

Paulo Victor Visintin

**DIMORFISMO SEXUAL EM AUTISMO: explorando diferenças de
expressão gênica como possível fator etiológico**

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

São Paulo

2023

Paulo Victor Visintin

**DIMORFISMO SEXUAL EM AUTISMO: explorando diferenças de
expressão gênica como possível fator etiológico**

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
Coorientadora: Dra. Tatiana Ferreira de
Almeida

São Paulo

2023

Visintin, Paulo Victor

Dimorfismo sexual em autismo : explorando diferenças de expressão gênica como possível fator etiológico / Paulo Victor Visintin. -- São Paulo, 2023.

xiii, 50 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: *Sexual dimorphism in autism: exploring gene expression differences as a possible etiological factor.*

1. Transtorno do espectro autista. 2. Expressão gênica. 3. Transcriptoma. 4. Desenvolvimento embrionário e fetal. 5. 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. Luciano Cesar Pontes de Azevedo

Paulo Victor Visintin

**DIMORFISMO SEXUAL EM AUTISMO: explorando diferenças de
expressão gênica como possível fator etiológico**

Presidente da banca: Profa. Dra. Karina Griesi Oliveira

BANCA EXAMINADORA

Membros titulares:

Profa. Dra. Patrícia Maria de Carvalho Aguiar

Dra. Ana Carolina Tahirá

Dra. Daniele Yumi Sunaga-Franze

Membros suplentes:

Profa. Dra. Patricia Severino

Dra. Anelisa Gollo Dantas

Aprovada em: 03/07/2023.

Agradecimentos

A Deus, pelas oportunidades a mim oferecidas até o presente momento.

Aos meus pais, Valcir e Eunice e demais familiares, por todo ensinamento, apoio e amor durante toda a minha caminhada. Sendo os pilares da minha construção como pessoa.

A minha orientadora, Dra. Karina Griesi Oliveira, por todo conhecimento, experiência e tempo já investido na minha formação. Sua dedicação e paciência em me instruir foi e está sendo fundamental para o meu aperfeiçoamento profissional e pessoal.

Aos amigos que ganhei no decorrer deste trabalho, em especial as minhas ex-colegas de grupo Camilly Erica e Bruna Garbes, pelas conversas e momentos de companheirismo que tornam a caminhada acadêmica mais enriquecedora.

Ao Instituto Israelita de Ensino e Pesquisa Albert Einstein, por disponibilizar uma estrutura física de excelência e ainda mais importante, por proporcionar o contato e amizade com colegas e profissionais incríveis.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão de bolsa de mestrado a mim concedida e indispensável para a minha participação neste projeto (número do processo: 2021/08293-3).

"O importante é não parar de questionar; a curiosidade tem sua própria razão de existir.
Não se pode deixar de se maravilhar ao contemplar os mistérios da eternidade, da
vida, da maravilhosa estrutura da realidade."

Albert Einstein

Sumário

Agradecimentos.....	v
Lista de figuras	ix
Lista de tabelas	x
Lista de abreviaturas	xi
Resumo	xii
1 INTRODUÇÃO	1
1.1 O transtorno do espectro autista	1
1.2 Dimorfismo sexual em transtorno do espectro autista e atuais teorias.....	2
1.3 Expressão gênica como possível mecanismo relacionado ao dimorfismo sexual em autistas	3
1.4 Objetivos.....	8
2 MÉTODOS.....	10
2.1 Bancos de dados.....	10
2.1.1 Células neuronais fetais <i>post-mortem</i> de indivíduos não diagnosticados com transtorno do espectro autista	10
2.1.2 Neurônios e células progenitoras neurais derivados de células-tronco pluripotentes induzidas de indivíduos não diagnosticados com transtorno do espectro autista.....	11
2.2 Análise de clusterização das amostras de tecido cerebral e seleção das janelas espaço-temporais homogêneas	11
2.3 Filtro de exclusão para genes com baixa expressão e com baixo coeficiente de variação	14
2.4 Análises de controle de qualidade das amostras	15
2.5 Análise de redes de coexpressão gênica: identificação de diferenças estruturais ou níveis de expressão entre homens e mulheres nos módulos de coexpressão.....	16
2.6 Análise de enriquecimento de genes associados ao transtorno do espectro autista	18
2.7 Análise da distribuição topológica dos genes e variantes de risco para o transtorno do espectro autista nos módulos de coexpressão de interesse	19
2.8 Análise de enriquecimento funcional	20
2.9 Análise de sobreposição dos módulos de interesse com módulos de outros estudos	20
2.10 Aspectos éticos	21

3 RESULTADOS	22
3.1 Análises de coexpressão gênica das amostras corticais de 12 a 13 semanas gestacionais.....	22
3.1.1 Diferenças estruturais e de expressão gênica entre os sexos.....	22
3.1.2 Identificação de módulos enriquecidos de genes associados ao transtorno do espectro autista	25
3.1.3 Análise da distribuição topológica dos genes associados ao transtorno do espectro autista nos módulos de coexpressão de interesse	27
3.1.4 Análise de enriquecimento funcional	30
3.1.5 Análise de sobreposição entre módulos <i>consensus</i> enriquecidos e módulos implicados previamente no transtorno do espectro autista.....	30
3.2 Análises de coexpressão gênica das amostras corticais de 16 a 21 semanas gestacionais.....	31
3.2.1 Diferenças estruturais e de expressão gênica entre os sexos.....	31
3.2.2 Identificação de módulos enriquecidos de genes associados ao transtorno do espectro autista	34
3.2.3 Análise da distribuição topológica dos genes de risco para o transtorno do espectro autista nos módulos de coexpressão de interesse	37
3.2.4 Análise de enriquecimento funcional	40
3.2.5 Análise de sobreposição entre módulos <i>consensus</i> enriquecidos e módulos implicados previamente no transtorno do espectro autista.....	40
3.3 Análises de coexpressão gênica de neurônios e células progenitoras neurais derivadas de células-tronco pluripotentes induzidas	41
4 DISCUSSÃO	42
5 CONCLUSÕES	47
6 REFERÊNCIAS.....	48
Abstract	
Apêndices	

Lista de figuras

Figura 1. Dendrogramas em <i>cluster</i> de módulos para amostras corticais de 12-13 semanas gestacionais	23
Figura 2. Análise de sobreposição entre módulos masculinos e femininos para amostras corticais de 12-13 semanas gestacionais	24
Figura 3. Análise de correlação dos níveis de expressão dos módulos <i>consensus</i> e os traços biológicos: idade, sexo e regiões cerebrais para amostras corticais de 12-13 semanas gestacionais	25
Figura 4. Análise de sobreposição entre módulos enriquecidos em cada sexo com módulos enriquecidos consensuais entre os sexos para amostras corticais de 12-13 semanas gestacionais	26
Figura 5. Análise da distribuição topológica entre genes SFARI e não SFARI, TADA e não TADA, e entre os genes mutados exclusivamente em um sexo ou em ambos nos módulos de coexpressão de interesse	29
Figura 6. Dendrogramas em <i>cluster</i> de módulos para amostras corticais de 16-21 semanas gestacionais	32
Figura 7. Análise de sobreposição entre módulos masculinos e femininos para amostras corticais de 16-21 semanas gestacionais	33
Figura 8. Análise de correlação dos níveis de expressão dos módulos <i>consensus</i> e os traços biológicos: idade, sexo e regiões cerebrais para amostras corticais de 16-21 semanas gestacionais	34
Figura 9. Análise de sobreposição entre módulos enriquecidos em cada sexo com módulos enriquecidos consensuais entre os sexos para amostras corticais de 16-21 semanas gestacionais	36
Figura 10. Valores <i>eigengene</i> do módulo <i>consensus tan</i> , enriquecido de genes associados ao transtorno do espectro autista e que se apresentam diferencialmente expressos entre os sexos	37
Figura 11. Análise da distribuição topológica entre genes SFARI e não SFARI, TADA e não TADA, e entre os genes mutados exclusivamente em um sexo ou em ambos nos módulos de coexpressão de interesse	39

Lista de tabelas

Tabela 1. Número de amostras totais e números de doadores totais, por sexo, referentes aos grupos amostrais disponíveis extraídos dos dados de neurônios e células progenitoras neurais derivados de células-tronco pluripotentes induzidas.....	11
Tabela 2. Número de amostras totais e números de doadores totais, por sexo, referentes aos grupos amostrais disponíveis extraídos dos dados de tecido cerebral <i>post-mortem</i>	14
Tabela 3. Número de genes correspondente a cada etapa dos filtros de qualidade, referentes aos grupos amostrais selecionados como viáveis para o presente estudo ..	15
Tabela 4. Número de amostras e números de doadores, por sexo, referentes aos grupos amostrais selecionados como viáveis para o presente estudo.....	16
Tabela 5. Valor do potencial (<i>power</i>) selecionado gerado pela função “ <i>pickSoftThreshold</i> ” para as análises F, M e <i>consensus</i> das amostras corticais selecionadas.....	18
Tabela 6. Módulos enriquecidos de genes SFARI e/ou enriquecidos de genes com variantes raras <i>de novo</i> de risco para o transtorno do espectro autista	26
Tabela 7. Análise de enriquecimento funcional dos módulos de interesse	30
Tabela 8. Análise de sobreposição entre os módulos enriquecidos de interesse e módulos de neurônios derivados de células-tronco pluripotentes induzidas descritos em outros estudos	31
Tabela 9. Módulos enriquecidos de genes SFARI e/ou enriquecidos de genes com variantes raras <i>de novo</i> de risco para o transtorno do espectro autista	35
Tabela 10. Análise de enriquecimento funcional para processos biológicos dos módulos <i>consensus</i>	40
Tabela 11. Análise de sobreposição entre módulos <i>consensus</i> enriquecidos e módulos de neurônios derivados de células-tronco pluripotentes induzidas de outros estudos ..	41

Lista de abreviaturas

CBC	Cerebelo (<i>Cerebellar cortex</i>)
CNVs	Variantes de número de cópias
Cons	<i>Consensus</i>
DI	Deficiência intelectual
F	Feminino (<i>Female</i>)
FPKM	Fragmentos por kilobase por milhão (<i>Fragments per kilobase per million</i>)
GO	Ontologia genética (<i>Gene ontology</i>)
iPSCs	Células-tronco pluripotentes induzidas (<i>Induced pluripotent stem cells</i>)
LBS	Sistema límbico (<i>Limbic system</i>)
M	Masculino (<i>Male</i>)
MSC	Córtex motor-somatossensorial primário (<i>Primary motor-somatosensory cortex</i>)
NPCs	Células progenitoras neurais (<i>Neural progenitor cells</i>)
PCA	Análise de componentes principais (<i>Principal component analysis</i>)
pcw	Semanas gestacionais (<i>Post conceptional week</i>)
PFC	Córtex pré-frontal (<i>Prefrontal cortex</i>)
REVEL	<i>Rare exome variant ensemble learner</i>
SFARI	Iniciativa de Pesquisa sobre Autismo da Fundação Simons (<i>Simons Foundation Autism Research Initiative</i>)
SNV	Variantes de nucleotídeo único
STC	Córtex temporal superior (<i>Superior temporal cortex</i>)
TEA	Transtorno do espectro autista
V1C	Córtex visual primário (<i>Primary visual cortex</i>)
WGCNA	Análise de redes de coexpressão gênica ponderada (<i>Weighted gene coexpression network analysis</i>)

Resumo

Introdução: O transtorno do espectro autista é caracterizado por comprometimento na comunicação social, somado a padrões repetitivos e restritos de comportamentos, interesses ou atividades. O transtorno do espectro autista é um transtorno geneticamente heterogêneo (com diferentes genes e padrões de herança podendo estar envolvidos em cada paciente) que afeta cerca de 1,5% da população e que apresenta um forte desvio de proporção sexual entre os afetados: há, em média, 4 homens para cada mulher diagnosticada. Embora haja algumas teorias para explicar este desvio de proporção sexual, os mecanismos fisiológicos por trás deste fenômeno não estão totalmente esclarecidos. Estudos de transcriptoma de pacientes autistas têm mostrado consistentemente a desregulação de módulos de coexpressão em células neuronais nesses pacientes. Considerando, portanto, que essa desregulação de expressão possa ser o fator desencadeante do transtorno do espectro autista, um mecanismo que poderia justificar o desvio de proporção sexual no transtorno seria a existência basal de uma regulação diferencial de expressão gênica cerebral entre indivíduos do sexo masculino e feminino. **Objetivos:** Analisar se os módulos de coexpressão gênica identificados em cérebros fetais, em janelas temporais e regiões já previamente implicadas no transtorno do espectro autista, diferem entre os sexos em termos de estrutura ou de relações de expressão; e analisar se a distribuição de genes que possuem mutações potencialmente patogênicas nos pacientes autistas dos diferentes sexos difere nas redes de coexpressão geradas. **Métodos:** Construímos redes de coexpressão gênica usando o método de análise de redes de coexpressão gênica ponderada a partir de dados de transcriptoma público de pessoas não diagnosticadas com transtorno do espectro autista, obtidos de tecidos cerebrais *post-mortem*. Os módulos de genes coexpressos identificados foram então analisados quanto ao seu enriquecimento para genes associados ao transtorno do espectro autista. **Resultados:** Em amostras corticais *post-mortem* de 12 a 13 semanas gestacionais de indivíduos não diagnosticados com transtorno do espectro autista identificamos 4 módulos enriquecidos de genes associados ao transtorno do espectro autista. Dentre eles, 3 módulos se apresentaram estruturalmente exclusivo num dado sexo. Ainda, os 2 módulos apresentaram sobreposição gênica com, no mínimo, um módulo já descrito em outro estudo com perfil de expressão mais elevado em indivíduos autistas. Quanto às amostras corticais *post-*

mortem de 16 a 21 semanas gestacionais, identificamos 15 módulos enriquecidos de genes associados ao transtorno do espectro autista. Dentre eles, um módulo apresentou níveis de expressão mais elevados em amostras masculinas. Ainda, 3 módulos apresentaram sobreposição gênica com, no mínimo, um módulo já descrito em outro estudo com perfil de expressão mais elevado em indivíduos autistas. **Conclusão:** Os resultados obtidos até o momento sugerem que as diferenças entre os sexos, tanto em nível estrutural como a nível do perfil de expressão gênica em módulos importantes para o transtorno do espectro autista, podem ser fatores que contribuem, ao menos em parte, para o dimorfismo sexual presente no transtorno do espectro autista, que se reflete na proporção diagnóstica mais prevalente em homens.

Descritores: Transtorno do espectro autista; Expressão gênica; Transcriptoma; Desenvolvimento embrionário e fetal; Redes reguladoras de genes

1 INTRODUÇÃO

1.1 O transtorno do espectro autista

O transtorno do espectro autista (TEA), um transtorno do neurodesenvolvimento com prevalência estimada de 1,5% em países desenvolvidos,⁽¹⁾ é caracterizado por prejuízos na comunicação social, acompanhados de padrões repetitivos e restritos de comportamentos, interesses ou atividades.⁽²⁾ Os sintomas relacionados a estas duas características que definem o diagnóstico de TEA podem ser muito variados entre os pacientes, podendo haver indivíduos seriamente comprometidos, inclusive apresentando deficiência intelectual (DI) comórbida (cerca de 30% dos casos de TEA),⁽³⁾ até indivíduos com um grau mais leve, que têm uma funcionalidade social próxima do normal. Devido a esta heterogeneidade clínica, se convencionou utilizar o termo “espectro” no nome do transtorno.

A etiologia desse transtorno ainda não é totalmente compreendida, mas sabe-se que fatores genéticos e ambientais podem contribuir isoladamente ou em conjunto para o fenótipo.^(2,4) No entanto, os estudos com gêmeos, os quais mostram taxas de concordância para TEA entre 60 e 90% em gêmeos monozigóticos e de 0 a 30% entre gêmeos dizigóticos, e estudos que calcularam o risco relativo entre familiares de um indivíduo autista, apontam que os fatores genéticos têm uma contribuição mais preponderante para o transtorno, sendo eles responsáveis por 38 a 90% da variância fenotípica.⁽⁴⁾ Com o desenvolvimento de tecnologias de triagem em larga escala, como *microarray* e o sequenciamento de nova geração, muito se avançou no conhecimento dos fatores genéticos relacionados a diversas condições, incluindo o TEA.⁽⁵⁻⁷⁾ Embora a heterogeneidade genética entre os indivíduos dificulte a compreensão completa da etiologia do transtorno, estes recentes estudos têm cada vez mais consolidado o conhecimento sobre a arquitetura genética do TEA: atualmente, acredita-se que um conjunto de pacientes (entre 20-30%) pode manifestar TEA em função de uma única mutação altamente penetrante; em outro grupo de pacientes, pode haver a combinação de algumas variantes de médio impacto funcional; finalmente, na maioria dos casos, a causa se daria por um acúmulo de muitos alelos de baixo risco.^(4,8)

1.2 Dimorfismo sexual em transtorno do espectro autista e atuais teorias

O desvio de proporção de afetados entre os sexos é uma característica observada em muitos transtornos mentais.⁽⁹⁾ Em TEA, esse desvio é bem marcante: observa-se uma prevalência diagnóstica cerca de quatro vezes maior em homens do que em mulheres.⁽³⁾ Interessante notar, no entanto, que entre autistas com DI esta prevalência diagnóstica é mais equilibrada entre os sexos, afetando, em média, 2 homens para cada mulher,⁽⁴⁾ enquanto entre autistas com fenótipos menos graves, essa proporção pode chegar a 10 homens para cada mulher afetada. Alguns estudos, inclusive, apontam diferenças neuroanatômicas entre homens e mulheres com TEA,⁽¹⁰⁾ dentre elas: anormalidade da conectividade cerebral, identificada no córtex pré-frontal (PFC – *prefrontal cortex*) de indivíduos homens com TEA,⁽¹¹⁾ bem como um maior volume do córtex temporal superior (STC – *superior temporal cortex*) bilateral nas mulheres em comparação aos homens nos grupos de indivíduos com TEA e indivíduos não afetados.⁽¹²⁾ Também se identificou diferenças sexuais morfométricas da substância cinzenta em regiões corticais e cerebelares que estão envolvidas no controle motor.⁽¹³⁾ Há ainda diferenças sexuais quanto ao padrão de alterações na organização das fibras e características microestruturais do corpo caloso em crianças de 3 a 5 anos com TEA.⁽¹⁴⁾

Outras evidências da literatura apontam ainda para outras variáveis potencialmente associadas ao dimorfismo sexual no TEA. Por exemplo, foram observados, em sangue de cordão umbilical, genes diferencialmente metilados entre os sexos, genes estes previamente associados a doenças psiquiátricas.⁽¹⁵⁾ Ainda, devemos considerar o impacto regulatório de proteínas oriundas de cromossomos sexuais, como por exemplo SRY (proteína ligada ao Y) e SOX3 (proteína ligada ao X), sobre genes autossômicos, um mecanismo potencialmente ligado ao viés masculino no TEA.⁽¹⁶⁾ Finalmente, foi observado que os estressores parentais pré-natais afetaram o comportamento da prole de formas diferentes entre os sexos.⁽¹⁷⁻¹⁹⁾ Este dimorfismo sexual quanto a vulnerabilidade a estressores pré-natais pode, em parte, nos ajudar a entender os mecanismos que levam a diferenças comportamentais entre os sexos e que, por consequência, podem contribuir para o subdiagnóstico de mulheres com TEA.

Há diversas teorias para explicar o dimorfismo sexual presente no TEA, entre elas temos duas principais. A primeira é o modelo de proteção feminino, também conhecido como “*Female protective model*”, a qual sugere que, em mulheres,

um maior número de variantes é necessário para a manifestação fenotípica do TEA (considerando-se um modelo poligênico) ou que a variante tenha alto efeito sobre o fenótipo (considerando um modelo monogênico).⁽²⁰⁾ Essa teoria encontra apoio em estudos que demonstram que as mulheres diagnosticadas com TEA têm maior número de variantes de número de cópias (CNV) deletérias em comparação aos homens, e que essas CNVs afetam uma maior quantidade de genes.⁽²¹⁻²⁴⁾ Além disso, observa-se também que a frequência de variantes de nucleotídeo único (SNV) *de novo* é maior em pacientes com TEA do sexo feminino.^(7,25,26) A segunda é a teoria do cérebro masculino extremo ou “*extreme male brain theory*”, a qual sugere que homens e mulheres diferem quanto a uma gama de habilidades. Este conjunto de habilidades formam um *continuum* em que em um extremo há habilidades sociais como a “empatia” (referindo-se à capacidade de intuir os estados emocionais dos outros) e no outro extremo há habilidades relacionadas a “sistematização”, ou reconhecimento de padrões. Com base em testes de inteligência social e reconhecimento de padrões, viu-se que mulheres tendem a performar melhor em teste de inteligência social enquanto os homens se saem melhor em testes que requerem sistematização ou reconhecimento de padrões.⁽²⁷⁾ A teoria propõe, de forma generalista, que homens e mulheres autistas tendem a apresentar habilidades mais relacionadas à extremidade sistematizadora do *continuum*, desta forma apresentando características estereotipadas de um “cérebro masculino extremo”. Ainda, estudos têm mostrado a existência de uma atividade esteroidogênica (progesterona, 17 α -hidroxi-progesterona, androstenediona e testosterona) fetal elevada em indivíduos autistas⁽²⁸⁾ bem como sugerem que uma exposição fetal elevada à testosterona pode estar relacionada ao risco aumentado de TEA em homens.^(27,29,30) No entanto, os mecanismos fisiológicos que poderiam levar a isso são pouco entendidos.

1.3 Expressão gênica como possível mecanismo relacionado ao dimorfismo sexual em autistas

A análise do transcriptoma de células neuronais tem sido usada nos estudos de TEA não apenas na tentativa de detecção de genes diferencialmente expressos individualmente, mas também para investigações em escala mais integrativa e global, identificando-se grupos/módulos de genes coexpressos nestas células e verificando-se se as relações de expressão de tais módulos têm associação com o

transtorno.^(6,31) Essa estratégia é interessante para estudar transtornos de etiologia heterogênea como TEA pois a captura de padrões de expressão de módulos de genes pode trazer informações não detectáveis em uma análise em nível gene a gene.⁽³²⁾ De fato, estudos em tecidos cerebrais *post-mortem* de autistas têm mostrado consistentemente que módulos de coexpressão relacionados à resposta imune (associados a células da glia) e módulos relacionados a sinapse e neurotransmissão (associados a neurônios) apresentam-se desregulados nestes pacientes.^(33–35) A desregulação deste módulo de genes relacionados a sinapse tem sido observada também em modelos neuronais derivados *in vitro* a partir de células-tronco pluripotentes induzidas (iPSC – *induced pluripotent stem cells*) de indivíduos com TEA, como demonstrado pelo nosso grupo⁽³⁶⁾ e por outros autores.^(37–39) Além disso, observa-se que um número significativamente maior de pacientes autistas têm variantes de perda de função nos genes desse módulo comparado a controles, reforçando seu envolvimento na etiologia do transtorno.

