Universidade de São Paulo
Faculdade de Ciências Farmacêuticas

Usando a Imunologia de Sistemas para Entender os Mecanismos Moleculares de Doenças e Vacinas

Helder Takashi Imoto Nakaya

São Paulo
Março/2019

Universidade de São Paulo
Faculdade de Ciências Farmacêuticas

Usando a Imunologia de Sistemas para Entender os Mecanismos Moleculares de Doenças e Vacinas

Helder Takashi Imoto Nakaya

Texto sistematizado apresentado à
Faculdade de Ciências Farmacêuticas da
Universidade de São Paulo para obtenção do Título de
Professor Livre-Docente junto ao
Departamento de Análises Clínicas e Toxicológicas.
Edital FCF/ATAc no 03/2019, de 18.02.2019.

São Paulo
Março/2019

We are drowning in a sea of data and starving for knowledge.

Today, biology is more about gathering data than hunting down new ideas.

Sydney Brenner

The saddest aspect of life right now is that

science gathers knowledge faster than society gathers wisdom.

Isaac Asimov

Esta tese é dedicada à minha amada esposa,

Ana Carolina Viegas Nakaya.

Agradecimentos

Aos colegas e amigos do Departamento de Análises Clínicas e Toxicológicas e de toda a Faculdade de Ciências Farmacêuticas da USP por todo apoio, conselhos e colaboração. Foram eles que tornaram minha vida acadêmica mais rica e permitiram minha evolução como docente.

Aos alunos e pós-docs do laboratório por toda alegria, orgulho e dedicação. Obrigado pelo voto de confiança nesta etapa tão importante de sua formação.

Aos pesquisadores do Brasil e do mundo que contribuíram enormemente com este trabalho.

A todos os funcionários da FCF pelo auxílio e carinho ao longo desses anos.

Às agências de fomento FAPESP, CAPES, CNPq, *Bill and Melinda Gates Foundation* e IMI pela concessão de auxílios à pesquisa, bolsas dos alunos e a bolsa de Produtividade em Pesquisa.

Resumo

Nakaya, H. I. Usando a Imunologia de Sistemas para Entender os Mecanismos Moleculares de Doenças e Vacinas. 2019. 72f. Texto Sistematizado (Livre docência). Faculdade de Ciências Farmacêuticas/Departamento de Análises Clínicas e Toxicológicas – Universidade de São Paulo, São Paulo, 2019.

Nas últimas décadas, dados relacionados com a saúde humana, desde informações clínicas e epidemiológicas até imagens médicas e experimentos ômicos, foram gerados e acumulados em uma quantidade sem precedentes na história. Um campo novo de pesquisa chamado "Imunologia de Sistemas" emergiu para tentar integrar, analisar, interpretar e predizer os mecanismos moleculares de doenças e vacinas. Esta tese mostra diversas aplicações da Imunologia de Sistemas no estudo de arboviroses, vacina da gripe, câncer, tuberculose, pneumonia, artrite, dentre outros. Também mostra o desenvolvimento de ferramentas computacionais amigáveis que permitem com que qualquer cientista, sem conhecimento prévio de bioinformática, possa realizar análises de Imunologia de Sistemas. Os achados das análises forneceram novas hipóteses e *insights* que, ao serem testados e validados experimentalmente, melhoram nosso entendimento sobre os processos imunológicos por trás da vacinação e de doenças humanas.

Palavras-chave: Imunologia, bioinformática, vacinologia, ferramentas computacionais, transcritoma, biologia de sistemas

Sumário

1. Introdução 8
1.1. Complexidade do sistema imunológico 8
1.2. Surgimento da Imunologia de Sistemas 12
1.3. RNAs não codificadores 17
2. Aplicação da Biologia de Sistemas no contexto da vacinação 21
2.1. Vacina contra gripe 21
2.2. Vacina experimental contra malária 24
2.3. Outras vacinas 27
2.4. RNAs não codificadores longos no contexto da vacinação 28
3. Aplicação da Biologia de Sistemas em doenças infecciosas 30
3.1. Arboviroses 30
3.2. Tuberculose 35
3.3. Pneumonia 38
3.4. Outras doenças infecciosas 42
4. Aplicação da Biologia de Sistemas em outras doenças humanas 42
4.1. Artrite idiopática juvenil 43
4.2. Câncer 47
5. Desenvolvimento de ferramentas computacionais de Imunologia de Sistemas 49
5.1. Ferramenta CEMiTool 50
5.2. Ferramenta MDP 55
5.3. Outras ferramentas computacionais 62
6. Discussão 63
7. Conclusão 64
8. Referências 65

1. Introdução

1.1. Complexidade do sistema imunológico

Patógenos e vacinas estimulam o sistema imune através de uma intricada e complexa rede de células especializadas e órgãos. A ativação deste sistema requer uma fina coordenação espacial e dinâmica de diversos processos biológicos, que vão desde o reconhecimento dos fatores derivados de patógenos até a intercomunicação e diferenciação das células. Embora um grande número de patógenos seja naturalmente eliminado pelo sistema imune, alguns conseguem se evadir das defesas do hospedeiro e se multiplicam, estabelecendo infecções que podem resultar em hospitalização e morte. A forma como as células do sistema imune respondem e interagem entre si e com os patógenos é o que determina a progressão da infecção. Até agora, pouco se sabe sobre a dinâmica da rede de células e moléculas envolvidas nesta resposta. Da mesma forma, temos um limitado conhecimento dos mecanismos moleculares envolvidos na imunidade induzida por vacinas.

