

Universidade de São Paulo

**Departamento de Análises Clínicas e Toxicológicas
Faculdade de Ciências Farmacêuticas**

Mestrado em Bioinformática

FERNANDO MARCON PASSOS

Usando Biologia de Sistemas para entender a Imunossenescência

**SÃO PAULO
2018**

Usando Biologia de Sistemas para entender a Imunossenescência

Dissertação apresentada ao Programa de
Pós-Graduação Interunidades em
Bioinformática da Universidade de São
Paulo - USP para a obtenção do título de
Mestre em Ciências.

Área de concentração: Bioinformática.

Orientador: Prof. Dr. Helder Takashi Imoto
Nakaya

São Paulo

2018

Agradecimentos

Agradeço primeiramente à minha família, por seu amor incondicional, suporte e fé inabalável em minhas capacidades.

Agradeço ao professor Helder I. Nakaya pela orientação e por todas as oportunidades oferecidas.

Agradeço à Liverpool School of Tropical Medicine por todo apoio financeiro.

Por fim, agradeço principalmente aos colegas do Laboratório de Biologia de Sistemas Computacionais (CSBL) pelo fato de terem sido pessoas incríveis, no qual sem eles este trabalho não seria possível.

Abstract

The immune system remodelling occurring with ageing, known as immunosenescence, contributes to increased susceptibility of the elderly to infectious diseases, cancer, autoimmunity, and decreased responses to vaccines. This remodelling is a poorly understood complex process that involves multiple factors. For instance, we still don't know about the molecular mechanisms involved in immunosenescence and how they can be caused by primary problems, such as early perturbations in biological processes or secondary problems, such as those caused by a failed response to achieve homeostasis, disturbed by primary problems.

Here we applied a Systems Biology framework to study the remodelling of the immune system by analysis of blood transcriptional profiles. We aim to understand how changes in the expression of immune components during life are related to immunosenescence. To this end, we developed novel methods tailored to identify and describe pathways and genes serving as biomarker of ageing. We applied these methods to a massive dataset from 1807 blood samples from healthy subjects gathered from public gene expression repositories.

We found 56 transcripts correlated with age, which suggests that signal transduction pathways and cytokines in regulating T lymphocytes, especially regulatory, are important to cell proliferation and senescence. Coexpression modules related to innate and adaptive immune system signalling were identified. Changes in global expression patterns in these modules reveal imbalances in signal transduction mechanisms regulating the interaction between the innate and adaptive immune system, probably due to the deregulation of TLR activation pathways and the production of cytokines induced by NF- κ B. We identified changes in these expression profiles around the age of 30 years and around the age of 55-60 years. These results suggest that transcriptional changes, that are characteristic of immunosenescence, occur at a young age and intensify with ageing.

The methodologies developed in this study allowed us to find transcriptional disorders matching morphological changes described in the literature and enabled us to identify new transcripts and biological processes not yet associated to immunosenescence. Such methodologies can be optimized and adapted not only to higher-quality transcriptomics assessment techniques, such as RNA-Seq, but also to other levels of biological information, such as metabolomics, proteomics, etc. We expect that our findings may assist in the comprehension of mechanisms involved in immunosenescence, and to aid in the creation of interventions capable of beneficial regulation of remodelling of the immune system during life, allowing the extension of a life with quality.

Resumo

O remodelamento do sistema imune que acontece durante o envelhecimento, conhecido como imunossenescência, contribui para o aumento da suscetibilidade dos idosos a doenças infecciosas, câncer, autoimunidade e respostas diminuídas às vacinas. Esse remodelamento é um processo complexo e envolve múltiplos fatores. Ainda compreendemos pouco sobre os mecanismos moleculares envolvidos na imunossenescência e como eles podem ser causados por problemas primários, processos biológicos que sofrem perturbações precoces, e quais deles tem origem secundária, ou seja, são causados pela tentativa de retornar à homeostase perdida devido aos problemas primários.

Neste contexto, este trabalho propõe a aplicação de paradigmas da biologia de sistemas no estudo do remodelamento do sistema imunológico através de seu perfil transcricional no sangue, com o intuito de compreender como o comportamento de seus componentes se alteram ao longo da vida e como tal dinâmica pode estar relacionada à imunossenescência. Para isso, desenvolvemos metodologias inovadoras tanto para a identificação de marcadores e processos biológicos relacionados à idade quanto para o estudo de seus comportamentos. Ademais, utilizamos quantidades massivas de dados provenientes de repositórios públicos, compreendendo 1807 amostras do sangue de indivíduos saudáveis entre 15 e 96 anos.

Foram encontrados 56 transcritos altamente correlacionados com a idade, que apontam a potencial importância de perturbações nas vias de transdução de sinais e citocinas na regulação de linfócitos T, especialmente regulatórias, levando ao aumento da proliferação celular e senescência. Módulos de coexpressão relacionados à sinalização do sistema imune inato e adaptativo foram identificados. Mudanças em padrões globais de expressão nestes módulos ocorridas durante a vida indicam um desbalanço em mecanismos de transdução de sinais que regulam a interação entre o sistema imune inato e adaptativo, provavelmente devido à desregulação de vias de ativação de TLRs e de produção de citocinas induzidas por NF- κ B. Identificamos

mudanças nestes perfis de expressão por volta dos 30 anos e por volta dos 55-60 anos. Estes resultados sugerem que as alterações transcricionais características da imunossenescência ocorrem ainda na juventude e se intensificam no início da velhice.

As metodologias desenvolvidas neste estudo permitiram encontrar perturbações a nível transcricional que condizem com as características morfológicas já descritas literatura e identificar novos transcritos e processos biológicos nunca antes relacionados à imunossenescência. Estas metodologias podem ser aperfeiçoadas e aplicadas em dados de transcritômica obtidos por meio do uso de tecnologias mais sensíveis, como RNA-Seq, mas também para outras camadas fenotípicas, como a metabolômica, a proteômica, a epigenômica, etc. Esperamos que nossos resultados auxiliem na compreensão dos mecanismos envolvidos na imunossenescência e subjacentes ao envelhecimento e futuramente tornar possível a criação de intervenções capazes de regular de forma benéfica o remodelamento do sistema imune durante a vida, permitindo a extensão de uma vida com qualidade.

SUMÁRIO

1. INTRODUÇÃO	8
1.1. O Envelhecimento na Sociedade e no Indivíduo	8
1.2. O Envelhecimento e seus Mecanismos Moleculares	9
1.3. O Remodelamento do Sistema Imunológico: Imunossenescência	14
1.4. Alterações no Sistema Imunológico Relacionadas à Imunossenescência	15
1.6. Biologia de sistemas	23
2. OBJETIVOS	25
2.1. Objetivo principal	25
2.2. Objetivos secundários	25
3. METODOLOGIA	27
3.1 Levantamento dos Estudos	28
3.2. Reanotação das Sondas nos Microarrays	28
3.3. Aquisição de dados e pré-processamento	29
3.3.1. Normalização e Controle de Qualidade	30
3.4. Criação de amostras representativas das idades: <i>AgeCollapsed</i>	31
3.5. Detecção de Sondas relacionadas à Idade: <i>AgingGenes</i>	32
3.4.2. Remoção de Sondas não Expressas e/ou Constantes	32
3.4.3. Seleção de Sondas Correlacionadas com a Idade	32
3.4.5. Análise de enriquecimento dos genes relacionados à idade	33
3.5. Análise de Transcritos Co-expressos ao longo da Vida: <i>AgingNet</i>	34
3.5.1. Similaridade entre Genes.	34
3.5.2. Construção da Rede de Co-Expressão e Detecção de Módulos	37
3.5.3. Análise de Enriquecimento de Vias dos Módulos da <i>AgingNet</i>	37
3.6. Detecção de Pontos de Mudanças em Módulos Relacionados à Idade	37
4. RESULTADOS	39
4.1. Levantamento de estudos, aquisição e pré-processamento	39
4.2. Reanotação das Plataformas	40
4.3. Detecção de Transcritos Relacionados à Idade	44
4.3. Rede de Co-Expressão do Envelhecimento: <i>AgingNet</i>	52
5. DISCUSSÃO	61
6. CONCLUSÃO	71
7. REFERÊNCIAS	72
8. ANEXOS	78
8.1 Tabelas Suplementares	78
8.2 Artigos publicados	79

1. INTRODUÇÃO

Desde os primórdios da humanidade, doenças infecciosas são um enorme desafio para a sobrevivência da espécie. Somente em meados do século 20 começamos a ser capazes de controlá-las em adultos e crianças, porém ainda de forma insatisfatória em idosos. Por volta dos 60 anos, as pessoas passam a ter uma maior susceptibilidade a infecções, particularmente às emergentes. Além disso, quando acometidas por alguma infecção, sofrem desfechos mais graves do que pessoas mais jovens. Um grande exemplo é a mortalidade por influenza nos Estados Unidos, onde cerca de 30.000 pessoas morrem a cada ano, sendo mais de 95% idosos (REID; TAUBENBERGER, 2003).

Desde a invenção da vacina, testemunhamos uma mudança radical na qualidade e potencial de vida dos seres humanos. Muitos fatores permitiram esse progresso além da vacinação, como a diminuição da mortalidade infantil, terapia por antibióticos, prevenção de doenças cardíacas e metabólicas, melhoramento das condições nutritivas e de higiene. A maior evidência disso é o aumento constante da expectativa média de vida no mundo. Estudos recentes indicam que tal crescimento permitirá que muitas crianças nascidas nos anos 2000 vivam até os 100 anos em países desenvolvidos (MANTON; VAUPEL, 1995).

Este aumento na expectativa de vida se reflete claramente na população idosa, que é a com maior crescimento atualmente. De acordo com as Nações Unidas, em 2015 havia em torno de 901 milhões de idosos, sendo que cerca de 125 milhões possuíam mais de 80 anos (UN, 2016). Além disso, essa população constituirá 25% da população mundial em 2030 e dobrará em 2050, passando para 2.1 bilhões, dos quais 434 milhões terão mais de 80 anos (UN, 2016).

1.1. O Envelhecimento na Sociedade e no Indivíduo

A diminuição da força de trabalho que ocorrerá em decorrência do envelhecimento global pode trazer consequências econômicas, políticas e sociais

graves. Com o aumento da proporção de idosos na população, aumenta também a razão entre o número de pessoas em idade produtiva (entre 15 e 64 anos) e o número de pessoas dependentes (idade menor que 15 e maior que 65 anos), um índice denominado razão de dependência. Por exemplo, em 2020, cada grupo de 100 pessoas em idade produtiva sustentará em torno de 43 indivíduos considerados dependentes. Em 2060, a projeção de dependentes sustentados passará de 43 para 66 indivíduos (“IBGE :: Instituto Brasileiro de Geografia e Estatística,” [s.d.]).

Os idosos vivem uma realidade debilitante em que diversos fatores fisiológicos, psicológicos e sociais contribuem para um cenário de exaustão de recursos, onde a capacidade reduzida de adaptação a novos desafios e o constante acúmulo de danos a nível sistêmico estabelecem um estado de fragilidade e, conseqüentemente, de vulnerabilidade. Essa situação é um processo dinâmico, oscilando entre envelhecimento saudável e patológico. Quando um indivíduo começa a tender para um cenário patológico observa-se uma grande incidência de doenças crônicas e neoplasias, tais como doença de Alzheimer, Parkinson, diabetes tipo 2, câncer, doenças cardíacas e artrite. Tais doenças, além de serem extremamente debilitantes para o indivíduo, estão entre as doenças mais dispendiosas para os sistemas de saúde (“Health and Economic Costs of Chronic Disease | About Chronic Disease | Chronic Disease Prevention and Health Promotion | CDC,” [s.d.]).

1.2. O Envelhecimento e seus Mecanismos Moleculares

O envelhecimento é um dos fenômenos biológicos mais complexos encontrados na natureza e se refere ao processo intrínseco, inevitável e irreversível de perda de viabilidade e aumento da vulnerabilidade relacionado à idade (Comfort, 1964), levando a um aumento na taxa de mortalidade logo após a maturidade (FLATT, 2012), conforme ilustrado na Figura 1.1.

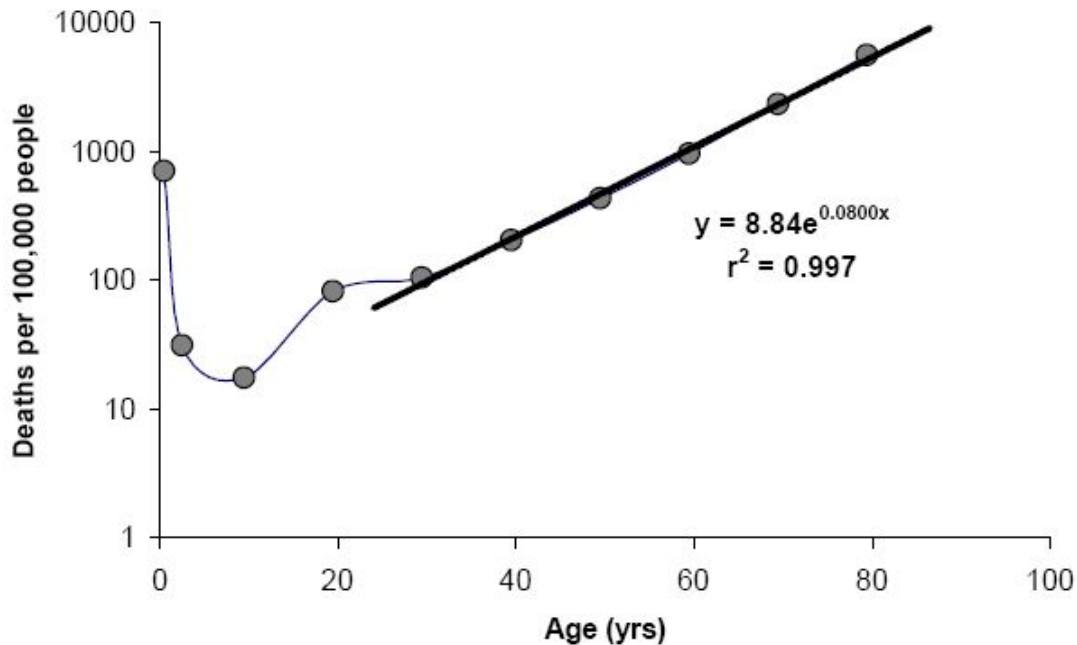


Figura 1.1. Taxa de Mortalidade (lei de Gompertz–Makeham). Taxas de mortalidade, expressas em número de mortes a cada 100.000 pessoas, em função da idade da população norte-americana de 2002. A linha preta representa a função de Gompertz extrapolada das taxas de mortalidade após a maturidade. Retirado de CDC/NCHS, *National Vital Statistics System, Mortality Data*.

Quando comparamos o aumento da expectativa máxima de vida com a expectativa média desde o início do século passado (Figura 1.2), percebemos que o aumento foi mínimo. Isso pode ser um sinal de que as intervenções desenvolvidas até então não atacaram diretamente a biologia do envelhecimento. No entanto, isso não implica necessariamente na inexistência de mecanismos biológicos do envelhecimento que sejam manipuláveis. Pelo contrário, existe um crescente volume de evidências experimentais, obtidas de uma grande variedade de organismos, que sugerem a existência de ao menos nove marcadores moleculares que caracterizam o envelhecimento e são conservados evolutivamente, conhecidos como hallmarks of aging (LÓPEZ-OTÍN et al., 2013), demonstrados na Figura 1.3.

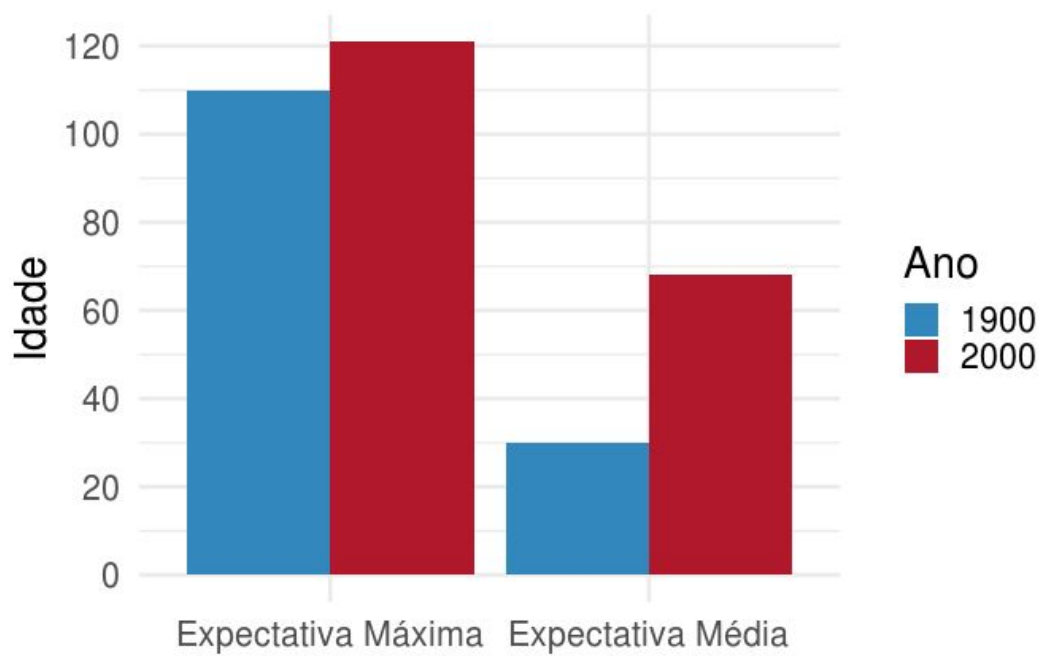


Figura 1.2. Aumento da Expectativa de Vida no Último Século. Gráfico de barras representando a expectativa de vida máxima e média nos anos 1900 e 2000 (vermelho).



Figura 1.3. Os Marcadores do Envelhecimento. A figura ilustra os nove marcadores do envelhecimento: instabilidade genômica, desgaste telomérico, alterações epigenéticas, perda de proteostase, desregulação da detecção de nutrientes, disfunção mitocondrial, senescência celular, exaustão de célula tronco, e alteração da comunicação intracelular. Retirada de (LÓPEZ-OTÍN et al., 2013).

Os marcadores do envelhecimento são fenômenos que se manifestam naturalmente durante o envelhecimento e, quando agravados experimentalmente,

aceleram ou retardam o processo de envelhecimento (LÓPEZ-OTÍN et al., 2013). Tais marcadores podem ser divididos em três categorias: primários, antagonistas e integrativos.

As principais características dos marcadores primários são o fato de sempre possuírem efeitos negativos, como, por exemplo, danos ao DNA, aneuploidia cromossomal, mutações no DNA mitocondrial, encurtamento dos telômeros, deriva epigenética e desequilíbrios da homeostase proteica (proteostase). Os marcadores antagonistas possuem efeitos opostos dependendo de sua intensidade e podem ser vistos como mecanismos desenvolvidos para proteger o organismo de danos ou escassez de nutrientes. Quando em baixo nível, podem mediar efeitos benéficos, mas quando suas atividades são exacerbadas ou crônicas, podem subverter seu propósito e gerar ainda mais danos. Um exemplo disso seria a senescência, a qual protege o organismo do desenvolvimento de câncer, mas quando em excesso, pode desencadear o envelhecimento (GORGOU LIS; HALAZONETIS, 2010). Um efeito similar também ocorre com espécies reativas de oxigênio (ROS, do inglês '*Reactive Oxygen Species*'), que participam do controle da sobrevivência e sinalização celular, mas podem produzir danos celulares quando em níveis exacerbados (HARMAN, 1965). A terceira categoria constitui os marcadores integrativos. Estes afetam diretamente a homeostase e a função dos tecidos e são representados pela exaustão de células-tronco e alterações na comunicação intracelular.

Foi proposta uma hierarquia entre os tipos de marcadores (Figura 1.4): os marcadores primários, cujas consequências prejudiciais se acumulam progressivamente com o tempo, podem ser considerados como os primeiros desencadeadores do processo de envelhecimento. Os marcadores antagônicos, que a princípio são benéficos e protetivos, tornam-se progressivamente deletérios em um processo que é parcialmente acelerado pelos marcadores primários. Finalmente, os marcadores integrativos surgem quando os danos acumulados pelos outros dois

marcadores não podem mais ser compensados pelos mecanismos de regulação homeostática dos tecidos ((LÓPEZ-OTÍN et al., 2013).

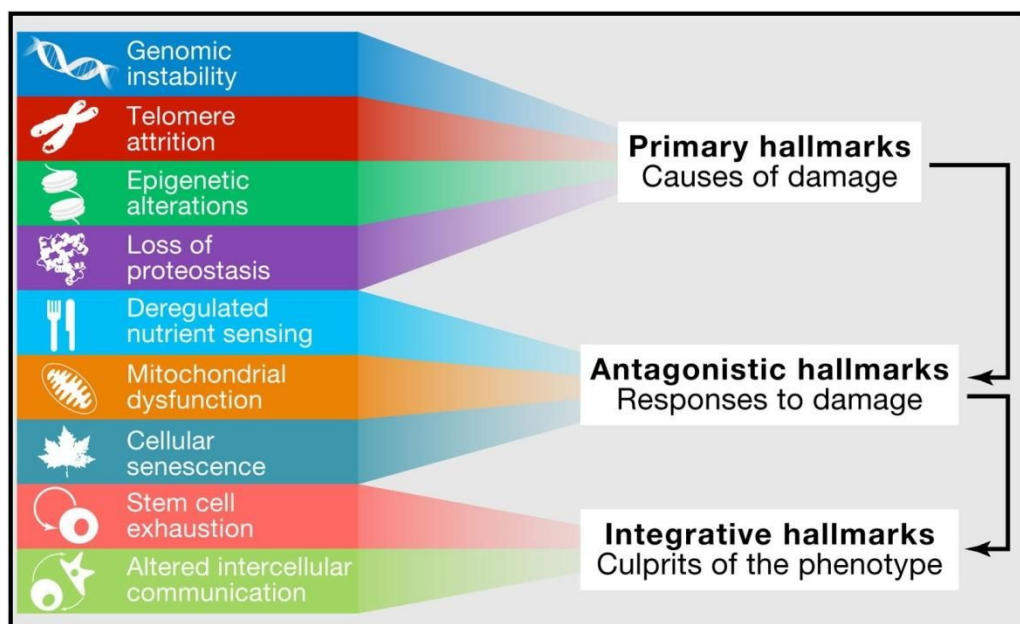


Figura 1.4. Interconexões funcionais entre os marcadores-chave do envelhecimento. Os nove marcadores moleculares do envelhecimento estão agrupados em 3 categorias: causas primárias de dano celular, respostas compensatórias e antagônicas ao dano, e aqueles que são os resultados dos dois outros grupos e responsáveis pelo declínio funcional associado ao envelhecimento. Retirada de (LÓPEZ-OTÍN et al., 2013).

1.3. O Remodelamento do Sistema Imunológico: Imunossenescência

O sistema imunológico mantém o estado saudável do organismo através da prevenção de infecções e transformações malignas, além de participar da homeostase tecidual. Sempre que o equilíbrio é perdido, desencadeia-se um processo inflamatório caracterizado pela secreção de uma miríade de mediadores inflamatórios (NIKOLICH-ŽUGICH, 2018). Durante o envelhecimento, diversas alterações nas células do sistema imunológico, no microambiente dos órgãos linfóides e não-linfóides, onde as células do sistema imune residem, e nos fatores circulantes que interagem tanto com células do sistema imune quanto com seu microambiente para assegurar respostas imunes apropriadas, assim como sua homeostase (NIKOLICH-ŽUGICH, 2018). Esses processos são altamente interconectados e acontecem simultaneamente

em um organismo, levando ao complexo remodelamento conhecido como imunossenescência (KUILMAN et al., 2010); (MUSZKAT et al., 2003).

Além disso, as doenças relacionadas ao envelhecimento compartilham algumas similaridades subjacentes, desencadeadas pelo desequilíbrio homeostático devido ao acúmulo de estresse e perturbações no sistema, tais como processos inflamatórios, problemas nos mecanismos de reparo e recuperação ao estado de homeostasia. Esse fenômeno é considerado um dos fatores que mais contribuem para o aumento da frequência de doenças malignas, morbidade e mortalidade entre a população de idosos (GINALDI et al., 2001).