Considerando, portanto, que a desregulação de expressão possa ser um fator desencadeante do TEA, um mecanismo que poderia justificar tanto o modelo de proteção feminino quanto o do cérebro masculino extremo seria uma regulação diferencial de expressão gênica cerebral entre indivíduos do sexo masculino e feminino. Esta diferença natural, eventualmente gerada pelos diferentes níveis de exposição à testosterona nos dois sexos, pode fazer com que as mulheres precisem de um maior número de variantes genéticas para ter uma desestabilização da regulação de expressão gênica suficiente para levar ao TEA; inversamente, o padrão de expressão gênica no cérebro masculino pode já estar mais próximo desse padrão desestabilizado que levaria ao transtorno.

No entanto, a escassez de amostras de tecido cerebral de pacientes autistas e os altos custos envolvidos na geração de células iPSC dificultam a realização de estudos com um número de indivíduos suficiente para se poder avaliar as diferenças entre pacientes de diferentes sexos. Ao mesmo tempo, para que se entenda qual diferença basal pode estar relacionada à maior susceptibilidade ao TEA no sexo masculino, é importante se avaliar o tecido neuronal de indivíduos não diagnosticados com TEA. Alguns trabalhos, de fato, já utilizaram dados de expressão gênica de cérebros fetais e adultos (neste último caso, de indivíduos sabidamente sem transtornos psiquiátricos) disponíveis na base de dados BrainSpan (<https://www.BrainSpan.org/>) na tentativa de avaliar essa questão.^(20,40,41)

Shi et al. verificaram quais genes são diferencialmente expressos entre os sexos (SBGs – *sex-biased genes*) em diversas regiões cerebrais em diferentes fases do desenvolvimento: pré-natal, infância, puberdade e vida adulta.⁽⁴¹⁾ Os autores mostraram que os SBGs variam entre estas 4 etapas do desenvolvimento, mostrando haver alterações dinâmicas ao longo da maturação do sistema nervoso. Em particular, chama a atenção que na fase pré-natal foi visto um maior número de genes com expressão aumentada em homens em praticamente todas as regiões do cérebro. Análises de anotação funcional, mostraram que esses genes estão principalmente envolvidos no controle do ciclo celular. Além disso, os genes com um perfil de expressão aumentado no sexo masculino mostraram-se enriquecidos de genes relacionados a transtornos psiquiátricos, incluindo TEA.

Usando dados de todas as áreas corticais disponíveis nos dados do BrainSpan de indivíduos de 13 a 40 anos e de 16 a 22 semanas gestacionais, Werling et al. buscaram por SBGs para avaliar duas hipóteses: se os genes relacionados a autismo estavam enriquecidos de SBGs (o que significaria que uma mutação patogênica nesses genes poderia ter diferentes consequências diretas para os diferentes sexos), ou se, embora os genes relacionados ao TEA possuam mesmos perfis de expressão em homens e mulheres, eles interagem com SBGs (assim, as diferentes consequências em cada um dos sexos de mutações patogênicas nos genes relacionados ao TEA seriam indiretas).⁽⁴⁰⁾ Os resultados do grupo apoiam a segunda hipótese: tanto em cérebro de adultos quanto em cérebros fetais, não houve enriquecimento de SBGs dentre genes relacionados ao TEA ou dentre genes com mutações patogênicas nesses pacientes; por outro lado, há enriquecimento de SBGs dentre os genes que fazem parte de módulos de coexpressão com valores aumentados em cérebro de autistas, os quais estão relacionados a astrócitos e microglia. Estes resultados sugerem que os próprios processos biológicos sexualmente dimórficos, como interações neurônio-gliais, podem estar modulando o impacto das variantes de TEA e resultando na disparidade de prevalências do TEA entre os sexos.

Ainda, Zhang et al. buscaram entender se as relações de coexpressão entre genes mutados em homens diferiam das relações encontradas em genes mutados em mulheres autistas.⁽²⁰⁾ Para isso, os autores partiram de dados de estudos de exomas e genomas de indivíduos com TEA para selecionar genes com mutações *de novo* potencialmente patogênicas e dividiram estes genes em 3 grupos: genes exclusivamente mutados em mulheres, genes mutados exclusivamente em homens, ou genes mutados

em ambos os sexos. Usando os dados de expressão de cérebros fetais do BrainSpan, e por análise de correlação, os autores encontraram que as três classes de genes (compartilhados entre ambos os sexos ou específicos de um dos sexos) tinham correlação significativa de expressão com um maior número de genes em cérebros femininos comparado aos masculinos. Esses resultados sugerem que pode haver efeitos compensatórios em mulheres em resposta a mutações em genes de risco para TEA, apoiando a hipótese do “efeito protetor feminino”.

Finalmente, Toledo et al.,⁽⁴²⁾ Utilizando dados de transcriptoma de cérebro fetal humano de idades gestacionais variando de 12 a 19 semanas pós-concepção (pcw – *post conceptional week*), investigaram as vias biológicas subjacentes às diferenças sexuais e demonstraram que há diferenças sexuais relevantes para o TEA, diferenças estas encontradas em módulos relacionados a processos biológicos como ciclo celular, diferenciação celular, metabolismo energético e organização da matriz extracelular.

Apesar das contribuições desses trabalhos para o entendimento das diferenças de expressão cerebrais entre os sexos que eventualmente podem estar relacionadas ao desvio de proporção sexual visto em TEA, alguns detalhes em suas metodologias requerem atenção. Todos estes trabalhos foram generalistas na escolha das idades ou regiões cerebrais usadas para a investigação da expressão gênica: Zhang et al.⁽²⁰⁾ e Shi et al.,⁽⁴¹⁾ embora tenham analisado as distintas regiões cerebrais separadamente, consideraram as amostras de 8 a 37 semanas gestacionais em conjunto; já Werling et al.⁽⁴⁰⁾ e Toledo et al.⁽⁴²⁾ embora tenham usado uma janela temporal mais restrita (16 a 22 semanas gestacionais e 12 a 19 semanas gestacionais, respectivamente), usaram todas as regiões corticais em conjunto em suas análises. É sabido que a regulação da expressão gênica é um processo bastante dinâmico ao longo do desenvolvimento⁽⁴³⁾ e diferencial entre regiões distintas do cérebro. Assim, ao se analisar em conjunto amostras de diferentes regiões e/ou tempos de desenvolvimento, é esperado que estas variáveis sejam os principais fatores direcionadores das diferenças de expressão gênica. Portanto, elas podem mascarar possíveis diferenças entre os sexos, pois estas serão comparativamente mais sutis em relação às diferenças observadas para tempos e regiões cerebrais. Ainda, para investigar esta questão, estes autores, à exceção de Toledo et al.,⁽⁴²⁾ usaram em seus trabalhos análises de expressão gênica diferencial ou análises simples de correlação de expressão entre pares de genes. Apesar de suas relevâncias, estas estratégias possuem uma limitação importante: não é

possível investigar com seu uso as complexas redes de interações de coexpressão entre os genes, o que é totalmente ignorado no primeiro tipo de análise ou olhado de forma matematicamente simplista na segunda. Por outro lado, a análise de módulos de coexpressão gênica⁽³²⁾ é uma estratégia mais interessante por ser uma análise sistêmica, ou seja, que contempla a complexidade dos sistemas biológicos. Esse tipo de abordagem permite investigar de forma ponderada (ou seja, atribuindo pesos) o quanto os genes de um determinado tecido ou tipo celular estão relacionados entre si em termos de co-regulação de expressão, reunindo os diversos genes expressos naquele tecido em módulos de acordo com essas relações. Além disso, estas análises atribuem valores de pertencimento de um gene a um determinado módulo, valores estes que refletem com quantos outros genes do mesmo módulo este gene tem relações de coexpressão. Dessa forma, é possível avaliar quais são os genes centrais (com mais relações de coexpressão) ou periféricos de um módulo. Finalmente, também pode-se analisar o nível de expressão que tais módulos/conjuntos de genes apresentam. Os cálculos matemáticos utilizados na construção dessas redes de coexpressão permitem eliminar ruídos e correlações espúrias entre pares de genes, o que em uma análise de correlação simples, apenas verificando-se a correlação de expressão entre um gene e outro, como usado em Zhang et al., seriam dadas como verdadeiras, de acordo com Langfelder et al.⁽³²⁾ Além disso, o tecido cerebral é composto por diferentes tipos celulares. Embora as análises de redes de coexpressão possam identificar módulos de coexpressão que são particulares de cada tipo celular, ajudando a entender as regulações que estão acontecendo em cada um deles, a investigação do transcriptoma de células progenitoras neurais (NPCs – *neural progenitor cells*), neurônios e astrócitos de forma isolada (obtidos por meio de processos de diferenciação *in vitro*) também pode contribuir de forma mais refinada para a questão das diferenças sexuais no desenvolvimento neuronal. Finalmente, frente ao crescente número de dados de exomas gerados de indivíduos autistas [Satterstrom et al.,⁽⁴⁴⁾ Fu et al.,⁽⁴⁵⁾ Zhou et al.,⁽⁴⁶⁾] não se faz inútil revisitar questões já abordadas previamente, uma vez que este maior volume de dados genéticos permite um melhor poder estatístico para as análises.

Em vista do exposto, neste trabalho, visamos investigar, em indivíduos não diagnosticados com TEA, se há diferenças entre os sexos nos padrões estruturais ou de níveis de expressão dos módulos de coexpressão gênica encontrados em células neuronais, as quais poderiam justificar a diferença de prevalência de TEA entre homens e mulheres. Para isso, procuramos analisar estes padrões em tecidos

oriundos de cérebro *post-mortem*, particularmente tecidos de regiões cerebrais e janelas temporais críticas para a fisiopatologia do TEA, bem como em células neuronais derivadas de iPSC, Utilizando a metodologia de redes ponderadas de coexpressão gênica.

Havendo quer sejam diferenças estruturais ou de níveis de expressão, a investigação de como os genes encontrados mutados em pacientes autistas dos diferentes sexos se distribuem nesses módulos pode contribuir para corroborar a hipótese de que padrões de regulação de expressão distintos entre os sexos estão relacionados com a diferença de proporção de afetados por TEA. Nesse sentido, nossa hipótese é de que genes mutados exclusivamente nas mulheres com autismo ou em ambos os sexos seriam genes que pudessem levar a uma perturbação de maior impacto para um módulo e, portanto, deveriam ser genes mais centrais nestes; por outro lado, os genes mutados exclusivamente em homens estariam distribuídos de forma randômica nos módulos, já que neste sexo haveria uma menor chance de “tamponamento” do efeito de uma mutação, quer seja pela constituição estrutural mais enxuta dos módulos, quer seja por suas relações de expressão já serem mais próximos daqueles que levam ao TEA.

1.4 Objetivos

1. Partindo de dados de transcriptoma de cérebros *post-mortem* de amostras de regiões corticais, cerebelares e límbicas que variam de 8 semanas gestacionais a 2 anos de idade, identificar grupos amostrais com perfil de expressão mais homogêneos entre si;
2. Identificar módulos de coexpressão gênica exclusivos de um determinado sexo, assim como módulos presentes em ambos os sexos (módulos *consensus*), realizando análises de coexpressão tanto com dados de transcriptômica de cérebro *post-mortem* como também de células neuronais derivadas de células-tronco pluripotentes induzidas;
3. Analisar se os módulos de coexpressão gênica diferem entre os sexos em termos de estrutura ou quanto à expressão gênica;
4. Identificar, dentre os módulos gerados em cada análise, quais estão enriquecidos de genes associados ao transtorno do espectro autista (genes SFARI e

genes encontrados mutados em pacientes autistas);

5. Analisar se a distribuição de genes já associados ao transtorno do espectro autista ou que possuem mutações potencialmente patogênicas exclusivamente em indivíduos do sexo masculino, ou do sexo feminino ou em ambos difere nas redes de coexpressão geradas;

6. Avaliar os módulos de interesse a nível funcional, ou seja, sua ontologia genética;

7. Avaliar se os módulos de interesse detectados em nosso estudo se sobrepõem aos módulos já descritos na literatura como sendo módulos desregulados em indivíduos autistas.

2 MÉTODOS

2.1 Bancos de dados

2.1.1 Células neuronais fetais *post-mortem* de indivíduos não diagnosticados com transtorno do espectro autista

Os dados de transcriptoma de cérebros *post-mortem* de indivíduos não diagnosticados com TEA foram coletados na base de dados do Atlas BrainSpan (<https://www.BrainSpan.org/>), desenvolvido por um consórcio composto por instituições como: *Allen Institute for Brain Science*, Universidade de Yale, *Harvard Medical School* e MIT HST/CSAIL, Universidade da Califórnia, dentre outras. O banco disponibiliza dados públicos de expressão gênica obtidos tanto por meio de RNA-seq quanto de *microarray*. As amostras deste banco foram obtidas de indivíduos de diversas etnias e variam de uma idade de oito semanas gestacionais até indivíduos adultos com até 40 anos. Além disso, as amostras cerebrais de cada indivíduo foram dissecadas de acordo com regiões anatômicas específicas e tiveram seus transcriptomas sequenciados separadamente. Assim, vale destacar que há dados de diferentes regiões pertencentes a um mesmo indivíduo.

Importante ressaltar que partimos de amostras de indivíduos não diagnosticados com TEA, uma vez que nosso objetivo é entender diferenças de expressão entre os sexos em condições não patológicas. Por tratar-se de amostras referentes ao período fetal do desenvolvimento, não é possível identificar se tais indivíduos desenvolveriam TEA, uma vez que o diagnóstico de TEA se dá num período mais tardio do desenvolvimento onde são avaliadas as características comportamentais destes indivíduos. Embora não haja como classificar diretamente as amostras como sendo de indivíduos autistas ou não autistas, destacamos que estas amostras de tecidos cerebrais passaram por uma triagem onde cérebros com aneuploidia ou outras alterações morfológicas importantes foram excluídas (<https://www.BrainSpan.org/static/download.html>).

2.1.2 Neurônios e células progenitoras neurais derivados de células-tronco pluripotentes induzidas de indivíduos não diagnosticados com transtorno do espectro autista

A fim de poder analisar células neurais de forma isolada, além de sanar a limitação descrita acima quanto a ausência diagnóstica de TEA das amostras oriundas de tecidos cerebrais fetais *post-mortem*, buscamos trabalhar com dados de transcriptoma de neurônios e NPCs derivados de iPSCs obtidas de indivíduos não autistas, sendo ambos os grupos amostrais (neurônios e NPCs) compostos de 15 amostras masculinas e 9 amostras femininas (Tabela 1), avaliadas no estudo de Hoffman et al. (GSE106589).⁽⁴⁷⁾

Tabela 1. Número de amostras totais e números de doadores totais, por sexo, referentes aos grupos amostrais disponíveis extraídos dos dados de neurônios e células progenitoras neurais derivados de células-tronco pluripotentes induzidas

Grupo amostral	Número de amostras femininas (número de doadores)	Número de amostras masculinas (número de doadores)
Neurônios derivados de iPSCs	9 (5)	15 (6)
NPCs derivadas de iPSCs	9 (5)	15 (6)

iPSCs: células-tronco pluripotentes induzidas; NPCs: células progenitoras neurais.

2.2 Análise de clusterização das amostras de tecido cerebral e seleção das janelas espaço-temporais homogêneas

Como já discutido anteriormente, as regiões anatômicas cerebrais e os períodos de desenvolvimento são provavelmente as principais variáveis direcionadoras das diferenças de expressão gênica que poderíamos encontrar caso analisássemos todas as amostras do banco BrainSpan em conjunto, mascarando possíveis diferenças entre os sexos. Assim, a fim de se entender como as amostras se agrupam e, eventualmente, usar grupos mais homogêneos, utilizamos a técnica estatística multivariada de análise de componentes principais (PCA – *principal component analysis*). Partimos de uma janela temporal ampla, considerando desde o período mais precoce do neurodesenvolvimento disponível no banco (8 semanas gestacionais) até a idade em que comumente ocorre a identificação dos sintomas dos

indivíduos autistas⁽²⁾ (de acordo com a 5ª edição do Manual de Diagnóstico e Estatístico de Transtornos Mentais, os sintomas do TEA são tipicamente reconhecidos até o segundo ano de vida). Além disso, consideramos analisar amostras de todas as distintas grandes regiões anatômicas cerebrais disponíveis: PFC e córtex motor-somatosensorial primário (MSC – *primary motor-somatosensory cortex*); córtices visual primário (V1C – *primary visual cortex*) e STC; cerebelo (CBC – *cerebellar cortex*); e sistema límbico (LBS – *limbic system*). A escolha final das janelas espaço-temporais selecionadas em nosso trabalho levou em consideração, portanto, (a) os agrupamentos reconhecidos por meio de análises multivariadas, (b) a representatividade de amostras masculinas e femininas para cada um dos grupos obtidos, (c) a disponibilidade de um número mínimo de amostras femininas, bem como masculinas, para cada grupo amostral, em que se possa atingir uma topologia de distribuição livre de escala dos dados para então aplicarmos o método de análise de redes de coexpressão gênica ponderada (WGCNA – *weighted gene coexpression network analysis*), descrito em detalhes na seção 2.5, e (d) a priorização de grupos amostrais pertencentes a uma janela temporal e região cerebral implicada na fisiopatologia do TEA com base no estudo de Willsey et al.⁽⁴⁸⁾ Para maior clareza das próximas seções, já apresentamos a seguir os resultados desta análise que culminaram com a escolha dos conjuntos amostrais a serem estudados efetivamente.

Inicialmente, com a análise de PCA, observamos dois grandes agrupamentos de amostras distintas quanto à faixa etária, sendo o primeiro grupo composto de amostras de 8 a 24 pcw e o segundo com amostras de 25 pcw a 2 anos (yrs – *years*) (Apêndice 1).

Em seguida, analisamos estes dois grupos separadamente. A análise de PCA com o grupo de 25 pcw a 2 yrs mostrou que as amostras estavam fortemente separadas por indivíduo, variável que se confunde com a variável idade, já que nesta janela temporal só havia amostras de cada idade de um único indivíduo (Apêndice 2A e 2B). Observamos também não haver uma linearidade temporal na distribuição das amostras ao longo dos eixos do primeiro e segundo componentes principais, sugerindo que de fato o principal direcionador das diferenças de expressão nessas amostras era a variável “indivíduo” e não a idade. Ainda, não seria possível o pareamento entre amostras femininas e masculinas de acordo com a idade, já que as amostras femininas se restringem a uma janela temporal mais avançada (1 e 2 anos de idade), enquanto as amostras masculinas correspondem a uma janela temporal com idades mais precoces

(Apêndice 2C). Quanto às regiões anatômicas, tanto amostras corticais como límbicas e cerebelares se distribuíram igualmente ao longo dos eixos dos componentes principais, tendo as amostras corticais apresentado pequenos agrupamentos (Apêndice 2D). Dessa forma, dadas as variáveis idade e indivíduo que se confundem nesse grupo amostral e ainda a diferença de idade entre os sexos, decidimos por não realizar as análises de redes de coexpressão com as amostras corticais de 25 pcw a 2 yrs.

Seguindo as análises com as amostras de 8 a 24 pcw, observamos três subagrupamentos: amostras de 8-9, 12-13 e 16-24 pcw (Apêndice 3A). Notadamente, as amostras de 8 e 9 pcw estavam representadas por um único indivíduo para cada idade e ainda somente amostras masculinas (Apêndice 3B e 3C). Foi vista também uma considerável separação de acordo com as regiões anatômicas das amostras, ficando as regiões corticais PFC/MSc e V1C/STC mais separadas das regiões CBC e LBS (Apêndice 3D).

Desta forma, as amostras corticais foram consideradas separadamente das amostras límbicas e cerebelares para novas análises de PCA de acordo com os agrupamentos encontrados por idade: (a) amostras de 8-9 pcw, grupo composto apenas por amostras masculinas, sendo, portanto, inviável de utilização para nosso objetivo; (b) amostras de 12-13 pcw, sendo 42 amostras femininas (referentes a 3 doadores mulheres) e 21 amostras masculinas (referentes a 2 doadores homens); (c) amostras de 16-24 pcw, sendo 18 amostras femininas (referentes a 3 doadores mulheres) e 49 amostras masculinas (referentes a 5 doadores homens). Finalmente, a análise de PCA das amostras pertencentes às regiões CBC e LBS de 8-24 pcw revelaram uma clara distinção entre as amostras de acordo com estas duas regiões (Apêndice 4). Com isso, subdividimos estas amostras em (d) amostras da região CBC de 8 a 24 pcw, contendo apenas 8 amostras de cada sexo, número amostral menor do que o número mínimo de 15 amostras recomendado para se realizar a análise WGCNA, inviabilizando desta forma a utilização desse grupo amostral; e (e) amostras da região LBS de 8 a 24 pcw, contendo 17 amostras femininas e 23 masculinas (Tabela 2).

Tabela 2. Número de amostras totais e números de doadores totais, por sexo, referentes aos grupos amostrais disponíveis extraídos dos dados de tecido cerebral *post-mortem*

Grupo amostral	Número de amostras femininas (número de doadores)	Número de amostras masculinas (número de doadores)
Córtex 8 - 9 pcw	0 (0)	11 (2)
Córtex 12 - 13 pcw	42 (3)	21 (2)
Córtex 16 - 24 pcw	18 (3)	49 (5)
Cerebelo 8 - 24 pcw	8 (6)	8 (6)
Límbico 8 - 24 pcw	17 (6)	23 (9)

pcw: semanas gestacionais.

Uma visão geral dos resultados de nossa análise de agrupamento de acordo com os perfis de expressão das amostras encontra-se esquematizada em forma de fluxograma no apêndice 5. Em função do número de amostras e pareamento entre os sexos, os grupos amostrais corticais de 12-13 pcw e 16-24 pcw assim como as amostras límbicas de 8-24 pcw são adequados para a análise de coexpressão gênica. Contudo, no presente trabalho, decidimos focar nos grupos amostrais corticais de 12-13 pcw e 16-24 pcw levando em consideração que há evidências de que esta janela temporal e região cerebral está implicada na fisiopatologia do TEA,⁽⁴⁸⁾ na qual um número maior de genes associados ao TEA são expressos. Importante mencionar que a janela temporal de 16-24 pcw é composta por células neurais mais maduras (maior presença de neurônios) se comparadas às amostras de 12-13 pcw (maior presença de NPCs).⁽³⁶⁾ Esta diferença quanto a população celular reforça a importância de avaliarmos tais grupos amostrais separadamente.

