo $M_{\text{Brain1-turquoise}}$ em 3 estudos de sangue. Para este módulo em particular, de fato, o modelo mais interessante provavelmente seria o de diferenciação direta. Infelizmente, os dados de estudos com células neuronais derivadas deste modelo não puderam ser usados em nossas análises dada a baixa quantidade de amostras. Consideramos, no entanto, que este tipo de modelo celular, por seu método mais fácil de obtenção do que por diferenciação a partir de iPSC, que requer o passo adicional de reprogramação, é um ótimo candidato que merece ser investigado assim que dados de transcriptoma mais robustos estiverem disponíveis.

Com relação aos módulos relacionados à tradução, observamos índices de preservação moderada para os 3 módulos ($M_{\text{NPC10-blue}}$, $M_{\text{Neur19-greenyellow}}$ e $M_{\text{Brain15-midnightblue}}$) com consistência desse resultado em urina e sangue. Nota-se, no entanto, que estes índices variaram, sendo, por exemplo, mais elevados nos estudos GSE31189 (urina) e GSE46510 (sangue). Estas variações podem ser espúrias, ou por outro lado, podem ter relação com a maneira com a qual as células foram isoladas nos diferentes estudos. De fato, para o estudo GSE31189, uma parte das amostras foram coletadas via cateter e outra parte como jato médio. Além disso, o processamento das amostras deste estudo foi diferente em relação aos outro estudo de urina, visto que as amostras foram centrifugadas e separadas em *pellet* celular e sobrenadante antes do armazenamento a -80°C , o que pode ter gerado alterações nos níveis de expressão para essas amostras. O autor deste estudo discute que processamento de células de urina através de diferentes protocolos pode apresentar erros técnicos e consequentemente resultados diferentes.⁽⁷⁸⁾ Em relação aos estudos de sangue, também foram revisados os protocolos de processamento das amostras e verificamos que todos realizaram a extração de RNA com protocolo semelhante e a partir de sangue total, com exceção do estudo GSE131793, no qual a extração foi realizada a partir de células mononucleares de sangue periférico (PBMC). A única peculiaridade do estudo GSE46510, o estudo onde se encontrou maior índice de preservação dentre aqueles realizados com sangue, é que este foi feito apenas com amostras de mulheres.

Curiosamente, estes módulos de tradução possuem sobreposição significativa com um módulo identificado como desregulado em sangue de pacientes autistas em uma metanálise realizada por Griesi-Oliveira et al.⁽⁵²⁾ e Tylee et al.⁽⁶⁴⁾ (dados não mostrados). Embora essa sobreposição já tivesse sido apontada em um estudo prévio do nosso grupo, importante ressaltar que no presente trabalho, pelo tipo de análise aplicada, demonstramos também que os padrões de co-expressão entre estes genes se mantêm de forma moderada nos dois tecidos.⁽⁵²⁾ Além disso, interessante notar que nossas análises corroboram outros trabalhos que já haviam demonstrado que há preservação de módulos de co-expressão neuronais em células de sangue,^(58,79) mas também adicionam as células epiteliais da urina como outro modelo celular que pode refletir estes módulos.

Finalmente, com relação aos módulos de resposta imune, todos mostraram preservação moderada em células epiteliais de urina. Já em sangue, embora os módulos $M_{\text{Brain}3\text{-brown}}$ e $M_{\text{Brain}6\text{-red}}$ tenham mostrado moderada preservação em 3 dos 5 estudos, observamos uma preservação moderada mais consistente para o módulo $M_{\text{Brain}2\text{-blue}}$, que mostrou este resultado nos cinco estudos. Novamente, observamos que esses índices variam sendo mais elevados nos estudos GSE31189 (urina) e GSE46510 e GSE129166 (sangue). Novamente, essas diferenças podem ser apenas ao acaso ou se dever a alguma particularidade da forma de isolamento e extração dessas células. No caso deste módulo, não há uma consistência nas formas de extração ou composição de amostras semelhante entre os estudos GSE129166 e GSE46510 e que se mostre diferente em comparação aos outros três estudos. No entanto, acreditamos ser importante citar que o estudo GSE46510 foi realizado apenas amostras de mulheres, diferentes dos demais que utilizaram amostras de ambos os sexos.