1.4. Alterações no Sistema Imunológico Relacionadas à Imunossenescência

As células do sistema imune são constantemente renovadas a partir de progenitoras hematopoiéticas pluripotentes (HSC) que residem na medula óssea. Essas células se diferenciam nos órgãos linfáticos e são capazes de se mover livremente. Na presença de sinalizadores inflamatórios, elas interagem com órgãos para capturarem resíduos celulares ou elementos invasores.

Devido a diversos fatores (e.g., encurtamento dos telômeros) decorrentes do envelhecimento, há uma diminuição na capacidade de renovação de células imunes levando à uma diminuição da quantidade total do tecido hematopoiético e de seus progenitores mielóides. O timo passa por um processo de involução fisiológica, com uma redução de volume e substituição por tecido adiposo nas partes funcionais, como o córtex e a medula. Esse processo começa cedo na vida e se completa por volta dos 40-50 anos.

1.4.1 Alterações no Sistema Imune Adaptativo

As células T são maturadas no timo, onde se diferenciam em subtipos celulares de acordo com seus receptores, e então migram para o sistema periférico. Os principais subtipos são denominados CD4+ e CD8+, e demonstram algumas mudanças durante o envelhecimento relativo às suas proporções, no qual o número de células

CD8+ tende a aumentar. A inserção de células T naíve diminui gradualmente, assim como suas divisões subsequentes, representadas pelo decaimento de TRECs (círculos de excisão do TCR) nos linfócitos T periféricos. Essa redução pode ser consequência tanto da involução tímica quanto da estimulação crônica de antígenos (HAYNES et al., 2003), o que poderia ajudar explicar a habilidade reduzida dos idosos de resistirem a novas infecções (VALLEJO, 2005). Além disso, as células T naíve em idosos mostram múltiplas alterações, como o encurtamento dos telômeros, produção reduzida de IL-2 e habilidade diminuída de diferenciarem-se em células efetoras. A perda no número e função de células T naíve pode ser compensada através da expansão de T CD8+, CD45RO+ e CD25+, que são capazes de produzir IL-2, e com uma capacidade protetiva humoral para vacinação com a expansão de células de memória efetoras (SCHWAIGER et al., 2003).

Células T regulatórias (Tregs) são um subconjunto caracterizado por uma alta expressão de CD25 e FOXP3, um fator de transcrição para a função e diferenciação de células Treg. Em idades mais avançadas o número de linfócitos CD4+ FOXP3+ aumentam, sendo que o acúmulo dessas células em idosos possui um papel importante na reativação de infecções crônicas e a mudança na proporção de Th17/Treg pode causar alterações na resposta imune, podendo levar ao desenvolvimento de doenças autoimunes e inflamatórias (LAGES et al., 2008).

O reservatório de células B também é influenciado pelo envelhecimento, levando a alterações tanto quantitativas quanto qualitativas na imunidade humoral (COLONNA-ROMANO et al., 2008). Células B são importantes apresentadoras de antígenos e que podem ser reguladores-chave no desenvolvimento de células T (COLONNA-ROMANO et al., 2008).

Em idosos, há um nível reduzido de IgM e IgD, certamente conectadas à transformação das células B naíve para células de memória (WEKSLER; SZABO, 2000). Durante a imunossenescência ocorre um aumento do nível de IgA e IgG. Esse desbalanço pode ser responsável pela redução de placas de Peyer ao nível da mucosa

gastrointestinal com respeito a IgA2, enquanto o aumento IgA1 pode ser secundário à deficiência da atividade do subconjunto de células T supressoras e consequente hiperfunção dos linfócitos B (VENTURA, 1991) .

1.4.2 Alterações no Sistema Imune Inato

A imunidade inata possui um papel crucial na defesa do organismo e sofre alterações importantes durante o envelhecimento, começando com a redução das barreiras epiteliais, gastrointestinais e da mucosa respiratória (NOMELLINI; GOMEZ; KOVACS, 2008). A alta incidência de eventos infecciosos em idosos pode ser secundário a alterações no sistema fagocitário (ANTONACI et al., 1984). Os déficits em idosos são causados principalmente por eventos infecciosos, particularmente nos sistemas gastrointestinais e respiratórios. O número reduzido de células plasmáticas na medula óssea de idosos (PRITZ et al., 2015) pode causar uma falta de produção de anticorpos, uma habilidade reduzida para responder a vírus e bactérias (BUFFA et al., 2013) e uma resposta alterada a vacinas, como demonstrado previamente em pacientes idosos imunizados com a vacina contra hepatite B (ROSENBERG et al., 2013).

As células dendríticas, uma das principais responsáveis pelo primeiro reconhecimento de patógenos na pele, mostra claras disfunções mitocondriais que interferem com seu papel protetivo (SIMON; HOLLANDER; MCMICHAEL, 2015). Em particular, existe uma perda na absorção de antígenos e da função apoptótica. Quando se compara a capacidade de absorção de antígenos das células dendríticas plasmocitóides (PCD) de idosos em relação aos jovens, é possível observar uma redução na habilidade de induzir a proliferação e estimular a secreção de IFN-gama em células CD4+ e CD8+ [(PRAKASH et al., 2013).

Os macrófagos são células com a função de processar e apresentar antígenos às células T, e são capazes de produzir citocinas pró-inflamatórias como TNF-alfa, IL-1, IL-6 e IL-9. Durante a imunossenescência há uma diminuição dos precursores de macrófagos, mesmo que o número de monócitos permaneça inalterado ([Della Bella et](#)

[al. 2007](#)). O encurtamento dos telômeros resulta em uma redução da produção de GS-CSF, e também de citocinas como TNF-alfa e IL-6 ([Davalos et al. 2010](#)). A função fagocítica parece ser reduzida, enquanto que a quimiotaxia parece permanecer conservada, especialmente na presença de certos fatores estimulantes do complemento C5a. A presença desses fatores reduz a produção de linfócitos derivados de quimiotáticos (LDCF) e quimiotoxinas. Nesse caso, os mecanismos inibitórios parecem estar relacionados a prostaglandinas que são produzidas em altas quantidade durante a imunossenescência, e que exerce uma ação inibitória ([Ventura et al. 1994](#)). A produção reduzida de LDCF pode estar relacionada à pouca porcentagem de linfócitos envolvidos na síntese de citocinas.

O número de neutrófilos parece ser preservado nos idosos, no entanto a expressão do receptor CD16 Fc gamma é reduzida, fazendo com que tanto a geração de superóxidos mediados pelo receptor Fc e a fagocitose sejam implicadas. Isso sugere que o declínio da resposta efetora de receptores Fc seja particularmente importante para a disfunção de neutrófilos dos idosos ([Butcher et al. 2001](#)). Em idosos, a resposta reduzida dessas células à *Streptococcus Aureus* é de fundamental importância clínica, pois esse evento aumenta a susceptibilidade à infecções pulmonares. O mesmo foi observado em ratos idosos, onde a migração de neutrófilos para os pulmões é reduzida, aumentando o risco de infecções pulmonares e recorrências ([Chen et al. 2014](#)). Além disso, recentemente foi descrito uma alteração na destruição mediada por patógenos de armadilhas extracelulares de neutrófilos (NETs), o que também implica no aumento de infecções aos idosos ([Brinkmann and Zychlinsky 2007](#)).

A alta incidência de doenças imunoproliferativas em idosos sugere a possibilidade de ocorrer uma deficiência nos mecanismos importantes para a supervisão imune, tal como a atividade de células NK. Essas células intervêm na eliminação de tumores, nas células infectadas por vírus e também na regulação imunológica inata e adaptativa, através da produção de citocinas e quimiocinas ([Brinkmann and Zychlinsky 2007](#)). Estudos recentes apontam que uma alta toxicidade

de NK está associada com uma longevidade, enquanto que baixa função NK está associada a um aumento de infecções, arteriosclerose e doenças neurodegenerativas. As células NK, por produzirem lise celular, poderiam causar a liberação de perforinas e granzimas, que, por sua vez, ativam caspases e provocam apoptose de células alvo. Durante a imunossenescência ocorre uma redução de linfocinas importantes para o processo de ativação de linfócitos, como a IL-2, e também para a morte de linhagens celulares resistentes à NK em resposta à IL-2. Isso contribui para o déficit de sua função, mesmo que na presença de um número normal de células NK ([Borrego et al. 1999](#)). Além disso, existe uma redistribuição de células NK com diminuição de células CD56, caracterizadas por uma alta densidade de antígenos de superfícies CD56. Em contraste, existe um aumento em células NK CD56-CD16 [64]. Isso resulta em uma redução na secreção de IFN para os idosos, em relação aos jovens ([Krishnaraj 1997](#)).

1.5 Inflamação Crônica durante o Envelhecimento: Inflammaging

A senescência é o fenômeno em que células entram em um estado permanente de não-proliferação como resposta a um estresse (e.g., encurtamento dos telômeros), desenvolvendo algumas alterações fenotípicas características. Essas características compreendem mudanças na organização da cromatina, alteração da expressão gênica e secreção de várias citocinas pró-inflamatórias, quimiocinas, fatores de crescimento e proteases, dependentes de fatores de estresse (CAMPISI; D'ADDA DI FAGAGNA, 2007).

Presume-se que a senescência celular contribui para o envelhecimento, pois sua quantidade aumenta com o envelhecimento (BAKER et al., 2016; VAN DEURSEN, 2014). Tal acúmulo pode refletir tanto um aumento na taxa de geração de células senescentes quanto uma diminuição da sua taxa de remoção, como consequência de uma resposta imune atenuada. Esse balanço requer um sistema eficiente de substituição celular que envolve, simultaneamente, a remoção de células senescentes e a mobilização de progenitores para o restabelecimento do número celular. Em

organismos idosos, esse sistema pode se tornar ineficiente uma vez que a capacidade regenerativa das células progenitoras pode se exaurir e, eventualmente, resultar no acúmulo de células senescentes (STAHL; BROWN, 2015).

Células senescentes manifestam alterações em seu secretoma, denominado senescence-associated secretory phenotype (SASP), no qual são particularmente enriquecidas em citocinas pró-inflamatórias e metaloproteinases de matriz (MMPs) ((KUILMAN et al., 2010), (RODIER; CAMPISI, 2011)). O SASP possui atividades parácrinas envolvidas no recrutamento de leucócitos e na migração de células imunes inatas para a vizinhança de regiões tumorais (MANTOVANI; ALLAVENA; SICA, 2004), (XUE et al., 2007)). Além disso, a produção de fatores de crescimento promove a proliferação celular através das proteínas GROs (growth-regulated oncogenes) e AREG (amphiregulin), assim como a formação de novos vasos sanguíneos através de proteínas como VEGF (vascular endothelial growth factor) (COPPÉ et al., 2008). Isso sugere que senescência celular e o SASP podem contribuir não só para - suprimir o desenvolvimento de câncer, mas também para promover o reparo de tecidos danificados.

Caso a secreção de SASP se torne crônica, fenótipos nocivos característicos do envelhecimento podem se desenvolver. Um exemplo disso é o caso do modulador SFRP1 (secreted frizzled related protein 1) bifásico, que pode estimular ou inibir a sinalização de Wnt dependendo do contexto fisiológico (ELZI et al., 2012). A sinalização crônica de Wnt pode induzir tanto células-tronco quando células diferenciadas à senescência (LIU et al, 2007).

Os componentes do SASP também regulam múltiplos aspectos da inflamação. As citocinas IL-6 e IL-8 ((COPPÉ et al., 2008); (ACOSTA et al., 2008); (KUILMAN et al., 2010)), algumas variedades de MCPs (monocyte chemoattractant proteins), MIPs (macrophage inflammatory proteins), e GM-CSF (granulocyte/macrophage colony-stimulating factor), podem promover a inflamação crônica (ADAMS et al, 2009; [\(COPPÉ et al., 2010\)](#); [\(FREUND et al., 2010\)](#); [\(DAVALOS et al., 2010\)](#); TCHKINIA et al,

2010). A inflamação crônica é a causa, ou ao menos um importante contribuinte em muitas das doenças relacionadas ao envelhecimento, tanto degenerativas quanto hiperplásicas, caracterizando o fenômeno denominado Inflammaging (FULOP et al., 2017).

Deve-se ressaltar que a senescência não é uma propriedade generalizada de todos os tecidos de organismos idosos (WANG et al., 2009), além disso, trabalhos recentes indicam que imunossenescência não é acompanhada por uma deterioração inevitável e progressiva da função imune, mas sim um processo complexo envolvendo uma série de alterações de reorganização e desenvolvimento, em que algumas características da função imune são reduzidas, outras aumentadas, e outras até mesmo intocadas ([COPPÉ et al., 2010](#)).

1.6 O Fenótipo Imunológico de Risco (IRP)

Todas alterações que acontecem na imunossenescência mencionadas acima levam ao desenvolvimento do fenótipo imunológico de risco (IRP, do inglês *Immune Risk Phenotype*), caracterizado pela inversão da proporção de células CD4+/CD8+ (aumento de CD8+ e diminuição de CD4+), redução de linfócitos B, soropositividade para o Citomegalovírus (CMV), aumento da produção de citocinas pró-inflamatórias, diminuição da responsividade à IL-2 e diferenciação tardia de células CD8+CD28-CD27- (WIKBY et al., 2005). Os marcadores do IRP foram delineados em dois estudos longitudinais denominados OCTO e NONA (WIKBY et al., 2005), onde foi avaliado seu poder preditivo da sobrevivência de idosos sobre os próximos 4 anos. Combinado com métricas de comprometimento cognitivo, o IRP foi capaz de prever 58% das mortes observadas (WIKBY et al., 2005).

Esse fenótipo possui um aumento característico na produção de IL-1, IL-4, IL-6 e IFN-gama, citocinas que regulam a diferenciação de células B. O aumento da secreção de citocinas, correlacionada ao impacto da infecção por CMV, é um dos principais

responsáveis pelo envelhecimento mal-sucedido. Além disso, observa-se uma capacidade proliferativa reduzida em células T específicas à infecção por CMV associadas ao encurtamento dos telômeros, o que leva à instabilidade genômica, ativação de processos de reparo de danos ao DNA e desencadeia a parada do ciclo celular e/ou apoptose. Como consequência, nem todas as células T de memória se diferenciam da mesma maneira, podendo ocorrer uma expansão de um subgrupo celular específico, o que leva ao agravamento no cenário clínico com aumento de doenças infecciosas e neoplasias (WEISKOPF; WEINBERGER; GRUBECK-LOEBENSTEIN, 2009).

De forma geral, as alterações observadas no sistema imunológico podem provocar deficiências na imunidade inata e adaptativa, levando à uma maior susceptibilidade dos idosos à autoimunidade, câncer e reduzindo sua capacidade defensiva contra microorganismos invasores e agentes patogênicos. É evidente o papel fundamental que imunossenescência exerce no processo de envelhecimento, fazendo com que seja imprescindível o estudo do comportamento do sistema imunológico ao longo da vida para identificar mecanismos subjacentes ao seu remodelamento no qual podem influenciar o passo e a qualidade do processo de envelhecimento. Além disso, também é importante distinguir relações de causa-e-efeito entre os mecanismos do sistema imune que se alteram na imunossenescência, identificando-se quais refletem problemas primários, ou seja, mecanismos que alteram a homeostase devido à sua resposta à algum estresse, e quais são secundários: mecanismos que se alteram devido à uma tentativa de se retornar à homeostase, que foi perdida devido aos problemas primários.

Através da compreensão dos mecanismos envolvidos na imunossenescência e da dinâmica entre processos que são perturbados durante a vida, pode-se identificar alvos biomoleculares passíveis de serem manipulados através de intervenções farmacológicas, que poderiam amenizar e/ou regular as características nocivas de um sistema imunológico danificado. Melhoras significativas no sistema imunológico poderiam até mesmo influenciar de forma positiva o próprio processo de

envelhecimento, aumentando as chances de idosos continuarem autônomos pelo maior período de vida possível, com qualidade de vida e livre de doenças.

1.6. Biologia de sistemas

O sistema imune é constituído por diversos componentes que interagem entre si através de sinalizadores, e a eficiência de sua resposta frente à um desafio depende diretamente da integridade de seus componentes e da qualidade da comunicação entre eles. Conforme discutido, a imunossenescência representa o complexo e contínuo remodelamento desse sistema ao longo da vida, podendo ocorrer alterações concomitantes em diversos níveis, tanto em seus componentes quanto nas suas interações. Tal complexidade faz com que métodos reducionistas tradicionais, caracterizados pelo estudo de sistemas através de seus componentes isolados, não sejam a melhor metodologia para lidar com o problema em questão. A perspectiva holística da biologia de sistemas, por sua vez, torna-se mais adequada neste cenário pois é capaz de lidar com informações provenientes de todos seus componentes simultaneamente, possibilitando uma melhor compreensão do comportamento e das características globais desse fenômeno (KITANO, 2002).

Os métodos da biologia de sistemas envolvem análise computacionais que integram dados gerados por vários tipos de tecnologias e diferentes níveis de complexidade biológica. Com o advento das tecnologias *high throughput*, tornou-se possível avaliar perturbações em praticamente todos os níveis moleculares de um sistema biológico, como o genoma, epigenoma, transcritoma, proteoma e metaboloma.

Em estudos de transcriptômica, análises realizadas para identificar grupos de genes com comportamento similares entre si têm o potencial de desvendar diversos mecanismos moleculares envolvidos na regulação de doenças, respostas à vacinas e câncer. Essas análises baseiam-se na premissa de que genes que sofrem perturbações em conjunto ou com perfis de expressão similares entre si tendem a fazer parte de um mesmo programa transcricional, ou seja, estão relacionados à uma mesma função molecular. Tradicionalmente, tais processos biológico são identificados através

de análises de diferença de expressão, onde se identifica grupos de genes que sofreram perturbações no seu nível de expressão entre classes fenotípicas diferentes, ou através de análises de co-expressão, no qual são identificados grupos de genes com perfil de expressão similares, através da clusterização de genes altamente correlacionados entre si e então enriquecidos para vias metabólicas descritas na literatura (ZHANG; HORVATH, 2005).

Muitos estudos fizeram uso da transcriptômica para identificar genes e processos biológicos envolvidos com o envelhecimento, como pode ser evidenciado pelo *GenAge* (DE MAGALHÃES; TOUSSAINT, 2004), um banco de dados que fornece uma lista de genes no qual foram identificados como relacionados aos processos envolvidos no envelhecimento tanto em humanos quanto em animais modelos. Contudo, mesmo com a quantidade gigantesca de dados gerados, estes estudos utilizaram coortes com poucas amostras ou com comparações conduzidas em fatias arbitrárias de faixas etárias. Tais desenhos experimentais podem levar a conclusões enviesadas ou limitadas às comparações conduzidas em cada artigo. Ademais, a discretização temporal pode mascarar mudanças concomitantes e progressivas nos processos biológicos causadores da imunossenescência. A única exceção encontrada é o estudo realizado por (IRIZAR et al., 2015), pois estudam aspectos do sistema imune de uma forma contínua. No entanto, se restringem ao estudos de transcritos não-codificadores de proteínas em células dendríticas.

Assim sendo, este trabalho propôs-se utilizar paradigmas da biologia de sistemas para estudar o remodelamento do sistema imunológico através de seu perfil transcricional no sangue, com o intuito de compreender como o comportamento de seus componentes se alteram ao longo da vida e como tal dinâmica pode estar relacionada à imunossenescência. Para tal, será desenvolvido metodologias para identificar possíveis marcadores e processos biológicos relacionados à idade em um contexto de séries temporais, com o intuito de identificar e analisar seus comportamentos e como eles podem estar relacionados à imunossenescência.

2. OBJETIVOS

2.1. Objetivo principal

Encontrar programas transcricionais relacionados à imunossenescência.

2.2. Objetivos secundários

2.2.1. Encontrar transcritos com perfil de expressão altamente relacionados à idade, denominados de *AgingGenes*.

2.2.2. Identificar rede, seus módulos, de transcritos co-expressos ao longo da vida, denominada *AgingNet*.

2.2.3. Avaliar os processos biológicos relacionados aos *AgingGenes* e aos módulos da *AgingNet* e como poderiam estar associados ao fenótipo da imunossenescência.

2.2.4. Identificar momentos da vida no qual os *AgingGenes* e *AgingNet* sofrem mudanças nos seus perfis transcricionais.

3. METODOLOGIA

Dados de transcriptômica por *microarray* do sangue de indivíduos saudáveis foram obtidos em bancos de dados públicos e pré-processados usando as linguagens de programação R e Bash. Todas as plataformas foram reanotadas para as versões mais atuais do genoma humano e banco de dados de RNAs não-codificadores. Em seguida, utilizou-se todas as amostras para criar um novo conjunto de dados constituído em amostras representativas de cada idade e transcritos em comum à todas as plataformas. Cada transcrito foi tratado como uma série temporal nas seguintes análises, sendo que cada ponto no tempo representa uma idade.

Primeiramente, identificou-se transcritos altamente correlacionados com a idade (*AgingGenes*). Cada transcrito foi classificado de acordo com seu tipo de correlação, podendo ser linear positiva, negativa ou não-linear. Em seguida, realizou-se uma análise de co-expressão para construir uma rede (*AgingNet*) em que seus nós são representados por transcritos e suas relações representadas pela similaridade entre seus perfis de expressão. Uma vez construída a rede, identificou-se subgrupos de genes com perfis de expressão similares ao longo das idades.

Todas as classes do *AgingGenes*, assim como os módulos da *AgingNet*, foram submetidos à uma análise enriquecimento de vias, a fim de elucidar os possíveis mecanismos moleculares dos quais estariam representando. Além disso, foi feita uma análise de Detecção de Pontos de Mudança (*Change-point Detection*) para identificar em quais momentos durante a vida tais programas moleculares apresentam mudanças em seu comportamento transcricional.

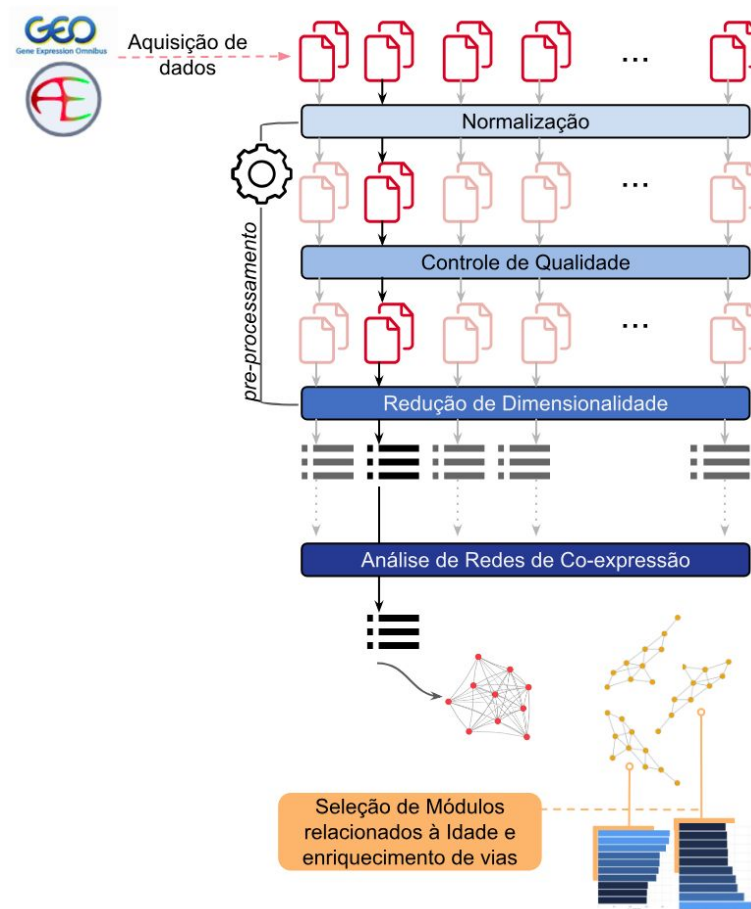


Figura 3.1. Fluxograma de análise. Resumo dos processos e análises realizadas no trabalho.