2.3 Filtro de exclusão para genes com baixa expressão e com baixo coeficiente de variação

Para cada uma das janelas espaço-temporais selecionadas, realizamos uma análise separadamente por sexo para exclusão dos genes com expressão de fragmentos por kilobase por milhão (FPKM – *fragments per kilobase per million*) < 1 em mais do que 50% das amostras (Tabela 3). Em seguida, usando os dados de expressão logaritmizados (em log2), retivemos apenas os genes com coeficiente de variação (CV) separadamente para cada sexo (desvio padrão/média dos valores de expressão $\geq 0,1$). Com isso, trabalhamos com genes cuja expressão varia mais amplamente entre as amostras, podendo então trazer informações biológicas relevantes. Consideramos então o conjunto união desses dois grupos de genes filtrados por sexo

separadamente. Finalmente excluímos os genes que com valor de expressão zerados em todas as amostras de um dado sexo. Este filtro nos permite trabalhar apenas com genes cuja expressão está presente em ambos os sexos.

Tabela 3. Número de genes correspondente a cada etapa dos filtros de qualidade, referentes aos grupos amostrais selecionados como viáveis para o presente estudo

Grupo amostral	Genes totais (n)	Após remover genes com baixa expressão (n)	Após remover genes com baixo CV (n)	Após remover genes não expressos em um dos sexos (n)
Córtex 12 - 13 pcw	52.376	14.880	9.433	9.430
Córtex 16 - 21 pcw	52.376	15.835	12.104	12.103

pcw: semanas gestacionais; CV: coeficiente de variação.

2.4 Análises de controle de qualidade das amostras

Uma vez estabelecido o grupo de genes a serem analisados na fase anterior, utilizamos então o gráfico do tipo *boxplot* para melhor visualizar a distribuição dos dados de cada amostra para identificar e excluir as amostras possivelmente *outliers*. Como critério de exclusão, consideramos como *outlier* a amostra, de um dado sexo, com mediana do valor de expressão maior ou menor que 1,5 desvios padrão (DP) em relação a média das medianas das amostras do dado sexo, isso para cada janela espaço-temporal separadamente. Dentre as amostras corticais de 12-13 pcw, quatro amostras femininas de 12 pcw (39_12pcw_F, 40_12pcw_F, 64_12pcw_F e 74_12pcw_F) e duas amostras masculinas de 13 pcw (89_13pcw_M e 114_13pcw_M) foram removidas de acordo com o critério mencionado acima. Desta forma, um total de 57 amostras, sendo 38 femininas e 19 masculinas, foram selecionadas para a construção das redes de coexpressão deste grupo amostral (Tabela 4).

Ao avaliarmos as amostras corticais de 16 a 24 pcw, primeiramente consideramos haver pareamento entre os sexos para as idades de 16 pcw [masculino (M)] e 17 pcw [feminino (F)] e entre 19 pcw (F) e 21 pcw (M). Porém, as amostras de 24 pcw (M) (n=11) foram excluídas por não haver amostras de idade próxima entre as amostras femininas. Em seguida, na etapa de exclusão das amostras *outliers*, removemos 3 amostras femininas (174_19pcw_F, 175_19pcw_F e 177_19pcw_F) e 7 amostras masculinas (122_16pcw_M, 123_16pcw_M, 126_16pcw_M, 133_16pcw_M, 143_16pcw_M, 158_16pcw_M e 193_21pcw_M), por possuírem uma mediana dos

valores de expressão com desvio de $\pm 1,5$ desvio-padrão do valor médio das medianas de todas as amostras do dado sexo (Tabela 4).

Tabela 4. Número de amostras e números de doadores, por sexo, referentes aos grupos amostrais selecionados como viáveis para o presente estudo

Grupo amostral	Número de amostras femininas (número de doadores)	Número de amostras masculinas (número de doadores)
Córtex 12 - 13 pcw	38 (4)	19 (2)
Córtex 16 - 21 pcw	15 (3)	31 (4)

pcw: semanas gestacionais.

2.5 Análise de redes de coexpressão gênica: identificação de diferenças estruturais ou níveis de expressão entre homens e mulheres nos módulos de coexpressão

Usamos o pacote de análise WGCNA, construído na linguagem R, para analisar os padrões de correlação entre genes nas amostras de transcriptoma selecionadas. O pacote WGCNA analisa inicialmente a correlação de expressão gênica entre todos os pares de genes de um dado, porém elevando estes valores de correlação a uma determinada potência para que se evidencie pares de genes altamente correlacionados de fato⁽³²⁾ (<https://youtu.be/6h2UMpPn2Kk>). Essa potência é escolhida após testes sucessivos de diferentes valores, por meio da função “*pickSoftThreshold*”, até que se encontre o valor de potência que leva o dado a seguir uma topologia de distribuição livre de escala, isto é, em que apenas uma pequena parte dos genes tem muitas conexões (*hubs*). Encontrado este valor, calcula-se as correlações de expressão entre os genes, e estas são utilizadas para detectar os módulos (conjuntos) de genes altamente correlacionados. O pacote ainda permite verificar uma série de outros atributos, os quais são importantes para nossas análises. O primeiro deles é o primeiro componente principal (calculado pela função “*moduleEigengenes*”): este é um valor que sumariza os valores de expressão de todos os genes de um dado módulo para cada indivíduo. Estes valores, chamados de *eigengenes*, são então utilizados para identificar se há módulos com valores de expressão diferentes de acordo com uma dada variável (no caso do presente trabalho, entre homens e mulheres). Um segundo atributo, obtido pela função “*signedKME*”, é o valor de conectividade ou pertencimento de um gene a todos os módulos detectados no dado. O gene é então atribuído ao módulo com o qual ele apresenta o maior valor de conectividade (valor *kME*); além disso, dentro de um

mesmo módulo, quanto maior o valor de kME de um gene, mais central ele é neste módulo (*hub gene*), ou seja, este gene apresenta relações de coexpressão com mais genes do módulo. O terceiro atributo que a ferramenta permite estudar são as propriedades topológicas dos módulos, permitindo a identificação de módulos com mesma estrutura em diferentes blocos de dados. Para isso utilizamos a função “*blockwiseConsensusModules*” que realiza a construção das redes de forma individual para os diferentes conjuntos de dados, comparando-as ao final a fim de detectar apenas os módulos consensuais entre os conjuntos de dados distintos.

De acordo com nossos objetivos, realizamos então a construção de três redes distintas. Inicialmente, analisamos separadamente amostras femininas das masculinas, gerando os módulos masculinos (módulos M) e módulos femininos (módulos F), com a finalidade de compará-los estruturalmente e identificar possíveis módulos exclusivos de um dos sexos. Em seguida, para identificar módulos presentes em ambos os sexos e avaliar diferenças de expressão gênica entre os grupos, analisamos os dados de expressão femininos e masculinos com a função “*blockwiseConsensusModules*”, gerando somente módulos consensuais entre os sexos, os quais foram denominados como “módulos *consensus*”.

Ao final, o algoritmo identifica cada módulo por cores diferentes. As cores dos módulos são atribuídas de acordo com seus tamanhos e, portanto, um módulo que foi denominado, por exemplo, como *blue* em um conjunto de dados não é igual necessariamente ao módulo *blue* identificado em um outro conjunto de dados. Genes cujos valores de kME demonstram que estes não possuem conectividade a nenhum módulo são alocados em um módulo denominado de Grey: ou seja, embora se utilize a denominação de “módulo”, este é um conjunto de genes que não possuem relações de coexpressão significativas com nenhum módulo detectado e nem entre si. Desta forma, enquanto nos módulos Grey (F) e Grey (M) temos genes que não apresentam relações de coexpressão gênica no dado sexo, no módulo Grey (*consensus*) podemos ter ou (a) genes que não apresentam valores de coexpressão gênica em pelo menos um dos sexos, ou ainda, (b) genes que não apresentam relações de coexpressão gênica em nenhum sexo.

Na construção dos módulos feminino e masculino, selecionamos valores de potência (*Power*) que atendessem a uma topologia livre de escala ($SFT.R.sq \geq 0.8$, aproximadamente) adequado às duas análises. Ainda, a escolha da potência para as diferentes análises considerou o objetivo de gerarmos módulos cuja conectividade

média (mean.k.) fosse o menos discrepante possível (Tabela 5).

Tabela 5. Valor do potencial (*power*) selecionado gerado pela função “*pickSoftThreshold*” para as análises F, M e *consensus* das amostras corticais selecionadas

Grupo amostral	Análise	Power	SFT.R.sq	slope	truncated.R.sq	mean.k	median.k	max.k
Córtex 12-13	F	23	0.770	-1.64	0.908	42.6	18.2	308
pcw	M	30	0.887	-1.22	0.960	90.9	44.8	545
Córtex 16-21	F	26	0.890	-1.45	0.955	79.4	42.0	514
pcw	M	19	0.790	-1.32	0.958	76.9	53.4	352

pcw: semanas gestacionais.

Para identificarmos quais módulos se apresentam estruturalmente semelhantes em ambos os sexos e quais são exclusivos de um único sexo, analisamos a sobreposição gênica dos módulos femininos e masculinos por meio da função *userListEnrichment* do WGCNA. Analisamos também a sobreposição gênica entre os módulos femininos e masculinos com os módulos *consensus* a fim de validar quais módulos são, de fato, exclusivos de um sexo e quais se encontram representados (mesmo que parcialmente) em ambos os sexos. Finalmente, para comparação das diferenças de expressão gênica, foi realizada uma análise de correlação entre os valores *eigengenes* obtidos para os módulos *consensus* com a variável sexo (o valor 1 foi atribuído para as amostras femininas e o valor 2 para as masculinas), idade e região cerebral (o valor 1 foi atribuído para as amostras corticais V1C/STC e o valor 2 para PFC/MS).

2.6 Análise de enriquecimento de genes associados ao transtorno do espectro autista

Para avaliar o possível enriquecimento de genes associados ao TEA nos módulos de genes coexpressos que apresentaram diferença de expressão ou de estrutura entre os sexos, utilizamos a lista completa de genes (n=1010) presentes no banco de dados da Iniciativa de Pesquisa sobre Autismo da Fundação Simons (SFARI – *Simons Foundation Autism Research Initiative*) (<https://gene.sfari.org/>). Este banco de dados compila genes já associados ao TEA, atribuindo a eles um *score* de evidência de associação. A curadoria desses dados é feita por meio de análise criteriosa da literatura científica publicada e revisada por pares por pesquisadores especialistas que buscam,

identificam e extraem sistematicamente informações sobre estudos genéticos do TEA em humanos e organismos experimentais. Cabe ressaltar que a inclusão de um gene como gene SFARI não se baseia apenas no critério de ter se identificado mutações raras *de novo* em tal gene: as evidências de sua associação ao transtorno podem vir de diversos estudos, como estudos de associação, funcionais ou animais.

Também verificamos o enriquecimento e distribuição topológica de 185 genes identificados como fortemente associados ao risco para TEA no trabalho de Fu et al.⁽⁴⁵⁾ Neste trabalho, a partir do sequenciamento de exoma de 63.237 indivíduos, os pesquisadores identificaram uma lista de 185 genes com forte associação ao TEA ($FDR < 0.5$) por meio de análises estatísticas que consideram a recorrência em um mesmo gene de variantes *de novo* de truncamento de proteínas (PTVs), variantes de sentido trocado (*missense*) e CNVs em indivíduos com TEA, ou a ocorrência mais frequente desse tipo de mutação em um determinado gene numa coorte de casos comparativamente a controles (TADA – *transmission and de novo association analysis*). Para facilitar a compreensão ao longo do texto, denominamos esta lista de 185 genes de genes TADA.

2.7 Análise da distribuição topológica dos genes e variantes de risco para o transtorno do espectro autista nos módulos de coexpressão de interesse

Utilizando a função “*signedKME*” do WGCNA, identificamos, os valores de pertencimento (valor kME) de cada gene ao módulo ao qual ele foi alocado: quanto maior este valor, mais central o gene está topologicamente no módulo, ou seja, maior o número de genes com os quais ele apresenta correlação de expressão. Assim, esta função nos permitiu investigar, nos módulos de interesse, a distribuição dos genes SFARI, bem como dos genes de risco para o TEA segundo a lista de Fu et al.⁽⁴⁵⁾

Para analisar a distribuição topológica nos módulos de coexpressão de interesse de genes para os quais pudéssemos avaliar o *status* mutacional de acordo com o sexo, utilizamos a lista de genes com variantes raras (frequência populacional $<1\%$) *de novo* identificadas nos 16.877 casos de autismo sequenciados no estudo de Zhou et al.⁽⁴⁶⁾ Dentre os genes dos módulos de interesse, avaliamos a distribuição topológica daqueles que possuíam variantes consideradas potencialmente patogênicas. Foram consideradas potencialmente patogênicas: (a) mutações de perda de função

(mutações de código de parada prematuro, de mudança de quadro de leitura ou de alteração de sítio de *splicing*) em genes classificados com $pLI \geq 0,8$; ou (b) mutações de troca de aminoácido (*missense*) que possuam uma pontuação *Rare exome variant ensemble learner* (REVEL) acima de 0,5, (sensibilidade: 0,754; especificidade: 0,891). O índice pLI destina-se a medir o quão restrito ou intolerante um gene ou transcrito é para variantes de perda de função, em que, numa escala de 0 a 1, o valor 1 é atribuído para um maior grau de intolerância. Já o índice REVEL para uma variante *missense*, varia de 0 a 1, com pontuações mais altas para uma maior probabilidade da variante ser causadora da doença.⁽⁴⁹⁾ Os genes com variantes raras assim selecionados foram separados em 3 categorias: mutados exclusivamente em homens, mutados exclusivamente em mulheres ou mutados em ambos (estas duas últimas categorias foram consideradas em conjunto também). A dispersão dos valores de kME destes 3 conjuntos foi comparada Utilizando-se o teste estatístico de Kruskal-Wallis.

2.8 Análise de enriquecimento funcional

A análise de enriquecimento funcional dos módulos de interesse (ou seja, qual a função dos genes em cada módulo) foi realizada pela ferramenta DAVID (<https://david.ncifcrf.gov/home.jsp>), Observando-se a categoria de processos biológicos do *Gene ontology* (categoria: GOTERM_BP_FAT). Utilizamos como *background* da análise funcional a lista de genes considerados como expressos em cada grupo amostral analisado (córtex 12-13 pcw = 14.880 genes; córtex 16-21 pcw = 15.835 genes).

2.9 Análise de sobreposição dos módulos de interesse com módulos de outros estudos

Finalmente, verificamos se algum dos módulos enriquecidos de genes associados ao TEA, que se mostraram diferencialmente expressos ou estruturalmente diferentes entre homens e mulheres, se tratava de algum módulo já identificado previamente como diferencialmente expressos em autistas por outros trabalhos, Utilizando para isso também uma análise de enriquecimento realizada com a função *userListEnrichment* do WGCNA.

2.10 Aspectos éticos

O presente projeto possui registro no Sistema Gerenciador de Projetos de Pesquisa (SGPP) de número 4482-20 e se encontra em conformidade com o Comitê de Ética em Pesquisa em Seres Humanos do Hospital Israelita Albert Einstein se enquadrando na resolução 510/2016 do Conselho Nacional de Saúde.

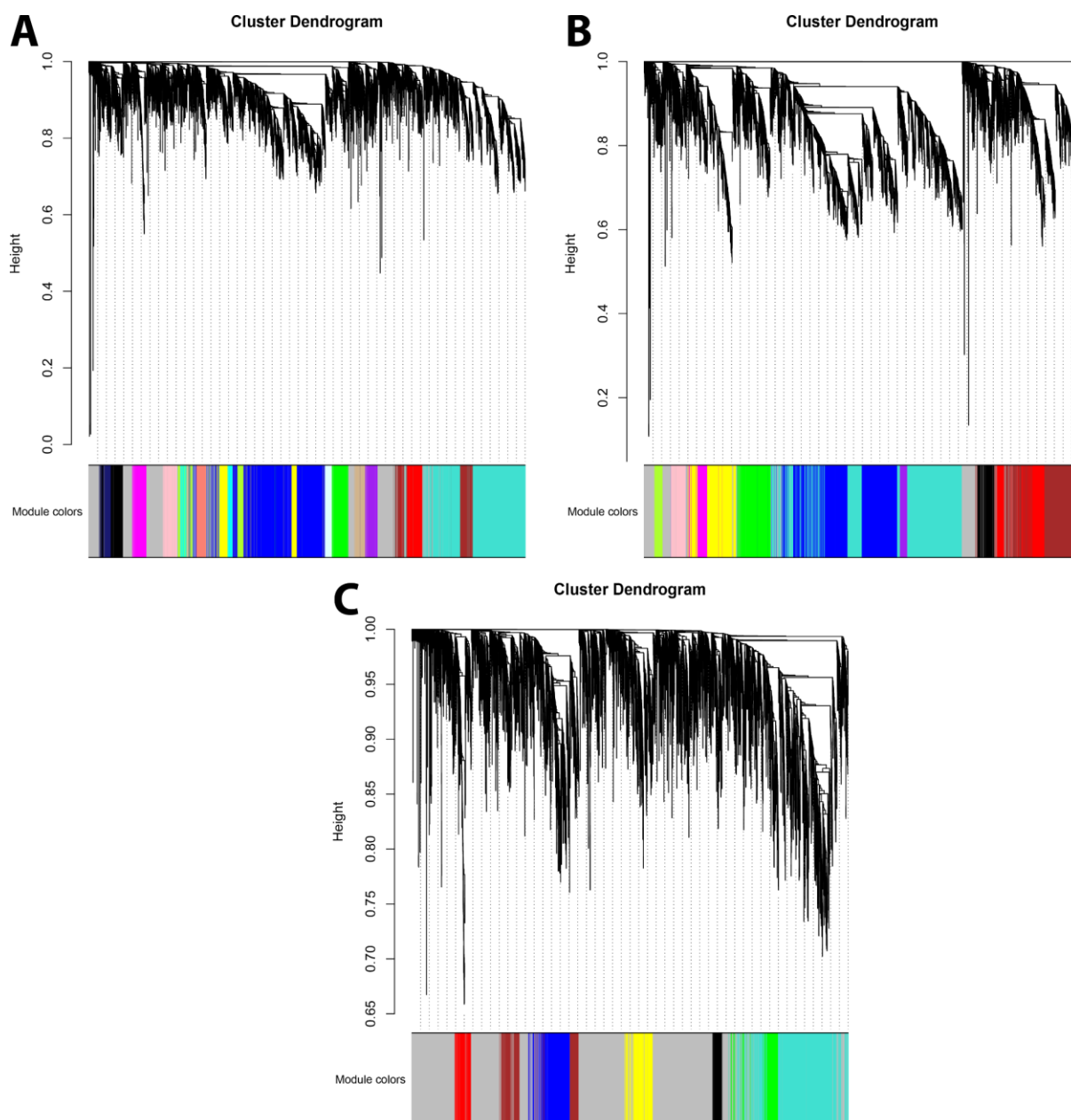
A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) contemplou o aluno de mestrado Paulo Victor Visintin com uma bolsa de mestrado para o presente projeto (número do processo: 2021/08293-3).

3 RESULTADOS

3.1 Análises de coexpressão gênica das amostras corticais de 12 a 13 semanas gestacionais

3.1.1 Diferenças estruturais e de expressão gênica entre os sexos

A fim de entender quais módulos de genes coexpressos nas amostras corticais da janela de 12-13 pcw estão presentes nos dois sexos ou apenas em um deles, ou seja, para entender diferenças estruturais gerais, geramos redes de coexpressão separadamente para amostras femininas e masculinas e, em seguida, para a identificação das estruturas modulares existentes em ambos os sexos (módulos *consensus*), utilizamos a função do pacote WGCNA chamada “*blockwiseConsensusModules*”. Os módulos obtidos para as amostras femininas, masculinas e *consensus* estão representados nos dendrogramas em *cluster* na figura 1, sendo 17 módulos femininos, 11 masculinos e 8 módulos *consensus*.

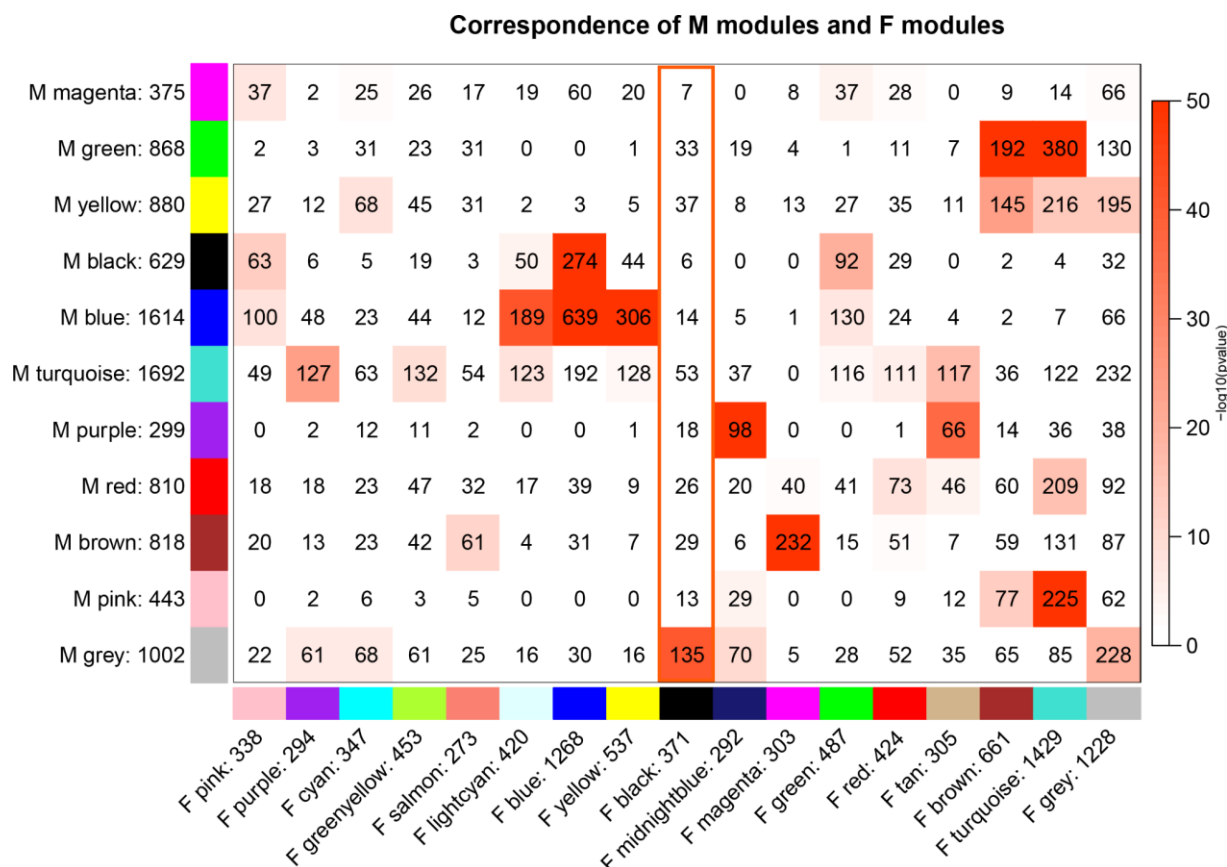


Resultados da análise de dendrograma em *cluster* no *software* R Studio. (A) Módulos presentes em amostras femininas; (B) Módulos presentes em amostras masculinas; (C) Módulos *consensus*, presentes em amostras femininas e masculinas.