No geral, observamos preservação moderada dos módulos de tradução e resposta imune em células epiteliais de urina e sangue, apontando estes tecidos como potenciais fontes para análise dos níveis de expressão dos genes desses módulos como forma de distinguir indivíduos com TEA/DI de controles. No entanto, é importante ressaltar que esta preservação é moderada, o que pode inviabilizar a observação dessas diferenças. Ainda assim, faz-se importante como próximo passo uma validação nesse sentido, comparando a expressão destes genes nesses grupos a fim de se entender se de fato esta estratégia poderá identificar indivíduos com TEA/DI. Assim como neste trabalho, realizado inteiramente com dados disponíveis

publicamente, a validação, pelo menos para o tecido sanguíneo, também pode ser realizada utilizando-se dados de expressão gênica de células de sangue de pacientes autistas disponíveis nestes mesmos bancos (não há até o momento dados de transcriptoma de células epiteliais da urina de autistas). Embora análises de módulos de co-expressão gênica em células sanguíneas de indivíduos com TEA já tenham sido realizadas por outros grupos, como inclusive citamos aqui anteriormente, uma alternativa a ser testada é considerar para a análise não do dado de expressão de todo o transcriptoma, mas apenas os genes que fazem parte dos M_{NPC10} -blue, M_{Neur19} -greenyellow ou $M_{Brain15}$ -midnightblue que estão expressos no sangue. Os níveis de expressão desses genes poderão então ser sumarizados em um único valor para cada indivíduo, através de análise de componente principal, e então os valores obtidos para pacientes comparados aos dos controles. Esta validação se mostra relevante pois, de fato, se os níveis de expressão dos genes desses módulos nesses indivíduos puderem distinguir pacientes dos transtornos em questão de controles, esta investigação abre caminho para o potencial de aplicação clínica desta estratégia.

6 CONCLUSÕES

1. Foram obtidos dados de expressão de todos os tecidos de interesse tais como células do sangue, células-tronco de papila dérmica, células sanguíneas de cordão umbilical e células epiteliais derivadas da urina com exceção de células neuronais derivadas destes tipos celulares através de protocolos de diferenciação direta pois não atendeu aos critérios de elegibilidade estabelecidos. Dessa forma, obteve-se o total de 13 estudos elegíveis para as etapas anteriores que incluíram 2 estudos de papila dérmica, 2 de urina, 5 de sangue e 4 de cordão umbilical;

2. Na análise de preservação de módulos de expressão dos tecidos de interesse com dados de células neuronais e cérebro *post-mortem* foram identificados a preservação moderada de dois módulos de interesse, ou seja, módulos desregulados em células neuronais de indivíduos com deficiência intelectual e/ou transtorno do espectro autista em células do sangue e urina associados a tradução e resposta imune. Este resultado sugere que a análise de expressão destes genes em células epiteliais de urina ou de sangue poderiam refletir a mesma desregulação observada nas células neuronais e de cérebro *post-mortem* de pacientes com deficiência intelectual e/ou transtorno do espectro autista, tendo assim um potencial de aplicação para diagnóstico destes transtornos.