3.1 Levantamento dos Estudos

Estudos de *microarray* foram obtidos de dois repositórios, Gene Expression Omnibus (GEO, (EDGAR; DOMRACHEV; LASH, 2002) e *ArrayExpress* (KOLESNIKOV et al., 2015). A fim de estudarmos o comportamento do sistema imune, utilizamos dados de sangue total (*whole blood*) e PBMC (*peripheral blood mononuclear cells*), provenientes de indivíduos saudáveis com variadas idades. Os critérios de inclusão para que um estudo fosse selecionado foram que houvesse no mínimo 10 amostras controles/saudáveis, presença de informação sobre idade para cada amostra, disponibilidade de dados brutos e informações sobre a sequência das sondas da plataforma utilizada.

3.2. Reanotação das Sondas nos Microarrays

Microarrays são constituídos de diversas sondas de curto comprimento que testam diversas partes de um mesmo transcrito, que podem ou não ser de um gene codificador de proteína. No entanto, devido ao fato de muitas plataformas terem sido desenvolvidas anos atrás, a anotação funcional de cada transcrito pode estar desatualizada. Por isso, foi realizada a reanotação das plataformas selecionadas utilizando versões mais recentes de bancos de dados, como o *Gencode* (HARROW et al., 2012), *Noncode* (ZHAO et al., 2016), *LNCipedia* (VOLDERS et al., 2013) e *MiTranscriptome* (IYER et al., 2015) (Figura 3.2). Através de intersecções entre coordenadas genômicas, sondas de cada plataforma foram confirmadas ou designadas à novos transcritos, como genes codificadores, RNAs não codificadores, dentre outros.

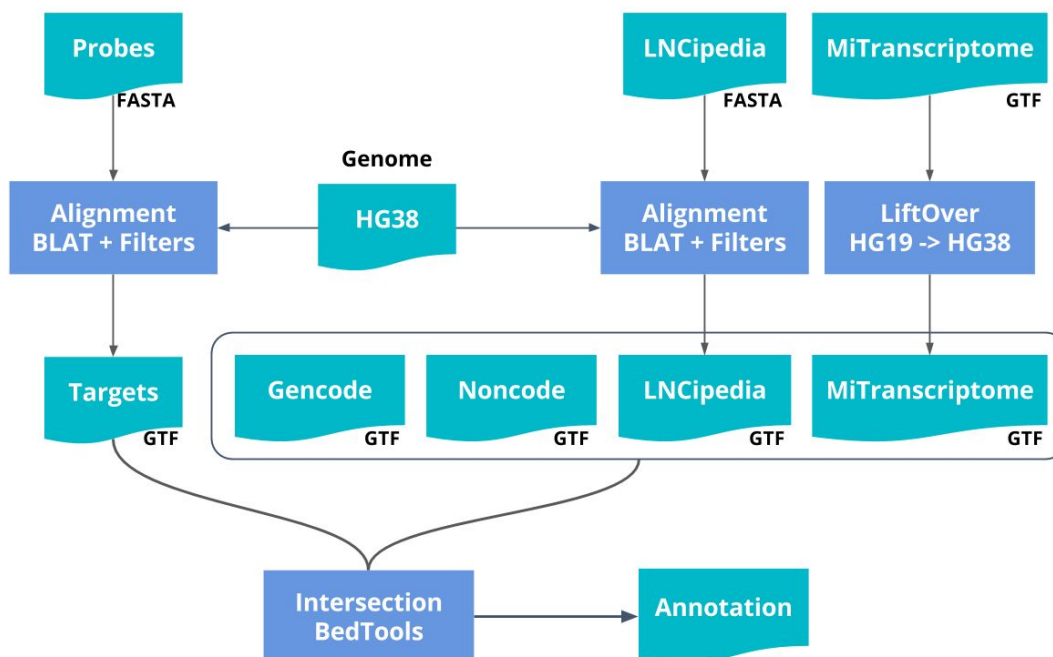


Figura 3.2. Fluxograma sobre a reanotação. Fluxograma representando as etapas do processo de reanotação das sondas de *microarray*. Arquivos estão representados pelas formas em azul claro e programas e processos por formas em azul escuro. Adaptado de Carvalho-Bürger, 2017.

3.3. Aquisição de dados e pré-processamento

A aquisição dos dados foi realizada via scripts em R e *bash*, tanto manualmente (para os dados do *ArrayExpress*) quanto através pacotes do projeto Bioconductor (e.g., GEOquery, (DAVIS; MELTZER, 2007)) para os dados provenientes do GEO.

3.3.1. Normalização

Uma vez obtidos os dados brutos e seus respectivos metadados, com características clínicas e experimentais, amostras controles foram normalizadas utilizando uma abordagem de "barcodes", denominada *Universal exPression Code* (PICCOLO et al., 2013)UPC, (PICCOLO et al., 2013). O UPC é um método aplicado individualmente à cada amostra e estima a probabilidade das sondas estarem expressas (ativas), baseando-se na hipótese de que valores de expressão são originários de duas populações: genes inativos, em que a medida de expressão representa a variação de fundo, e genes ativos, constituído pela variação de fundo

mais um sinal. Como pode ser observado no gráfico a seguir (Figura 3.3), após uma amostra ser normalizada por UPC, os valores de suas sondas podem variar entre 0 e 1, onde apenas uma pequena parcela das *probes* está ativa (valor de expressão > 0.5).

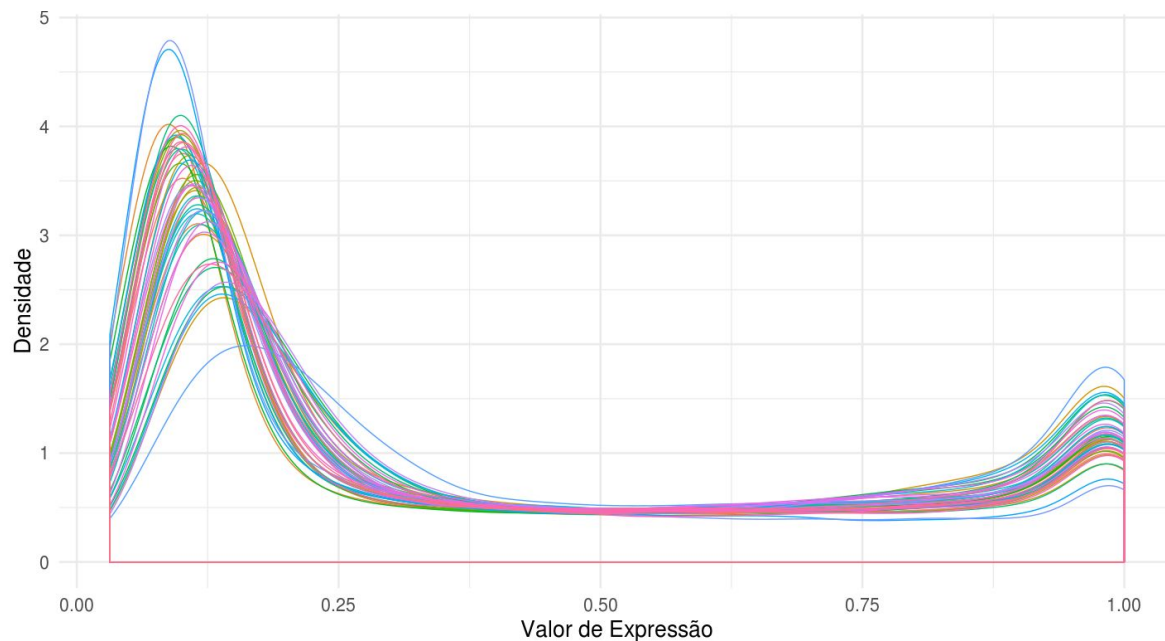


Figura 3.3. Densidade dos valores de expressão do estudo GSE46097. O eixo horizontal representa o nível de expressão e o vertical a densidade de sondas. Cada linha representa uma amostra de um estudo.

3.3.2. Controle de Qualidade

A presença de efeito de lote e identificação de possíveis amostras *outliers*, foram avaliados pela análise de componentes principais (PCA), que permite avaliar as distâncias relativas das amostras no plano de suas componentes principais. Com o PCA é possível detectar *outliers*, ou seja, amostras que possuem um distanciamento elevado em relação às outras, e também agrupamentos de amostras que possivelmente refletem variáveis técnicas, como data do experimento ou *scanner* utilizado. Foram removidas amostras com grande quantidade de variação em relação às outras amostras, e com variação não explicada pelas variáveis fenotípicas.

3.4. Criação de amostras representativas das idades: *AgeCollapsed*

Todos os estudos foram transformados em um novo conjunto de dados denominado *ageCollapsed*, no qual todas as amostras com mesma idade, independentemente do estudo de origem, foram combinadas em uma única amostra representativa de tal idade. Uma vez agregado os estudos através de seus transcritos em comum, corrigiu-se os valores de expressão para efeito de lote (no caso, os estudos) através do pacote *Combat* (JOHNSON; LI; RABINOVIC, 2007) do R. Em seguida, combinou-se as amostras de mesma idade através do valor mediano de expressão de cada transcrito. Utilizou-se a mediana por ser uma métrica menos suscetível à valores discrepantes. Uma vez colapsadas as amostras por idade, as amostras representativas das idades foram reordenadas de forma crescente e cada transcrito tratado como uma série temporal, como exemplificado na figura 3.7. Todas as análises subsequentes foram feitas sobre o *ageCollapsed*.

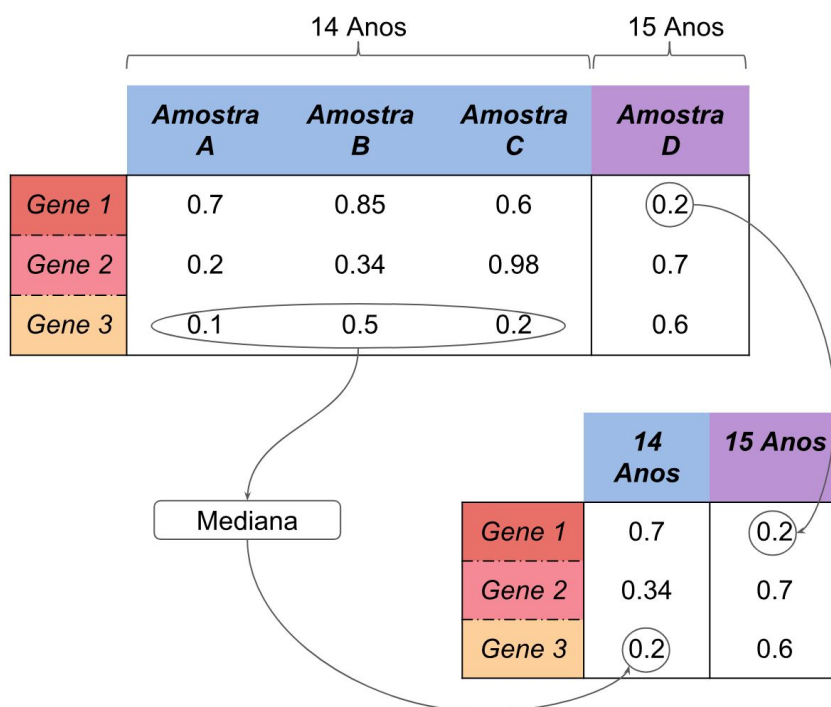


Figura 3.4. Criação de amostras representativas das idades. Esquema ilustrando o colapsamento de amostras com a mesma idade através da mediana de cada gene.

3.5. Detecção de Transcritos altamente relacionados à Idade: *AgingGenes*

A primeira análise realizada foi a identificação de transcritos altamente relacionados com a idade. Para isso, aplicou-se uma transformação de média móvel com uma janela temporal de 5 anos no *ageCollapsed*, para reduzir o ruído das séries temporais (transcritos). Essa transformação foi feita através do pacote *TTR* ("CRAN - Package TTR," [s.d.]) do R. Em seguida, reduziu-se a dimensionalidade através de dois procedimentos com o intuito de utilizar apenas as sondas mais importantes para a análise. Por fim, selecionou-se as sondas altamente correlacionadas com a idade.

3.5.1. Remoção de Sondas não Expressas e/ou Constantes

Como a normalização por UPC nos dá a probabilidade de um transcrito estar sendo expresso, todas que possuíam um valor menor ou igual a 0.5 foram classificados como não-expressos, conforme sugestão de (PICCOLO et al., 2013) Caso um transcrito nunca tenha sido expresso, ou seja, seu valor de expressão foi menor ou igual a 0.5 para todas as idades, o mesmo foi removido. Além disso, transcritos com desvio-padrão (sd) próximo a zero ($sd < 0.01$) também foram removidos.

3.5.2. Concordância das Correlações entre os Gêneros: *SexGenes*

Previamente à qualquer análise, avaliou-se a concordância entre os gêneros da relação dos perfis de expressão gênica e a idade. Para tal, construiu-se dois conjunto de dados semelhantes ao *ageCollapsed* porém cada qual constituído apenas por amostras de cada sexo. Calculou-se a correlação (Spearman) entre a expressão de cada transcrito com a idade e identificou-se quais transcritos minimamente correlacionados com a idade (com valor absoluto do Coeficiente de correlação de Spearman (Rho) > 0.35) possuíam discrepâncias na direção das correlações entre os sexos. Ou seja, transcritos correlacionados positivamente com a idade em homens mas negativamente em mulheres, e vice-versa. Tais transcritos foram denominados de *SexGenes* e removidos das análises futuras.

3.5.3. Seleção de Transcritos Correlacionadas com a Idade

Transcritos correlacionados à idade foram identificadas usando correlação de Spearman ($Rho > 0.75$) e *Maximal Information Coefficient* (MIC, $MIC > 0.75$), que é uma forma de detecção de relações não-lineares entre duas variáveis, presente no pacote *minerva* (RESHEF et al., 2011) do R, e que será melhor explicada na Sessão 3.6 da Metodologia. Transcritos que passaram por esses critérios foram selecionadas para as análises seguintes e denominados de *AgingGenes*. Os *AgingGenes* possuem três classes de transcritos: transcritos com correlação linear positiva (*pos*, $Rho > 0.75$), linear negativa (*neg*, $Rho < -0.75$) e não-linear (*mic*, $MIC > 0.75$). Como o MIC também pode identificar transcritos correlacionados linearmente com a idade, apenas os identificados exclusivamente pelo MIC foram definidos como não-lineares.

Rosa - LinearNeg; verde - mic. Análise de Enriquecimento dos *AgingGenes*

Cada classe de *AgingGenes* foi submetido à uma análise de sobreposição de vias (ORA, *Over Representation Analysis*) do banco de dados Reactome (vs. 2016, (CROFT et al., 2011) através da ferramenta do pacote *enrichR* (KULESHOV et al., 2016) do R.

Mesmo que os *AgingGenes* estejam separados em classes conforme seu perfil transcricional, partimos do pressuposto que cada classe representa alterações transcricionais em diferentes processos de um mesmo sistema, possuindo, então, algum tipo de relação ao nível de processos biológicos. Aproveitando-se da característica hierárquica existente entre as vias do Reactome, desenvolveu-se uma metodologia que permite um olhar global sobre as estruturas hierárquicas enriquecidas entre os *AgingGenes*. Para tal, todas as vias com valor de enriquecimento (*Combined Score*) maior do que 0 foram previamente selecionadas e combinadas em uma estrutura de árvore por meio de suas relações hierárquicas. Em seguida, removeu-se todas as vias “filhas” (nó terminal de uma estrutura em árvore) com *Combined Score* < 5. Tal procedimento leva à criação de diversas estruturas (árvores) de enriquecimento com os mais variados tamanhos (número de componentes/vias na estrutura). Após selecionar apenas as grandes estruturas, constituídas por mais de 6 componentes,

criou-se redes do pacote *visNetwork* (“visNetwork: Network Visualization using ‘vis.js’ Library,” [s.d.]) do *R* para representá-las e auxiliar a compreensão.

3.6. Análise de Transcritos Co-expressos ao longo da Vida: *AgingNet*

A análise de co-expressão consiste em criar uma rede representa as relações existente entre transcritos. Nesta rede, denominada *AgingNet*, seus nós são representados por transcritos e suas inter-relações pela similaridade entre seus perfis de expressão ao longo da vida. Além disso, pode-se identificar grupos de transcritos altamente co-expressos, denominados de módulos de co-expressão, que possam representar processos biológicos específicos. Resumidamente, a similaridade entre todos pares de transcritos foi calculada pela métrica DTW-MIC sobre o conjunto de dados *ageCollapsed*, criando-se então uma matriz de similaridade. Em seguida, clusterizou-se a matriz através do algoritmo HDBSCAN (CAMPELLO; MOULAVI; SANDER, 2013) para identificar módulos de co-expressão.

3.6.1. Similaridade entre Genes.

Para determinar a similaridade do perfil de expressão entre pares de transcritos ao longo da vida, foi utilizada uma métrica desenvolvida para séries temporais, denominada DTW-MIC ((RICCADONNA et al., 2016)). Tal métrica é uma combinação de duas outras, *Dynamic Time Warping* (DTW, (ITAKURA, 1975)) e *Maximal Information Coefficient* (MIC, (ALBANESE et al., 2018)).

A DTW é uma medida de distância entre duas sequências, que leva em consideração deslocamentos temporais. O algoritmo DTW usa programação dinâmica para encontrar um alinhamento ótimo entre as duas séries, através de uma distorção não-linear dos eixos do tempo. A magnitude desta distorção é refletida no valor de dissimilaridade, e a similaridade entre as formas das curvas têm maior impacto para o DTW do que o valor da distância ponto-a-ponto entre as duas séries temporais, como é o caso da distância euclidiana (Figura 3.5). Para obter a medida de similaridade (DTWs) a partir da distância de dissimilaridade (DTWd), usou-se a transformação

DTWs = $1/(1 + \text{DTWd})$, onde DTWd é a distância normalizada de duas séries temporais, calculada através do pacote *dtw* (<http://dtw.r-forge.r-project.org/>).

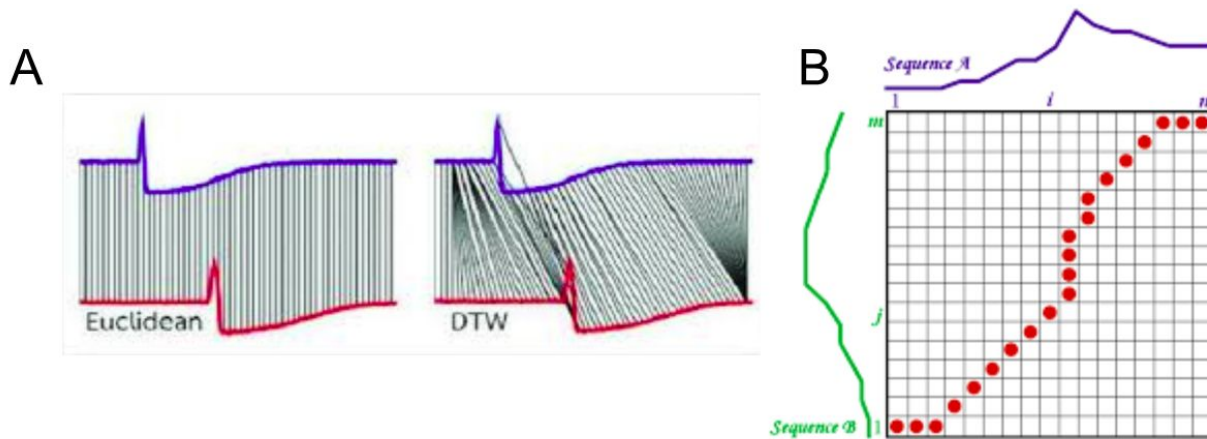


Figura 3.5. Algoritmo DTW (A) ilustra como a distância Euclidiana difere da DTW ao calcular a distância entre duas séries temporais com mesmo perfil de expressão, porém deslocadas no tempo. (B) mostra a matriz de distância gerada pela programação dinâmica durante o processo de otimização para encontrar a melhor distorção entre as duas séries temporais. Ilustração retirada de (CRIEL; TSIPORKOVA, 2006).

A medida MIC, é membro de uma família de estatísticas conhecida como *Maximal Information-based Nonparametric Exploration* (MINE) e desenvolvida para explorar as relações entre duas variáveis em conjuntos de dados multidimensionais. As duas características distintivas do MIC são generalidade, sendo possível capturar relações de diferentes naturezas, e equitabilidade, que é a propriedade de penalizar níveis similares de ruído da mesma maneira, independentemente da natureza da relação entre as variáveis como representado na Figura 3.6.

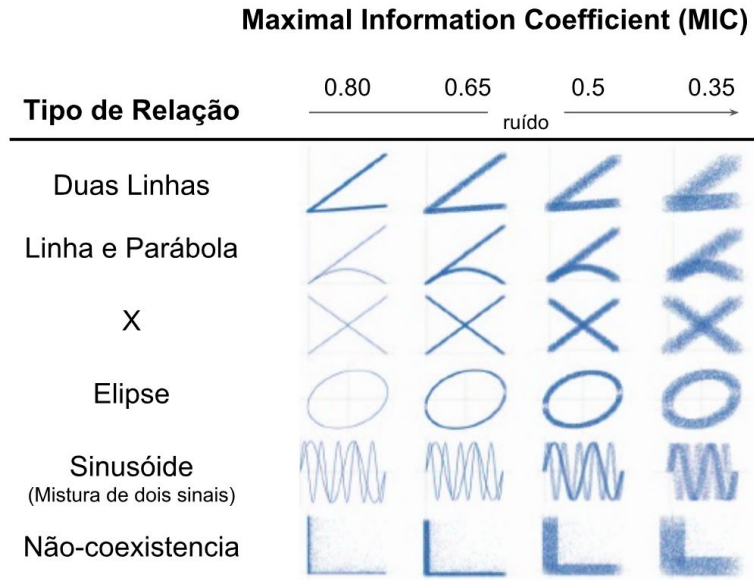


Figura 3.6. Propriedades de Detecção do Coeficiente de Informação Máxima (MIC). Imagem adaptada de (RESHEF et al., 2011), mostrando os tipos de relações que podem ser encontradas entre duas variáveis e a influência que o nível de ruído tem sobre o MIC.

Conforme mencionado anteriormente, neste trabalho foi utilizado uma composição das medidas DTW com MIC, denominada DTW-MIC (RICCADONNA et al., 2016) e definida em como:

$$DTW - MIC(T_1, T_2) = \frac{1}{\sqrt{2}} \sqrt{DTW_s(T_1, T_2)^2 + MIC(T_1, T_2)^2}.$$

Essa nova definição combina as contribuições provindas tanto pela DTW quanto pela MIC, ou seja, leva em consideração tanto deslocamentos temporais quanto correlações não-lineares. Conforme evidenciado através de simulações em (RICCADONNA et al., 2016), tal combinação é mais efetiva que o *Coeficiente de Correlação de Pearson*, e também do que DTW e MIC se consideradas separadamente.

Uma vez calculada a similaridade para cada par de transcrito, obtemos uma matriz de similaridade onde o número de colunas e linhas é igual ao número de transcritos e cada célula dessa matriz representa a similaridade entre duas sondas.

3.6.2. Detecção de Módulos e Sub-Módulos de Co-Expressão

A detecção de módulos de co-expressão foi feita clusterizando-se a matriz de similaridade dos transcritos através de um algoritmo baseado em densidade, denominado HDBSCAN (“Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms [R package dbscan version 1.1-2],” [s.d.]). Esse algoritmo agrupa pontos em um dado espaço que estão muito próximos uns aos outros (pontos com muitos vizinhos) e define como *outlier* os isolados, situados em regiões de baixa densidade (CAMPELLO; MOULAVI; SANDER, 2013). Uma vez detectado os módulos, separou-se os transcritos em sub-módulos conforme seu perfil de expressão.

3.6.3. Análise de Enriquecimento de Vias dos Módulos da *AgingNet*

O mesmo procedimento de criação de estruturas de enriquecimento desenvolvidos para os *AgingGenes* (descritos na Seção 3.5.4) foi aplicado a cada sub-módulo, com uma alteração: foram removidas todas as vias “filhas” com *Combined Score* < 10. Além disso, calculou-se a correlação entre a idade e o perfil de expressão mediana dos transcritos integrantes de cada via enriquecida. Vias correlacionadas ($|\text{Rho}| > 0.35$) foram classificadas conforme sua tendência (positiva e negativa), sendo essa informação agregada à visualização das estruturas de enriquecimento.