Figura 1. Dendrogramas em *cluster* de módulos para amostras corticais de 12-13 semanas gestacionais

Comparando-se os 17 módulos femininos aos 11 módulos masculinos gerados, por meio de análise de sobreposição dos genes constituintes, identificamos que a maioria dos módulos apresentam sobreposição gênica considerável com um ou mais módulos do sexo oposto, indicando que, quanto ao aspecto estrutural, tais módulos são relativamente presentes em ambos os sexos (Figura 2). No entanto, o módulo *F_black* apresentou sobreposições menos robustas com relação aos módulos do sexo oposto, Observando-se um número maior de genes deste módulo alocado no módulo masculino

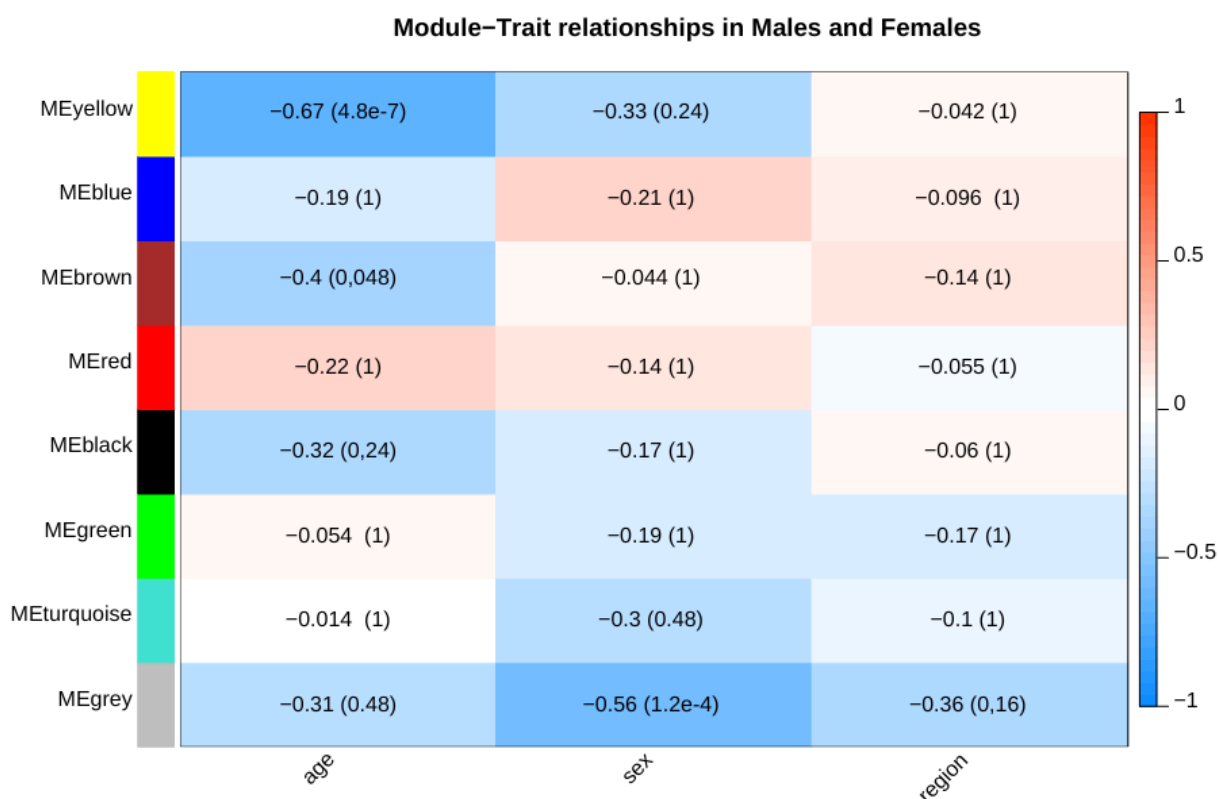
grey. Ainda, este módulo não apresenta sobreposição robusta com nenhum módulo *consensus* (Apêndice 6), sugerindo, que estes genes não apresentam relações de coexpressão entre si no sexo oposto.



Resultados da análise de sobreposição no *software* R Studio. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e módulos masculinos. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Figura 2. Análise de sobreposição entre módulos masculinos e femininos para amostras corticais de 12-13 semanas gestacionais

Quanto às diferenças de expressão gênica, dentre os 8 módulos *consensus* obtidos, nenhum apresentou níveis de expressão significativamente correlacionados com o sexo ($p < 0,05$) (Figura 3). Porém, 2 módulos apresentaram perfis de expressão correlacionados com a idade, sendo os valores de correlação > 0 correspondente aos módulos com expressão aumentada com o avançar da idade. Por fim, nenhum módulo apresentou correlação com a região cerebral. Os valores de correlação < 0 correspondem à maior expressão nos córtices V1C e STC, comparado aos córtices PFC e MSC.



Resultados de análises de correlação no *software* R Studio. Uma correlação acima de 0, em cor vermelha, correspondente aos módulos com expressão aumentada com o avançar da idade, módulos com níveis de expressão maiores em homens, bem como maior expressão gênica nos córtices pré-frontal e motor-somatossensorial primário.

Figura 3. Análise de correlação dos níveis de expressão dos módulos *consensus* e os traços biológicos: idade, sexo e regiões cerebrais para amostras corticais de 12-13 semanas gestacionais

3.1.2 Identificação de módulos enriquecidos de genes associados ao transtorno do espectro autista

Para identificarmos, dentre os módulos obtidos, quais podem trazer informações relevantes para responder pelo dimorfismo sexual no TEA, avaliamos então qual a proporção de genes associados ao TEA em cada um deles. Inicialmente, através da lista de gene SFARI, identificamos 2 módulos femininos (módulos “F”), 1 módulo masculino (módulos “M”) e 1 módulo *consensus* (módulos “C”) enriquecidos destes genes, conforme apontado na tabela 6.

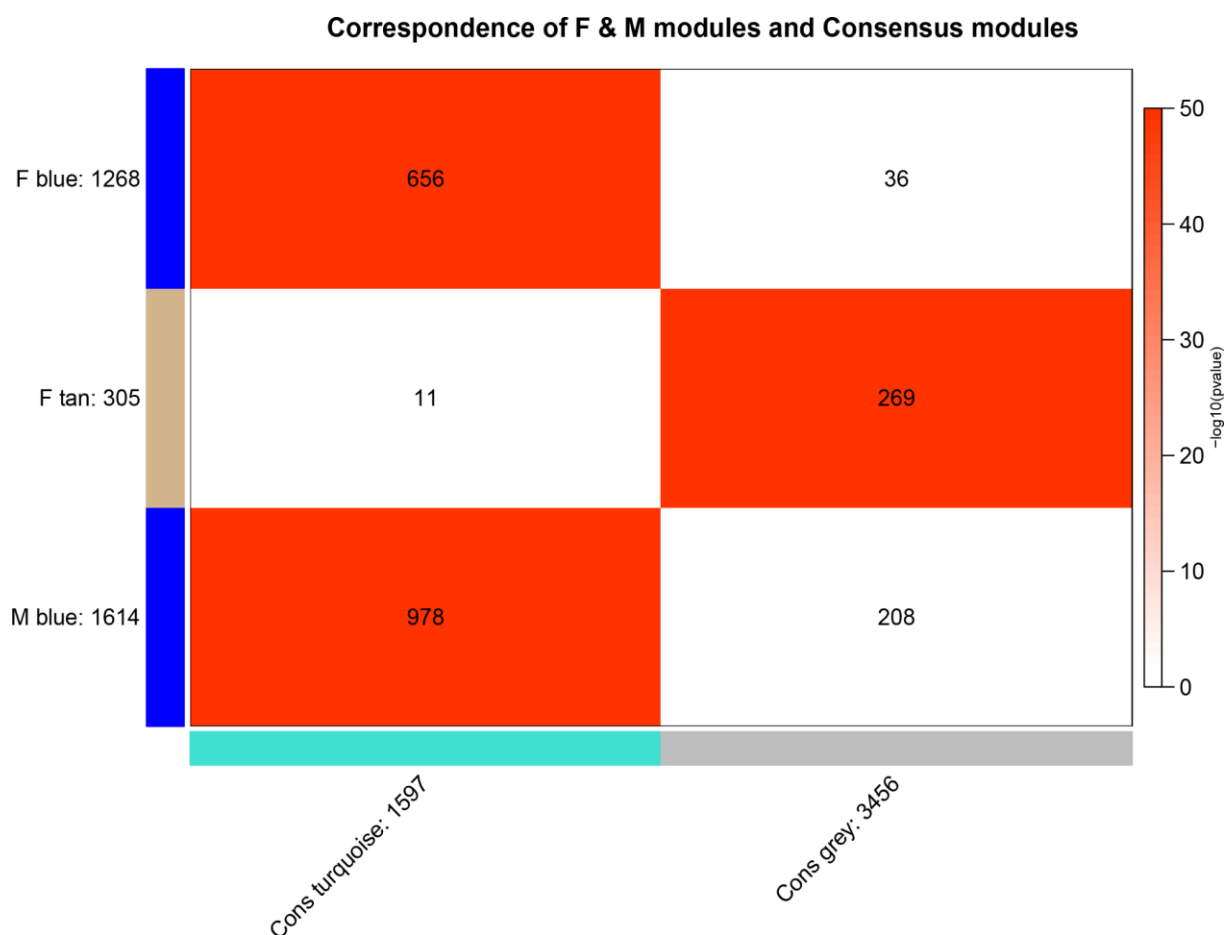
Ainda, identificamos 3 módulos enriquecidos para a lista de genes TADA,⁽⁴⁵⁾ sendo todos estes módulos também enriquecidos de genes SFARI (*F_blue*, *M_blue* e *C_turquoise*).

Tabela 6. Módulos enriquecidos de genes SFARI e/ou enriquecidos de genes com variantes raras *de novo* de risco para o transtorno do espectro autista

Módulo	Genes codificantes	Genes SFARI	OR (Valor-p)	Genes TADA	OR (Valor-p)
<i>F_blue</i>	1028	107	1,60 (0.0029)	36	3,31 (6.78e-6)
<i>F_tan</i>	241	37	2,39 (0.00042)	n/s	n/s
<i>M_blue</i>	1229	166	2,48 (3.14e-16)	53	5,55 (1.92e-14)
<i>C_turquoise</i>	1259	146	1,94 (1.48e-8)	49	4,53 (2.52e-11)

OR: razão de chances.

Para facilitar a visualização das sobreposições somente entre os módulos enriquecidos de genes associados ao TEA nas 3 diferentes redes geradas, apresentamos na figura 4 tais relações.



Resultados da análise de sobreposição no *software* R Studio. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e masculinos com módulos *consensus* enriquecidos de genes TEA. A escala de cores corresponde ao valor de valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Figura 4. Análise de sobreposição entre módulos enriquecidos em cada sexo com módulos enriquecidos consensuais entre os sexos para amostras corticais de 12-13 semanas gestacionais

A sobreposição dos módulos *F_blue* e *M_blue* com o mesmo módulo consenso (*Cons_turquoise*) indica que suas estruturas estão preservadas em ambos os

sexos. Dessa forma, podemos considerar que o módulo *Cons_turquoise*, enriquecido de genes associados ao TEA, não possui diferenças estruturais e nem de níveis de expressão (Figura 3) entre os sexos. Por outro lado, observa-se que o módulo *F_tan* não apresenta sobreposição com nenhum módulo consenso, conforme apontamos no apêndice 6 (há apenas uma sobreposição robusta com o conjunto *Cons_grey*), apesar de apresentar sobreposições discretas com dois módulos da rede masculina (*M_turquoise* e *M_purple*; Figura 2), os quais, por sua vez, não apresentaram enriquecimento de genes associados ao TEA. Este resultado parece indicar que embora os genes do módulo *F_tan* tenham também relações de coexpressão no sexo masculino, estas relações são diferentes a ponto destes genes não emergirem como módulo comum na análise consenso, sugerindo, assim, uma diferença estrutural de expressão gênica entre os sexos.

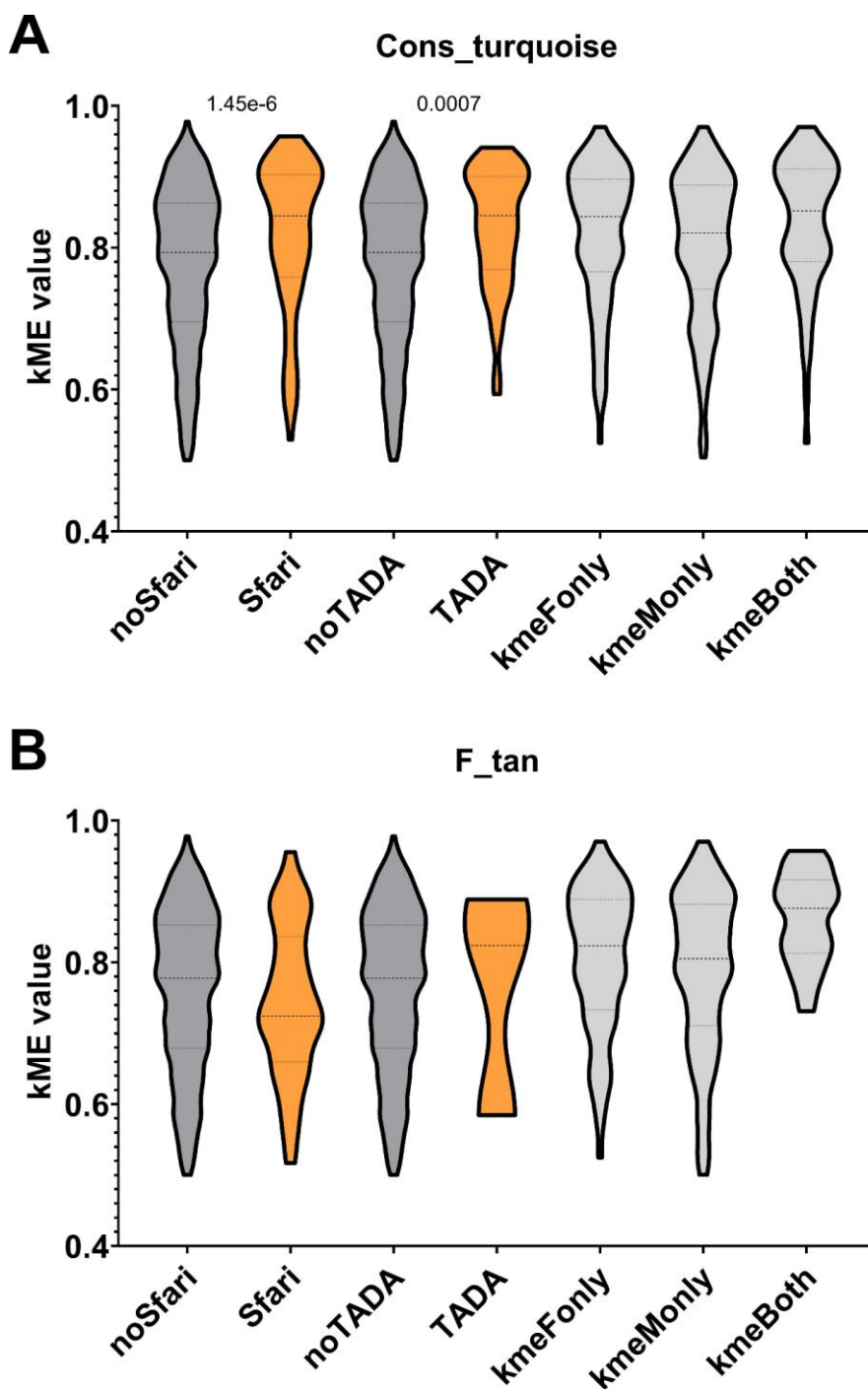
Em conjunto, estes dados sugerem que, se por um lado parece não haver módulos com diferença no nível de expressão gênica entre os sexos na janela de 12-13 semanas gestacionais, há indícios de uma diferença estrutural nas redes masculina e feminina de interesse para este estudo, caracterizada por um conjunto de genes associados ao TEA que possuem relações de coexpressão distintas no sexo feminino e masculino.

3.1.3 Análise da distribuição topológica dos genes associados ao transtorno do espectro autista nos módulos de coexpressão de interesse

A fim de lançar luz sobre de que modo as diferenças na organização das redes masculina e feminina podem se relacionar à diferença de prevalência de TEA entre os sexos, analisamos então a distribuição topológica, dentro dos módulos de interesse, de genes que possuem mutações potencialmente patogênicas em indivíduos com o transtorno, agrupando-os em três categorias: genes identificados com mutações potencialmente patogênicas em ambos os sexos (*both*), exclusivamente em mulheres (*Fonly*) ou exclusivamente em homens (*Monly*). A distribuição topológica dos genes SFARI e TADA também foi observada para fins comparativos. Esta análise é particularmente interessante para módulos apresentando níveis de expressão gênica desbalanceados entre os sexos, pois a distribuição topológica dos genes mutados exclusivamente em homens, em mulheres ou em ambos poderia se revelar desigual já

que as mutações teriam que impactar a regulação do módulo de forma diferente para cada sexo para levar a um mesmo nível limiar para a ocorrência do fenótipo.

Apesar do módulo *Cons_turquoise* não se apresentar diferencialmente expresso entre os sexos, optamos por ainda assim avaliar estas distribuições topológicas. Observamos que os genes SFARI e TADA possuem uma posição mais central no módulo (valores de kMEs mais altos), comparativamente a genes sem associação ao TEA, porém não observamos diferença significativa quanto à distribuição topológica dos genes *Fonly*, *Monly* e *both* (Figura 5A). Já o módulo *F_tan* não apresentou diferença significativa quanto à distribuição topológica dos genes SFARI e TADA nem quanto às comparações entre genes *Fonly*, *Monly* e *both*.



Resultados da análise da distribuição topológica no *software GraphPad Prism*. (A) Comparação dos valores de kME do módulo *Cons_turquoise*; (B) Comparação dos valores de kME do módulo *F_tan*. As comparações se dão entre os genes SFARI e não SFARI, entre os genes TADA e não TADA, e entre os genes mutados exclusivamente para cada sexo (*kmeFonly* e *kmeMonly*) e estes também comparados aos genes mutados em ambos os sexos (*kmeboth*).

Figura 5. Análise da distribuição topológica entre genes SFARI e não SFARI, TADA e não TADA, e entre os genes mutados exclusivamente em um sexo ou em ambos nos módulos de coexpressão de interesse

3.1.4 Análise de enriquecimento funcional

Para identificar as principais funções dos módulos enriquecidos de genes associados ao TEA selecionados, fizemos uma análise de enriquecimento de ontologia genética (GO – *gene ontology*). Identificamos que o módulo *Cons_Turquoise* é funcionalmente relacionado a sinalização celular e ainda o módulo *F_tan* se apresenta enriquecido para funções ligadas ao desenvolvimento do sistema nervoso, conforme descrito na tabela 7.

Tabela 7. Análise de enriquecimento funcional dos módulos de interesse

Módulo	Função	Count	% genes	Valor-p	Benjamini
Cons_turquoise	<i>regulation of signaling</i>	328	24.4	7.90e-08	2.80e-04
	<i>regulation of cell communication</i>	324	24.1	7.70e-08	2.80e-04
	<i>regulation of signal transduction</i>	272	20.2	2.50e-05	1.60e-02
	<i>intracellular signal transduction</i>	237	17.6	1.40e-04	4.90e-02
	<i>vesicle-mediated transport</i>	154	11.4	8.30e-06	7.20e-03
	<i>nervous system development</i>	73	28.1	1.80e-07	1.70e-04
F_tan	<i>cellular response to chemical stimulus</i>	61	23.5	1.40e-04	2.60e-02
	<i>cell surface receptor signaling pathway</i>	55	21.2	2.00e-04	3.00e-02
	<i>cell development</i>	55	21.2	2.50e-04	3.40e-02
	<i>neurogenesis</i>	52	20	5.40e-06	2.60e-03

Count: contagem de leituras.

3.1.5 Análise de sobreposição entre módulos *consensus* enriquecidos e módulos implicados previamente no transtorno do espectro autista

Consultamos a literatura em busca de outros estudos com indivíduos autistas que tivessem gerado redes de coexpressão a partir de amostras que se aproximam das regiões cerebrais e janela temporal de desenvolvimento analisados neste presente estudo. Neste sentido, analisamos a sobreposição gênica entre os módulos de interesse encontrados na janela de 12-13 semanas gestacionais e módulos que se apresentam diferencialmente expressos em modelos neuronais derivados de iPSC de indivíduos autistas (Tabela 8).

O módulo *Cons_turquoise* se apresentou significativamente sobreposto aos módulos enriquecidos de genes de risco para o TEA descritos por Schafer et al.,⁽³⁸⁾ a saber: *brown_Schafer*, *lightcyan_Schafer* e *lightyellow_Schafer* (este

último, com expressão aumentada em indivíduos com TEA). Ainda, o módulo *Cons_turquoise* se sobrepõe aos módulos *blue_Mariani* e *brown_Mariani*, descritos por Mariani et al.,⁽³⁷⁾ ambos com expressão aumentada em autistas. Por fim o módulo *Cons_turquoise* também se sobrepõe ao módulo *MNeu1-turquoise_Griesi*, reportado por nosso grupo em trabalho anterior⁽³⁶⁾ como um módulo enriquecido de genes SFARI e com aumento de expressão em autistas. O módulo *F_tan* apresentou sobreposição significativa com os módulos *brown_Mariani*, de Mariani et al.⁽³⁷⁾ e o módulo *MNeu2-tan_Griesi*, reportado por nosso grupo em trabalho anterior⁽³⁶⁾ como um módulo enriquecido de genes SFARI, porém sem diferenças de expressão em indivíduos com TEA.