7 REFERÊNCIAS

1. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. *MMWR Surveill Summ.* 2018;67(6):1–23.
2. Coll-Tane M, Krebbers A, Castells-Nobau A, Zweier C, Schenck A. Intellectual disability and autism spectrum disorders 'on the fly': insights from *drosophila*. *Dis Model Mech.* 2019;12(5):dmm039180.
3. Rylaarsdam L, Guemez-Gamboa A. Genetic causes and modifiers of autism spectrum disorder. *Front Cell Neurosci.* 2019;13:385.
4. Thurm A, Farmer C, Salzman E, Lord C, Bishop S. State of the field: differentiating intellectual disability from autism spectrum disorder. *Front Psychiatry.* 2019;10:526.
5. Mulas F, Rojas M. [Intellectual developmental disability overlapping with autism spectrum disorder and attention deficit-hyperactivity disorder]. *Medicina (B Aires).* 2018;78 Suppl 2:63–8. Spanish.
6. Selau T, da Silva MA, de Mendonça Filho EJ, Bandeira DR. Evidence of validity and reliability of the adaptive functioning scale for intellectual disability (EFA-DI). *Psicol Reflex Crit.* 2020;33(1):26.
7. Santos-Cortez RL, Khan V, Khan FS, Mughal ZU, Chakchouk I, Lee K, et al. Novel candidate genes and variants underlying autosomal recessive neurodevelopmental disorders with intellectual disability. *Hum Genet.* 2018;137(9):735–52.
8. Caballero Pérez V, López Pisón FJ, Miramar Gallart MD, González Álvarez A, García Jiménez MC, García Iñiguez JP, et al. Phenotype in patients with intellectual disability and pathological results in array CGH. *Neurologia.* 2017;32(9):568–78.
9. Masi A, DeMayo MM, Glozier N, Guastella AJ. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neurosci Bull.* 2017;33(2):183–93.
10. Mukherjee SB. Autism spectrum disorders — diagnosis and management. *Indian J Pediatr.* 2017;84(4):307–14.
11. Li D, Karnath HO, Xu X. Candidate biomarkers in children with autism spectrum disorder: a review of MRI studies. *Neurosci Bull.* 2017;33(2):219–37.
12. Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. *Lancet.* 2018;392(10146):508–20.
13. Gyawali S, Patra BN. Autism spectrum disorder: trends in research exploring etiopathogenesis. *Psychiatry Clin Neurosci.* 2019;73(8):466-75.
14. Aspromonte MC, Bellini M, Gasparini A, Carraro M, Bettella E, Polli R, et al. Characterization of intellectual disability and autism comorbidity through gene panel sequencing. *Hum Mutat.* 2019;40(9):1346–63.
15. Golden CE, Buxbaum JD, De Rubeis S. Disrupted circuits in mouse models of autism spectrum disorder and intellectual disability. *Curr Opin Neurobiol.* 2018;48:106–12.
16. Li C, Zhou H, Wang T, Long S, Du X, Xu X, et al. Performance of the autism spectrum rating scale and social responsiveness scale in identifying autism spectrum disorder among cases of intellectual disability. *Neurosci Bull.* 2018;34(6):972–80.

17. Kasilingam N, Waddington H, Van Der Meer L. Early intervention for children with autism spectrum disorder in New Zealand: what children get and what parents want. *Intl J Disabil Dev Educ.* 2021;68(4):521-37.
18. Lai DC, Chiang CH, Hou YM, Liu JH, Yao SF, Guo HR, et al. Predictors of effectiveness of early intervention on children with intellectual disability: a retrospective cohort study. *BMC Pediatr.* 2014;14(1):170.
19. Ronski M, Sevcik RA, Barton-Hulsey A, Whitmore AS. Early intervention and AAC: what a difference 30 years makes. *Augment Altern Commun.* 2015;31(3):181–202.
20. Sweet RA, Henteleff RA, Zhang W, Sampson AR, Lewis DA. Reduced dendritic spine density in auditory cortex of subjects with schizophrenia. *Neuropsychopharmacology.* 2009;34(2):374–89.
21. Bruel AL, Vitobello A, Tran Mau-Them F, Nambot S, Sorlin A, Denommé-Pichon AS, et al. Next-generation sequencing approaches and challenges in the diagnosis of developmental anomalies and intellectual disability. *Clin Genet.* 2020;98(5):433-44.
22. Han JY, Lee IG. Genetic tests by next-generation sequencing in children with developmental delay and/or intellectual disability. *Clin Exp Pediatr.* 2020;63(6):195–202.
23. Martínez F, Caro-Llopis A, Roselló M, Oltra S, Mayo S, Monfort S, et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J Med Genet.* 2017;54(2):87–92.
24. Griesi-Oliveira K, Sertie AL. Autism spectrum disorders : an updated guide for genetic counseling. *Einstein (Sao Paulo).* 2017;15(2):233–8.
25. Harripaul R, Noor A, Ayub M, Vincent JB. The use of next-generation sequencing for research and diagnostics for intellectual disability. *Cold Spring Harb Perspect Med.* 2017;7(3):a026864.
26. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci.* 2015;16(9):551–63.
27. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 2016;17(1):9–18.
28. Poot M, van der Smagt JJ, Brilstra EH, Bourgeron T. Disentangling the myriad genomics of complex disorders, specifically focusing on autism, epilepsy, and schizophrenia. *Cytogenet Genome Res.* 2011;135(3-4):228–40.
29. Srivastava AK, Schwartz CE. Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neurosci Biobehav Rev.* 2014;46(Pt 2):161–74.
30. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol.* 2011;7(1):e1001057.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(1):559.
32. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4(1):e17.
33. Parikshak NN, Gandal MJ, Geschwind DH, Angeles L, Angeles L, Angeles L. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet.* 2015;16(8):441–58.