3.7. Detecção de Pontos de Mudanças em Módulos Relacionados à Idade

Para analisar o comportamento dos processos identificados neste trabalho, aplicou-se uma transformação por média móvel com janela de 5 anos à cada transcrito do *ageCollapsed*. Em seguida, foi calculado o perfil de expressão de cada classe de *AgingGenes* e sub-módulos da *AgingNet* através da mediana dos valores de expressão

de seus transcritos por idade, obtendo-se uma série temporal de expressão condensada para cada classe/sub-módulo.

Posteriormente, realizou-se uma análise de detecção de pontos sobre o perfil de expressão de cada módulo para identificar mudanças importantes em seu comportamento ao longo da vida. Esse teste foi feito através do pacote *cpm* (MATTHEWS; FOULKES, 2015) do *R*, com o método de Cramer-von-Mises no qual detecta mudanças estatisticamente significativas ($p\text{-valor} < 0.01$) tanto na média quanto na variância entre fases temporais.

4. RESULTADOS

4.1. Levantamento e Aquisição de Dados

As pesquisas realizadas nos bancos de dados utilizando as palavras-chaves “human”, “whole blood”, “pbmc” e “peripheral blood” resultaram em 87 potenciais estudos. Após aplicados os critérios de inclusão definidos na Seção 3.1, apenas 12 estudos foram de fato selecionados para análise, como mostrados na Tabela 1.

Tabela 1 - Relação dos estudos selecionados para a análise.

Estudo	Base de Dados	Empresa	Plataforma	Tecido	N. Amostras*	Idade Mínima	Idade Máxima	N. Mulheres	N. Homens
GSE73089	GEO	Agilent	GPL4133	PBMC	120	40	97	64	56
E-TABM-305	Array Express	Illumina	A-MEXP-691	linfócitos	1240	14	81		
GSE15573	GEO	Illumina	GPL6102	PBMC	15	45	67	10	5
GSE48556	GEO	Illumina	GPL6947	PBMC	33	44	76	24	9
GSE54514	GEO	Illumina	GPL6947	sangue total	36	24	70	24	12
GSE12288	GEO	Affymetrix	GPL96	sangue total	112	37	68	28	84
GSE18781	GEO	Affymetrix	GPL570	sangue total	25	22	83	20	5
GSE42057	GEO	Affymetrix	GPL570	PBMC	42	46	80		
GSE22255	GEO	Affymetrix	GPL570	PBMC	20	45	73	10	10
GSE19314	GEO	Affymetrix	GPL570	PBMC	20	34	70	15	5
GSE16028	GEO	Affymetrix	GPL570	sangue total	109	23	64	59	50
GSE46097	GEO	Affymetrix	GPL571	sangue total	63	41	80	36	27

*Número de Amostras Controle. Células em branco nas colunas de gênero são informações ausentes ou não informativas, como 0 ou 1.

Os estudos selecionados são de 3 subtipos do tecido sanguíneo, realizados em 7 plataformas distintas de *microarray* (Figura 4.1), provenientes de 3 empresas (Agilent, Illumina e Affymetrix) e totalizando 1835 amostras.

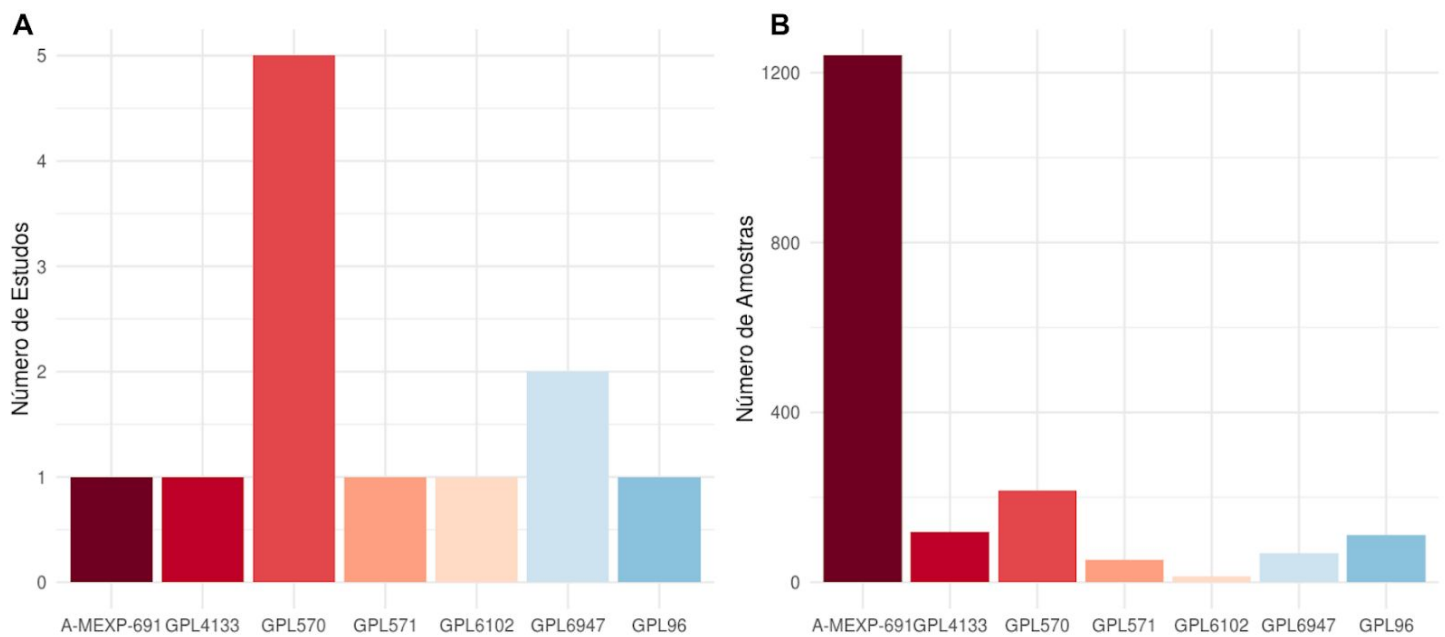


Figura 4.1. Quantidade de Dados Disponíveis por Plataforma. Gráficos de barras representando quantidade de estudos (A) e amostras (B) para cada plataforma.

4.2. Reanotação das Plataformas

A reanotação das plataformas identificou diversas classes de transcritos, sendo a maioria sondas representando transcritos codificadores de proteínas provenientes do *GenCode*. Além disso, encontramos mais de 10.000 transcritos não-codificadores de proteínas provenientes dos bancos de dados *Noncode*, *LNCipedia* e *MiTranscriptome* (Figura 4.2).

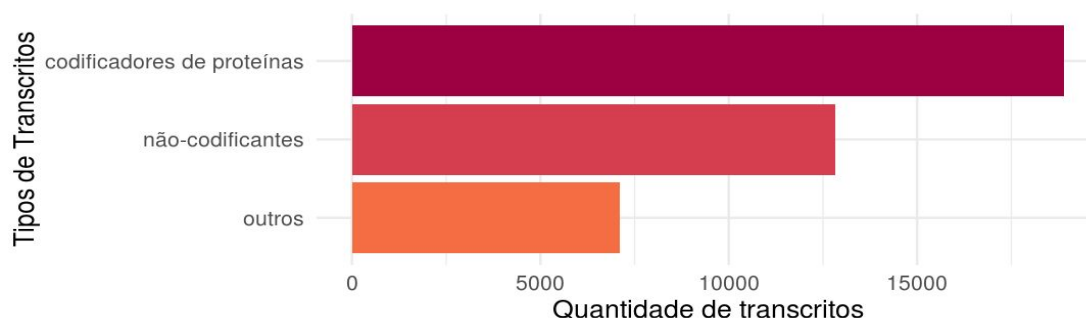


Figura 4.2. Quantidade dos Principais Tipos de Transcritos encontrados em todas as Plataformas.

Gráfico de barras representando a quantidade total de cada tipo de transcrito encontrado em todas as plataformas, após a reanotação. Transcritos classificados como “outros” se referem à pseudogenes, imunoglobulinas e RNAs mitocondriais.

Analisando-se a proporção das classes de transcritos em cada plataforma (Figura 4.3), nota-se uma grande variabilidade tanto na quantidade total quanto na proporção das classes de transcritos encontrados em cada plataforma.



Figura 4.3. Proporção dos Tipos de Transcritos por Plataforma. Gráficos de pizza representando a proporção dos tipos de transcritos encontrados em cada plataforma, após a reanotação. São mostrados a proporção de transcritos codificadores de proteínas (em vermelho), não-codificantes (em amarelo) e outros (verde).

4.3. Pré-Processamento e Criação do *ageCollapsed*

Após o pré-processamento e controle de qualidade, remaneceram 1807 amostras com idades entre 15 e 96 anos, sendo a maior parte delas concentradas entre os jovens e adultos (Figura 4.4). Essas amostras foram utilizadas para construir o conjunto de dados *ageCollapsed* (conforme descrito na Seção 3.4 da Metodologia). O *ageCollapsed* consiste em uma combinação de todos os estudos em um novo conjunto de dados, representando o valor de expressão dos transcritos para cada idade. Dessa forma, obtemos uma matriz de expressão onde os transcritos em comum à todas as plataformas (8348) são representados pelas linhas e as idades pelas colunas, com um total de 82 colunas, cada uma representando uma idade entre os 15 e 96 anos (com exceção da idade 95).

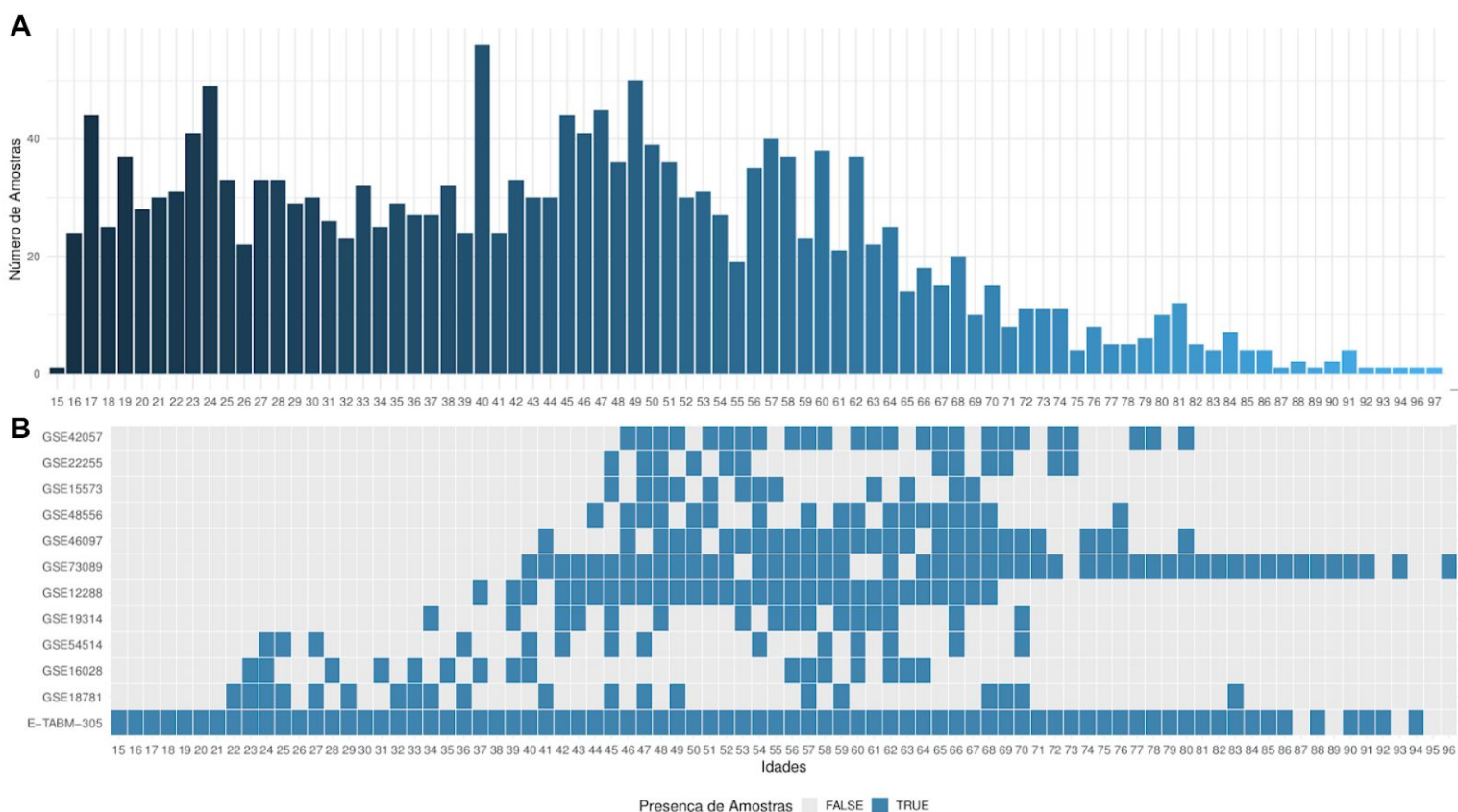


Figura 4.4. Representação da Distribuição e Representatividade das Idades. (A) *Histograma das idades de todas as amostras.* (B) *Gráfico ilustrando a representatividade de cada idade pelas amostras encontradas em todos os estudos.*

4.4. Avaliação da Concordância das Relações dos Transcritos com a Idade entre os Sexos

Previamente à qualquer análise, avaliou-se a influência do sexo no contexto de expressão gênica ao longo da vida com o intuito de remover transcritos com comportamentos discrepantes entre homens e mulheres. Mais especificamente, avaliou-se se os transcritos possuíam a mesma direção de correlação (Spearman) entre seu perfil de expressão e a idade para ambos os sexos. Para isso, criou-se dois conjuntos de dados, semelhantes ao *ageCollapsed*, contendo cada qual apenas amostras de cada gênero (1003 amostras de mulheres e 761 de homens). Em seguida, avaliou-se a direção dos transcritos correlacionados com a idade ($|\text{Rho}| > 0.35$) e identificou-se 38 transcritos com direções opostas, denominados *SexGenes*. Os *SexGenes*, que foram removidos das análises seguintes, estão listados na tabela suplementar (Anexo A - Tabela dos *SexGenes*) e representados pelos pontos de cor preta na Figura 4.5.

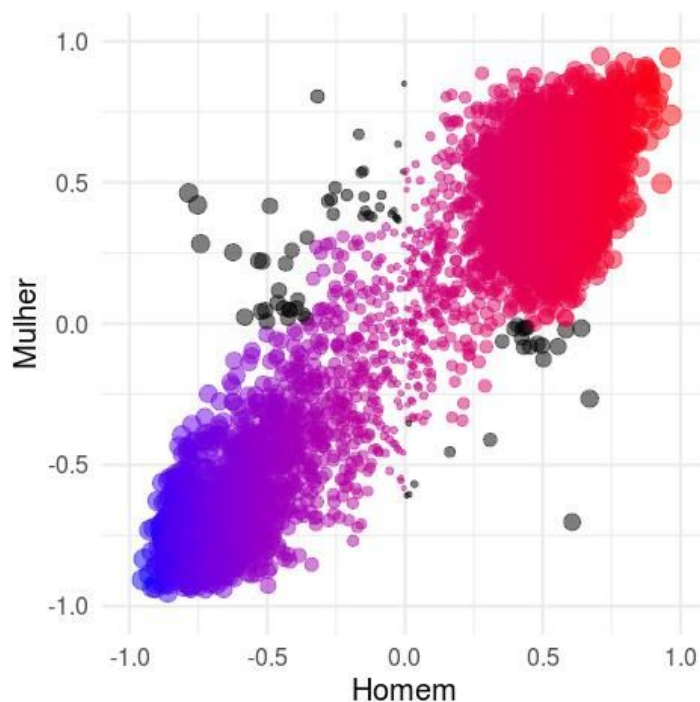


Figura 4.5. Concordância das Correlações entre os Sexos. Cada ponto representa um transcrito, os eixos representam os valores de correlações (Spearman) com a idade dos transcritos, para cada sexo, e as cores demonstram a intensidade da correlação.

4.5. Análises dos *AgingGenes* e *AgingNet*

Neste trabalho empreendeu-se essencialmente duas análises. A primeira, análise dos *AgingGenes*, tem como objetivo identificar transcritos altamente relacionados à idade e avaliar como suas funções biológicas podem estar relacionados à imunossenescência, levando em consideração seu comportamento ao longo da vida.

A segunda, análise de co-expressão, consiste em identificar quais transcritos interagem entre si ao longo da vida, quais processos biológicos possivelmente representam e como seu comportamento ao longo do tempo pode estar relacionado à imunossenescência. Tais interações entre transcritos foi representada através de uma rede de co-expressão, denominada de *AgingNet*.

4.5.1. Transcritos Relacionados à Idade: *AgingGenes*

Foram encontrados 56 *AgingGenes*, dentre os quais 10 possuem correlações lineares positivas (*pos*), 6 com correlações lineares negativas (*neg*) e 40 com correlações não-lineares (*mic*). Dentre os *mic*, identificou-se 3 subgrupos de transcritos com perfis de expressões distintos, sendo constituídos de 7, 25 e 8 transcritos cada e denominados de *micA*, *micB* e *micC*, respectivamente. As figuras à seguir (Figuras 4.6 A e B) ilustram os genes correspondentes a cada *AgingGenes* encontrados no banco de dados StringDB (SZKLARCZYK et al., 2015).

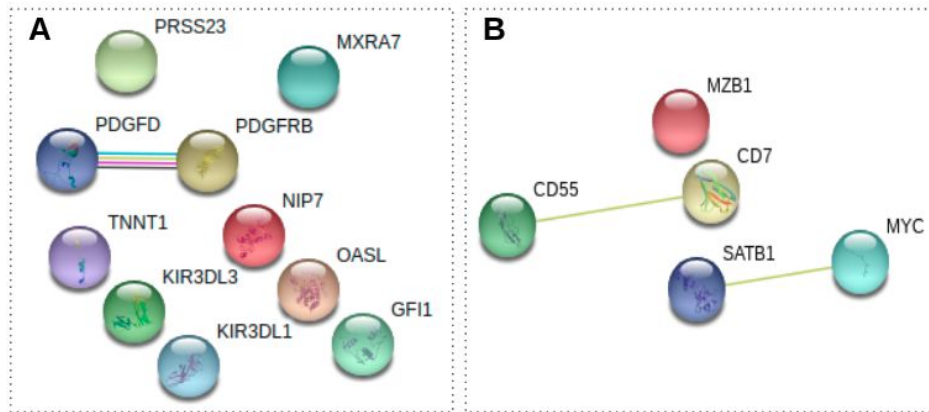


Figura 4.6 A. *AgingGenes* com Correlação Linear. (A) genes com correlações lineares positivas (*pos*) e (B) negativas (*neg*). As linhas representam interações entre genes detectadas pelo *String*.

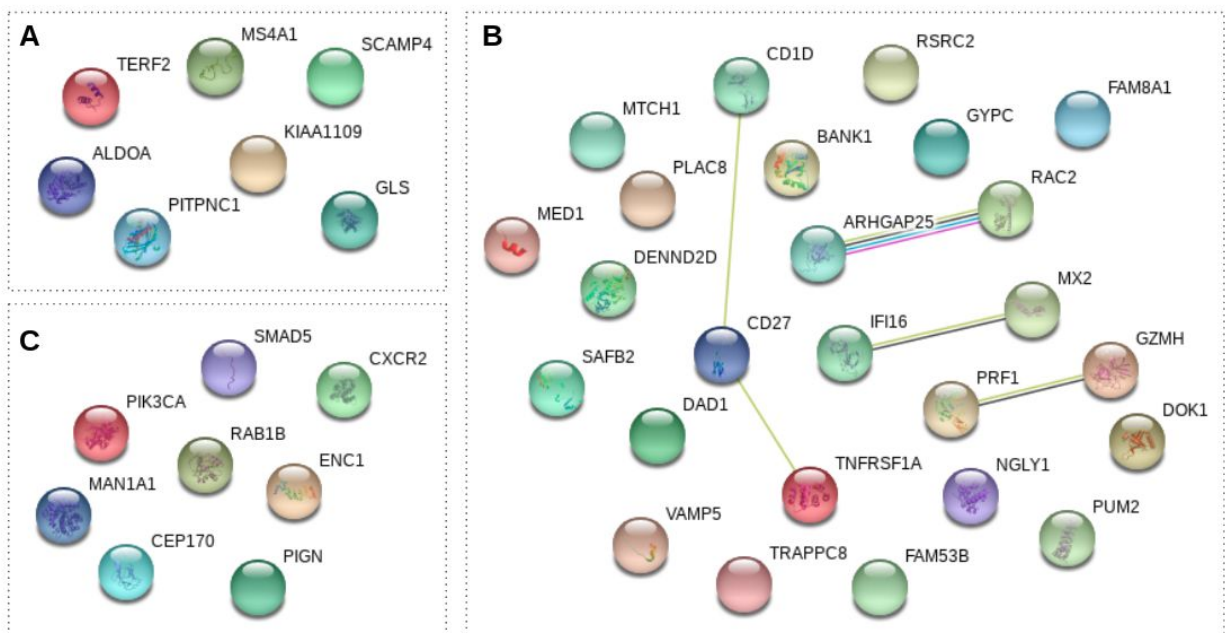


Figura 4.6 B. *AgingGenes* com Correlação Não-Linear (*mic*). A, B e C representam os genes com correlações não-lineares *micA*, *micB* e *micC*, respectivamente. As linhas representam interações entre genes detectadas pelo *String*.

De todos os *AgingGenes*, apenas 5 genes foram encontrados no *GenAge* e correspondem aos genes TERF2, MED1, PIK3CA, MYC e PDGFRB, com correlações não-lineares (*micA*, *micB*, *micC*) e correlações lineares negativas e positivas,

respectivamente. Interessante notar que cada um desses 5 genes se enquadra em um tipo distinto de correlação com a idade.

À seguir, para melhor entender em quais processos biológicos os *AgingGenes* estão envolvidos, realizou-se uma análise de enriquecimento de vias para cada classe de transcritos conforme descrito na Seção 3.5.4. Foram encontrados três grandes estruturas de enriquecimento, ilustradas nas Figuras 4.7 A, B e C.