A imunidade inata é crítica para a detecção inicial de patógenos e a ativação subsequente da imunidade adaptativa (1). Esta detecção geralmente ocorre através de receptores de reconhecimento de padrões (PRRs) presentes em células do sistema imune inato como células dendríticas, células *Natural Killer* (NK), neutrófilos, macrófagos, dentre outras (2). O conjunto de PRRs contém vários receptores diferentes que podem desencadear vias de sinalização específicas ou parcialmente em comum (3). Além das sinalizações desencadeadas por PRRs, a sinalização por citocinas e até microRNAs em exossomos também desempenham um papel importante na imunidade inata (4). Dependendo dos fatores de transcrição e de outros genes reguladores envolvidos nas vias de sinalização, as células

podem desencadear um processo inflamatório, promover fagocitose, liberar citocinas e quimiocinas importantes, se diferenciar ou simplesmente morrer por apoptose, necroptose ou piroptose. Caracterizar esta intricada rede de sinalização e regulação induzida por infecção ou vacinação nos ajudaria a entender melhor quais são os principais fatores que determinam uma efetiva resposta imune.

A imunidade adaptativa pode ser grosseiramente dividida em resposta humoral ou celular. Na imunidade humoral, o principal tipo celular é o linfócito B enquanto na mediada por células é o linfócito T. Estes linfócitos podem ser subdivididos em inúmeras populações dependendo de suas funções e marcadores de superfície. Por exemplo, para a produção de anticorpos (resposta humoral), os linfócitos B precisam ser ativados pelo antígeno para se diferenciarem em células secretoras de anticorpos. Porém, a comunicação deste linfócitos B com células do sistema imune inato e adaptativo (5) é crítica para o andamento deste processo que envolve a migração celular por diferentes locais (sangue periférico, linfonodo, medula óssea). Os linfócitos T, por sua vez, podem ser subdivididos em diversos grupos como células T CD8+ citotóxicas (TC1, TC2, TC17) e células T CD4+ *helpers* (que podem ser subdivididas em Th1, Th2, Th17, Th foliculares, Tregs, etc) (6). As vias de sinalização envolvidas na diferenciação linfócitos em células efetoras ou exaustas, assim como na formação da memória, são extremamente complexas (7). Enquanto várias peças deste quebra-cabeça já foram estudadas, ainda há uma miríade de outros fatores que podem influenciar os processos biológicos relacionados à imunidade adaptativa. É como se, usando as mesmas peças (ou acrescentando e tirando algumas outras) pudéssemos montar vários quebra-cabeças, um para cada perturbação biológica.

A resposta imune induzida pela vacinação utiliza praticamente as mesmas vias de sinalização da resposta imune contra uma infecção (8). Os antígenos e adjuvantes presentes

na vacina irão ativar as células do sistema imune e promover sua diferenciação em células efetoras e de memória, usando moléculas efetoras e fatores de transcrição de uma forma similar à infecção (9, 10). Vacinas contendo vírus atenuados como a da Febre Amarela e da gripe (*Live Attenuated Influenza Vaccine*) serão ainda mais similares já que, nestes casos, ocorrerá a infecção das células sem o aparecimento da doença (salvo casos raros com a vacina da pólio atenuada e da própria Febre Amarela). A resposta imune induzida por vacinas é, portanto, igualmente complexa e requer técnicas modernas de análise em larga escala que permitam a caracterização em paralelo dos diferentes componentes biológicos envolvidos. A Figura 1 resume os principais eventos envolvidos na resposta imune induzida pela vacinação.