34. Aldinger KA, Timms AE, MacDonald JW, McNamara HK, Herstein JS, Bammler TK, et al. Transcriptome data of temporal and cingulate cortex in the Rett syndrome brain. *Sci Data*. 2020;7(1):192.
35. Deng V, Matagne V, Banine F, Frerking M, Ohliger P, Budden S, et al. FXRD1 is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Hum Mol Genet*. 2007;16(6):640–50.
36. Lin P, Nicholls L, Assareh H, Fang Z, Amos TG, Edwards RJ, et al. Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q Genes in Rett syndrome. *BMC Genomics*. 2016;17(1):427.
37. Lockstone HE, Harris LW, Swatton JE, Wayland MT, Holland AJ, Bahn S. Gene expression profiling in the adult Down syndrome brain. *Genomics*. 2007;90(6):647–60.
38. Pelleri MC, Cattani C, Vitale L, Antonaros F, Strippoli P, Locatelli C, et al. Integrated Quantitative Transcriptome Maps of Human Trisomy 21 Tissues and Cells. *Front Genet*. 2018;9:125.
39. Olmos-Serrano JL, Kang HJ, Tyler WA, Silbereis JC, Cheng F, Zhu Y, et al. Down syndrome developmental brain transcriptome reveals defective oligodendrocyte differentiation and myelination. *Neuron*. 2016;89(6):1208–22.
40. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun*. 2014;5(1):5748.
41. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 2016;540(7633):423–7.
42. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011;474(7351):380–4.
43. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131(5):861–72.
44. Costamagna G, Andreoli L, Corti S, Faravelli I. iPSCs-based neural 3D systems: a multidimensional approach for disease modeling and drug discovery. *Cells*. 2019;8(11):1438.
45. Griesi-Oliveira K, Sunaga DY, Alvizi L, Vadasz E, Passos-Bueno MR. Stem cells as a good tool to investigate dysregulated biological systems in autism spectrum disorders. *Autism Res*. 2013;6(5):354–61. Erratum in: *Autism Res*. 2015;8(1):121.
46. Logan S, Arzua T, Canfield SG, Seminary ER, Sison SL, Ebert AD, et al. Studying human neurological disorders using induced pluripotent stem cells: from 2D monolayer to 3D organoid and blood brain barrier models. *Compr Physiol*. 2019;9(2):565–611.
47. Park GH, Noh H, Shao Z, Ni P, Qin Y, Liu D, et al. Activated microglia cause metabolic disruptions in developmental cortical interneurons that persist in interneurons from individuals with schizophrenia. *Nat Neurosci*. 2020;23(11):1352–64.
48. Wang M, Zhang L, Gage FH. Modeling neuropsychiatric disorders using human induced pluripotent stem cells. *Protein Cell*. 2020;11(1):45–59.
49. DeRosa BA, El Hokayem J, Artimovich E, Garcia-serje C, Phillips AW, Van Booven D, et al. Convergent pathways in idiopathic autism revealed by time course transcriptomic analysis of patient- derived neurons. *Sci Rep*. 2018;8(1):8423.

50. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, et al. FOXP1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell*. 2015;162(2):375–90.
51. Schafer ST, Paquola AC, Stern S, Gosselin D, Ku M, Pena M, et al. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat Neurosci*. 2019;22(2):243–55.
52. Griesi-Oliveira K, Fogo MS, Pinto BG, Alves AY, Suzuki AM, Morales AG, et al. Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. *Mol Psychiatry*. 2020 [Epub ahead of print].
53. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
54. Lombardo MV. Ribosomal protein genes in post-mortem cortical tissue and iPSC-derived neural progenitor cells are commonly upregulated in expression in autism. *Mol Psychiatry*. 2020. [Epub ahead of print].
55. Lombardo MV, Courchesne E, Lewis NE, Pramparo T. Hierarchical cortical transcriptome disorganization in autism. *Mol Autism*. 2017;8(1):29.
56. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018;359(6376):693–7.
57. Pinto BG. Transcriptômica de autismo e deficiências intelectuais: uma abordagem de biologia de sistemas para identificar especificidades e convergências [dissertação]. São Paulo: Universidade de São Paulo, Instituto de Matemática; 2020.
58. Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, Van den Berg L, et al. Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics*. 2010;11:589.
59. Amoh Y, Mii S, Aki R, Hamada Y, Kawahara K, Hoffman RM, et al. Multipotent nestin-expressing stem cells capable of forming neurons are located in the upper, middle and lower part of the vibrissa hair follicle. *Cell Cycle*. 2012;11(18):3513–7.
60. Hunt DP, Morris PN, Sterling J, Anderson JA, Joannides A, Jahoda C, et al. A highly enriched niche of precursor cells with neuronal and glial potential within the hair follicle dermal papilla of adult skin. *Stem Cells*. 2008;26(1):163–72.
61. Zhang M, Lin YH, Sun YJ, Li K, Lin Y, Sun YJ, et al. Pharmacological reprogramming of fibroblasts into neural stem cells by signaling-directed transcriptional activation. *Cell Stem Cell*. 2016;18(5):653–67.
62. Griesi-Oliveira K, Suzuki AM, Alves AY, Mafra AC, Yamamoto GL, Ezquina S, et al. Actin cytoskeleton dynamics in stem cells from autistic individuals. *Sci Rep*. 2018;8(1):11138.
63. Suzuki AM, Griesi-Oliveira K, de Oliveira Freitas Machado C, Vadasz E, Zachi EC, Passos-Bueno MR, et al. Altered mTORC1 signaling in multipotent stem cells from nearly 25% of patients with nonsyndromic autism spectrum disorders. *Mol Psychiatry*. 2015;20(5):551–2.
64. Tylee DS, Hess JL, Quinn TP, Barve R, Huang H, Zhang-James Y, et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: a combined-samples mega-analysis. *Am J Med Genet B Neuropsychiatr Genet*. 2017;174(3):181–201.

65. Wang L, Wang L, Huang W, Su H, Xue Y, Su Z, et al. Generation of integration-free neural progenitor cells from cells in human urine. *Nat Methods*. 2013;10(1):84–9.
66. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
67. Larriba Y, Rueda C, Fernandez MA, Peddada SD. Microarray data normalization and robust detection of rhythmic features. 2019;1986:207–25.
68. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–93.
69. Harvard University. Data analysis for genomics. Cambridge (MA): edX; c2021 [cited 2021 May 19]. Available from: <https://courses.edx.org/courses/HarvardX/PH525x/1T2014/course/#>
70. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 2018;19(4):575–92.
71. Chen J, Wang X, Hu B, He Y, Qian X, Wang W. Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. *PeerJ*. 2018;6:e4692.
72. Wei S, Chen J, Huang Y, Sun Q, Wang H, Liang X, et al. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. *J Cell Physiol*. 2020;235(3):2037–48.
73. Xu P, Yang J, Liu J, Yang X, Liao J, Yuan F, et al. Identification of glioblastoma gene prognosis modules based on weighted gene co-expression network analysis. *BMC Med Genomics*. 2018;11(1):96.
74. Yuan L, Qian K, Qian G, Zhu Y. Identification of biomarkers associated with pathological stage and prognosis of clear cell renal cell carcinoma by co-expression network analysis. *Front Physiol*. 2018;9:399.
75. Broadbent J, Sampson D, Sabapathy S, Haseler LJ, Wagner KH, Bulmer AC, et al. Gene networks in skeletal muscle following endurance exercise are coexpressed in blood neutrophils and linked with blood inflammation markers. *J Appl Physiol* (1985). 2017;122(4):752–66.
76. Mina E, van Roon-Mom W, Hettne K, van Zwet E, Goeman J, Neri C, et al. Common disease signatures from gene expression analysis in Huntington's disease human blood and brain. *Orphanet J Rare Dis*. 2016;11(1):97.
77. Neidlin M, Dimitrakopoulou S, Alexopoulos LG. Multi-tissue network analysis for drug prioritization in knee osteoarthritis. 2019;9(1):15176.
78. Urquidi V, Goodison S, Cai Y, Sun Y, Rosser CJ. A candidate molecular biomarker panel for the detection of bladder cancer. *Cancer Epidemiol Biomarkers Prev*. 2012;21(12):2149–58.
79. Pramparo T, Lombardo M V, Campbell K, Barnes CC, Marinero S, Solso S, et al. Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol*. 2015;11(12):841.