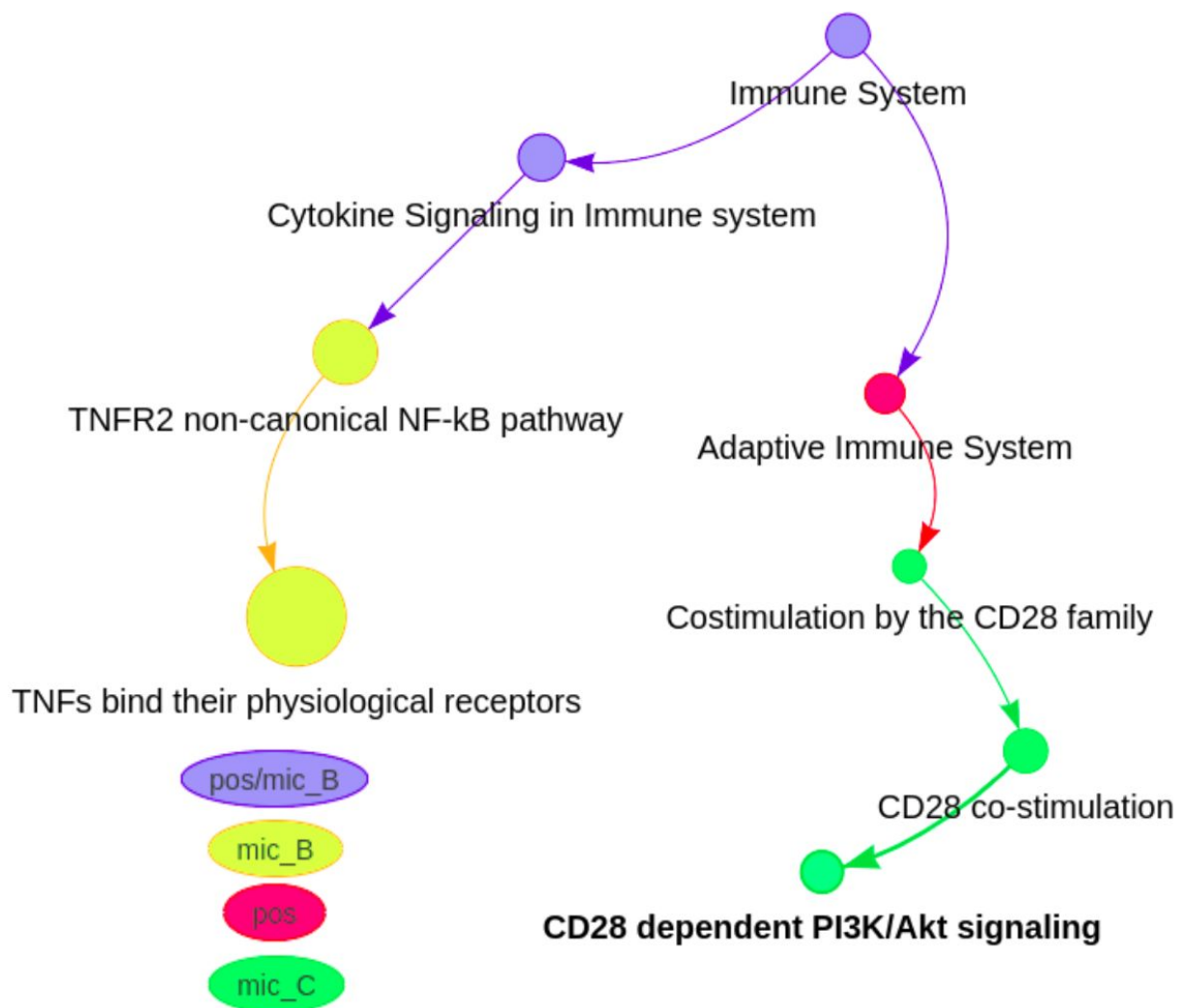


Figura 4.7 A. Vias do Reactome Enriquecidas para os *AgingGenes* (Estrutura 1). Cada círculo representa uma via enriquecida, sendo que suas dimensões refletem o valor de enriquecimento (*Combined Score*) e a cor representa em quais classes dos *AgingGenes* tal via foi enriquecida.



Figura 4.7 B. Vias do Reactome Enriquecidas para os *AgingGenes* (Estrutura 2). Cada círculo representa uma via enriquecida, sendo que suas dimensões refletem o valor de enriquecimento (*Combined Score*) e a cor representa em quais classes dos *AgingGenes* tal via foi enriquecida.

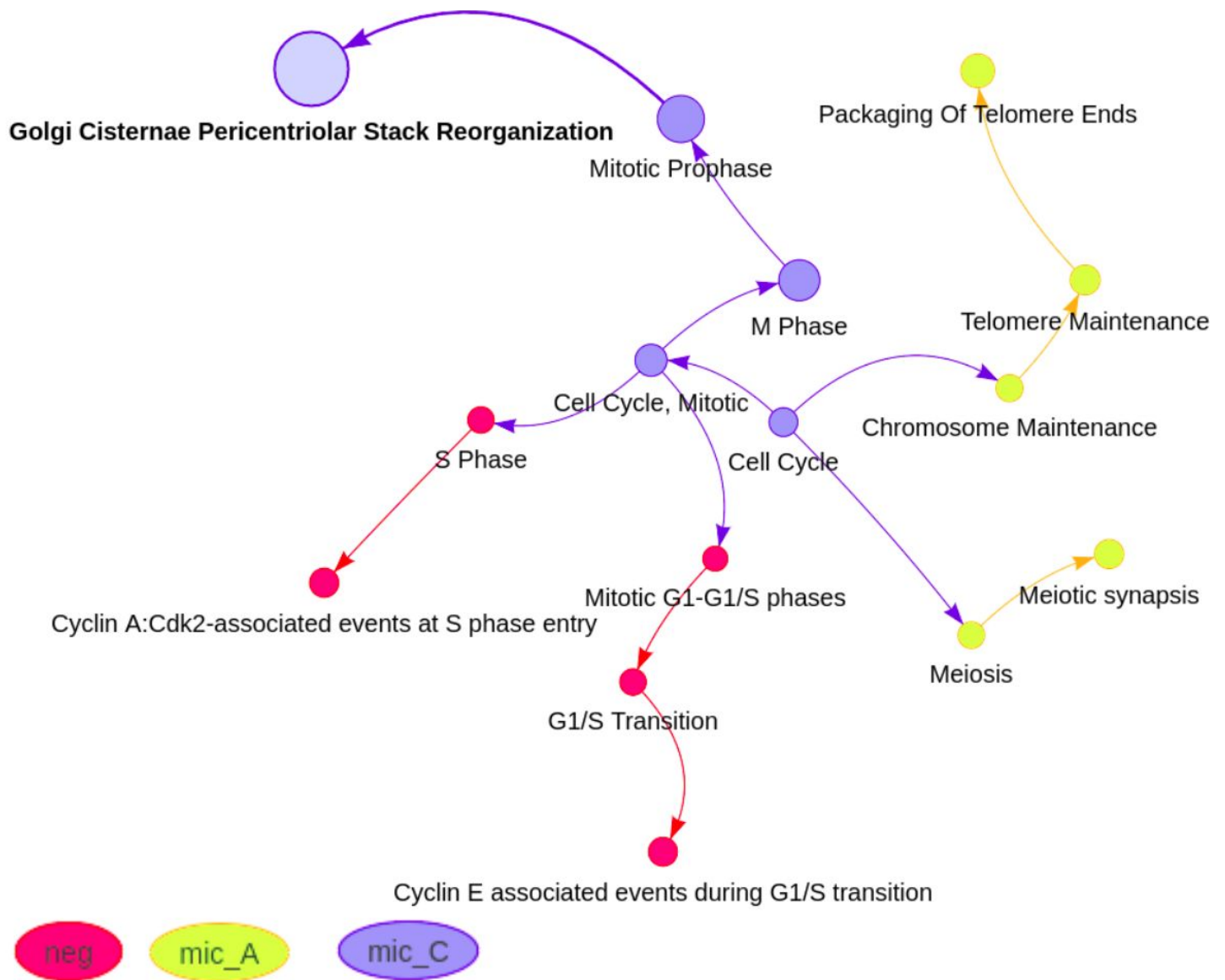


Figura 4.7 C. Vias do Reactome Enriquecidas para os *AgingGenes* (Estrutura 3). Cada círculo representa uma via enriquecida, sendo que suas dimensões refletem o valor de enriquecimento (*Combined Score*) e a cor representa em quais classes dos *AgingGenes* tal via foi enriquecida.

Uma vez explorado quais os possíveis processos biológicos no qual os *AgingGenes* estão relacionados, analisou-se seu comportamento ao longo da vida. Para isso, criou-se um perfil de expressão representativo para cada classe de *AgingGenes*, ilustrados nas Figuras 4.8 A e B. Tais perfis representam a expressão mediana de todos os transcritos para cada idade, separados pela sua classe. Simultaneamente foi feita a análise de detecção de mudanças de estado e as idades

onde ocorreram alterações estão sinalizadas através de triângulos, coloridos conforme o sub-módulo. Além disso, a dimensão de cada triângulo representa a intensidade de alteração, calculada através da razão (\log_2FC , \log_2 Fold-Change) entre a expressão da idade onde ocorreu a mudança de estado e a média de expressão das idades compreendidas entre 19 e 25 anos.

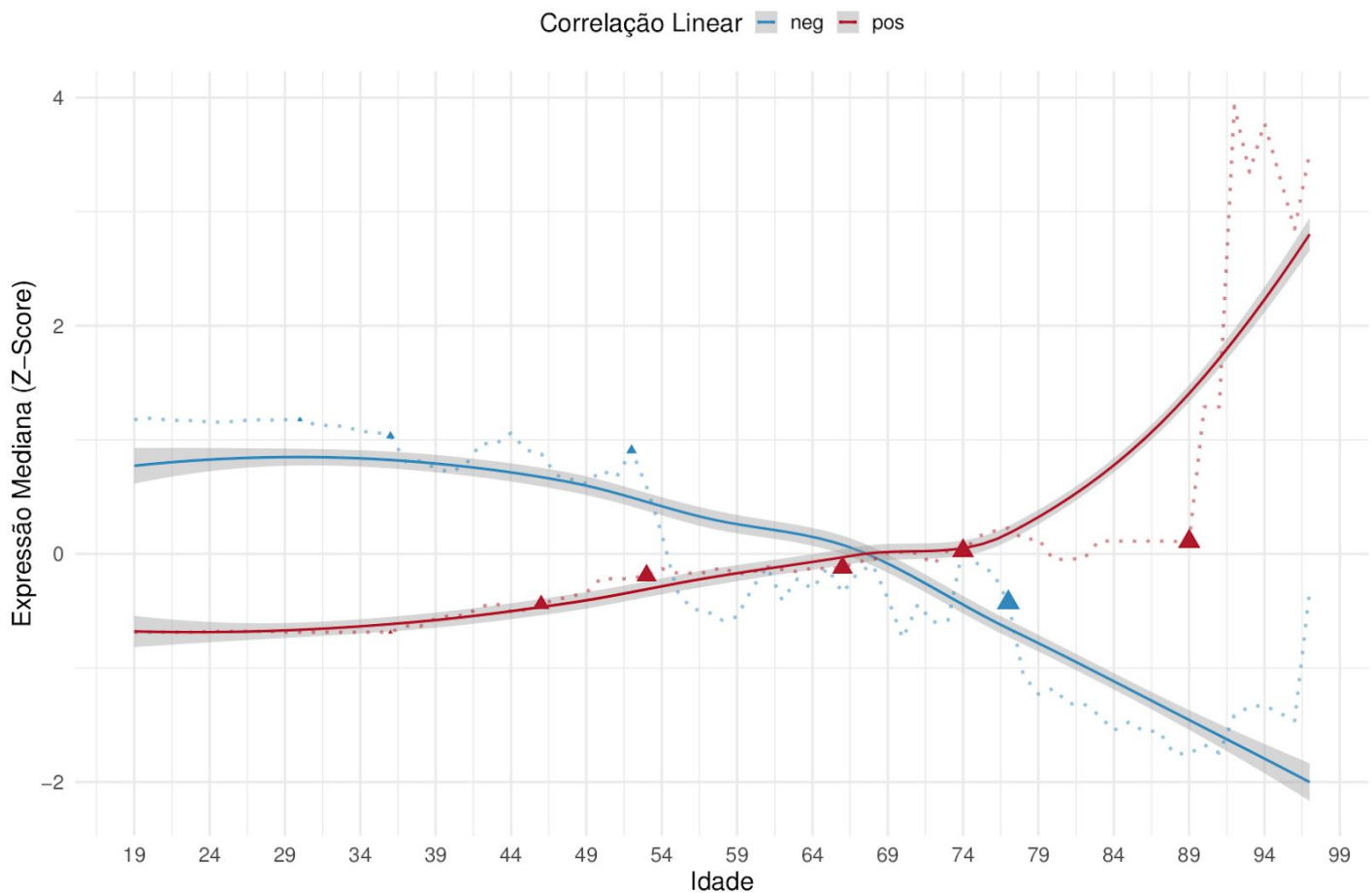


Figura 4.8 A. Perfil de Expressão dos *AgingGenes* com Correlações Lineares. A linha pontilhada representa o perfil mediano de expressão reescalado por z-score para, a linha completa representa a regressão *loess* utilizando todos os transcritos de cada classe e sua sombra representa a variabilidade. As cores representam cada tipo de correlação linear. Os triângulos representam momentos onde ocorreram alterações de estado e sua dimensão representa a intensidade da alteração ($|\log_2FC|$).

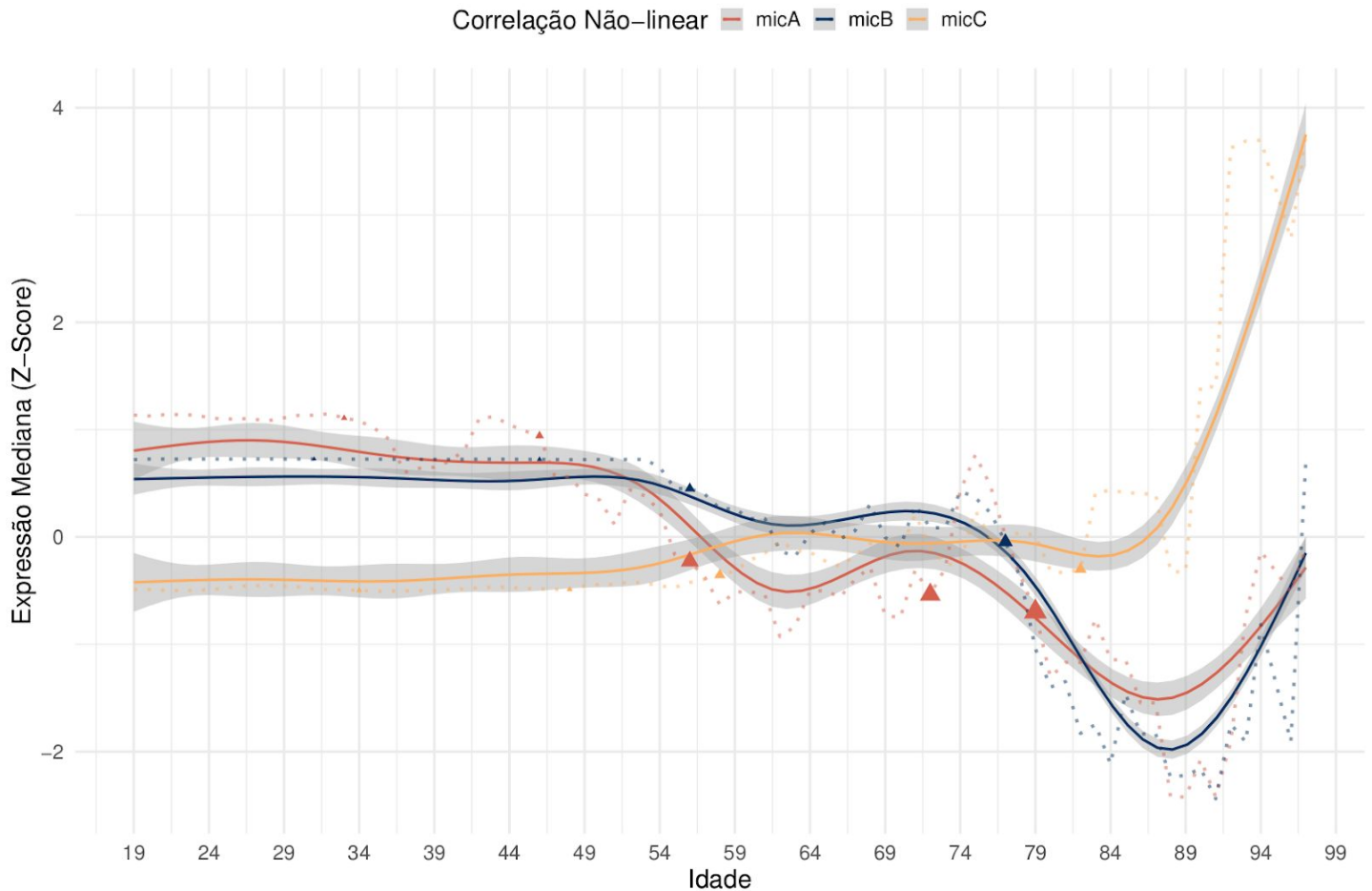


Figura 4.8 B. Perfil Mediano de Expressão para os *AgingGenes* com correlações não-lineares. A linha pontilhada representa o perfil mediano de expressão reescalado por z-score, a linha completa representa a regressão *loess* utilizando todos os transcritos de cada classe e sua sombra representa a variabilidade. As cores representam cada subgrupo dos *AgingGenes* não-lineares. Os triângulos representam momentos onde ocorreram alterações de estado e sua dimensão representa a intensidade da alteração ($|\log_2FC|$).

Ao avaliarmos o perfil de expressão mediano dos *AgingGenes* com correlação linear, nota-se um aumento mais acentuado e progressivo da classe *pos* à partir dos 35 anos, onde sofre sua primeira pequena mudança de estado aos 36 anos e segue com mudanças cada vez mais intensas até os 70 anos, onde sofre sua maior mudança. Já o perfil da classe *neg* apresenta um comportamento mais conturbado, começando a apresentar decaimentos logo aos 30 anos, sua primeira mudança de estado, e novamente, de forma mais intensa, aos 36, juntamente com a classe *pos*. A classe

micA, relacionada à manutenção de telômeros, aparenta ser a classe com maiores perturbações sofridas ao longo da vida. Após 10 anos desde às últimas alterações sofridas pelos *AgingGenes* de classe *pos* e *neg* (36 anos), as alterações parecem se intensificar levando à um estado de clara perturbação até os 56 anos, no qual todas as classes novamente sofrem mudanças de estado. A classe *micC* possui uma tendência positiva com a idade, sofrendo grandes alterações por volta dos 60 anos. Já a classe *micB* possui um perfil similar à classe *micA*, porém com alterações menos intensas (com exceção do período compreendido entre os anos 85 e 90).

A class ***micC***, que possui uma tendência positiva em relação à idade, está relacionado à proliferação celular, vias do ciclo celular e mitose. Além disso, está enriquecida tanto para vias de sinalização PI3K/Akt dependente de CD28 e quanto para sinalização por fatores de crescimento, como FGFR e EGFR. O comportamento dessas vias entram em acordo com a classe ***neg***, no qual indica uma diminuição da expressão de genes associados à entrada da fase estacionária S. Essa classe também está relacionada à sinalização por NOTCH1 em câncer. Ao mesmo tempo, a classe ***micA***, o perfil que sofre quedas mais abruptas durante a vida, está enriquecido para vias de manutenção de telômeros, como empacotamento das regiões terminais dos telômeros, e principalmente enriquecida para via de indução da senescência por estresses relacionados aos telômeros e danos ao DNA. A classe ***micB***, com tendências de queda na expressão mais sutis durante a vida, está enriquecido para vias de sinalização de *Tumor Necrosis Factor* (TNF), tanto por TNFR1 (vias listadas na tabela Anexo A - Enriquecimento dos *AgingGenes*) quanto por TNFR2, envolvido na sinalização de citocinas através da ativação de NF-κB pela via não-canônica. Sinalização de citocinas também foi enriquecidas para a classe *pos*. A classe *pos* também está relacionada à sinalização de PI3K/AKT em câncer e, juntamente com a classe *micC*, também está enriquecida para a via de sinalização constitutiva por PI3K aberrantes em câncer.

4.5.2 Rede de Co-Expressão do Envelhecimento: *AgingNet*

A *AgingNet* consiste em uma rede de co-expressão criada através do *ageCollapsed*, onde selecionou-se 3569 transcritos para constituírem seus nós e a similaridade DTW-MIC suas inter-relações. Esses transcritos foram selecionados através da correlação de seu perfil de expressão com a idade, no qual selecionou-se apenas os com correlações maiores do que de 0.35 (correlação linear: $|\text{Rho}| > 0.35$ e correlação não-linear: $\text{MIC} > 0.35$). À seguir identificou-se grupos de transcritos muito similares entre si através da clusterização pelo algoritmo *hdbscan*, encontrando-se 3 grandes módulos de co-expressão, denominados M0, M1 e M2. O M0 é um módulo “residual”, no qual representa transcritos que não são membros de nenhum *cluster*, sendo desconsiderado nas análises posteriores.

Os módulos restantes, M1 e M2, são constituídos de 1154 e 231 transcritos, respectivamente, e os mais importantes estão ilustrados nas Figuras 4.9 A e B. Para tal, calculou-se sua força (grau ponderado) na rede dos módulos e selecionou-se apenas os 100 transcritos com maior força. Importante notar que 5 *AgingGenes* estão presentes entre os 100 transcritos mais importantes dos dois módulos: os genes GLS, TERF2 e MYC presentes no módulo M1 e os genes GYPC e NGLY1 presentes no módulo M2. Além disso, há 9 transcritos entre eles que também estão presentes no banco de dados *GenAge*, sendo os genes RAE1, ARNTL, MYC e TERF2, presentes no módulo M1, e HDAC3, PRKDC, VCP, SP1 e FOXO1 no módulo M2.

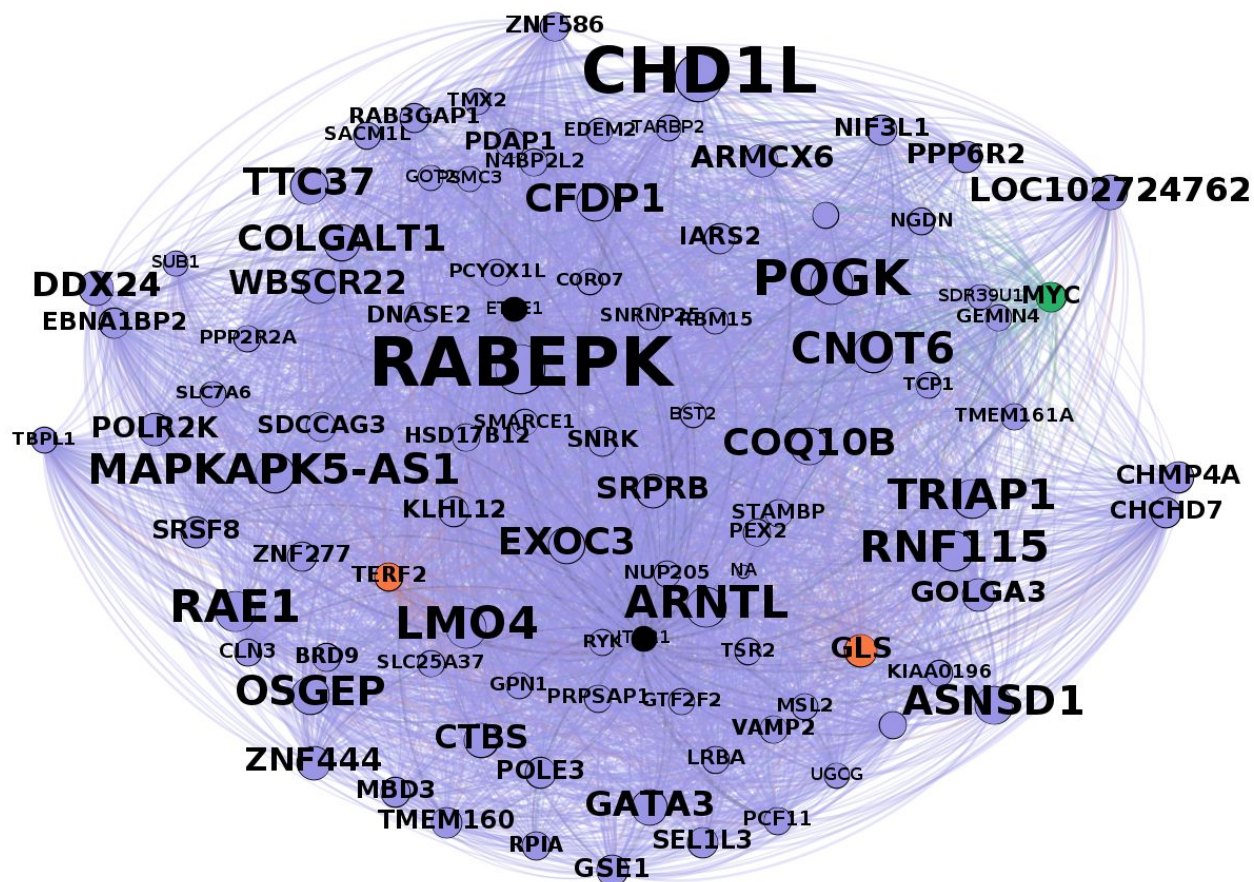


Figura 4.9 - A. Rede do Módulo M1. Cada circunferência representa um transcrito e as linhas seus valores de similaridade DTW-MIC. A dimensão das circunferências são proporcionais à sua força na rede e as cores seu tipo de correlação com a idade: em roxo estão representados transcritos com correlações lineares negativa, e em laranja e verde estão representados *AgingGenes* da classe *micA* e *neg*, respectivamente.