Tabela 8. Análise de sobreposição entre os módulos enriquecidos de interesse e módulos de neurônios derivados de células-tronco pluripotentes induzidas descritos em outros estudos

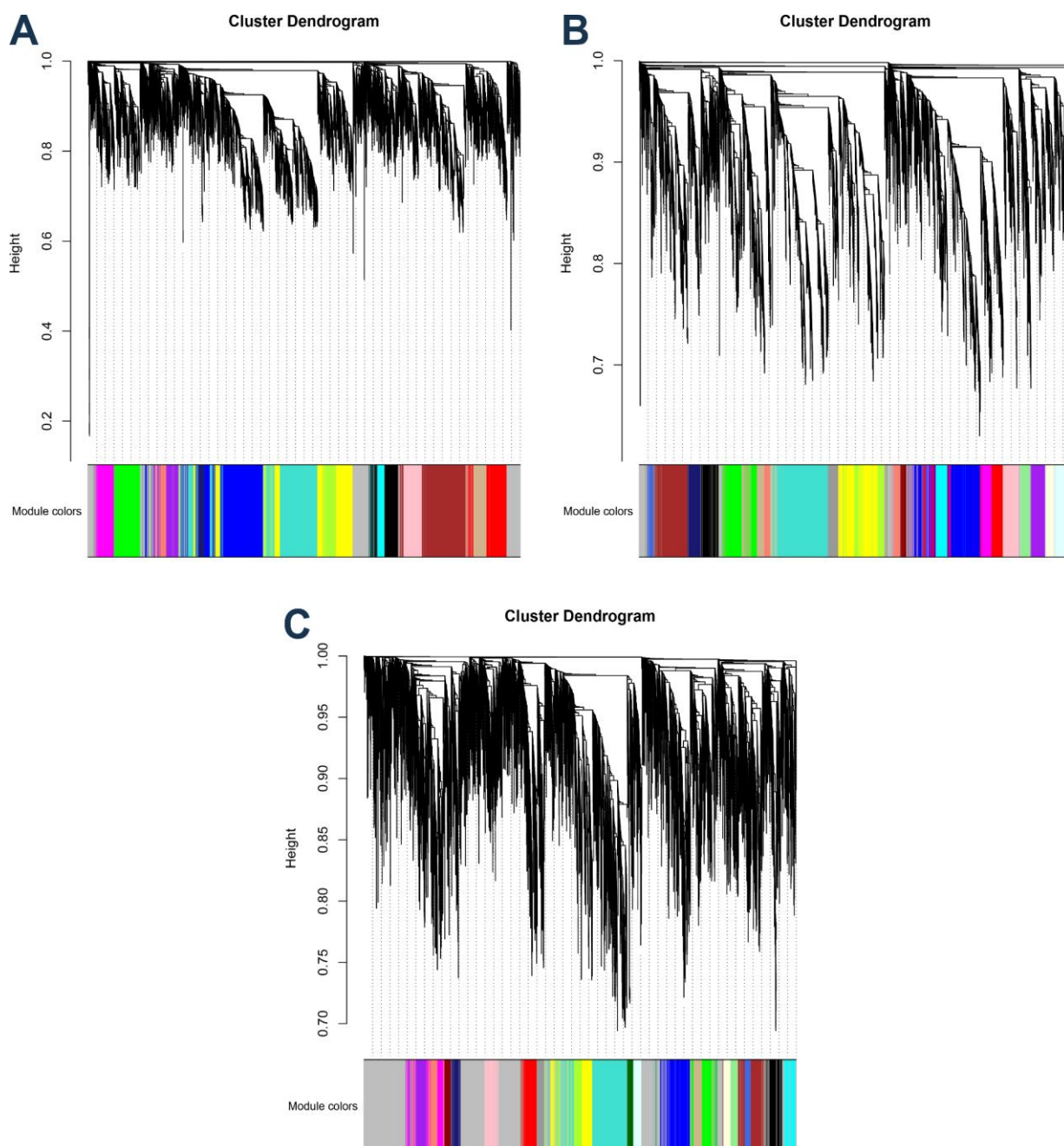
Módulo	Genes totais	Genes sobrepostos	Outros estudos	Valor-p corrigido
<i>Cons_turquoise*</i>	1597	307	lightyellow_Schafer**	6.14e-77
		607	brown_Schafer*	1.74e-42
		97	lightcyan_Schafer*	4.52e-28
		206	blue_Mariani**	9.98e-24
		124	MNeu1-turquoise_Griesi**	6.31e-16
		145	brown_Mariani**	3.15e-08
<i>F_tan*</i>	305	32	MNeu2-tan_Griesi*	7.00e-08
		38	brown_Mariani**	0.0024

*Módulo enriquecido de genes de risco para o TEA. **Módulo enriquecido de genes de risco para o TEA com expressão elevada em indivíduos com TEA.

3.2 Análises de coexpressão gênica das amostras corticais de 16 a 21 semanas gestacionais

3.2.1 Diferenças estruturais e de expressão gênica entre os sexos

Utilizamos para as amostras corticais da janela de 16-21 pcw a mesma sequência de análises utilizadas para as amostras corticais de 12-13 pcw. Os módulos de coexpressão obtidos para as amostras femininas e masculinas de forma separada, bem como, os módulos *consensus* estão representados nos dendrogramas em *cluster* na figura 6.

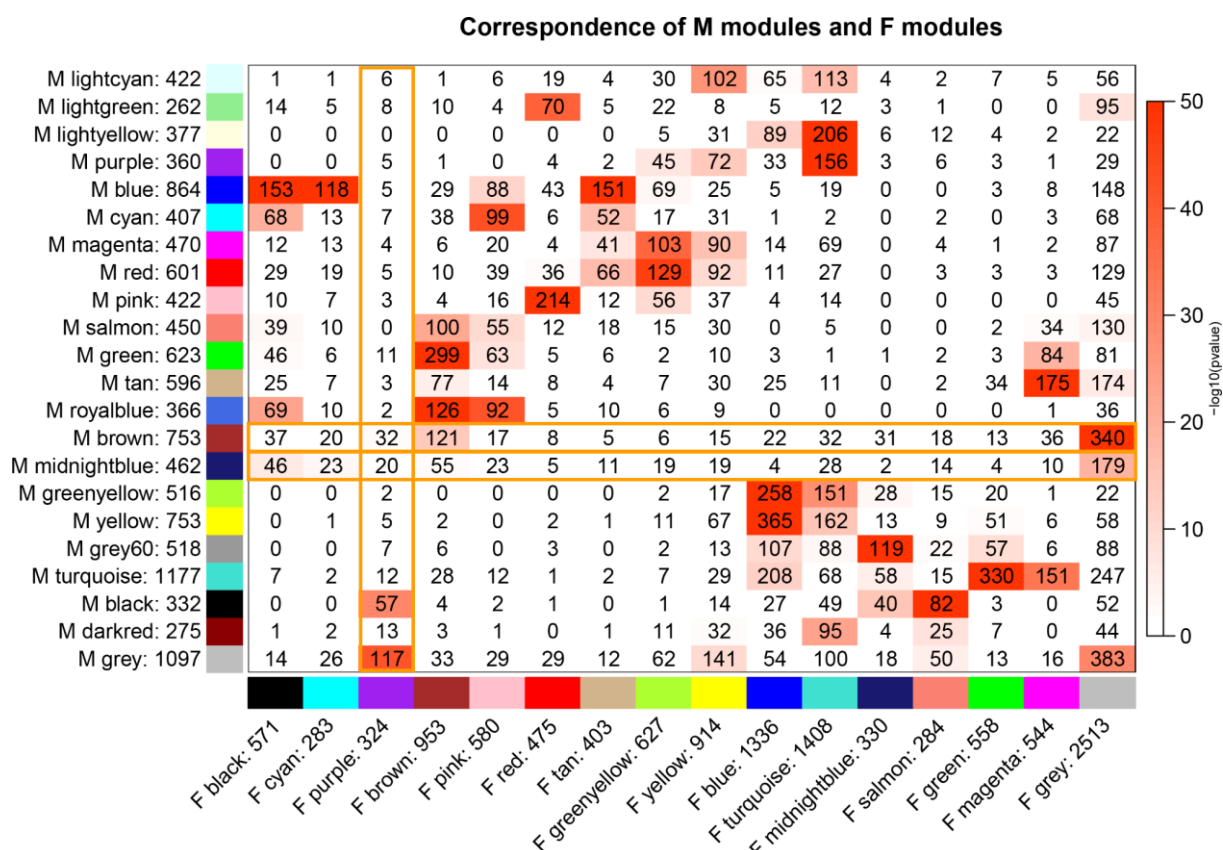


Resultados da análise de dendrograma em *cluster* no *software* R Studio. (A) Módulos presentes em amostras femininas; (B) Módulos presentes em amostras masculinas; (C) Módulos *consensus*, presentes em amostras femininas e masculinas

Figura 6. Dendrogramas em *cluster* de módulos para amostras corticais de 16-21 semanas gestacionais

Para este grupo amostral foram gerados 16 módulos femininos, 22 módulos masculinos e 23 módulos *consensus*. Através de análise de sobreposição dos genes constituintes dos módulos femininos e masculinos, identificamos que a maior parte dos módulos apresentam sobreposição gênica com um ou mais módulos do sexo oposto, indicando que, quanto ao aspecto estrutural, tais módulos são relativamente presentes em ambos os sexos (Figura 7). No entanto, 3 módulos apresentam sobreposições menos

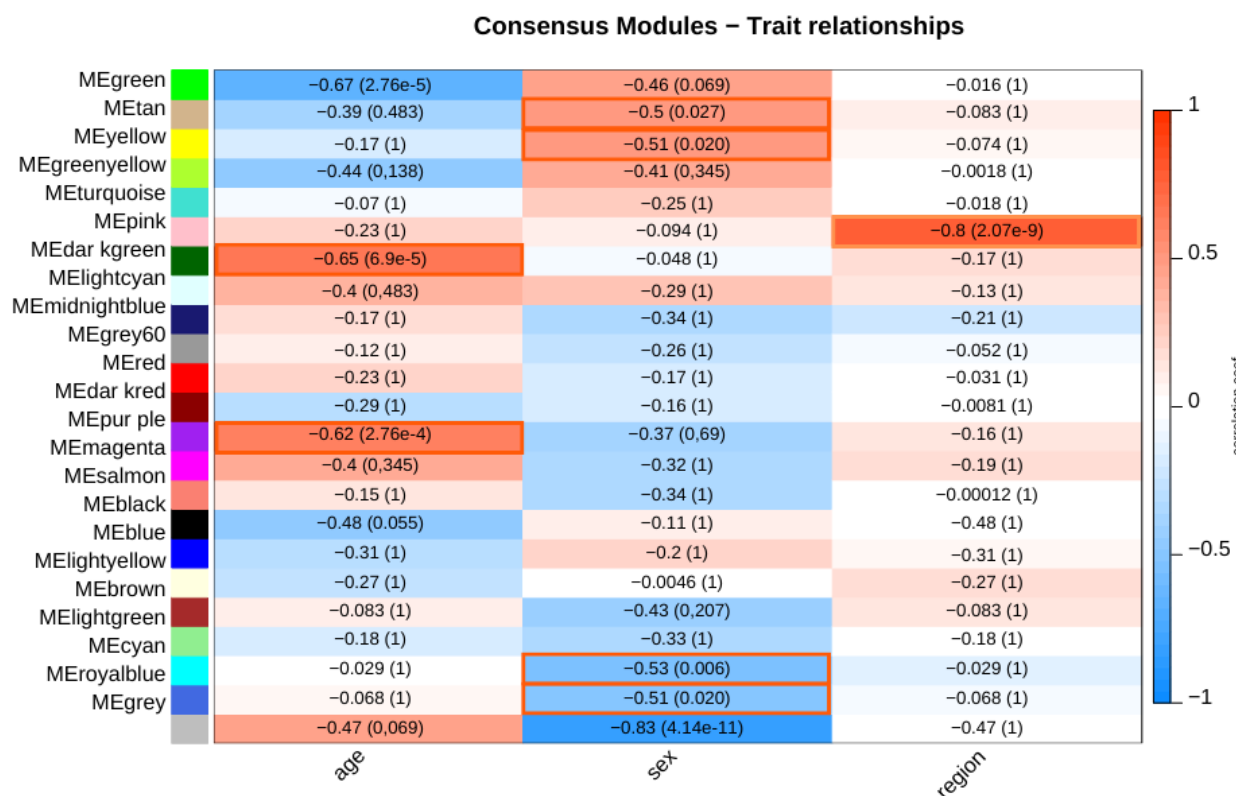
robustas com relação aos módulos do sexo oposto, a saber: os módulos *M_brown*, *M_midnightblue* e *F_purple*.



Resultados da análise de sobreposição no *software* R Studio. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e módulos masculinos. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Figura 7. Análise de sobreposição entre módulos masculinos e femininos para amostras corticais de 16-21 semanas gestacionais

Quanto às diferenças de expressão gênica, dentre os 23 módulos *consensus* obtidos, detectamos 4 cujos níveis de expressão mostraram correlação com o sexo ($p < 0,05$, considerando a correção para múltiplos testes), sendo os valores positivos relacionados a maior expressão no sexo masculino (Figura 8). Observamos também que 2 módulos apresentam um perfil de expressão correlacionado à idade ($p < 0,05$), sendo os valores de correlação > 0 correspondente aos módulos com expressão aumentada com o avançar da idade. Ainda, 1 módulo apresentou forte correlação com a região cerebral ($p < 0,05$), sendo o valor de correlação > 0 correspondente a maior expressão nos córtices PFC e MSC, comparado aos córtices V1C e STC.



Resultados de análises de correlação no *software* R Studio. Uma correlação acima de 0, em cor vermelha, correspondente aos módulos com expressão aumentada com o avançar da idade, módulos com níveis de expressão maiores em homens, bem como maior expressão gênica nos córtices pré-frontal e motor-somatossensorial primário.

Figura 8. Análise de correlação dos níveis de expressão dos módulos *consensus* e os traços biológicos: idade, sexo e regiões cerebrais para amostras corticais de 16-21 semanas gestacionais

3.2.2 Identificação de módulos enriquecidos de genes associados ao transtorno do espectro autista

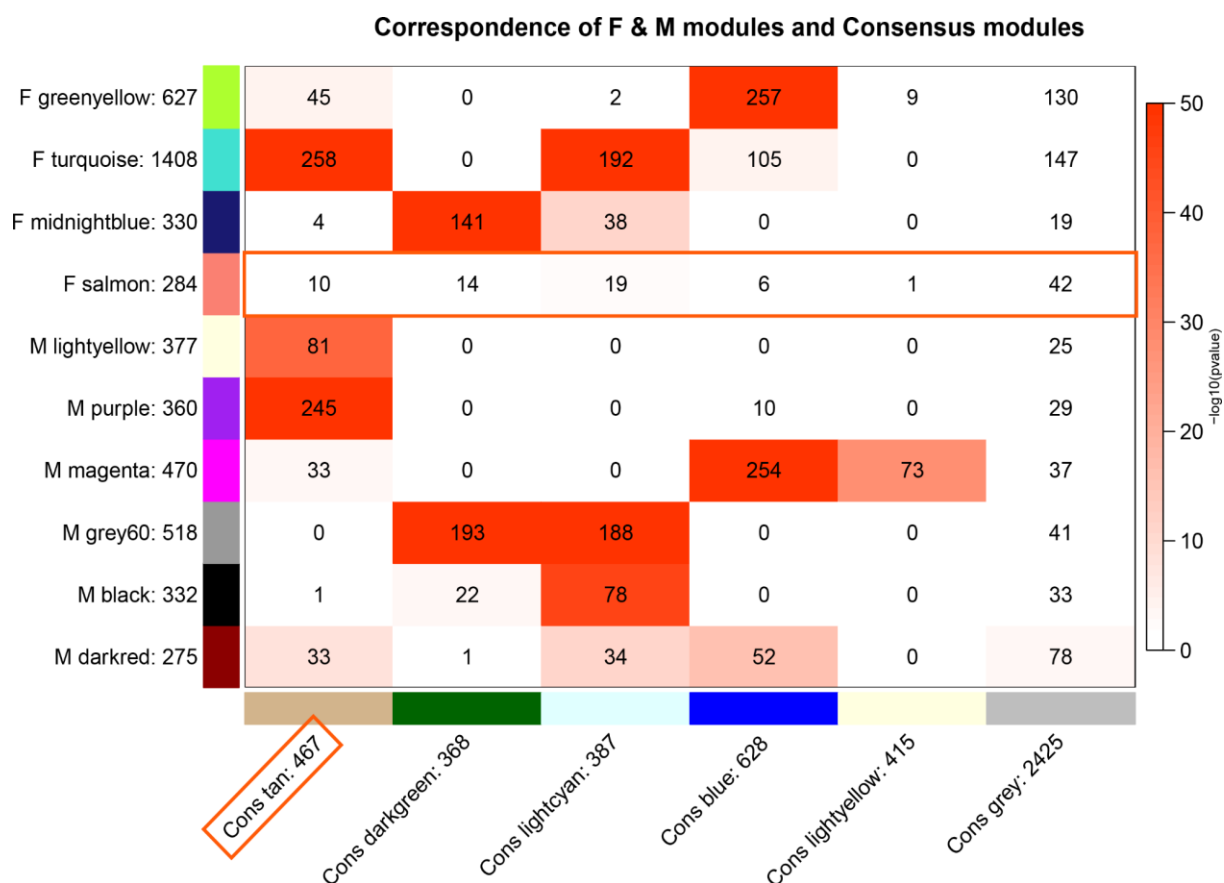
Dentre os módulos obtidos, identificamos 4 módulos femininos (módulos “F”), 6 módulos masculinos (módulos “M”) e 4 módulos *consensus* (módulos “C”) enriquecidos significativamente de genes SFARI ($p < 0.05$), conforme apontado na tabela 9. Já com relação aos genes TADA, identificamos 2 módulos femininos, 3 masculinos e 3 *consensus* enriquecidos por tais genes (Tabela 9).

Tabela 9. Módulos enriquecidos de genes SFARI e/ou enriquecidos de genes com variantes raras *de novo* de risco para o transtorno do espectro autista

Módulo	Genes codificantes	Genes SFARI	OR (Valor-p)	Genes TADA	OR (Valor-p)
F_greenyellow	450	55	1.97 (0.0006)	19	3.18 (0.0011)
F_midnightblue	275	36	2.11 (0.0044)	n/s	n/s
F_salmon	239	33	2.24 (0.0038)	n/s	n/s
F_turquoise	1060	146	2.45 (6.20e-16)	40	3.10 (6.11e-7)
M_black	272	44	2.74 (2.69e-6)	n/s	n/s
M_darkred	209	38	3.14 (8.66e-7)	10	3.50 (0.044)
M_grey60	391	47	1.92 (0.000131)	n/s	n/s
M_lightyellow	281	38	2.20 (0.0057)	n/s	n/s
M_magenta	334	60	3.16 (1.04e-10)	21	4.96 (8.84e-7)
M_purple	273	46	2.88 (4.97e-7)	14	3.84 (0.0022)
C_blue	482	76	2.73 (1.46e-10)	27	4.51 (9.06e-8)
C_darkgreen	296	41	2.26 (5.10e-4)	n/s	n/s
C_lightcyan	304	47	2.60 (4.28e-6)	n/s	n/s
C_tan	349	61	3.06 (2.40e-10)	17	3.68 (0.0007)
C_lightyellow	275	n/s	n/s	13	3.50 (0.0092)

OR: razão de chances.

Para facilitar a visualização das correspondências entre os módulos enriquecidos de genes associados ao TEA encontrados nas três análises, apresentamos na figura 9 tais relações e destacamos dentre os módulos *consensus* enriquecidos, o módulo *Cons_tan* que também apresenta diferença de expressão entre os sexos (*box* laranja), como já apontado na figura 8.



Resultados da análise de sobreposição no *software* R Studio. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e masculinos com módulos *consensus* enriquecidos de genes TEA. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

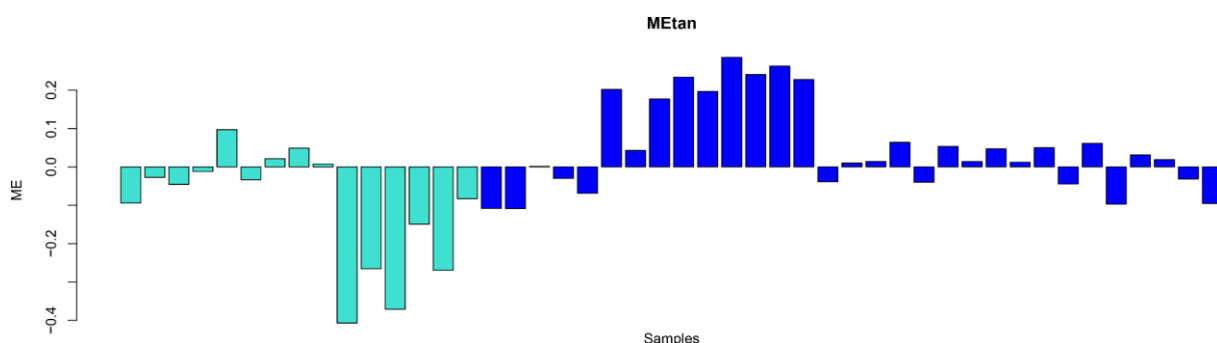
Figura 9. Análise de sobreposição entre módulos enriquecidos em cada sexo com módulos enriquecidos consensuais entre os sexos para amostras corticais de 16-21 semanas gestacionais

Observando os módulos *consensus* enriquecidos de genes associados ao TEA, constatamos que todos possuem sobreposição tanto com módulos femininos como com módulos masculinos enriquecidos, como esperado, com exceção do módulo *Cons_lightyellow*. Este módulo apresenta sobreposição com o módulo masculino enriquecido de genes associados ao TEA *M_magenta*, bem como com os módulos não enriquecidos *M_blue*, *M_cyan* e *M_red*, enquanto que em relação aos módulos femininos, este módulo apresenta uma sobreposição robusta apenas com módulos não enriquecidos (*F_tan* e *F_pink*), conforme mostrado nos apêndices 8 e 9.