Abstract

Introduction: Intellectual disability and autism spectrum disorders affect about 3% of the population and it can be found as single or combined conditions. A dysregulation of a network of co-expressed genes related to synapse has been consistently observed in neurons of autism spectrum disorder patients by several study groups including our group. Finding have suggested that the expression profile of these genes could be used as a biomarker for this disorder. **Purpose:** To identify more accessible cell models that can reflect these same co-expression modules dysregulated in neuronal cells of patients with autism spectrum disorder and/or intellectual disability. **Methods:** To identify if the modules of interest are preserved in any of these cellular models we used module preservation analyzes. We compared the transcriptome data from neuronal cells (derived from pluripotent stem cells - iPSCs or *post mortem* brain) with publicly available transcriptome data of blood, dermal papilla stem cells and urine derived epithelial cells, or neuronal cells derived from these cell types through direct differentiation protocols. **Results:** After searching for transcriptome data of the tissues of interest, a total of 13 studies were obtained for the analyzes (2 of dermal papilla, 2 of urine epithelial cells, 5 of blood, and 4 of umbilical cord). A module related to unregulated translation in neural progenitors derived from iPSC and the *post-mortem* brain were observed moderately preserved in four blood studies and two urine studies. All immune response modules dysregulated in neuronal cells derived from iPSC and *post-mortem* brain were observed with moderate preservation in two urine studies. In blood, two of these modules achieved moderate preservation in three studies and one module also had moderate preservation in five studies. The preservation consistency of this module in expression data from different studies indicates that the expression relationships of the genes in this module identified in neuronal cells and *post-mortem* brain were maintained in these other tissues. **Conclusion:** Our results suggest that the analysis of expression of these genes in urine or blood epithelial cells could reflect the same dysregulation observed in the neuronal cells and *post-mortem* brain of autism spectrum disorder/intellectual disability patients, therefore, reveling a potential application as a biomarker for these disorders.

Keywords: Autism spectrum disorder; Intellectual disability; Transcriptome; Gene regulatory networks

Apêndices

Apêndice 1. Resultado da análise de preservação de módulos de células progenitoras neurais nos tecidos testados

Análise de preservação - rede referência: NPC derivadas de iPSC*								
GSE	N° amostral	Tecido	Módulo	Moderadamente preservado (Z _{score})	log.psummary	Módulo	Altamente preservado (Z _{score})	log.psummary
GSE31324	12	papila	<i>brown</i>	5.4	-7.9			
			<i>purple</i>	8.1	-16.8			
			<i>greenyellow</i>	9.5	-24.3			
GSE44765	18	papila	<i>greenyellow</i>	3.6	-3.9	<i>yellow</i>	11	-26.9
			<i>purple</i>	7.4	14.0			
GSE28242	13	urina	<i>green</i>	3.6	-4.8			
			blue	4.4	-5.5			
			<i>yellow</i>	8.4	-18.0			
GSE31189	40	urina	blue	6	-10.5	<i>yellow</i>	10	-23.2
			<i>purple</i>	6.3	-11.0			
			<i>green</i>	7.3	-16.9			
GSE129166	106	sangue	<i>green</i>	2.6	-3.1	<i>yellow</i>	14	-49.1
			blue	3.9	-4.7			
			<i>magenta</i>	6.3	-10.0			
GSE126124	98	sangue	blue	3.6	-4.3	<i>yellow</i>	15	-30.6
			<i>green</i>	5.3	-7.4			
			<i>magenta</i>	5.6	-8.5			
GSE131793	18	sangue	<i>turquesa</i>	2.6	-2.8			
			<i>yellow</i>	2.9	-5.0			
			<i>purple</i>	3.8	-5.0			

continua...