O módulo M2 também possui a maioria dos transcritos com correlações negativas (162 transcritos) e 55 transcritos com correlações positivas e 23 não-lineares. Os 10 transcritos mais importantes dessa rede são WDR48, XKR8, MRPL24, TM7SF3, BTBD1, HDAC3, PRKDC, RAP1GDS1, COX16 e API5, todos correlacionados negativamente com a idade.

Ao analisarmos o perfil de expressão geral dos transcritos que constituem cada módulo foram identificados 4 subgrupos de transcritos em cada módulo com perfis de expressão distintos entre si, resultando em 8 sub-módulos. Os sub-módulos do M1 e M2 foram denominados de M1S1, ..., M1S4 e M2S1,..., M2S4, respectivamente. À seguir, cada sub-módulo foi resumido através da mediana da expressão de seus transcritos para cada idade, e construiu-se seu perfil de expressão geral ilustrado nas Figuras 4.10 A e B. Simultaneamente foi feita a análise de detecção de mudanças de estado e as idades onde ocorreram alterações, da mesma forma feita para os *AgingGenes*.

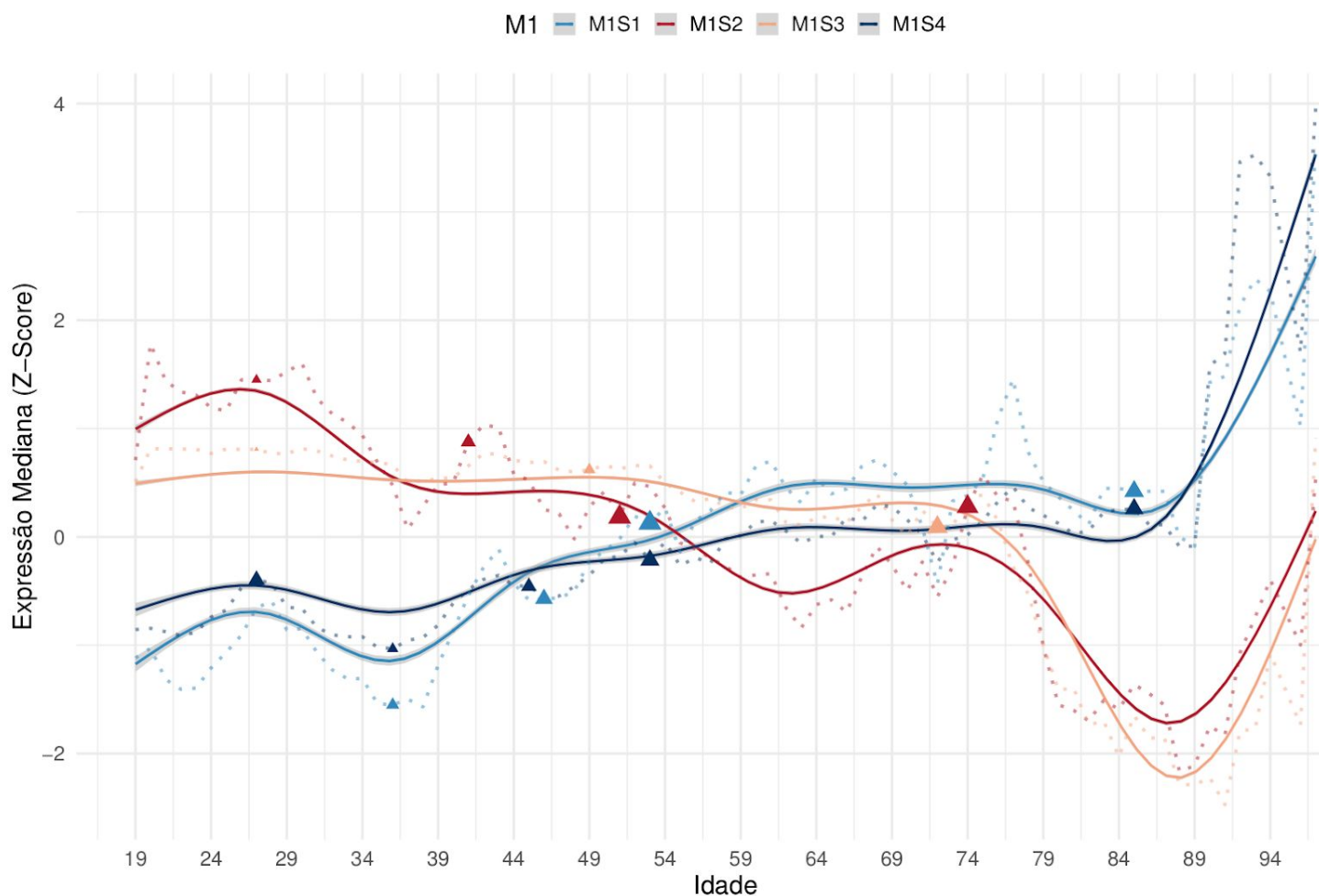


Figura 4.10 A - Perfis de Expressão dos Sub-módulos Constituintes do Módulo M1. As linhas pontilhadas representam o perfil mediano de expressão reescalado por z-score, a linha completa representa a regressão *loess* utilizando todos os transcritos de cada classe e sua sombra representa a variabilidade. As cores representam cada sub-módulo, os triângulos representam momentos onde ocorreram alterações de estado e sua dimensão representa a intensidade da alteração ($|\log_2FC|$).

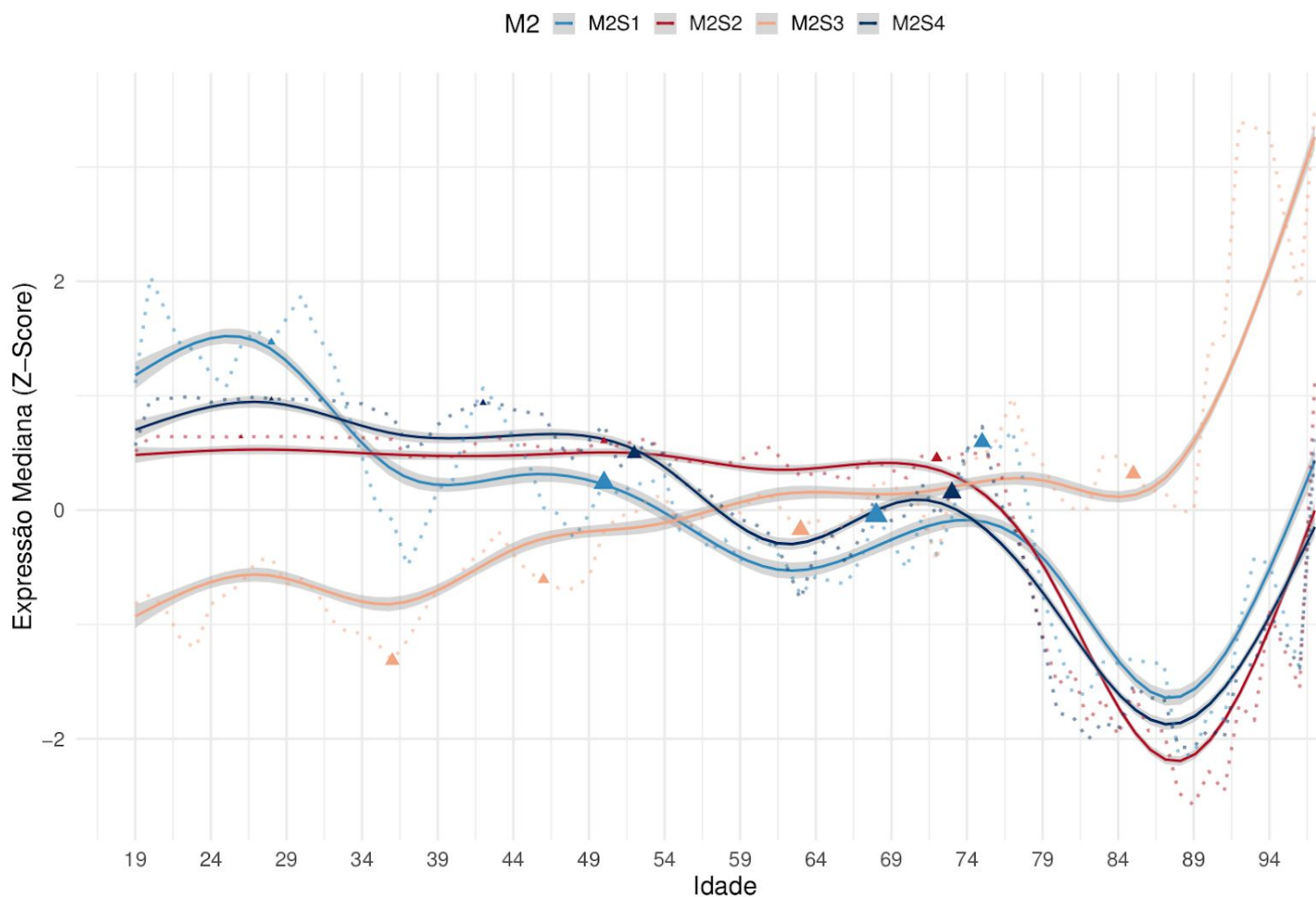


Figura 4.10 B - Perfis de Expressão dos sub-módulos que constituem o Módulo M2. As linhas pontilhadas representam o perfil mediano de expressão reescalado por z-score, a linha completa representa a regressão *loess* utilizando todos os transcritos de cada classe e sua sombra representa a variabilidade. As cores representam cada sub-módulo, os triângulos representam momentos onde ocorreram alterações de estado e sua dimensão representa a intensidade da alteração ($|\log_2FC|$).

Na análise de enriquecimento de vias, foi encontrado duas grandes estruturas intra-conectadas para o módulo M1 (Figura 4.11 A e B), enquanto que para o módulo M2 encontrou-se 6 estruturas menores e mais desconexas (Figura 4.11 C).

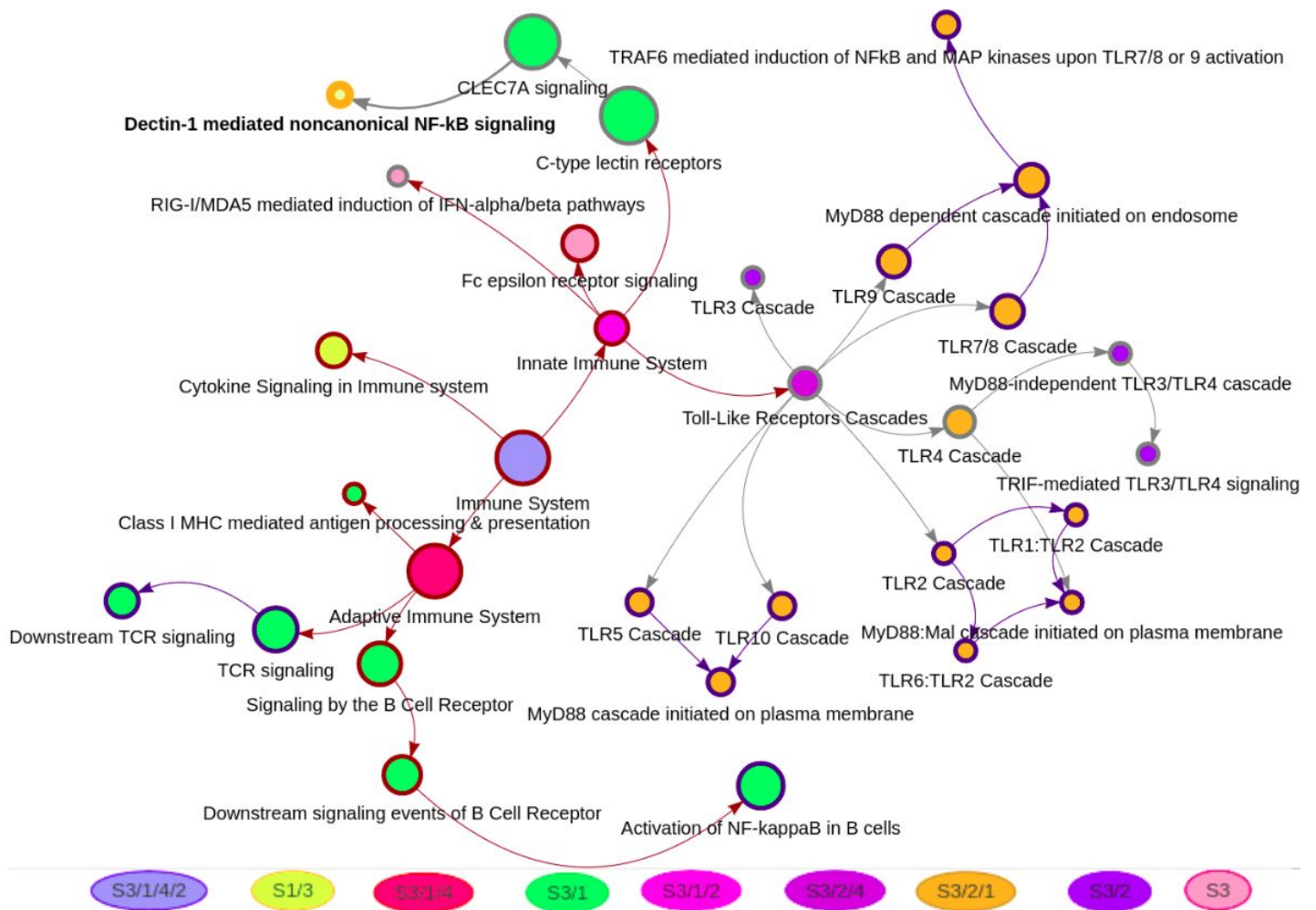


Figura 4.11 A - Enriquecimento de Vias dos Sub-módulos que constituem o Módulo M1 (Estrutura 1). A cor dos círculos representam o sub-módulo no qual uma dada via foi enriquecida, indicado pela legenda inferior, e a cor da borda destes círculos representam a correlação entre o perfil de expressão mediano da via do Reactome com a idade, sendo as cores cinzas, vermelhas e azuis representativas de correlações positivas, negativas e não-correlacionadas ($|Rho| < 0.35$), respectivamente. A dimensão dos círculos representam o valor de enriquecimento da via (*Combined Score*).

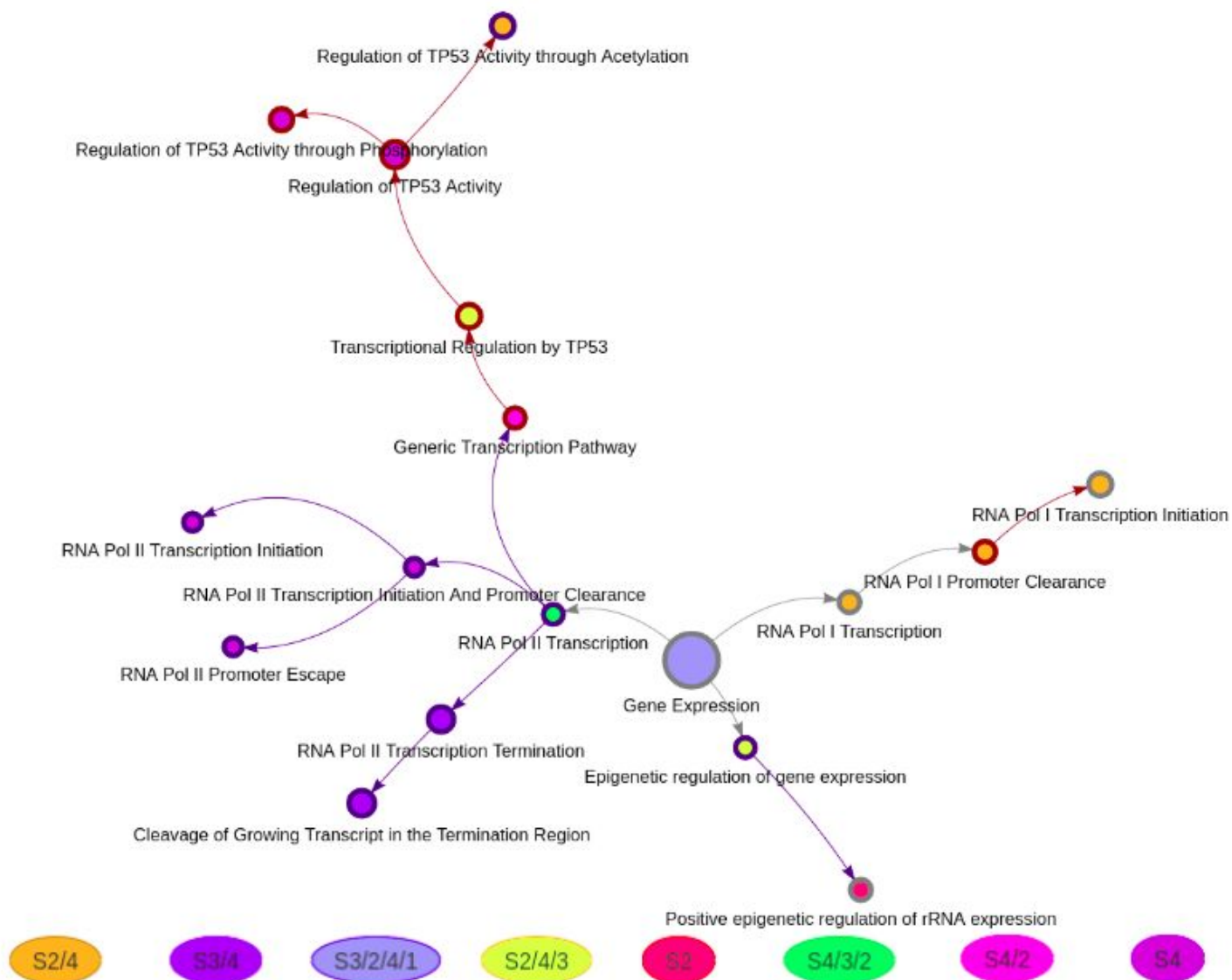


Figura 4.11 B - Enriquecimento de Vias dos Sub-módulos que constituem o Módulo M1 (Estrutura 2). A cor dos círculos representam o sub-módulo no qual uma dada via foi enriquecida, indicado pela legenda inferior, e a cor da borda destes círculos representam a correlação entre o perfil de expressão mediano da via do Reactome com a idade, sendo as cores cinzas, vermelhas e azuis representativas de correlações positivas, negativas e não-correlacionadas ($|\text{Rho}| < 0.35$), respectivamente. A dimensão dos círculos representam o valor de enriquecimento da via (*Combined Score*).

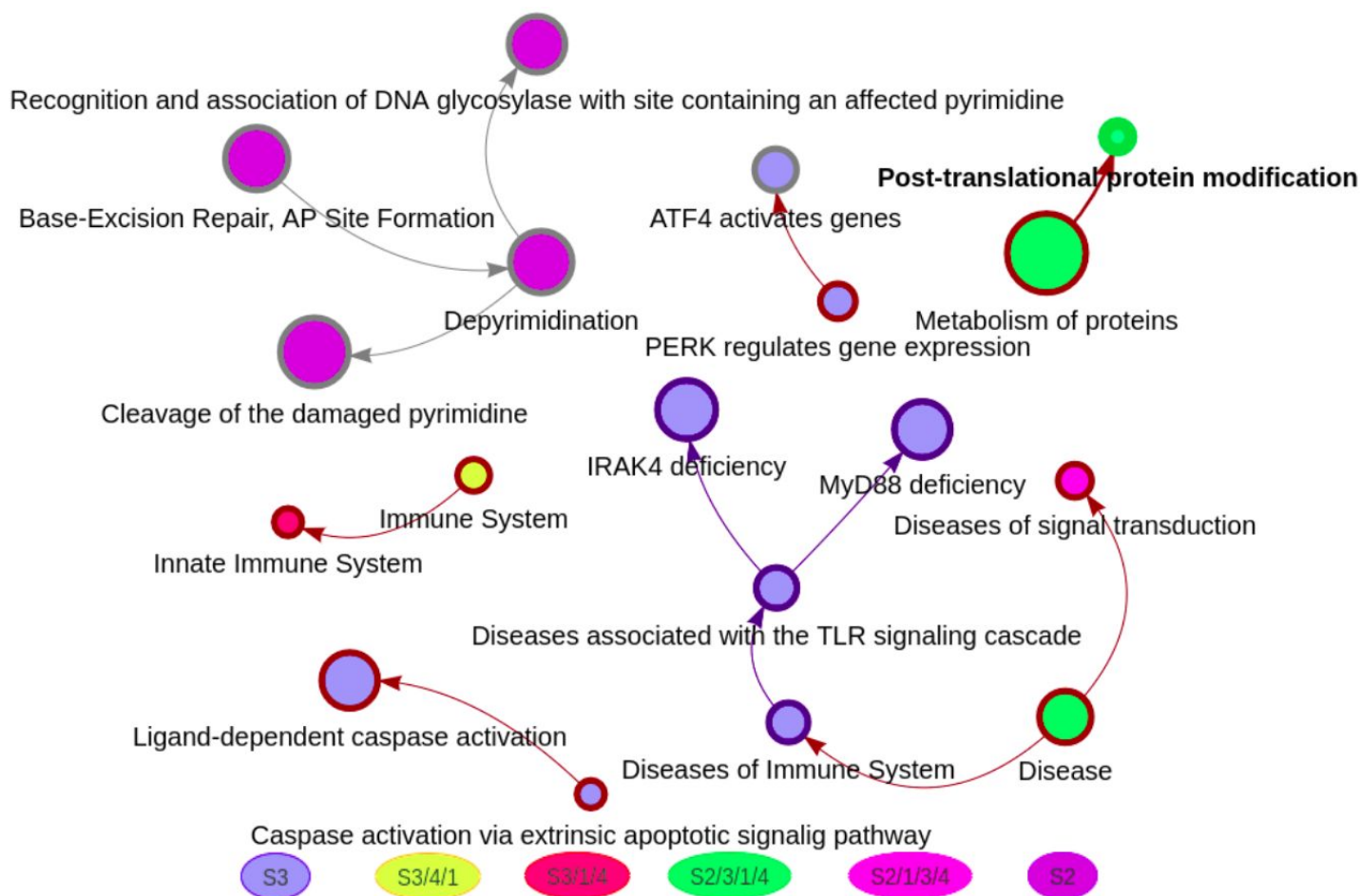


Figura 4.11 C - Enriquecimento de Vias dos Sub-módulos que constituem o Módulo M2. A cor dos círculos representam o sub-módulo no qual uma dada via foi enriquecida, indicado pela legenda inferior, e a cor da borda destes círculos representam a correlação entre o perfil de expressão mediano da via do Reactome com a idade, sendo as cores cinzas, vermelhas e azuis representativas de correlações positivas, negativas e não-correlacionadas ($|\text{Rho}| < 0.35$), respectivamente. A dimensão dos círculos representam o valor de enriquecimento da via (*Combined Score*).

5. DISCUSSÃO

Muito se descobriu nas últimas décadas sobre as alterações que ocorrem no sistema imune ao longo da vida, e como elas influenciam o envelhecimento saudável das pessoas. Essas alterações são agrupadas no que se denomina imunossenescência. Apesar de tal avanço, ainda existem diversas questões em aberto relacionadas à identificação de programas transcricionais específicos que sofrem alterações ao longo da vida e como a dinâmica entre esses processos podem ser responsáveis pelo remodelamento do sistema imune, levando ao declínio da função imune em idosos. Conforme explicado por (NIKOLICH-ŽUGICH, 2018), não há um entendimento compreensivo sobre quais das funções imunes em declínio são problemas primários, processos biológicos que sofrem perturbações cedo na vida, e quais delas são secundárias, ou seja, uma tentativa de retornar à homeostase que foi perdida devido aos problemas primários.

Neste contexto, este trabalho aplica métodos da biologia de sistemas no estudo do remodelamento transcricional que ocorre em sangue periférico com o aumento da idade. Com isso pretendemos aumentar nossa compreensão relacionada às alterações imunes observadas ao longo da vida e como estas podem estar relacionadas à imunossenescência. Para isso, desenvolvemos novas metodologias para identificar transcritos e processos biológicos que possam estar relacionados à imunossenescência ou envelhecimento. Fizemos o agrupamento de 1807 amostras de sangue coletadas de indivíduos saudáveis com idades entre 15 e 96 anos. Os dados de expressão oriundos de indivíduos com a mesma idade foram integrados em uma amostra representativa desta idade, criando-se um novo conjunto de dados, aqui denominado como ageCollapsed, que representa a evolução temporal nos níveis de expressão de 8348 transcritos anotados nas versões mais atuais do genoma humano (hg38).