O módulo *F_salmon* apresentou sobreposição robusta apenas com os módulos *Cons_pink* (Apêndice 8) e *M_black* (Figura 7), ambos não enriquecidos de genes SFARI ou TADA.

Destacamos ainda que o módulo *Cons_tan*, enriquecido de genes associados ao TEA e com diferença de expressão entre os sexos, apresentou níveis de

expressão gênica mais elevados em amostras masculinas (Figuras 9 e 11). O gráfico na figura 10 mostra a variabilidade dos valores *eigengene* entre as amostras de cada sexo deste módulo, resultando no índice de correlação de 0,5 (entre 0 e 1) entre o nível de expressão e o sexo.



Resultado da análise de expressão gênica diferencial no *software* R Studio. Amostras femininas (turquesa) e masculinas (azul).

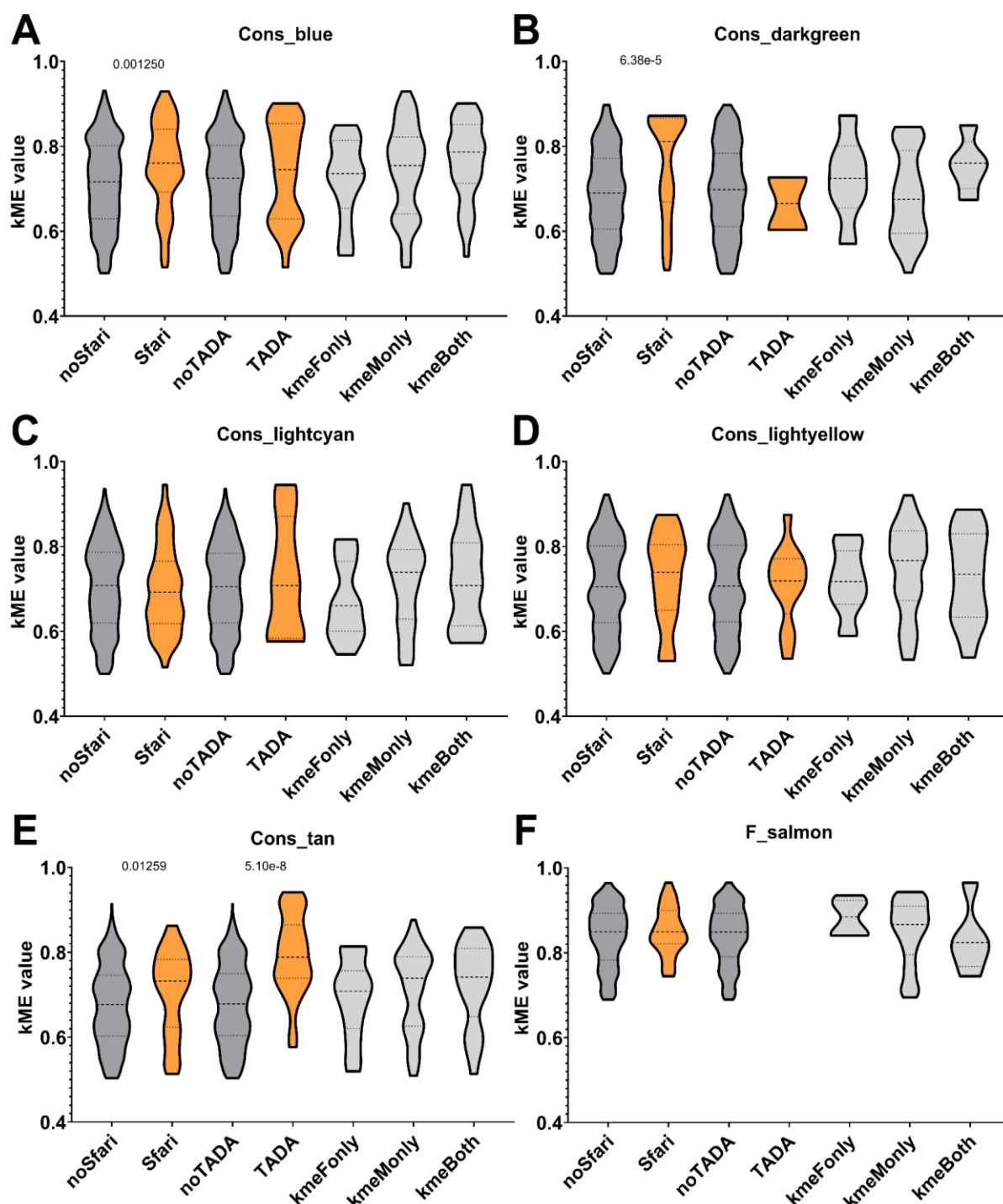
Figura 10. Valores *eigengene* do módulo *consensus tan*, enriquecido de genes associados ao transtorno do espectro autista e que se apresentam diferencialmente expressos entre os sexos

Estes dados sugerem que, na janela de 16-21 semanas gestacionais, há ao menos um módulo com diferenças de nível de expressão gênica (*Cons_tan*) no cérebro de homens comparativamente a mulheres que podem estar relacionadas à diferença de prevalência de TEA entre os sexos.

3.2.3 Análise da distribuição topológica dos genes de risco para o transtorno do espectro autista nos módulos de coexpressão de interesse

Assim como realizado para os módulos de 12-13 semanas gestacionais, analisamos a distribuição topológica de genes que possuem mutações potencialmente patogênicas em indivíduos com o transtorno dentro dos módulos de interesse identificados na rede gerada para as amostras de 16-21 semanas gestacionais. A distribuição topológica dos genes SFARI e TADA também foi observada para fins comparativos. O módulo *Cons_tan*, que se mostrou com expressão mais elevada em indivíduos do sexo masculino, apresentou também uma distribuição topológica dos genes SFARI e TADA mais central neste módulo comparativamente a genes sem associação ao TEA, porém esta distribuição não se mostrou diferente para os grupos de genes *Fonly*, *Monly* e *both* (Figura 11E). Dentre os demais módulos consenso

enriquecidos para genes relacionados ao TEA, apenas nos módulos *Cons_darkgreen* e *Cons_blue* os genes SFARI se apresentam mais centrais, porém não se repetindo esse padrão para os genes TADA (Figura 11A e B). Curiosamente, nenhum dos módulos de interesse investigados tiveram diferença significativa entre os genes *Fonly*, *Monly* e *both*.



Resultados da análise da distribuição topológica no software *GraphPad Prism*; (A) Comparação dos valores de kME do módulo *Cons_blue*; (B) Comparação dos valores de kME do módulo *Cons_darkgreen*; (C) Comparação dos valores de kME do módulo *Cons_lightcyan*; (D) Comparação dos valores de kME do módulo *Cons_lightyellow*; (E) Comparação dos valores de kME do módulo *Cons_tan*; (F) Comparação dos valores de kME do módulo *F_salmon*. As comparações se dão entre os genes SFARI e não SFARI, entre os genes TADA e não TADA, e entre os genes mutados exclusivamente para cada sexo (*kmeFonly* e *kmeMonly*) e estes também comparados aos genes mutados em ambos os sexos (*kmeboth*).

Figura 11. Análise da distribuição topológica entre genes SFARI e não SFARI, TADA e não TADA, e entre os genes mutados exclusivamente em um sexo ou em ambos nos módulos de coexpressão de interesse

3.2.4 Análise de enriquecimento funcional

Ao analisar o enriquecimento funcional dos módulos de interesse, identificamos que os módulos *Cons_tan* e *Cons_blue* se apresentaram significativamente enriquecidos de genes relacionados à reguladores da expressão gênica. Já o módulo *Cons_lightcyan* se mostra relacionado à sinalização celular e ao neurodesenvolvimento e o módulo *Cons_darkgreen*, à adesão celular (Tabela 10). Os módulos *Cons_lightyellow* e *F_salmon* não apresentaram resultados estatisticamente significante para enriquecimento funcional relacionado aos processos biológicos.

Tabela 10. Análise de enriquecimento funcional para processos biológicos dos módulos *consensus*

Módulo	Função	Count	% de genes	Valor-p	Benjamini
Cons_tan	<i>regulation of gene expression</i>	142	37.9	2.70E-08	2.10E-05
	<i>cellular macromolecule biosynthetic process</i>	140	37.3	2.00E-04	2.30E-02
	<i>regulation of macromolecule biosynthetic process</i>	137	36.5	2.70E-09	6.30E-06
	<i>RNA metabolic process</i>	137	36.5	6.50E-05	9.00E-03
	<i>regulation of cellular macromolecule biosynthetic process</i>	136	36.3	2.70E-09	6.30E-06
Cons_darkgreen	<i>cell-cell adhesion</i>	33	10.5	1.40E-06	1.90E-03
	<i>cell-cell adhesion via plasma-membrane adhesion molecules</i>	25	7.9	4.00E-12	1.50E-08
	<i>homophilic cell adhesion via plasma membrane adhesion molecules</i>	18	5.7	2.90E-09	5.70E-06
Cons_lightcyan	<i>regulation of signaling</i>	100	30.4	6.40E-07	4.90E-04
	<i>regulation of cell communication</i>	96	29.2	4.60E-06	2.40E-03
	<i>nervous system development</i>	79	24	4.90E-05	1.20E-02
	<i>intracellular signal transduction</i>	74	22.5	5.20E-05	1.20E-02
	<i>cell-cell signaling</i>	56	17	6.90E-07	4.90E-04
	<i>gene expression</i>	213	40.7	7.70E-06	1.80E-03
	<i>RNA metabolic process</i>	188	35.9	1.90E-06	4.50E-04
	<i>cellular macromolecule biosynthetic process</i>	186	35.6	1.10E-04	1.80E-02
	<i>regulation of gene expression</i>	185	35.4	1.50E-08	6.80E-06
	<i>regulation of cellular macromolecule biosynthetic process</i>	174	33.3	5.80E-09	3.50E-06

Count: contagem de leituras.

3.2.5 Análise de sobreposição entre módulos *consensus* enriquecidos e módulos implicados previamente no transtorno do espectro autista

Os módulos *Cons_darkgreen* e *Cons_lightcyan* se sobrepõem ao módulo *brown_Mariani* [Mariani et al.⁽³⁷⁾], sendo este um módulo com expressão

aumentada em autistas. Ainda, ambos se sobrepõem ao módulo *MNeu2-tan_Griesi*, um módulo enriquecido de genes SFARI, porém sem diferenças de expressão em indivíduos com TEA.⁽³⁶⁾ Já o módulo *Cons_blue* se sobrepõe ao módulo *lightcyan_Schafer* [Schafer et al.⁽³⁸⁾], um módulo enriquecido de genes de risco para o TEA.

Por fim, o módulo *Cons_lightyellow* se mostrou sobreposto a dois módulos de Schafer et al.,⁽³⁸⁾ a saber: *brown_Schafer*, um módulo enriquecido de genes de risco para o TEA e *lightyellow_Schafer*, um módulo enriquecido e que apresenta expressão aumentada em autistas (Tabela 11). Os módulos *Cons_tan* e *F_salmon*, também enriquecidos para genes associados ao TEA, não apresentaram sobreposição significativa com nenhum módulo previamente implicado no transtorno.

Tabela 11. Análise de sobreposição entre módulos *consensus* enriquecidos e módulos de neurônios derivados de células-tronco pluripotentes induzidas de outros estudos

Módulo	Genes totais	Genes sobrepostos	Outros estudos	Valor-p corrigido
Cons_darkgreen*	296	36	MNeu2-tan_Griesi*	9.51E-09
		63	brown_Mariani**	8.54E-13
		18	MNeu14-red_Griesi*	0.0095
Cons_lightcyan*	304	48	MNeu2-tan_Griesi*	1.13E-16
Cons_blue*	482	63	brown_Mariani**	1.01E-11
		88	lightcyan_Schafer*	1.17E-56
Cons_lightyellow*	275	65	lightyellow_Schafer**	3.86E-09
		188	brown_Schafer*	4.57E-11

*Módulo enriquecido de genes de risco para o TEA. **Módulo enriquecido de genes de risco para o TEA com expressão elevada em indivíduos com TEA.

3.3 Análises de coexpressão gênica de neurônios e células progenitoras neurais derivadas de células-tronco pluripotentes induzidas

Buscamos trabalhar com dados de transcriptoma de neurônios e NPCs derivados de iPSCs obtidas de indivíduos não autistas, contudo, devido ao baixo número amostral (abaixo de 15 amostras) necessário para a construção das redes com amostras femininas, optamos por não prosseguir com as análises para os neurônios e NPCs derivados de iPSCs. Ressaltamos que tentamos construir os módulos partindo de apenas 9 amostras femininas porém, de fato, o dado não assumiu uma distribuição seguindo o modelo de topologia livre de escala ($SFT.R.sq \geq 0.8$) mesmo ao se elevar os índices de correlação de expressão dos diferentes pares de genes a potências tão altas quanto 30 (função “*pickSoftThreshold*”, ver detalhes do método na seção 2.5).

4 DISCUSSÃO

Visando melhor compreender os mecanismos ainda não esclarecidos que contribuem para o forte desvio de proporção sexual entre os indivíduos autistas, e cientes de que estudos da literatura têm mostrado consistentemente a desregulação de módulos de coexpressão gênica em células neuronais em pacientes autistas,^(33–39) formulamos como hipótese para este trabalho a possibilidade da existência basal de uma regulação diferencial de expressão gênica cerebral entre os sexos, quer seja estrutural ou de nível de expressão. Assim sendo, usando dados de transcriptoma de amostras de regiões cerebrais e janelas temporais críticas para a fisiopatologia do TEA, identificamos módulos de coexpressão gênica enriquecidos de genes associados ao transtorno e verificamos se estes apresentavam níveis de expressão ou relações de coexpressão diferentes entre os sexos.

Os resultados obtidos para as amostras de tecidos corticais de cérebro fetal *post-mortem* de 12 a 13 semanas gestacionais apontaram um módulo comum entre os dois sexos (*Cons_turquoise*) enriquecido de genes com evidência de associação ao TEA. No entanto, este módulo, que é enriquecido de genes relacionados à regulação de sinalização celular, não apresentou diferença significativa quanto aos níveis de expressão entre os sexos. Por outro lado, este módulo mostra sobreposição significativa com módulos com expressão aumentada em indivíduos com TEA descritos por outros estudos na literatura.^(36–38) Assim, estes resultados sugerem que este módulo pode estar envolvido na fisiopatologia do transtorno, porém seus padrões de expressão entre os sexos não respondem pelo desvio de proporção sexual em TEA. De fato, não observamos uma distribuição topológica diferenciada dos genes mutados exclusivamente em homens, em mulheres ou em ambos os sexos dentro deste módulo. Finalmente, é interessante notar que os genes SFARI e TADA possuem uma posição central neste módulo, ou seja, um impacto na sua regulação seria de fato deletéria para o desenvolvimento do sistema nervoso.

Nesta mesma janela temporal, detectamos um módulo feminino (*F_tan*) que também é enriquecido de genes com evidência de associação ao TEA. Apesar deste módulo ter sobreposições com módulos masculinos (estes, porém, sem enriquecimento de genes relacionados ao TEA), as correlações de expressão dentro destes conjuntos de genes parecem ser diferentes a ponto de estes não emergirem como

módulo comum na análise consenso. Estes resultados sugerem que a maneira como estes genes do módulo *F_tan* são co-regulados, particularmente os genes associados ao TEA, agiria de forma protetiva no sexo feminino em relação à susceptibilidade de desenvolvimento do fenótipo. Análises de redes de coexpressão com mais amostras dos dois sexos seriam importantes para compreender se o aparecimento do módulo *F_tan* e seu enriquecimento de genes associados ao TEA foi apenas um achado espúrio ou se trata-se de fato de um módulo estruturalmente diferente entre os sexos.

Já em amostras de 16 a 21 semanas gestacionais, identificamos 5 módulos consenso enriquecidos de genes relacionados ao TEA. Dentre estes, os módulos *Cons_darkgreen*, *Cons_lightcyan* e *Cons_blue* não se mostraram diferencialmente expressos e nem com evidência de diferença estrutural entre homens e mulheres. Além do enriquecimento de genes associados ao TEA, estes 3 módulos apresentaram sobreposição significativa com módulos previamente identificados como desregulados em autistas na literatura. De forma similar ao que discutimos para o módulo *Cons_tan* da janela de 12-13pcw, estes resultados sugerem que os 3 módulos acima destacados possuem uma relevância para fisiopatologia do TEA, mas não respondem pela desproporção de casos de homens e mulheres afetadas. A ausência de diferença na distribuição topológica dos genes exclusivamente mutados em homens, mulheres ou em ambos também suporta essa conclusão.

Um quarto módulo enriquecido de genes associados ao TEA, o módulo *Cons_lightyellow*, apresentou sobreposição com o módulo masculino enriquecido *M_magenta*, porém, dentre os módulos femininos este módulo apresentou sobreposição mais volumosa com o módulo feminino *F_pink*, um módulo não enriquecido para genes com variantes raras *de novo*. Isto indica que este módulo, embora consensual entre os sexos e enriquecido para genes de risco para o TEA, apresenta relações de coexpressão dos genes com variantes raras *de novo* nas amostras masculinas que não se refletem nas mulheres. Este resultado apoia nossa hipótese de que a desproporção entre homens e mulheres afetados por TEA pode ter relação com módulos estruturalmente diferentes entre os sexos.

Finalmente, o quinto módulo enriquecido de genes associados ao TEA, o *Cons_tan*, um módulo funcionalmente associado à regulação de expressão gênica, mostrou-se com níveis de expressão mais elevados nos homens. Curiosamente, não encontramos sobreposição gênica significativa entre este módulo e módulos detectados em células neuronais derivadas de iPSC descritos até o momento na

literatura como sendo desregulados em indivíduos autistas.⁽³⁶⁻³⁸⁾ Importante frisar, porém, que os dados de transcriptoma derivados de tecidos cerebrais *post-mortem* são constituídos, naturalmente, por uma população celular neuronal e por células gliais, ao passo que nas amostras de neurônios derivados de iPSCs temos uma população celular mais pura de neurônios. Ainda, os fenômenos relativos à regulação da expressão gênica podem se dar de formas diferentes em neurônios e células gliais. Portanto, estas diferenças intrínsecas aos dados de expressão de diferentes populações celulares podem em certo nível comprometer as análises de sobreposição entre os módulos que partiram de cérebro *post-mortem* neste estudo e os módulos de neurônios derivados de iPSC da literatura. Assim, se o módulo *Cons_turquoise* for relacionado a um tipo celular (p.e., astrócitos) que está pouco representado na cultura de células derivadas de iPSC, este módulo sequer emergiria na análise de redes do transcriptoma destas células. Não detectamos, porém, enriquecimento de marcadores de astrócitos e oligodendrócitos neste módulo (dado não mostrado), mas ainda assim, este resultado não invalida a possibilidade explicitada. Futuros estudos de transcriptômica de outros tipos celulares, tais como astrócitos e oligodendrócitos, derivados de iPSC de pacientes autistas podem ajudar a elucidar esta questão.

Ainda, não observamos nenhuma diferença na distribuição topológica dos genes mutados exclusivamente em homens, mulheres ou em ambos os sexos no módulo *Cons_tan*, ao contrário do que prevíamos em nossa hipótese, particularmente para módulos que se mostraram diferencialmente expressos entre os sexos. No entanto, é importante ressaltar que testamos apenas a distribuição de genes nos quais foram detectadas, em pacientes autistas, variantes *de novo* potencialmente patogênicas, as quais estão mais associadas a casos monogênicos. Uma vez que não há dados disponíveis publicamente de variantes herdadas e frequentes, não foi possível avaliarmos nossa questão de trabalho dentro da perspectiva de um modelo poligênico, ou seja, avaliar se o número de variantes por indivíduo em um dado módulo é maior entre as mulheres do que nos homens. Alguns consórcios internacionais, como a iniciativa MISSNG (<https://research.mss.ng/>), disponibiliza dados de sequenciamento de genomas de pacientes autistas mediante submissão de projeto para o banco. Como perspectiva de continuidade deste trabalho, seria interessante utilizar estes dados para testar a hipótese acima.

Apesar dos resultados obtidos até o momento indicarem a presença de módulos de interesse que possuem diferenças tanto em nível estrutural, bem como,

no nível de expressão gênica diferencial entre os sexos, os grupos amostrais de cérebro *post-mortem* utilizados, embora sejam compostos por um número considerável de amostras (córtex 12-13 pcw: n=57; córtex 16-21 pcw: n=46), apresentam uma variabilidade baixa em termos de indivíduos doadores [córtex 12-13 pcw: n=6 (4 mulheres e 2 homens); córtex 16-21 pcw: n=7 (3 mulheres e 4 homens)]. Buscamos também utilizar dados de transcriptoma de neurônios, bem como NPCs derivadas de iPSCs de indivíduos não diagnosticados com TEA para a construção de módulos de coexpressão gênica. Porém ambos os grupos amostrais (neurônio e NPCs) são compostos por somente 9 amostras femininas cada, o que impossibilitou atingirmos uma topologia de distribuição livre de escala dos dados, invalidando assim o uso destas amostras. Assim, uma validação em amostras de tamanho mais amplo se faz importante para suportar os achados deste estudo.

Vale registrar o desafio que tivemos no decorrer do projeto com relação a utilização de um critério objetivo para a exclusão de amostras *outliers* dos grupos amostrais selecionados para este estudo. Inicialmente, para auxiliar na tomada de decisão de quais amostras eventualmente excluir em caso de grande dissimilaridade de perfil de expressão, plotamos alguns tipos de gráficos que quantificam a similaridade de um grupo amostral, sendo estes: PCA, identificando as variáveis idade, região cerebral, sexo e doador da amostra; *boxplot*, para visualizar como as amostras se distribuem; e a clusterização hierárquica (por meio da função R *hclust*), uma análise de agrupamento hierárquico das amostras por meio de um cálculo de dissimilaridades dos objetos agrupados. Ao observarmos tais gráficos não ficou muito claro quais amostras são, consistentemente, *outliers*. Decidimos então adotar como critério a exclusão das amostras com a mediana do valor de expressão maior ou menor que 1,5 DP em relação a média das medianas das amostras do dado sexo, conforme descrito na seção 2.4.