...continuação

GSE4607	121	sangue	<i>turquesa</i>	2.5	-2.3	<i>yellow</i>	14	-50.8
			<i>green</i>	3	-3.0			
			blue	4.8	-6.3			
			<i>magenta</i>	5.5	-8.7			
GSE46510	151	sangue	<i>magenta</i>	2.6	-2.7	<i>yellow</i>	16	-63.8
			<i>green</i>	4.3	-6.0			
			<i>purple</i>	4.8	-6.1			
			blue	5;7	-9.0			
GSE60403	11	cordão umbilical	<i>magenta</i>	4.3	-5.0	<i>yellow</i>	10	-25.5
GSE51546	15	cordão umbilical						
GSE48354	34	cordão umbilical	<i>yellow</i>	5.6	-8.4			
GSE123302	219	cordão umbilical	<i>magenta</i>	3.6	-2.5			
			<i>yellow</i>	7.8	-4.5			

*Em negrito, encontram-se destacados os módulos de interesse, ou seja, os módulos detectados como diferencialmente expressos em TEA/DI. **Valor em log que sumariza os valores de Z_{scores} individuais. Quanto maior esse valor, maior evidência de preservação.
Baseado em: *Griesi-Oliveira K, Fogo MS, Pinto BG, Alves AY, Suzuki AM, Morales AG, et al. Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. Mol Psychiatry. 2020. [Epub ahead of print].⁽⁵²⁾

Apêndice 2. Resultado da análise de preservação de módulos de células neuronais nos tecidos testados

Análise de preservação - rede referência: neurônios derivados de IPSC*								
GSE	N° amostral	Tecido	Módulo	Moderadamente preservado (Z _{score})	log.psummary	Módulo	Altamente preservado (Z _{score})	log.psummary
GSE31324	12	papila	<i>midnightblue</i>	6.9	-12.4	<i>brown</i>	14	-74.8
			<i>pink</i>	7.8	-18.8			
GSE44765	18	papila	<i>purple</i>	3	-4.0	<i>brown</i>	14	-81.3
			<i>pink</i>	4.5	-6.0			
			<i>royalblue</i>	4.6	-7.2			
GSE28242	13	urina	greenyellow	2.7	-2.9			
			<i>royalblue</i>	6.9	-13.6			
GSE31189	40	urina	greenyellow	5.6	-16.1			
			<i>royalblue</i>	7.4	16.8			
GSE129166	106	sangue	greenyellow	3.9	-3.3	<i>royalblue</i>	10	-29.6
			<i>red</i>	4.6	-8.0			
			<i>salmon</i>	5.8	-10.7			
			<i>brown</i>	4.6	-6.5			
			<i>yellow</i>	6	-16.5			
GSE126124	98	sangue	greenyellow	3.9	-5.7	<i>royalblue</i>	11	-33.1
			<i>red</i>	5	-6.6			
			<i>yellow</i>	8.3	-27.8			
GSE131793	18	sangue				<i>brown</i>	13	-61.0
GSE4607	121	sangue	<i>purple</i>	2.4	-3.6	<i>brown</i>	11	-41.0
			<i>black</i>	2.6	-2.5	<i>royalblue</i>	12	-43.9
			<i>red</i>	3.4	-3.5			
			greenyellow	4.3	-5.5			
			<i>yellow</i>	5	-9.6			
GSE46510	151	sangue	<i>black</i>	2.5	-2.5	<i>royalblue</i>	12	-42.1

continua...