Com essa quantidade massiva de dados, dado tanto pela alta cobertura do genoma humano quanto pela grande quantidade de amostras, foi encontrado

transcritos altamente relacionados à idade, denominados de *AgingGenes*, e processos biológicos do sistema imune alterados com a idade, representados por uma rede de co-expressão denominada *AgingNet*. Por meio do perfil de expressão dos *AgingGenes* e dos componentes da *AgingNet*, identificou-se momentos de perturbações transcricionais importantes através de análises de detecção de pontos de mudanças.

AgingGenes

Através do perfil de expressão dos *AgingGenes* (Figuras 4.8 A e B) nota-se dois momentos importantes durante a vida, na qual ocorram alterações transcricionais em todas as classes. A primeira fase se refere às idades compreendidas entre os 30-36 anos, momento em que surgem as primeiras mudanças, mesmo que sutis. Essas mudanças se intensificam continuamente e por volta dos 45-55 anos, segunda fase, todas as classes de *AgingGenes* sofrem alterações novamente em seu perfil de expressão, porém mais intensas. Também existem grandes alterações transcricionais após os 89 anos. No entanto, elas podem refletir tanto alterações biológicas autênticas que acontecem em idades avançadas, quanto artefatos técnicos causados pela baixa quantidade de amostras de indivíduos com essas idades. Por este motivo, as alterações observadas em indivíduos com idade superior a 89 anos foram desconsideradas.

Uma característica notável no comportamento dos *AgingGenes* é a aparente concordância entre o período na vida onde se apresentam as primeiras mudanças de estado, primeira fase, com a mudança abrupta na projeção de mortalidade dada pela lei de Gompertz–Makeham (ilustrada na Figura 1.1), no qual indica que a taxa de mortalidade humana aumenta exponencialmente logo após a maturação, por volta dos 30 anos. Essa taxa é composta de duas funções de mortalidade, representado fatores independentes da idade, como mortes desencadeadas por acidentes, e dependentes, como morte natural. A mudança para estado exponencial nessa taxa pode ser provocado por fatores dependentes da idade, podendo estar relacionadas à perturbações fisiológicas que acontecem precocemente durante a vida, tais como

acúmulos de danos, alterações na homeostase e infecções, levando a perturbações transcricionais concomitantes de diversos sistemas do corpo humano, inclusive no sistema imune, e desencadeando a perda de função e proteção observada em idosos.

A segunda característica notável no perfil de expressão do *AgingGenes* é a aparente inflexão da expressão de todas as classes de *AgingGenes* durante a segunda fase, entre as idades de 55 e 65 anos, e provavelmente desencadeadas pelas diversas mudanças de estados ocorridas para todas as classes de *AgingGenes* desde os 46 anos. Essa característica também foi notada em (IRIZAR et al., 2015), onde esta inflexão foi verificada por volta dos 54 anos e sugeriu-se que o período entre 49.3 e 55.6 anos possam ser críticos durante a vida.

De forma geral, as primeiras alterações a surgirem entre os *AgingGenes* durante a primeira fase estão relacionados principalmente à problemas nas vias de transdução de sinais, como sinalização por NOTCH1, *Tumor Necrosis Factor* (TNF), PI3K/Akt e fatores de crescimento. A primeira via a ser alterada (*neg*) está relacionada à sinalização por NOTCH1, que regula a expressão de genes com papéis importantes no desenvolvimento de linfócitos T e está envolvido na função supressora de células regulatórias, onde a superexpressão de seus ligantes podem induzir células T regulatórias (Tregs) a exercerem suas funções supressivas (NAKAMURA; KITANI; STROBER, 2001)). Estas funções supressivas são mediadas por TGF- β , no qual são secretadas e apresentadas na membrana, inibindo a função de APCs, o principal foco da supressão mediada por Treg ((DU et al., 2006; GREGG et al., 2004; HUBER et al., 2004)). Além disso, TGF- β pode induzir a diferenciação de células CD4+ *naïve* em Tregs, facilitando a expansão de populações periféricas de Treg ((DU et al., 2006; GREGG et al., 2004; HUBER et al., 2004)). A classe *neg* também está enriquecida para vias de entrada na fase estacionária (S) do ciclo celular, que, com sua tendência negativa com a idade, indica um aumento de replicação celular. Tal fato é corroborado pela classe *micC* que possui uma tendência positiva com a idade e grandes alterações por volta dos 60 anos. Ela está relacionada à proliferação celular através de vias de mitose, sinalização por fatores de crescimento epitelial (EGFR) e fibroblasto (FGFR) e,

juntamente com a classe *pos*, também para sinalização de PI3K/Akt dependente de CD28 e sinalização por PI3K aberrantes em câncer. Akt regula o crescimento celular, contribuindo para a proliferação celular através da fosforilação dos inibidores p21 e p27 e sua cascata de sinalização é ativada por receptores de células B, T, citocinas, entre outros, que induzem a produção de PIP3 pela PI3K. Desequilíbrios nas vias de fatores de crescimento e sinalização por PI3k/Akt podem levar ao desenvolvimento de doenças autoimunes (PATEL; MOHAN, 2005).

Perturbações nas vias de sinalização por TNF, enriquecidas na classe *micB*, também podem levar ao aumento da proliferação celular e angiogênese. Essa classe possui tendências mais sutis de queda na expressão durante a vida, e está relacionada à vias de sinalização tanto por TNFR1 quanto por TNFR2 envolvido na sinalização de citocinas através da ativação de NF- κ B pela via não-canônica. O TNF exerce papéis regulatórios sobre células imunes, desencadeando a transcrição de genes responsáveis por inflamação, proliferação, diferenciação e apoptose. Para combater uma infecção, por exemplo, o TNF facilita a proliferação de clones de células imunológicas, estimula a diferenciação e recrutamento de células imunes *naïve* e, assim que a infecção é eliminada, orquestra a destruição de clones imunológicos supérfluos para reduzir a inflamação e danos teciduais. No entanto, durante o desenvolvimento de autoimunidade, progenitores anormais de células T, e outros tipos celulares, proliferam e começam a maturação no timo e uma variedade de defeitos na via de sinalização TNFR2/NF κ B são encontrados nessas doenças (FAUSTMAN; DAVIS, 2013). Esses defeitos incluem polimorfismos e super-regulação no gene TNFR2, assim como diminuição dos receptores TNFR2 (FAUSTMAN; DAVIS, 2013). A sinalização por TNFR2 parece oferecer papéis protetivos em diversas doenças, incluindo doenças autoimunes, doenças do coração, doenças neurodegenerativas e desmielinizantes, e doenças infecciosas (FAUSTMAN; DAVIS, 2013). A ativação de NF κ B pela via clássica é revertida para a não-canônica em situações onde TRAF2/3 ou IAP são bloqueados. Em conjunto, o TNF e TRAF3 são importantes em células T ativadas, sendo que o NF- κ B induzido por TNF é importante na inflamação, uma vez

que NF- κ B é um transativador global de inúmeras citocinas pró-inflamatórias, quimiocinas e seus receptores, e um regulador crítico da ativação e função de leucócitos.

A classe *micA* tem o perfil que sofre quedas mais abruptas durante a vida, sendo enriquecida para vias de manutenção de telômeros, como empacotamento das regiões terminais dos telômeros, e principalmente enriquecida para via de indução da senescência por estresses relacionados aos telômeros e danos ao DNA, podendo ser um reflexo à grande proliferação celular supracitada.

O enriquecimento geral dos *AgingGenes* levam à crer que há uma desregulação nos processos de transdução de sinais que começam logo após os 30 anos e se intensificam por volta dos 55 anos, relacionados à sinalização de linfócitos T, especialmente na regulação de da atividade de *Treg*. Isso pode levar à um aumento da proliferação celular de células danificadas e auto-reativas e indução de células senescentes devido ao mau funcionamento da manutenção de telômeros. A má sinalização de Tregs pode levar à um acúmulo de linfócitos e outras células imunológicas supérfluas no corpo, podendo desencadear doenças autoimunes e aumento de inflamação crônica, também pelo acúmulo de células senescentes, contribuindo para o *inflammaging*.

Análise de Co-Expressão: AgingNet

O segundo objetivo deste trabalho consiste em identificar programas transcricionais existentes ao longo da vida e, principalmente, entender como seu comportamento pode estar relacionado à imunossenescência. Foram identificados dois grande módulos constituídos de transcritos altamente co-expressos ao longo da vida, sendo que cada um dos módulos, M1 e M2, possuem 4 sub-grupos de transcritos com perfis de expressão distintos.

Da mesma forma que os *AgingGenes*, o perfil de ambos os módulos são marcados por duas grandes fases, onde ocorrem as primeiras alterações em seus

comportamentos por volta dos 30 anos, seguidas com perturbações mais intensas por volta dos 50 anos, correspondendo à fase de inflexão da expressão supra-mencionada.

Módulo M1

O módulo M1 possui basicamente duas grandes estruturas de vias, a primeira relacionada à expressão gênica e regulação de TP53 e a segunda relacionada à vias de sinalização do sistema imune, tanto do sistema adaptativo (como vias de sinalização à jusante de receptores de células B e T) quanto do sistema inato, representado pelas vias de sinalização de citocinas e cascatas de *Toll-Like Receptors* (TLRs). De modo geral, as vias de expressão gênica possuem uma correlação negativa com a idade, principalmente as de transcrição pela RNA Polimerase II (RNA-Pol II). Vias de transcrição por RNA polimerase I (RNA-Pol I) também foi enriquecidas, no entanto não apresentou nenhuma correlação com a idade. Apesar da via de expressão gênica possuir uma correlação negativa, a expressão das vias de regulação epigenética aumentam com a idade, principalmente a regulação da expressão de RNAs ribossomais. A RNA-Pol II transcreve principalmente RNAs mensageiros e/ou não-codificadores de proteínas (miRNAs, snRNAs e snoRNAs). Nessa mesma grande estrutura, também encontra-se vias enriquecidas para regulação da atividade de TP53, tanto por fosforilação quanto por acetilação, e estão respectivamente correlacionadas positivamente e negativamente com a idade. A fosforilação de p53 inibe sua interação com Mdm2 e, portanto, previne a ubiquitinação Mdm2-mediada e sua degradação (REED; QUELLE, 2014). Uma vez estabilizada, o p53 pode sofrer alterações pós-traducionais, tais como acetilação, levando à um aumento da ligação do p53 ao DNA e promotores p53-específicos (REED; QUELLE, 2014). Além disso, a via de acetilação de p53 é elevada em respostas à stress e seu nível de expressão correlaciona com a ativação e estabilização de p53 (REED; QUELLE, 2014). As possíveis consequências da modulação do p53 em uma célula é a parada do ciclo celular e reparo de DNA, senescência e apoptose, dependendo da quantidade de p53 ativado e de quais sítios foram acetilados (REED; QUELLE, 2014). O fato da atividade

de RNA Polimerase II diminuir com a idade e as vias de regulação de TP53 por fosforilação e acetilação possuírem correlações discrepantes, pode sugerir que há uma perturbação nas transdução de sinais para as vias de resposta à estresse e reparo de DNA.

Conforme mencionado anteriormente, a segunda grande estrutura de enriquecimento encontrada no módulo M1 sugere um forte enriquecimento de vias para sinalização à jusante de receptores de células B (BCRs), T (TCRs), lectina tipo-C (CLRs), Fcε e TLRs. Apesar das vias gerais do sistema imune possuírem uma correlação positiva com a idade, vias mais específicas de sinalização possuem uma correlação negativa, e estão majoritariamente relacionadas à modulação de NF-κB. Nas vias do sistema adaptativo, percebe-se que há um aumento ao longo da vida na expressão de genes relacionados à degradação no processamento de antígenos do grupo MHC classe I e correlações negativas com a idade tanto na modulação de NF-κB pelos BCRs quanto na sinalização por TCRs, o que pode sugerir uma diminuição da regulação de genes responsivos à NF-κB e aumento na proliferação de células T, através da via TCR-AKT-mTOR. Ademais, a não ativação do TCR permite um aumento da atividade de p53, descrito acima, e consequentemente a atividade de p21, resultando em parada do ciclo-celular e/ou apoptose.

Analisando as vias do sistema inato, percebe-se que as vias de modulação de NF-κB por TLRs (com exceção dos TLR3 e TLR4) possuem correlações negativas com a idade, seja por vias de vias de MyD88 e TRAF6 ou por sinalização não-canônica de Dectin-1. Todos os TLRs, com exceção do TLR3, podem iniciar a sinalização à jusante através do MyD88 ou MyD88:MAL. Nessa via, dependente de MyD88, a ativação do TLR (como TLR7 e TLR8) leva ao recrutamento de receptores IL-1 associados à família IRAK, que é seguido pela ativação de TRAF6. TRAF6 induz a TAK1 à se auto-fosforilar e, uma vez ativada, pode mediar a indução do fator de transcrição NF-κB ou MAPK, tais como JNK, p38 e ERK. Isso resulta na translocação do NF-κB ativo e MAPKs para o núcleo e a iniciação da transcrição gênica apropriada, levando à produção de muitas citocinas pró-inflamatórias e peptídeos antimicrobianos. Um outro adaptador, conhecido

como MyD88-adaptor-like (MAL), também participa da sinalização por TLR2 e 4. A ativação dos receptores TLR1:TLR2 ou TLR6:TLR2, através de MyD88 e TIRAP/MAL ativam IRAK e TRAF6, que por conseguinte aumentam a expressão de citocinas pró-inflamatórias através de NF- κ B (fase inicial). Nas vias de sinalização do sistema adaptativo, receptores das células B e T também possuem correlações lineares negativas com a idade, sugerindo que há alterações desfavoráveis na sinalização entre o sistema imune inato e adaptativo e possível alteração na capacidade de montagem de respostas imunológicas através de processos inflamatórios frente a um desafio.

Ao mesmo tempo, há um aumento na sinalização de receptores Fc ϵ (Fc ϵ RI), que em humanos controlam a ativação de células mastóides e basófilos, participando na apresentação de antígenos mediada por Imunoglobulina E (IgE). O agregamento desse receptor induz múltiplas vias de sinalização que controlam diversas respostas efectoras, incluindo a secreção de mediadores alérgicos e indução de transcrição de genes de citocinas, resultando na secreção de moléculas como IL-6, TNF- α e fatores estimuladores de colônia granulócitos-macrófagos (TURNER; KINET, 1999). Fc ϵ RI é, portanto, central para a indução e manutenção de respostas alérgicas e pode conferir proteção fisiológica em infecções parasíticas. No entanto, a sinalização por esse receptor aumenta de forma significativa com a vida, o que pode aumentar de forma exacerbada concentrações sorológicas de IL-6, uma citocina extremamente relacionada com o envelhecimento e apelidada de “Citocida dos Gerontólogos”. Existem fortes evidências que sugerem que concentrações sorológicas de IL-6 aumentam com a idade (MAGGIO et al., 2006) e também que a produção excessiva ou redução da remoção de radicais livres de oxigênio, que estimula a produção de IL-6, possam ser importante para esse aumento (MAGGIO et al., 2006). Além disso, permanece uma questão fundamental se níveis altos de IL-6 tem o propósito de resolver uma resposta inflamatória que é inapropriadamente longa ou se uma desregulação primária na produção de IL-6 é responsável pelo estado crônico pró-inflamatório, que possui um impacto negativo na qualidade da saúde.

Módulo M2

O módulo M2 possui estruturas de enriquecimento menores do que o M1, por isso permitiu-se a presença de vias de sinalização com poucos componentes enriquecidos. Mas, da mesma forma que o módulo M1, a expressão mediana das vias gerais do sistema imune (inato e adaptativo) possuem uma correlação positiva com a idade. Nesse módulo, também foi detectado duas grandes estruturas de vias de sinalização, relacionados à doenças de transdução de sinais no sistema imune e reparo de DNA, de forma análoga aos mecanismos realçados pelos *AgingGenes*,

Apesar de não estarem correlacionadas com a idade, há diversas vias altamente enriquecidas para maquinarias de reparo de DNA, especialmente reparo por excisão de base e dipirimidinação, podendo indicar que mecanismos de reparo possam estar sendo ativados em diversos momento durante vida, não sendo necessariamente algo progressivo, detectável por correlações lineares.

A via de doenças de transdução de sinais possui forte correlação positiva com a idade, ao mesmo tempo que vias de deficiência em MyD88 e TNFR1 em TLR2 e/ou TLR4 diminuem ao longo da vida, compactuando com os achados no módulo M1 sobre o TLR2. Tal fato sugere que há uma possível diminuição na transdução de sinais desses dois TLRs devido à possíveis problemas de acúmulo de mutações em genes relacionados à produção e/ou modulação de MyD88, IRAK6 e TRAF6, membros do eixo de sinalização desse dois TLRs, uma vez que vias de doenças relacionadas à sinalização através de MyD88 e IRAK6 foram enriquecidas e possuem correlação negativa com a idade. Da mesma forma, ambos os genes TICAM1 e TRAF6, que se associam na cascata de ativação de NF- κ B possuem uma expressão com tendência negativa ao longo da vida. Outra opção também seria que possíveis mutações/repressões ao gene Ly96, codificador da proteína MD2 e associada ao receptor TLR4, levem à perturbações na perfeita configuração de TLR4, uma vez que seu perfil de expressão tende a diminuir com a idade, com valor de correlação $Rho = -0.82$.

Ademais, vias enriquecidas de produção de ceramida mediada por TNFR1 possuem correlações negativas com a idade. Essa via intercede tanto na sobrevivência celular através do NF-kB quanto na morte celular, através da ativação de caspases. Uma vez que o gene TRAF2, participante da modulação de NF-kB no eixo TNFR1-IKK, possui uma correlação negativa com a idade e a via de ativação de caspases por ligantes foi enriquecida, com uma correlação positiva com a idade, pode-se pressupor um possível aumento na modulação de morte celular programada.

Por fim, também houveram vias enriquecidas de regulação da expressão gênica por PERK enriquecida e com correlação positiva. O PERK, codificada pelo gene EIF2AK3 ($Rho = 0.225$), é uma proteína localizada na membrana do retículo endoplasmático (ER) e mantida inativa através da ligação da porção situada no lúmen com uma chaperona do ER denominada BiP. Essa chaperona também se liga à proteínas desdobradas e, portanto, se dissocia do PERK quando proteínas desdobradas se acumulam no ER. Quando ativa, o PERK pode induzir a parada da tradução, fazendo com que a tradução de proteínas direcionadas para o ER sejam sub-reguladas, desencadeando a depleção de Ciclina D1, que por sua vez desencadeia a parada do ciclo celular na fase G1. Baseando no perfil de expressão do submódulo 3 (M2S3), no qual está relacionado a tal via, pode-se sugerir que desequilíbrios na proteostase, um dos *hallmarks* do envelhecimento, possam acontecer desde cedo na vida, por volta dos 35 anos, e tendem a se intensificar continuamente até em idades mais avançadas, por volta dos 80 anos, onde ficam extremamente perturbadas.

6. CONCLUSÃO

As metodologias desenvolvidas neste estudo permitiram encontrar perturbações a nível transcricional que condizem com as características morfológicas já descritas na literatura e, principalmente, identificar novos transcritos e processos biológicos nunca antes relacionados à imunossenescência.

Ademais, descobriu-se que existem pelo menos dois grandes momentos durante a vida no qual o sistema imune sofre alterações que podem levar ao estabelecimento da imunossenescência. A primeira delas reside logo no início da vida, por volta dos 30 anos, onde se nota perturbações relacionadas à transdução de sinais envolvidas principalmente na sinalização de linfócitos, controle de células T regulatórias e proliferação celular. Alterações continuam a acontecer até meados dos 55 anos, onde se intensificam, ocasionando uma ruptura clara no perfil de expressão dos componentes aqui estudados, denominada de ponto de inflexão transcricional. Essas alterações podem ser um dos responsáveis diretos pelo declínio fisiológico observado em idosos.

Tais metodologias podem ser adaptadas para dados de transcriptômica de maior qualidade, como RNA-Seq, no qual seria possível identificar novos transcritos de forma mais precisa, surgimento de isoformas correlacionadas com a idade e estudar a dinâmica de subpopulações celulares do sistema imunológico ao longo da vida. Mais ainda, também podem ser aplicada para outros níveis de informações biológicas, tais como metabolômica e proteômica, para melhor entender as relações entre os componentes do sistema imune. Com isso, espera-se que os achados deste estudo tenham o potencial de auxiliar a compreensão dos mecanismos envolvidos na imunossenescência e subjacentes ao envelhecimento, e futuramente tornar possível a criação de intervenções capazes de regular de forma benéfica o remodelamento do sistema imune durante a vida, permitindo a extensão da vida com qualidade.

8. REFERÊNCIAS

ACOSTA, J. C. et al. Control of senescence by CXCR2 and its ligands. **Cell Cycle**, v. 7, n. 19, p. 2956–2959, 13 Oct. 2008.

ALBANESE, D. et al. A practical tool for maximal information coefficient analysis. **GigaScience**, v. 7, n. 4, p. 1–8, 1 Apr. 2018.

ANTONACI, S. et al. Non-specific immunity in aging: deficiency of monocyte and polymorphonuclear cell-mediated functions. **Mechanisms of Ageing and Development**, v. 24, n. 3, p. 367–375, Mar. 1984.

BAKER, D. J. et al. Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. **Nature**, v. 530, n. 7589, p. 184–189, 11 Feb. 2016.

BUFFA, S. et al. A novel B cell population revealed by a CD38/CD24 gating strategy: CD38(-)CD24 (-) B cells in centenarian offspring and elderly people. **Age**, v. 35, n. 5, p. 2009–2024, Oct. 2013.

Carvalho-Bürguer, M - Análise Transcricional de RNAs Não Codificadores Longos em Pacientes com Dengue. Dissertação (Mestrado em Bioinformática) - Interunidades em Bioinformática, Universidade de São Paulo. São Paulo: 2017, 86p.

CAMPELLO, R. J. G. B.; MOULAVI, D.; SANDER, J. Density-Based Clustering Based on Hierarchical Density Estimates. In: PEI, J. et al. (Eds.). . **Advances in Knowledge Discovery and Data Mining**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. v. 7819p. 160–172.

CAMPISI, J.; D'ADDA DI FAGAGNA, F. Cellular senescence: when bad things happen to good cells. **Nature Reviews. Molecular Cell Biology**, v. 8, n. 9, p. 729–740, Sep. 2007.

COLONNA-ROMANO, G. et al. B cell immunosenescence in the elderly and in centenarians. **Rejuvenation Research**, v. 11, n. 2, p. 433–439, Apr. 2008.

COPPÉ, J.-P. et al. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. **PLoS Biology**, v. 6, n. 12, p. 2853–2868, 2 Dec. 2008.

COPPÉ, J.-P. et al. The senescence-associated secretory phenotype: the dark side of tumor suppression. **Annual review of pathology**, v. 5, p. 99–118, 2010.

CRAN - Package TTR. Disponível em:

<<https://cran.r-project.org/web/packages/TTR/index.html>>. Acesso em: 1 nov. 2018.

CRIEL, J.; TSIPORKOVA, E. Gene Time Expression Warper: a tool for alignment, template matching and visualization of gene expression time series. **Bioinformatics**, v. 22, n. 2, p. 251–252, 15 Jan. 2006.

CROFT, D. et al. Reactome: a database of reactions, pathways and biological processes. **Nucleic Acids Research**, v. 39, n. Database issue, p. D691-7, Jan. 2011.

DAVALOS, A. R. et al. Senescent cells as a source of inflammatory factors for tumor progression. **Cancer metastasis reviews**, v. 29, n. 2, p. 273–283, Jun. 2010.

DAVIS, S.; MELTZER, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. **Bioinformatics**, v. 23, n. 14, p. 1846–1847, 15 Jul. 2007.

DE MAGALHÃES, J. P.; TOUSSAINT, O. GenAge: a genomic and proteomic network map of human ageing. **FEBS Letters**, v. 571, n. 1–3, p. 243–247, 30 Jul. 2004.

Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms [R package dbscan version 1.1-2]. Disponível em: <https://cran.r-project.org/web/packages/dbscan/index.html>. Acesso em: 1 nov. 2018.