De modo geral, em amostras corticais derivadas de cérebro fetal *post-mortem* de indivíduos não diagnosticados com TEA, pertencentes a uma janela temporal e regiões cerebrais críticas para fisiopatologia do TEA, há módulos enriquecidos de genes associados ao TEA. Dentre eles, há módulos que apresentaram relações de coexpressão diferente entre os sexos ou que se mostraram diferentes em relação ao nível de expressão gênica, particularmente apresentando níveis de expressão mais elevados em amostras masculinas. Sendo assim, os resultados obtidos até o momento sugerem que as diferenças entre os sexos, em nível estrutural como a nível do perfil de expressão gênica em módulos importantes para o TEA, podem ser fatores que

contribuem, ao menos em parte, para o dimorfismo sexual presente no TEA, que se reflete na proporção diagnóstica mais prevalente em homens. Entretanto, mais pesquisas são necessárias para entender melhor o papel da regulação diferencial da expressão gênica no transtorno do espectro autista nas diferenças sexuais.

5 CONCLUSÕES

1. O módulo *Cons_turquoise*, presente em amostras corticais de 12 à 13 semanas gestacionais, é comum a ambos os sexos, mas não apresenta diferença significativa nos níveis de expressão entre eles, sugerindo que seus padrões de expressão não são responsáveis pelo viés sexual observado no transtorno do espectro autista;

2. Outro módulo oriundo de amostras corticais de 12 à 13 semanas gestacionais, o módulo *F_Tan*, específico em amostras femininas, também está enriquecido com genes associados ao transtorno do espectro autista e pode estar atuando de forma protetora em fêmeas;

3. Três módulos presentes em amostras corticais de 16 à 21 semanas gestacionais (*Cons_darkgreen*, *Cons_lightcyan* e *Cons_blue*) não mostraram níveis de expressão diferencial ou diferenças estruturais entre homens e mulheres, indicando sua relevância para a fisiopatologia do transtorno do espectro autista, mas não para o viés sexual;

4. O módulo *Cons_yellow* (córtex de 16 à 21 semanas gestacionais) apresenta relações de coexpressão, dos genes com variantes raras *de novo*, nas amostras masculinas que não se refletem nas mulheres. Isso sugere que a desproporção entre homens e mulheres afetados por transtorno do espectro autista pode estar relacionada a módulos genéticos estruturalmente diferentes entre os sexos;

5. As diferenças sexuais na expressão gênica podem contribuir para o viés sexual observado no transtorno do espectro autista, mas ainda não está claro como essas diferenças ocorrem e como elas se relacionam com a fisiopatologia do transtorno do espectro autista.

6 REFERÊNCIAS

1. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The changing epidemiology of autism spectrum disorders. *Annu Rev Public Health*. 2017;38(1):81–102.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington (VA): American Psychiatric Association; 2013.
3. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al.; EdS1; PhD-7. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveill Summ*. 2020;69(4):1–12.
4. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci*. 2015;16(9):551–63.
5. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA*. 2015;314(9):895-903.
6. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010;466(7304):368-72.
7. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012;74(2):285–99.
8. Griesi-Oliveira K, Sertié AL. Autism spectrum disorders: an updated guide for genetic counseling. *einstein (São Paulo)*. 2017;15(2):233–8.
9. Gilks WP, Abbott JK, Morrow EH. Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet*. 2014;30(10):453–63.
10. Lai MC, Lombardo MV, Suckling J, Ruigrok AN, Chakrabarti B, Ecker C, et al.; MRC AIMS Consortium. Biological sex affects the neurobiology of autism. *Brain*. 2013;136(Pt 9):2799–815.
11. Zeestraten EA, Gudbrandsen MC, Daly E, de Schotten MT, Catani M, Dell'Acqua F, et al.; MRC AIMS Consortium. Sex differences in frontal lobe connectivity in adults with autism spectrum conditions. *Transl Psychiatry*. 2017;7(4):e1090.
12. Schaer M, Kochalka J, Padmanabhan A, Supekar K, Menon V. Sex differences in cortical volume and gyrification in autism. *Mol Autism*. 2015;6(1):42.
13. Supekar K, Menon V. Sex differences in structural organization of motor systems and their dissociable links with repetitive/restricted behaviors in children with autism. *Mol Autism*. 2015;6(1):50.
14. Nordahl CW, Iosif AM, Young GS, Perry LM, Dougherty R, Lee A, et al. Sex differences in the corpus callosum in preSchool-aged children with autism spectrum disorder. *Mol Autism*. 2015;6(1):1–11.
15. Maschietto M, Bastos LC, Tahira AC, Bastos EP, Euclydes VL, Brentani A, et al. Sex differences in DNA methylation of the cord blood are related to sex-bias psychiatric diseases. *Sci Rep*. 2017 ;7:44547.
16. Tahira AC, Barbosa AR, Feltrin AS, Gastaldi VD, de Toledo VH, de Carvalho Pereira JG, et al. Putative contributions of the sex chromosome proteins SOX3 and SRY to neurodevelopmental disorders. *Am J Med Genet B Neuropsychiatr Genet*. 2019;180(6):390–414.

17. Verstraeten BS, McCreary JK, Falkenberg EA, Fang X, Weyers S, Metz GA, et al. Multiple prenatal stresses increase sexual dimorphism in adult offspring behavior. *Psychoneuroendocrinology*. 2019;107:251–60.
18. Chan JC, Nugent BM, Bale TL. Parental advisory: maternal and paternal stress can impact offspring neurodevelopment. *Biol Psychiatry*. 2018;83(10):886–94.
19. Bale TL, Epperson CN. Sex differences and stress across the lifespan. *Nat Neurosci*. 2015;18(10):1413–20.
20. Zhang Y, Li N, Li C, Zhang Z, Teng H, Wang Y, et al. Genetic evidence of gender difference in autism spectrum disorder supports the Female-protective effect. *Transl Psychiatry*. 2020;10(1):4.
21. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*. 2011;70(5):898–907.
22. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*. 2011;70(5):886–97.
23. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, et al. A higher mutational burden in Females supports a “Female protective model” in neurodevelopmental disorders. *Am J Hum Genet*. 2014;94(3):415–25.
24. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70(5):863–85.
25. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012;485(7397):242–5.
26. De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE, et al.; UK10K Consortium. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209–15.
27. Baron-Cohen S. The extreme male brain theory of autism. *Trends Cogn Sci*. 2002;6(6):248–54.
28. Baron-Cohen S, Auyeung B, Nørgaard-Pedersen B, Hougaard DM, Abdallah MW, Melgaard L, et al. Elevated fetal steroidogenic activity in autism. *Mol Psychiatry*. 2015;20(3):369–76.
29. Auyeung B, Baron-Cohen S, Ashwin E, Knickmeyer R, Taylor K, Hackett G, et al. Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychol Sci*. 2009;20(2):144–8.
30. Baron-Cohen S, Lombardo MV, Auyeung B, Ashwin E, Chakrabarti B, Knickmeyer R. Why are autism spectrum conditions more prevalent in males? *PLoS Biol*. 2011;9(6):e1001081.
31. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4(1):e17.
32. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
33. 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.
34. 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.

35. 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.
36. 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*. 2021;26(5):1589-605.
37. 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.
38. 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.
39. 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.
40. Werling DM, Parikshak NN, Geschwind DH. Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat Commun*. 2016;7:10717.
41. Shi L, Zhang Z, Su B. Sex biased gene expression profiling of human brains at major developmental stages. *Sci Rep*. 2016;6(1):21181.
42. de Toledo VH, Feltrin AS, Barbosa AR, Tahira AC, Brentani H. Sex differences in gene regulatory networks during mid-gestational brain development. *Front Hum Neurosci*. 2022;16:955607.
43. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478(7370):483–9.
44. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al.; Autism Sequencing Consortium; iPSYCH-Broad Consortium. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*. 2020;180(3):568–584.e23.
45. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S, et al.; Autism Sequencing Consortium (ASC); Broad Institute Center for Common Disease Genomics (Broad-CCDG); iPSYCH-BROAD Consortium. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat Genet*. 2022;54(9):1320–31.
46. Zhou X, Feliciano P, Shu C, Wang T, Astrovskaya I, Hall JB, et al.; SPARK Consortium. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat Genet*. 2022;54(9):1305–19.
47. Hoffman GE, Hartley BJ, Flaherty E, Ladrán I, Gochman P, Ruderfer DM, et al. Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat Commun*. 2017;8(1):2225.
48. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013;155(5):997-1007.
49. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-85.

Abstract

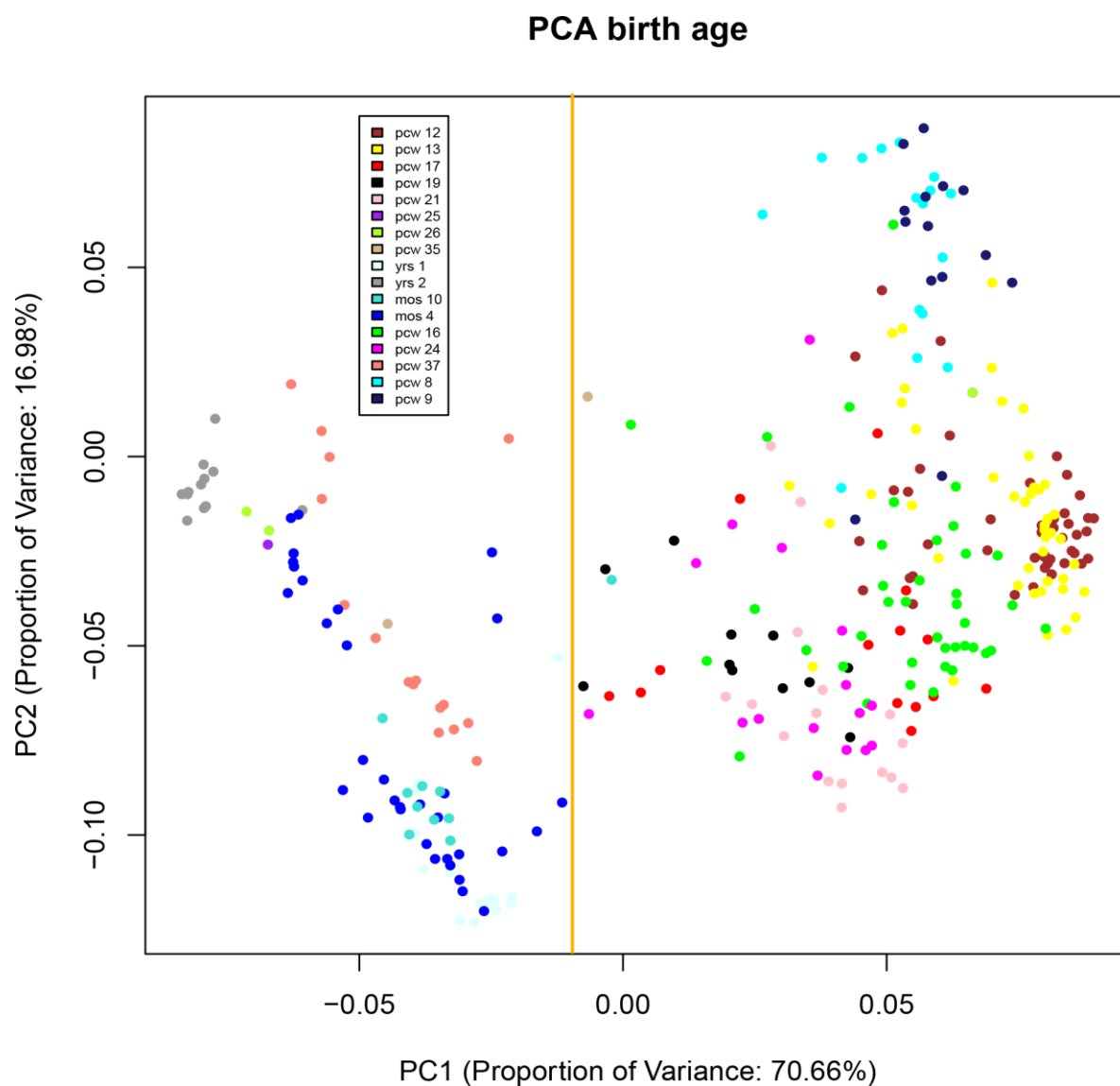
Introduction: Autism spectrum disorder is characterized by impairment in social communication, combined with repetitive and restrictive patterns of behavior, interests or activities. Autism spectrum disorder is a genetically heterogeneous disorder (with different genes and inheritance patterns potentially involved in each patient) that affects about 1.5% of the population and exhibits a strong sex ratio deviation among those affected: there are, on average, 4 men for every woman diagnosed. Although there are some theories to explain this sex ratio deviation, the physiological mechanisms behind this phenomenon are not fully understood. Transcriptome studies of autistic patients have consistently shown dysregulation of co-expression modules in neuronal cells in these patients. Therefore, considering that this dysregulation of expression may be the triggering factor for autism spectrum disorder, a mechanism that could justify the sex ratio deviation in the disorder would be the existence of a basal differential regulation of gene expression in the brain between male and Female individuals. **Purpose:** To analyze whether the co-expression gene modules identified in fetal brains, in temporal windows and regions previously implicated in autism spectrum disorder, differ between sexes in terms of structure or expression relationships; and to analyze whether the distribution of genes that have potentially pathogenic mutations in autistic patients of different sexes differs in the generated co-expression networks. **Methods:** We constructed gene co-expression networks using the weighted gene co-expression network analysis method from publicly available transcriptome data from individuals without autism spectrum disorder, obtained from post-mortem brain tissues. The identified co-expressed gene modules were then analyzed for their enrichment of genes associated with autism spectrum disorder. **Results:** In post-mortem cortical samples from 12 to 13 gestational weeks of individuals without autism spectrum disorder, we identified 4 enriched gene modules associated with autism spectrum disorder. Among them, 3 modules were structurally unique in a given sex. Moreover, 2 modules showed gene overlap with at least one module already described in another study with a higher expression profile in autistic individuals. Regarding post-mortem cortical samples from 16 to 21 gestational weeks, we identified 15 enriched gene modules associated with autism spectrum disorder. Among them, one module had higher expression levels in male samples. Additionally, 3 modules showed gene overlap with at least one module already described in another study with a

higher expression profile in autistic individuals. **Conclusion:** The results obtained so far suggest that differences between sexes, both in terms of structure and gene expression profile in modules important for autism spectrum disorder, may be factors that contribute, at least in part, to the sexual dimorphism present in autism spectrum disorder, which is reflected in the more prevalent diagnostic ratio in men.

Keywords: Autism spectrum disorder; Gene expression; Transcriptome; Embryonic and fetal development; Gene regulatory networks

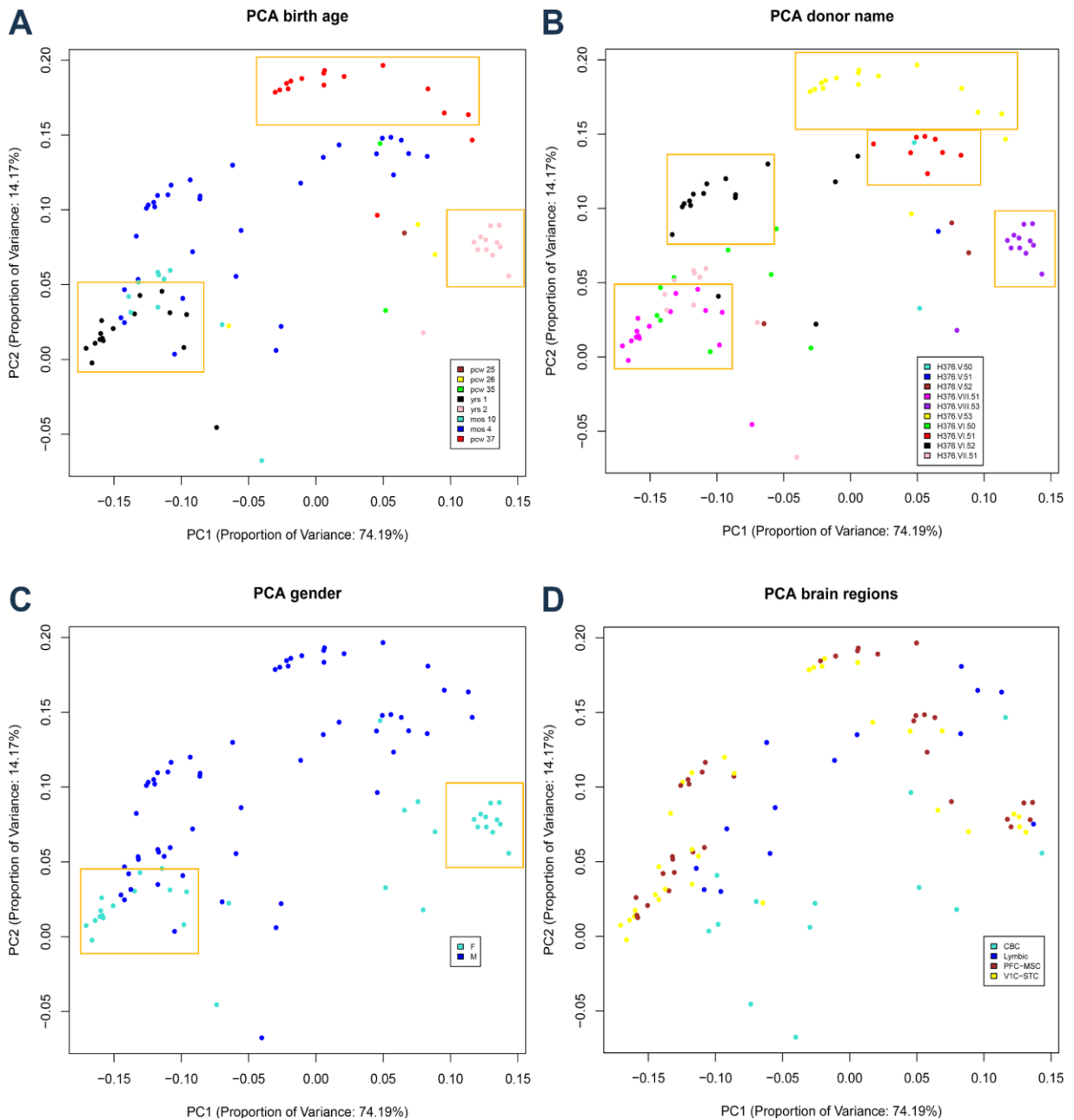
Apêndices

Apêndice 1. Análise de componentes principais de amostras corticais, cerebelares e límbicas de 8 semanas gestacionais a 2 anos



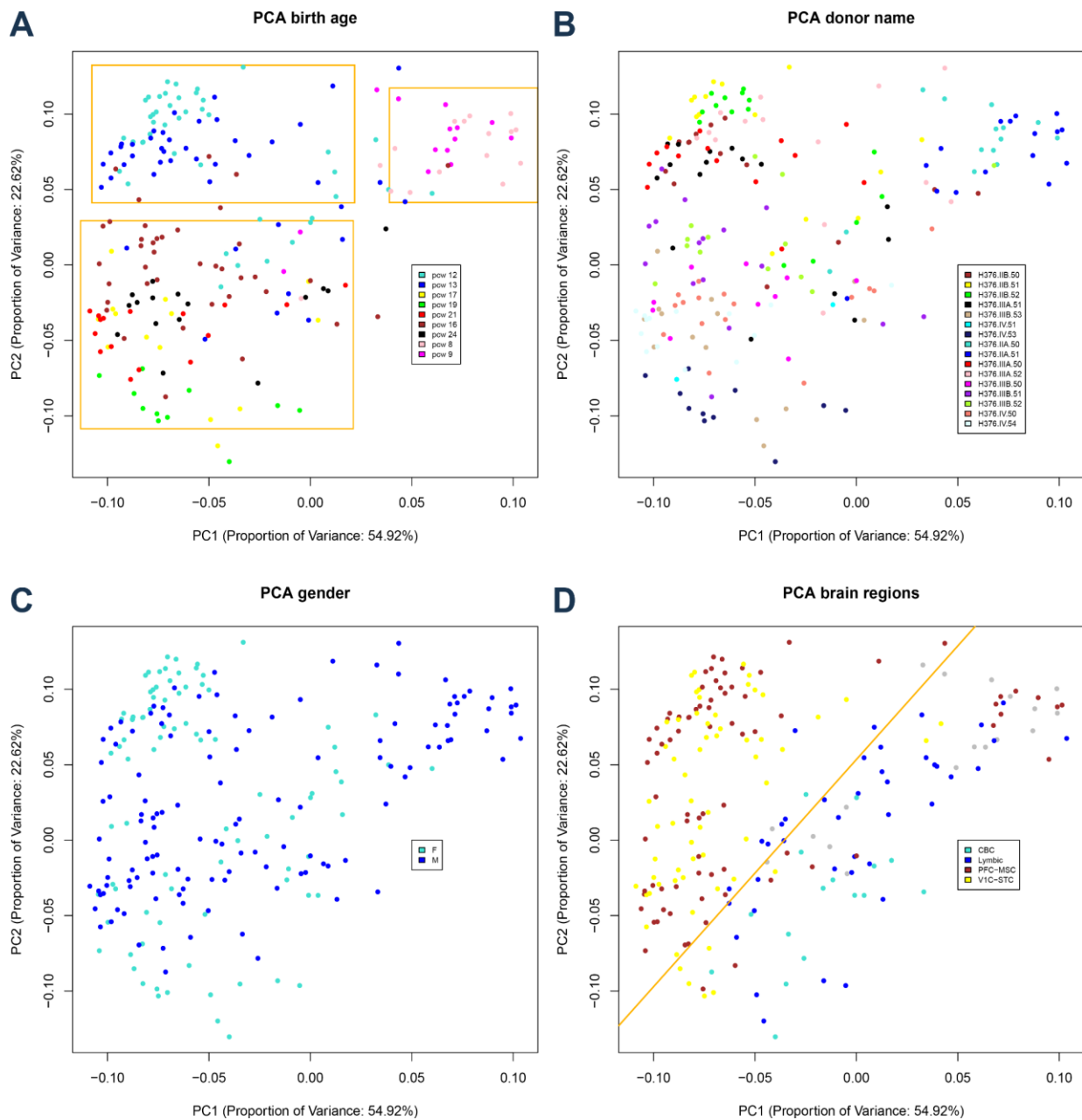
Resultado da análise de componente principal no *software* R. Amostras de regiões cerebrais do córtex pré-frontal e motor-somatosensorial primário (PFC/MS), córtex visual primário e córtex temporal superior (V1C/STC), cerebelo (CBC) e região límbica (LBS).