...continuação

			<i>salmon</i>	2.5	-2.8			
			<i>red</i>	3.2	-3.3			
			<i>brown</i>	4	-5.3			
			greenyellow	5.6	-10.6			
GSE60403	11	cordão umbilical	<i>red</i>	2.2	-1.9	<i>brown</i>	16	-96.1
			<i>royalblue</i>	2.6	-3.6			
			<i>yellow</i>	5	-10.3			
GSE51546	15	cordão umbilical						
GSE48354	34	cordão umbilical	<i>royalblue</i>	2.9	-4.0	<i>brown</i>	28	-306.8
			<i>cyan</i>	7.6	-15.1			
GSE123302	219	cordão umbilical	<i>cyan</i>	3.4	-4.1	<i>brown</i>	13	-49.5
			<i>yellow</i>	3.6	-8.1			

*Em negrito, encontram-se destacados os módulos de interesse, ou seja, os módulos detectados como diferencialmente expressos em TEA/DI. **Valor em log que sumariza os valores de Z_{scores} individuais. Quanto maior esse valor, maior evidência de preservação.

Baseado em: *Griesi-Oliveira K, Fogo MS, Pinto BG, Alves AY, Suzuki AM, Morales AG, et al. Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. Mol Psychiatry. 2020. [Epub ahead of print].⁽⁵²⁾

Apêndice 3. Resultado da análise de preservação de módulos de cérebro *post-mortem* nos tecidos testados

Análise de preservação - rede referência: cérebro <i>post-mortem</i> *								
GSE	N° amostral	Tecido	Módulo	Moderadamente preservado (Z _{score})	log.psummary	Módulo	Altamente preservado (Z _{score})	log.psummary
GSE31324	12	papila	<i>brown</i>	2.4	-2.4			
			<i>midnightblue</i>	4.8	-9.6			
			<i>red</i>	5.9	-14.9			
GSE44765	18	papila	<i>midnightblue</i>	2.1	-1.8			
			<i>turquesa</i>	2.2	-3.2			
			<i>red</i>	2.9	-4.6			
			<i>yellow</i>	3	-5.5			
GSE28242	13	urina	<i>blue</i>	3.3	-7.3			
			<i>brown</i>	4.4	-8.4			
			<i>midnightblue</i>	5	-12.5			
			<i>red</i>	7.2	-14.1			
GSE31189	40	urina	<i>red</i>	2.5	-2.9			
			<i>brown</i>	3.2	-5.7			
			<i>blue</i>	4	-12.2			
			<i>midnightblue</i>	9.6	-45.4			
GSE129166	106	sangue	<i>midnightblue</i>	3.5	-7.0			
			<i>brown</i>	3.8	-14.6			
			<i>turquesa</i>	4	-14.7			
			<i>blue</i>	6.2	-21.1			
GSE126124	98	sangue	<i>turquesa</i>	2.3	-2.3	violeta	11	-59.3
			<i>red</i>	2.5	-6.3			
			<i>blue</i>	3.2	-6.9			
			<i>midnightblue</i>	4.8	-11.4			
GSE131793	18	sangue	<i>blue</i>	2.8	-3.3	<i>red</i>	14	-79.7

continua...

...continuação

			<i>green</i>	6.8	-18.7	violeta	19	-203
GSE4607	121	sangue	midnightblue	3.7	-8.6			
			brown	3.8	-14.3			
			blue	3.8	-11.7			
			red	4.4	-13.4			
			<i>greenyellow</i>	7.9	-23.0			
GSE46510	151	sangue	brown	2.2	-9.1			
			turquesa	2.2	-9.8			
			blue	5	-18.8			
			midnightblue	6.9	-25.1			
GSE60403	11	cordão umbilical	brown	2.3	-4.4			
			blue	4.4	-11.3			
			<i>yellow</i>	5.2	-15.4			
GSE51546	15	cordão umbilical						
GSE48354	34	cordão umbilical	red	2.5	-4.6			
			violeta	3.8	-7.4			
GSE123302	219	cordão umbilical	blue	3.1	-5.9	<i>greenyellow</i>	10	-37.4
			<i>yellow</i>	5	-7.9			

*Em negrito, encontram-se destacados os módulos de interesse, ou seja, os módulos detectados como diferencialmente expressos em TEA/DI. **Valor em log que sumariza os valores de Z_{scores} individuais. Quanto maior esse valor, maior evidência de preservação.

Baseado em: *Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al.; PsychENCODE Consortium, Peters MA, Gerstein M, Liu C, Iakoucheva LM, Pinto D, Geschwind DH. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science. 2018;362(6420):eaat8127. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder.⁽⁵³⁾