DU, W. et al. TGF-beta signaling is required for the function of insulin-reactive T regulatory cells. **The Journal of Clinical Investigation**, v. 116, n. 5, p. 1360–1370, May 2006.

EDGAR, R.; DOMRACHEV, M.; LASH, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. **Nucleic Acids Research**, v. 30, n. 1, p. 207–210, 1 Jan. 2002.

ELZI, D. J. et al. Wnt antagonist SFRP1 functions as a secreted mediator of senescence. **Molecular and Cellular Biology**, v. 32, n. 21, p. 4388–4399, Nov. 2012.

FAUSTMAN, D. L.; DAVIS, M. TNF receptor 2 and disease: autoimmunity and regenerative medicine. **Frontiers in immunology**, v. 4, p. 478, 23 Dec. 2013.

FLATT, T. A new definition of aging? **Frontiers in genetics**, v. 3, p. 148, 23 Aug. 2012.

FREUND, A. et al. Inflammatory networks during cellular senescence: causes and consequences. **Trends in Molecular Medicine**, v. 16, n. 5, p. 238–246, 3 May 2010.

FULOP, T. et al. Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? **Frontiers in immunology**, v. 8, p. 1960, 2017.

GINALDI, L. et al. Immunosenescence and infectious diseases. **Microbes and Infection**, v. 3, n. 10, p. 851–857, Aug. 2001.

GORGOLIS, V. G.; HALAZONETIS, T. D. Oncogene-induced senescence: the bright and dark side of the response. **Current Opinion in Cell Biology**, v. 22, n. 6, p. 816–827, Dec. 2010.

GREGG, R. K. et al. A sudden decline in active membrane-bound TGF-beta impairs both T regulatory cell function and protection against autoimmune diabetes. **Journal of Immunology**, v. 173, n. 12, p. 7308–7316, 15 Dec. 2004.

HARMAN, D. The free radical theory of aging: effect of age on serum copper levels. **Journal of Gerontology**, v. 20, p. 151–153, Apr. 1965.

HARROW, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. **Genome Research**, v. 22, n. 9, p. 1760–1774, Sep. 2012.

HAYNES, L. et al. CD4 T cell memory derived from young naive cells functions well into old age, but memory generated from aged naive cells functions poorly. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 25, p. 15053–15058, 9 Dec. 2003.

Health and Economic Costs of Chronic Disease | About Chronic Disease | Chronic Disease Prevention and Health Promotion | CDC. Disponível em: <<https://www.cdc.gov/chronicdisease/about/costs/index.htm>>. Acesso em: 2 oct. 2018.

HUBER, S. et al. Cutting edge: TGF-beta signaling is required for the in vivo expansion and immunosuppressive capacity of regulatory CD4+CD25+ T cells. **Journal of Immunology**, v. 173, n. 11, p. 6526–6531, 1 Dec. 2004.

IBGE :: Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://ww2.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2013/default.shtm>. Acesso em: 2 oct. 2018.

IRIZAR, H. et al. Age gene expression and coexpression progressive signatures in peripheral blood leukocytes. **Experimental Gerontology**, v. 72, p. 50–56, Dec. 2015.

ITAKURA, F. Minimum prediction residual principle applied to speech recognition. **IEEE transactions on acoustics, speech, and signal processing**, v. 23, n. 1, p. 67–72, Feb. 1975.

IYER, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. **Nature Genetics**, v. 47, n. 3, p. 199–208, Mar. 2015.

JOHNSON, W. E.; LI, C.; RABINOVIC, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. **Biostatistics**, v. 8, n. 1, p. 118–127, Jan. 2007.

KITANO, H. Systems biology: a brief overview. **Science**, v. 295, n. 5560, p. 1662–1664, 1 Mar. 2002.

KOLESNIKOV, N. et al. ArrayExpress update--simplifying data submissions. **Nucleic Acids Research**, v. 43, n. Database issue, p. D1113-6, Jan. 2015.

KUILMAN, T. et al. The essence of senescence. **Genes & Development**, v. 24, n. 22, p. 2463–2479, 15 Nov. 2010.

KULESHOV, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. **Nucleic Acids Research**, v. 44, n. W1, p. W90-7, 8 Jul. 2016.

- LAGES, C. S. et al. Functional regulatory T cells accumulate in aged hosts and promote chronic infectious disease reactivation. **Journal of Immunology**, v. 181, n. 3, p. 1835–1848, 1 Aug. 2008.
- LÓPEZ-OTÍN, C. et al. The hallmarks of aging. **Cell**, v. 153, n. 6, p. 1194–1217, 6 Jun. 2013.
- MAGGIO, M. et al. Interleukin-6 in aging and chronic disease: a magnificent pathway. **The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences**, v. 61, n. 6, p. 575–584, Jun. 2006.
- MANTON, K. G.; VAUPEL, J. W. Survival after the age of 80 in the United States, Sweden, France, England, and Japan. **The New England Journal of Medicine**, v. 333, n. 18, p. 1232–1235, 2 Nov. 1995.
- MANTOVANI, A.; ALLAVENA, P.; SICA, A. Tumour-associated macrophages as a prototypic type II polarised phagocyte population: role in tumour progression. **European Journal of Cancer**, v. 40, n. 11, p. 1660–1667, Jul. 2004.
- MATTHEWS, G. J.; FOULKES, A. S. MixMAP : AnR Package for Mixed Modeling of Meta-Analysis Values in Genetic Association Studies. **Journal of statistical software**, v. 66, n. 3, p. 1–20, 2015.
- MUSZKAT, M. et al. Local and systemic immune response in nursing-home elderly following intranasal or intramuscular immunization with inactivated influenza vaccine. **Vaccine**, v. 21, n. 11–12, p. 1180–1186, Mar. 2003.
- NAKAMURA, K.; KITANI, A.; STROBER, W. Cell contact-dependent immunosuppression by CD4(+)CD25(+) regulatory T cells is mediated by cell surface-bound transforming growth factor beta. **The Journal of Experimental Medicine**, v. 194, n. 5, p. 629–644, 3 Sep. 2001.
- NIKOLICH-ŽUGICH, J. The twilight of immunity: emerging concepts in aging of the immune system. **Nature Immunology**, v. 19, n. 1, p. 10–19, Jan. 2018.
- NOMELLINI, V.; GOMEZ, C. R.; KOVACS, E. J. Aging and impairment of innate immunity. **Contributions to microbiology**, v. 15, p. 188–205, 2008.
- PATEL, R. K.; MOHAN, C. PI3K/AKT signaling and systemic autoimmunity. **Immunologic Research**, v. 31, n. 1, p. 47–55, 2005.
- PICCOLO, S. R. et al. Multiplatform single-sample estimates of transcriptional activation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 44, p. 17778–17783, 29 Oct. 2013.
- PRAKASH, S. et al. Impaired secretion of interferons by dendritic cells from aged subjects to influenza : role of histone modifications. **Age**, v. 35, n. 5, p. 1785–1797, Oct.

2013.

PRITZ, T. et al. Plasma cell numbers decrease in bone marrow of old patients. **European Journal of Immunology**, v. 45, n. 3, p. 738–746, Mar. 2015.

REED, S. M.; QUELLE, D. E. p53 acetylation: regulation and consequences. **Cancers**, v. 7, n. 1, p. 30–69, 23 Dec. 2014.

REID, A. H.; TAUBENBERGER, J. K. The origin of the 1918 pandemic influenza virus: a continuing enigma. **The Journal of General Virology**, v. 84, n. Pt 9, p. 2285–2292, Sep. 2003.

RESHEF, D. N. et al. Detecting novel associations in large data sets. **Science**, v. 334, n. 6062, p. 1518–1524, 16 Dec. 2011.

RICCADONNA, S. et al. DTW-MIC Coexpression Networks from Time-Course Data. **Plos One**, v. 11, n. 3, p. e0152648, 31 Mar. 2016.

RODIER, F.; CAMPISI, J. Four faces of cellular senescence. **The Journal of Cell Biology**, v. 192, n. 4, p. 547–556, 21 Feb. 2011.

ROSENBERG, C. et al. Age is an important determinant in humoral and T cell responses to immunization with hepatitis B surface antigen. **Human vaccines & immunotherapeutics**, v. 9, n. 7, p. 1466–1476, Jul. 2013.

SCHWAIGER, S. et al. IL-4-producing CD8⁺ T cells with a CD62L⁺⁺(bright) phenotype accumulate in a subgroup of older adults and are associated with the maintenance of intact humoral immunity in old age. **Journal of Immunology**, v. 170, n. 1, p. 613–619, 1 Jan. 2003.

SIMON, A. K.; HOLLANDER, G. A.; MCMICHAEL, A. Evolution of the immune system in humans from infancy to old age. **Proceedings. Biological Sciences / the Royal Society**, v. 282, n. 1821, p. 20143085, 22 Dec. 2015.

STAHL, E. C.; BROWN, B. N. Cell therapy strategies to combat immunosenescence. **Organogenesis**, v. 11, n. 4, p. 159–172, 20 Nov. 2015.

SZKLARCZYK, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. **Nucleic Acids Research**, v. 43, n. Database issue, p. D447–52, Jan. 2015.

TURNER, H.; KINET, J.-P. Signalling through the high-affinity IgE receptor FcεRI. **Nature**, v. 402, n. 6760supp, p. 24–30, 25 Nov. 1999.

UN. **World population ageing 2015 highlights**. [s.l.] United Nations, 2016.

VALLEJO, A. N. CD28 extinction in human T cells: altered functions and the program of T-cell senescence. **Immunological Reviews**, v. 205, p. 158–169, Jun. 2005.

VAN DEURSEN, J. M. The role of senescent cells in ageing. **Nature**, v. 509, n. 7501, p.

439–446, 22 May 2014.

VENTURA, M. T. Evaluation of IgA1-IgA2 levels in serum and saliva of young and elderly people. **Allergologia et Immunopathologia**, v. 19, n. 5, p. 183–185, Oct. 1991.

visNetwork: Network Visualization using “vis.js” Library. Disponível em:

<<http://datastorm-open.github.io/visNetwork/>>. Acesso em: 1 nov. 2018.

VOLDERS, P.-J. et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. **Nucleic Acids Research**, v. 41, n. Database issue, p. D246-51, Jan. 2013.

WEISKOPF, D.; WEINBERGER, B.; GRUBECK-LOEBENSTEIN, B. The aging of the immune system. **Transplant International**, v. 22, n. 11, p. 1041–1050, Nov. 2009.

WEKSLER, M. E.; SZABO, P. The effect of age on the B-cell repertoire. **Journal of Clinical Immunology**, v. 20, n. 4, p. 240–249, Jul. 2000.

WIKBY, A. et al. An immune risk phenotype, cognitive impairment, and survival in very late life: impact of allostatic load in Swedish octogenarian and nonagenarian humans.

The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, v. 60, n. 5, p. 556–565, May 2005.

XUE, W. et al. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. **Nature**, v. 445, n. 7128, p. 656–660, 8 Feb. 2007.

ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. **Statistical Applications in Genetics and Molecular Biology**, v. 4, p. Article17, 12 Aug. 2005.

ZHAO, Y. et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. **Nucleic Acids Research**, v. 44, n. D1, p. D203-8, 4 Jan. 2016.

9. ANEXOS

9.1 Anexo A - Tabelas Suplementares

Todas as tabelas suplementares estão disponíveis através do link
<https://drive.google.com/open?id=1LweUTN9Fe7n4mbLCCeid3wPq7ySGtj7T>

Inflammation induced by influenza virus impairs human innate immune control of pneumococcus

Simon P. Jochems^{1,8*}, Fernando Marcon^{2,8}, Beatriz F. Carniel¹, Mark Holloway¹, Elena Mitsi¹, Emma Smith¹, Jenna F. Gritzfeld¹, Carla Solórzano¹, Jesús Reiné¹, Sherin Pojar¹, Elissavet Nikolau¹, Esther L. German¹, Angie Hyder-Wright^{1,3}, Helen Hill^{1,3}, Caz Hales^{1,3}, Wouter A. A. de Steenhuijsen Pijters^{4,5,6}, Debby Bogaert^{4,5,6}, Hugh Adler¹, Seher Zaidi¹, Victoria Connor^{1,3}, Stephen B. Gordon^{1,7}, Jamie Ryland¹, Helder I. Nakaya^{1,2*} and Daniela M. Ferreira^{1*}

Colonization of the upper respiratory tract by pneumococcus is important both as a determinant of disease and for transmission into the population. The immunological mechanisms that contain pneumococcus during colonization are well studied in mice but remain unclear in humans. Loss of this control of pneumococcus following infection with influenza virus is associated with secondary bacterial pneumonia. We used a human challenge model with type 6B pneumococcus to show that acquisition of pneumococcus induced early degranulation of resident neutrophils and recruitment of monocytes to the nose. Monocyte function was associated with the clearance of pneumococcus. Prior nasal infection with live attenuated influenza virus induced inflammation, impaired innate immune function and altered genome-wide nasal gene responses to the carriage of pneumococcus. Levels of the cytokine CXCL10, promoted by viral infection, at the time pneumococcus was encountered were positively associated with bacterial load.

Pneumonia is a major global health problem; it kills more children under 5 years of age than any other disease¹. The burden of disease is aggravated by old age, chronic lung disease, immunosuppression and viral co-infection. Secondary pneumonia following pandemic and seasonal influenza virus infection is a significant cause of mortality worldwide².

Nasopharyngeal colonization by *Streptococcus pneumoniae* (pneumococcus) is common, with 40–95% of infants and 10–25% of adults colonized at any given time³. Such pneumococcal carriage is important as the prerequisite of infection⁴, the primary reservoir for transmission⁵ and the predominant source of immunizing exposure and immunological boosting in both children and adults^{6,7}.

Immune dysregulation caused by infection with respiratory viruses such as influenza leads to increased carriage load⁸. Increased carriage load has been associated with pneumonia incidence and severity, as well as with within-household *S. pneumoniae* transmission^{9–11}. The mechanisms and markers associated with this pathogen synergy have been difficult to study in human subjects due to the rapidly progressing nature of the disease.

One safe way to simulate influenza infection in the nose is using live attenuated influenza vaccine (LAIV), consisting of cold-adapted influenza viruses. LAIV has been shown to affect subsequent susceptibility to *S. pneumoniae* and to lead to increased carriage load in mouse models of infection and in vaccinated children^{12,13}. Furthermore, LAIV administration prior to *S. pneumoniae* challenge led to a 50% increase in *S. pneumoniae* acquisition, as detected by molecular methods, as well as a tenfold increase in nasopharyngeal bacterial load¹⁴.

In mouse models of pneumococcal carriage, recruitment of neutrophils and monocytes to the nasopharynx dependent on the T_H17 subset of helper T cells mediates immunological control and clearance^{15–17}. Influenza virus infection promotes type I interferons, which interfere with recruitment of these phagocytes, although interferon- γ (IFN- γ) is postulated to impair phagocytosis by macrophages through downregulation of the scavenger receptor MARCO (macrophage receptor with collagenous structure)^{18–20}. However, the precise immune mechanisms and gene regulators involved in the control and clearance of pneumococcal carriage in humans have not been revealed²¹. Moreover, how these mechanisms are altered during human influenza virus infection remains largely unknown.

Systems-biology approaches have allowed the identification of immune mechanisms associated with protection from infectious diseases and with robust immune responses during vaccination^{22–28}. Here, we applied systems biology to nasal samples collected in the setting of human challenge with LAIV and *S. pneumoniae*, to emulate nasal effects of influenza infection on *S. pneumoniae* carriage. We identified the key cellular mechanisms that control newly acquired pneumococcal carriage in humans, and how they are disrupted following nasal influenza infection.

Results

LAIV-induced inflammation leads to increased pneumococcal carriage load and acquisition. In a double-blinded controlled randomized clinical trial, we administered LAIV ($n=55$) 3 days prior to *S. pneumoniae* inoculation (day 0). To verify the requisite topical

¹Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK. ²Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil. ³Royal Liverpool and Broadgreen University Hospital, Liverpool, UK. ⁴Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK. ⁵Department of Paediatric Immunology and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands. ⁶Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands.

⁷Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi. ⁸These authors contributed equally: Simon P. Jochems and Fernando Marcon *e-mail: simon.jochems@lstm.ac.uk; hnaikaya@usp.br; daniela.ferreira@lstm.ac.uk

Calcium/calmodulin-dependent kinase kinase 2 regulates hematopoietic stem and progenitor cell regeneration

Luigi Racioppi^{1,2,4}, William Lento^{1,4}, Wei Huang¹, Stephanie Arvai¹, Phuong L Doan¹, Jeffrey R Harris¹, Fernando Marcon³, Helder I Nakaya³, Yaping Liu¹ and Nelson Chao¹

Hematopoietic stem and progenitor cells (HSPCs) are predominantly quiescent in adults, but proliferate in response to bone marrow (BM) injury. Here, we show that deletion of Ca^{2+} /calmodulin (CaM)-dependent protein kinase kinase 2 (CaMKK2) promotes HSPC regeneration and hematopoietic recovery following radiation injury. Using Camkk2-enhanced green fluorescent protein (EGFP) reporter mice, we found that Camkk2 expression is developmentally regulated in HSPC. Deletion of Camkk2 in HSPC results in a significant downregulation of genes affiliated with the quiescent signature. Accordingly, HSPC from Camkk2 null mice have a high proliferative capability when stimulated *in vitro* in the presence of BM-derived endothelial cells. In addition, Camkk2 null mice are more resistant to radiation injury and show accelerated hematopoietic recovery, enhanced HSPC regeneration and ultimately a prolonged survival following sublethal or lethal total body irradiation. Mechanistically, we propose that CaMKK2 regulates the HSPC response to hematopoietic damage by coupling radiation signaling to activation of the anti-proliferative AMP-activated protein kinase. Finally, we demonstrated that systemic administration of the small molecule CaMKK2 inhibitor, STO-609, to irradiated mice enhanced HSPC recovery and improved survival. These findings identify CaMKK2 as an important regulator of HSPC regeneration and demonstrate CaMKK2 inhibition is a novel approach to promoting hematopoietic recovery after BM injury. *Cell Death and Disease* (2017) 8, e3076; doi:10.1038/cddis.2017.474; published online 5 October 2017

Hematopoietic stem and progenitor cells (HSPCs) reside in specialized bone marrow (BM) niches that provide signals to ensure blood production and maintain the long-term hematopoietic stem cell (LT-HSC) pool. Extensive studies of the niche have identified several cell types such as osteoblasts,¹ endothelial cells,² osteomacs,³ regulatory T cells⁴ and sympathetic neurons⁵ as contributors of the physiologic microenvironment.^{6,7} These cells engage HSPC through both physical contacts and soluble paracrine signaling molecules including CXC chemokine ligand 12 (CXCL12), stem cell factor (SCF), non-canonical and canonical Wnt ligands, and epidermal growth factor^{8,9} to control niche retention and self-renewal. Although these molecules may trigger calcium transients, the role of calcium-dependent cascades in the mechanism regulating HSPC regeneration has not been elucidated.^{10–13}

Calmodulin (CaM) is the primary intercellular calcium sensor and binding to free cytosolic Ca^{2+} causes conformational changes that facilitate its interaction with the multi-functional Ser/Thr kinases Ca^{2+} /CaM-dependent protein kinase I, IV (CaMKI and CaMKIV, respectively) and Ca^{2+} /CaM-dependent protein kinase kinase 1 (CaMKK1) and CaMKK2 to activate Ca^{2+} /CaM-dependent signaling cascades.^{14,15} CaMKK2 activation permits phosphorylation of CaMKI, CaMKIV and the adenosine monophosphate

activated protein kinase (AMPK).¹⁶ The expression of CaMKK2 is relatively cell type restricted and outside the brain it is found in osteoblasts,¹⁷ macrophages¹⁸ and myeloid progenitors.¹⁹

Here, we demonstrate CaMKK2 functions as a critical kinase that regulates the regeneration of HSPC. CaMKK2 deficiency downregulates genes affiliated with stem cell quiescence and causes a HSPC hyper-proliferative phenotype *in vitro* and accelerates hematopoietic recovery following radiation injury *in vivo*. Mechanistically, we demonstrate CaMKK2 is required to link radiation injury with AMPK activation and p53 accumulation. Importantly, the transient inhibition of CaMKK2 with the small molecule CaMKK2 kinase inhibitor STO-609²⁰ improves survival and hematopoietic regeneration.

Results

CaMKK2 expression is enriched in HSPC *in vivo*. We used CaMKK2-enhanced green fluorescent protein (EGFP) mice to determine the location and phenotype of Camkk2-expressing cells within the BM microenvironment. The analysis of reporter bone sections by immunofluorescence revealed activity in individual single cells throughout the BM (Figure 1A). A subset of EGFP-positive cells was located

¹Division of Hematological Malignancies and Cellular Therapy, Department of Medicine, Duke University Medical Center, Durham, NC, USA; ²Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy and ³Department of Pathophysiology and Toxicology, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

*Corresponding author: L. Racioppi, Division of Hematological Malignancies and Cellular Therapy, Department of Medicine, Duke University Medical Center, 905 S. LaSalle Street, 2006 GSRB1, Durham, NC 27710, USA. Tel: +1 919 681 6236; Fax: +1 919 613 1319; E-mail: luigi.racioppi@dm.duke.edu

⁴These authors contributed equally to this work.

Received 02.7.17; revised 04.8.17; accepted 09.8.17; Edited by H-U Simon

SOFTWARE

Open Access



CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses

Pedro S. T. Russo^{1†}, Gustavo R. Ferreira^{1†}, Lucas E. Cardozo¹, Matheus C. Bürger¹, Raul Arias-Carrasco², Sandra R. Maruyama³, Thiago D. C. Hirata¹, Diógenes S. Lima¹, Fernando M. Passos¹, Kiyoshi F. Fukutani³, Melissa Lever¹, João S. Silva³, Vinicius Maracaja-Coutinho² and Helder I. Nakaya^{1*}

Abstract

Background: The analysis of modular gene co-expression networks is a well-established method commonly used for discovering the systems-level functionality of genes. In addition, these studies provide a basis for the discovery of clinically relevant molecular pathways underlying different diseases and conditions.

Results: In this paper, we present a fast and easy-to-use Bioconductor package named CEMiTool that unifies the discovery and the analysis of co-expression modules. Using the same real datasets, we demonstrate that CEMiTool outperforms existing tools, and provides unique results in a user-friendly html report with high quality graphs. Among its features, our tool evaluates whether modules contain genes that are over-represented by specific pathways or that are altered in a specific sample group, as well as it integrates transcriptomic data with interactome information, identifying the potential hubs on each network. We successfully applied CEMiTool to over 1000 transcriptome datasets, and to a new RNA-seq dataset of patients infected with *Leishmania*, revealing novel insights of the disease's physiopathology.

Conclusion: The CEMiTool R package provides users with an easy-to-use method to automatically implement gene co-expression network analyses, obtain key information about the discovered gene modules using additional downstream analyses and retrieve publication-ready results via a high-quality interactive report.

Keywords: Co-expression modules, Gene networks, Modular analysis, Leishmaniasis, Transcriptomics

Background

Cellular processes are controlled by a host of interacting molecules whose activity and levels are frequently co-regulated or co-expressed. Detecting the groups (i.e. modules) of co-expressed genes in a myriad of biological conditions has generated important insights in brain evolution [1], coronary artery disease [2], and macrophage activation [3], among many other biological conditions.

Following evidence that genes interact with each other in a scale-free fashion [4], Zhang and Horvath developed an R package named WGCNA (Weighted

Gene-Coexpression Network Analysis) that identifies co-expressed gene modules [5]. Although tutorials and examples are available for using the package, following its workflow verbatim is time-consuming and tiresome. Moreover, users are often required to manually select parameters and to filter the input genes prior running WGCNA. This hinders workflow automation and can impact reproducibility since different researchers may utilize different parameters, obtaining distinct results for the same data set. More importantly, WGCNA is limited in terms of the functional analyses available for the package users.

After identifying co-expressed gene modules, researchers are often interested in performing functional and integrative analyses. Over-representation analysis (ORA) can be used to reveal if a set of co-expressed genes is enriched for

* Correspondence: hnakaya@usp.br

[†]Equal contributors

¹Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, SP 05508-900, Brazil
Full list of author information is available at the end of the article