Apêndice 2. Análise de componentes principais de amostras das regiões corticais, cerebelares e límbicas numa janela temporal de 25 semanas gestacionais a 2 anos



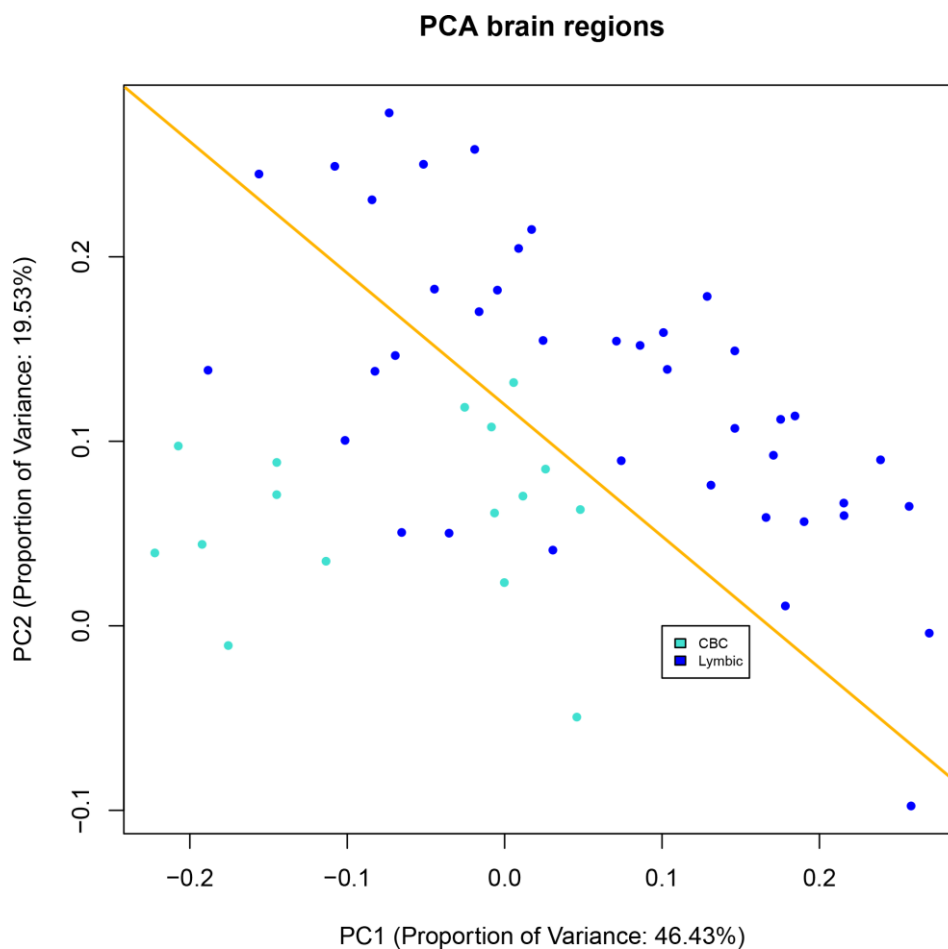
Resultado da análise de componente principal no *software* R. (A) Agrupamentos de amostras por idade [semanas gestacionais (pcw); meses (mos); anos (yrs)]. (B) Agrupamentos de amostras por doador da amostra. (C) Agrupamentos de amostras por sexo [feminino (F); masculino (M)]. (D) Agrupamentos de amostras por região cerebral [córtices pré-frontal e motor-somatosensorial primário (PFC/MSC); córtices visual primário e temporal superior (V1C/STC); cerebelo (CBC); sistema límbico (LBS)].

Apêndice 3. Análise de componentes principais de amostras das regiões corticais, cerebelares e límbicas numa janela temporal de 8-24 semanas gestacionais



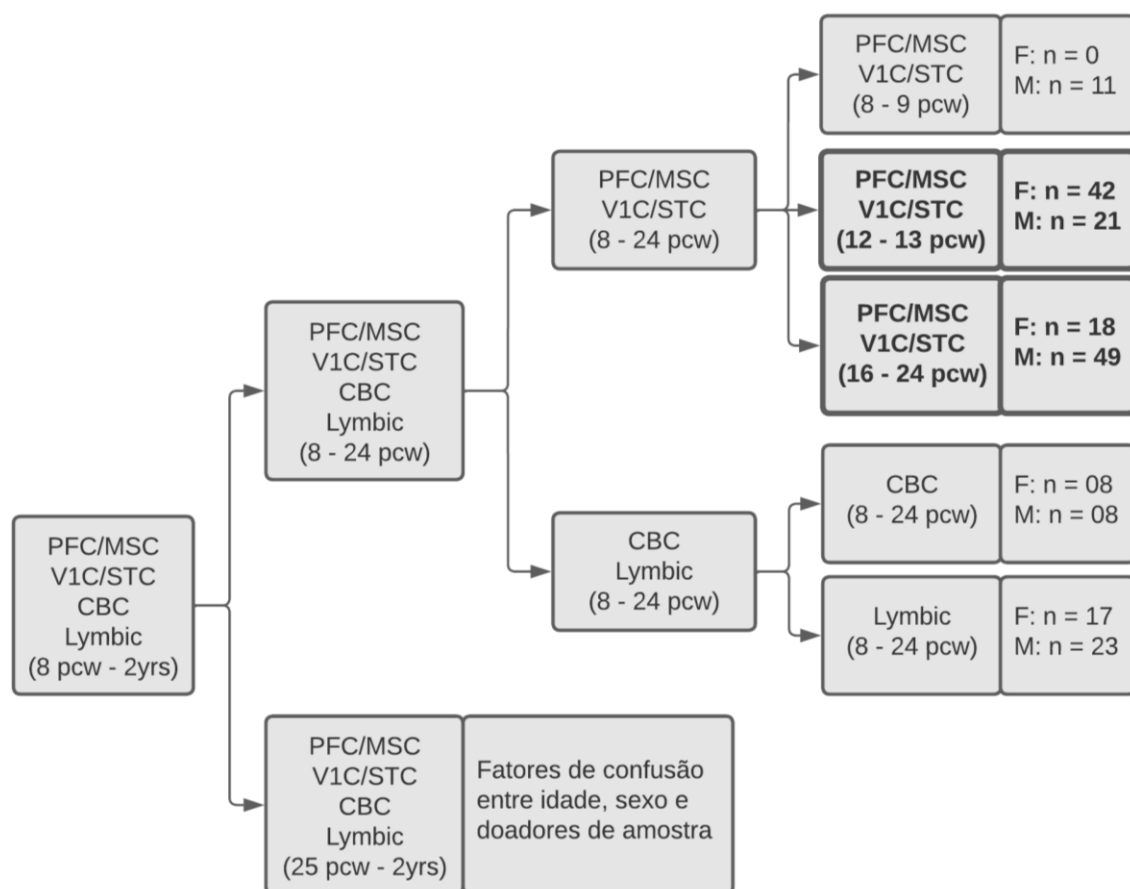
Resultado da análise de componente principal no *software* R. (A) Agrupamentos de amostras por idade [semanas gestacionais (pcw)]. (B) Agrupamentos de amostras por doador da amostra. (C) Agrupamentos de amostras por sexo [feminino (F); masculino (M)]. (D) Agrupamentos de amostras por região cerebral [córtices pré-frontal e motor-somatossensorial primário (PFC/MSc); córtices visual primário e temporal superior (V1C/STC); cerebelo (CBC); sistema límbico (LBS)].

Apêndice 4. Análise de componentes principais de amostras cerebelares e límbicas numa janela temporal de 8 a 24 semanas gestacionais



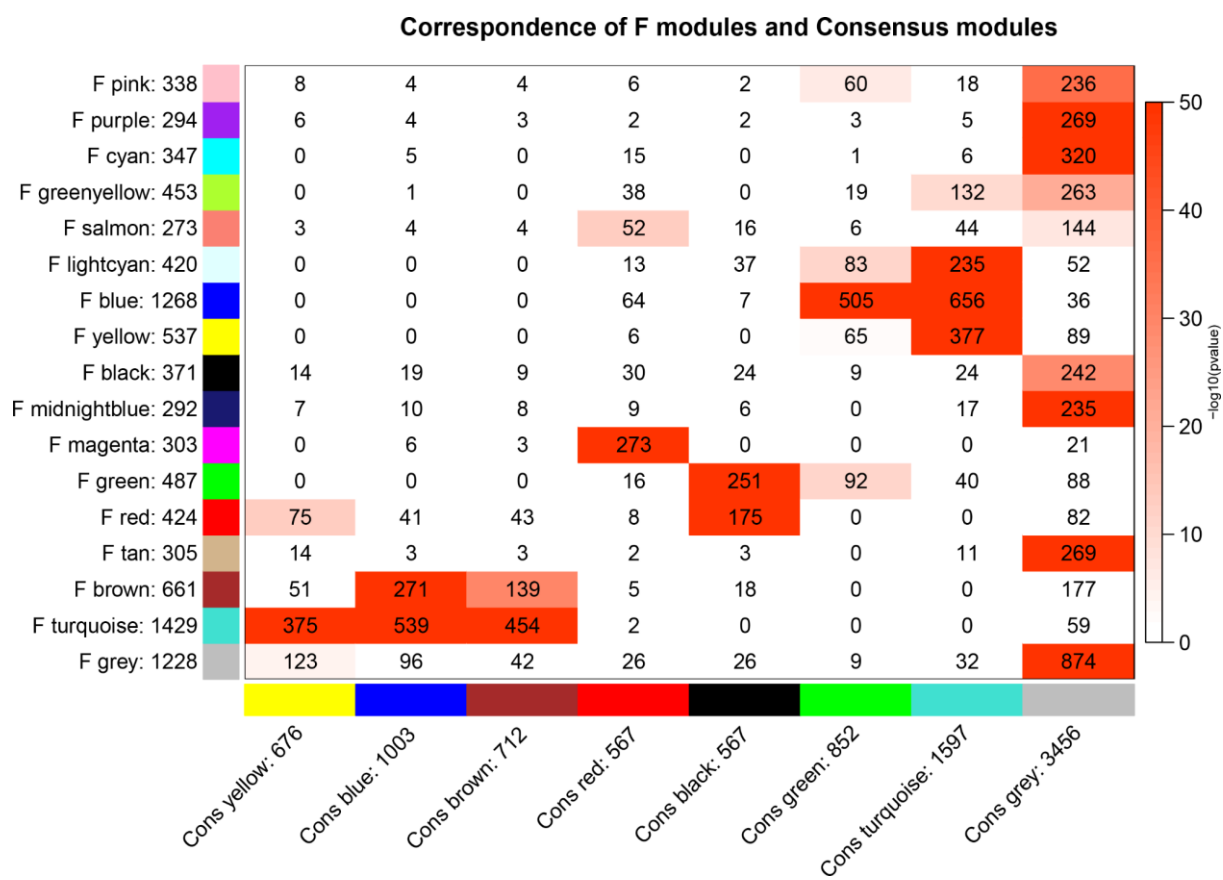
Resultado da análise de componente principal no *software* R. Destacou-se a separação entre as amostras de regiões cerebelares das amostras de regiões límbicas.

Apêndice 5. Grupos de amostras encontrados após diferentes análises de componente principal



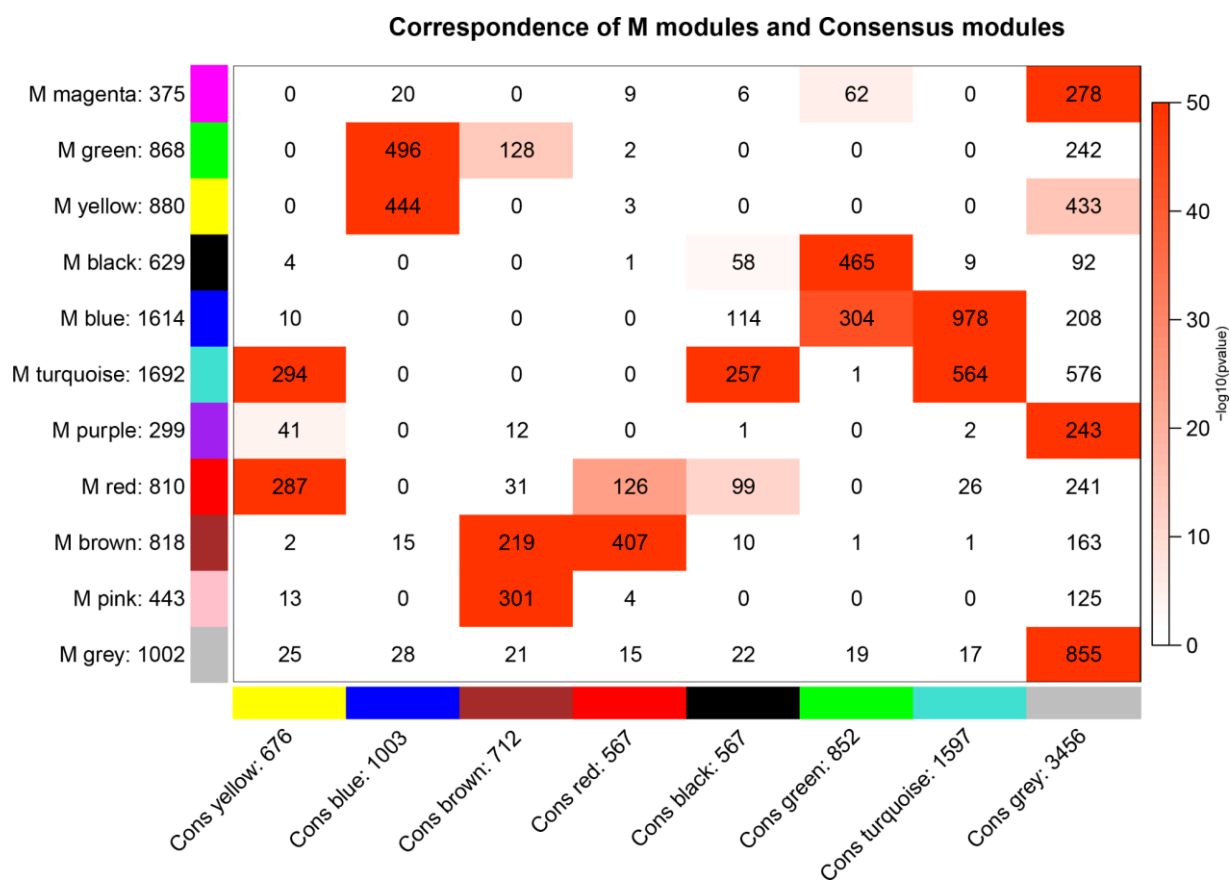
Destaque para os dois grupos amostrais selecionados: córtex de 12-13 pcw e córtex de 16-24 pcw

Apêndice 6. Análise de sobreposição entre módulos femininos e módulos *consensus* de amostras corticais de 12-13 semanas gestacionais



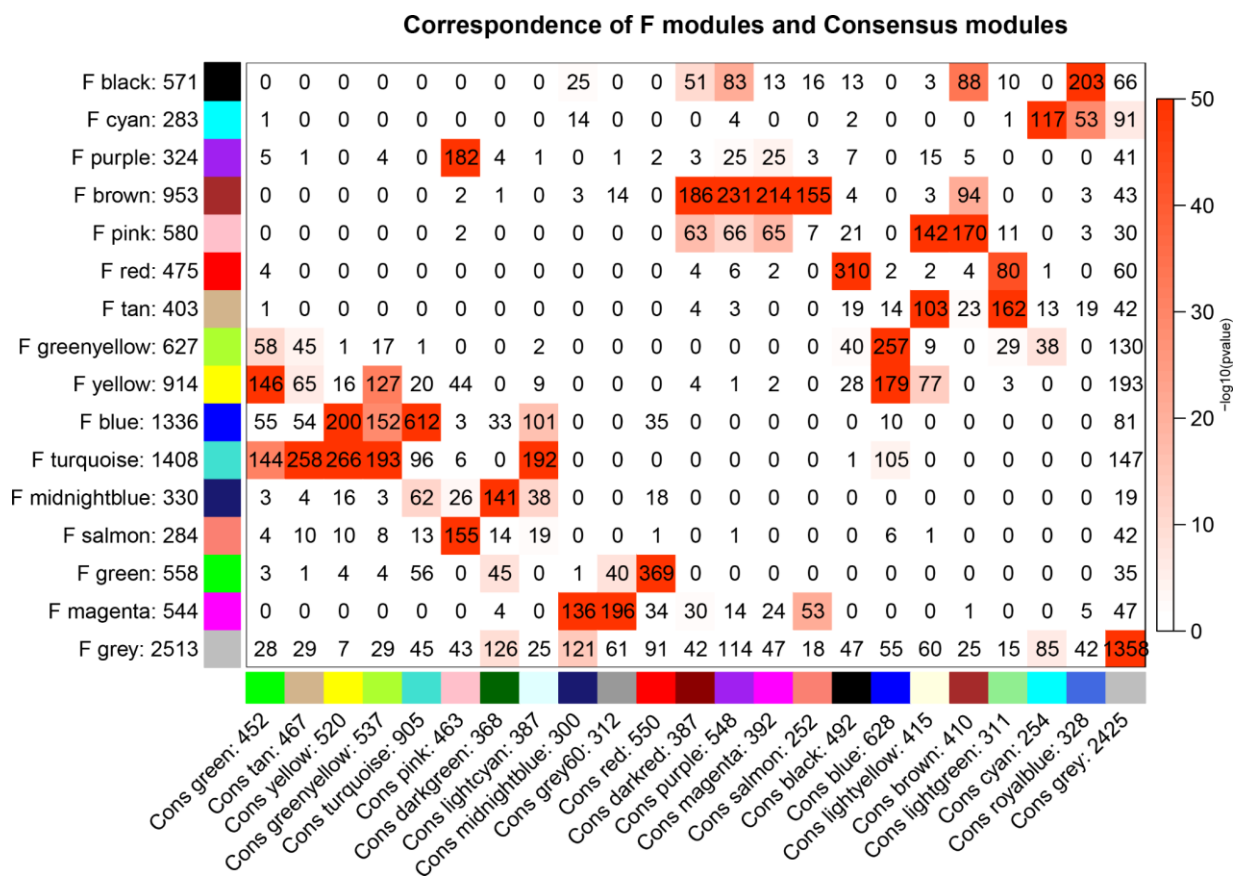
Resultados da análise de sobreposição no *software* R. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e módulos *consensus*. A escala de cores corresponde ao valor-p (em log₁₀) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Apêndice 7. Análise de sobreposição entre módulos masculinos e módulos *consensus* de amostras corticais de 12-13 semanas gestacionais



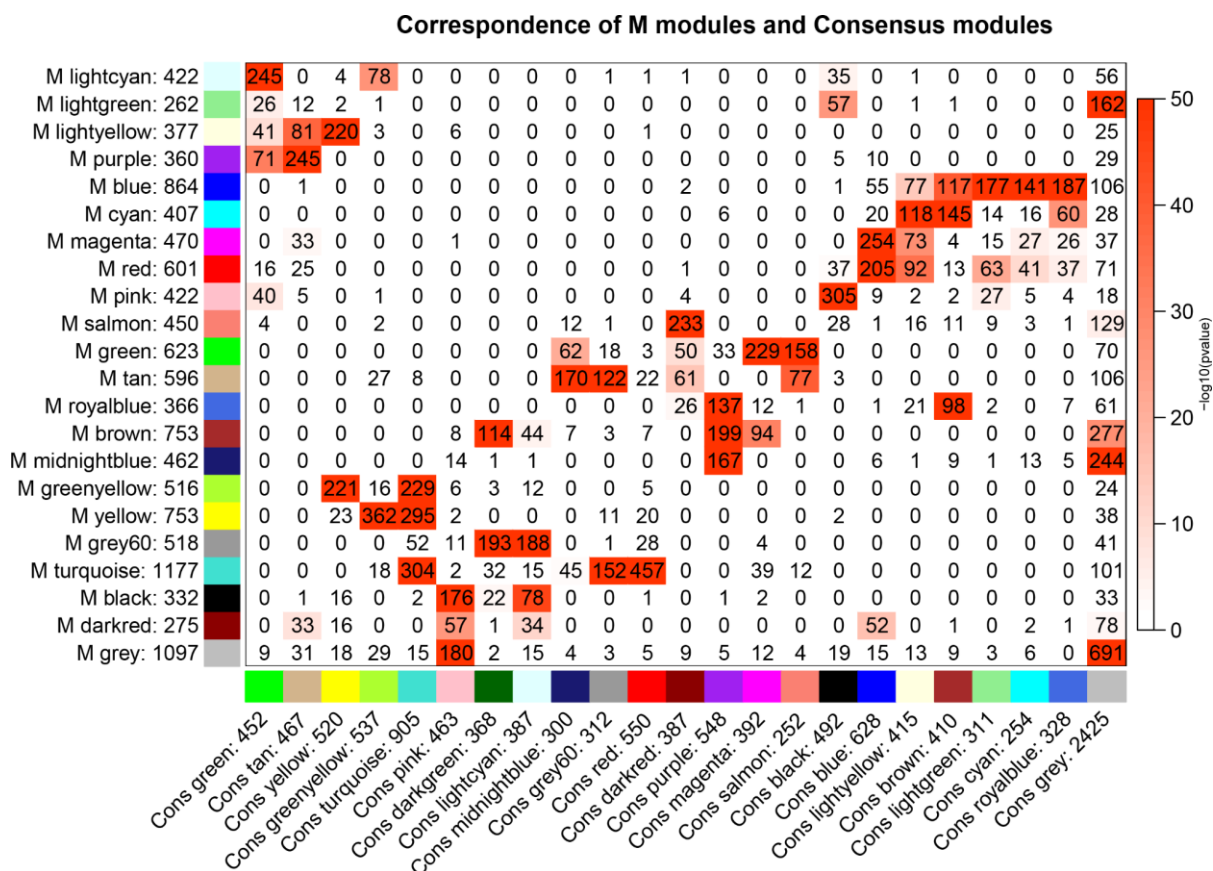
Resultados da análise de sobreposição no *software* R. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e módulos *consensus*. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Apêndice 8. Análise de sobreposição entre módulos femininos e módulos *consensus* de amostras corticais de 16-21 semanas gestacionais



Resultados da análise de sobreposição no *software* R. Matriz ilustrando o número de genes sobrepostos entre módulos femininos e módulos *consensus*. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.

Apêndice 9. Análise de sobreposição entre módulos masculinos e módulos *consensus* de amostras corticais de 16-21 semanas gestacionais



Resultados da análise de sobreposição no *software* R. Matriz ilustrando o número de genes sobrepostos entre módulos masculinos e módulos *consensus*. A escala de cores corresponde ao valor-p (em log10) correspondente ao teste de Fisher, considerando ainda o número de genes que compõem cada módulo.