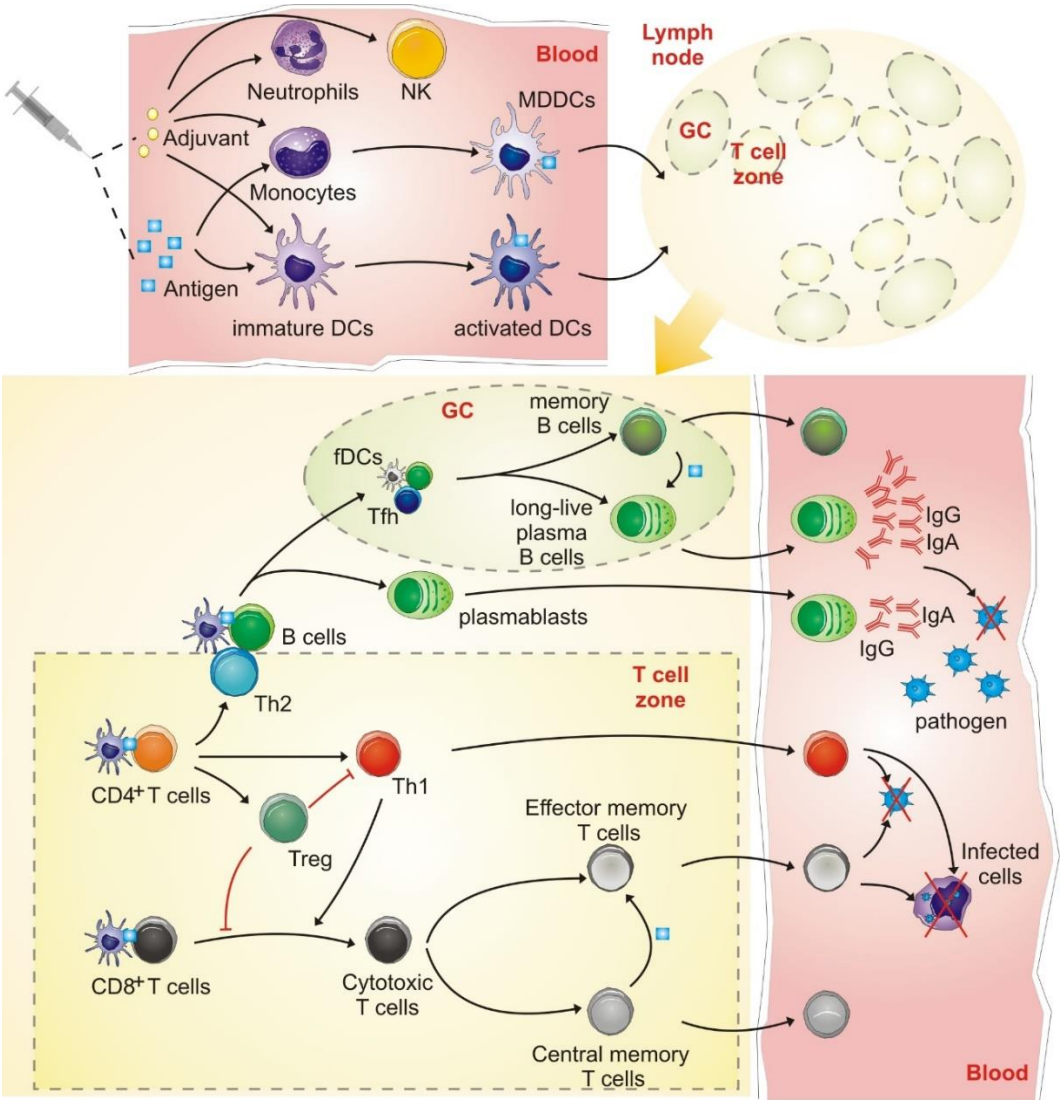

Figura 1. Principais efetores da resposta imune induzida por vacina (inativada). Os efetores imunes induzidos pela vacina incluem linfócitos T CD8+ citotóxicos e anticorpos produzidos por linfócitos B. A geração e manutenção de respostas específicas de células B e CD8+ é suportada por linfócitos T *helper* CD4+. Células do sistema imune inato, como células dendríticas e monócitos, desempenham um papel fundamental na regulação da quantidade, qualidade e longevidade da resposta imune adaptativa induzida pela vacina. DC, célula dendrítica; fDC, célula dendrítica folicular; Tfh, célula T *helper* CD4+ foliculares; Treg, células T CD4+ CD25+ reguladoras; NK, células *Natural Killers*; GC, centro

germinativo. Figura extraída de meu capítulo “Nakaya, H. (2013) Systems Biology of Infectious Diseases and Vaccines, in Computational Systems Biology” (11).

O objetivo nesta seção é apenas mostrar a complexidade de nosso sistema imunológico (11) e a necessidade de se utilizar ferramentas experimentais e métodos analíticos que permitam capturar e descrever toda esta complexidade. Não cabe a esta tese contemplar todos os detalhes conhecidos da imunologia moderna em relação à infecção e vacinação.

1.2. Surgimento da Imunologia de Sistemas

As tecnologias de larga escala permitem hoje que cientistas estudem genomas inteiros, assim como todo o conjunto de transcritos (transcritoma), proteínas (proteoma) e metabólitos (metaboloma) de células e tecidos. Com a diminuição dos custos relacionados a estas tecnologias, muitos laboratórios são agora capazes de produzir um volume colossal de dados. Há 10 anos atrás, o sequenciamento completo de um genoma humano custava 10 milhões de dólares. Em 2014, este custo caiu para apenas 1000 dólares (12). Algo similar aconteceu para as técnicas que medem o transcritoma, metaboloma, proteoma, etc. Como consequência, hoje temos acesso a uma quantidade imensa de dados de genomas completos, transcritomas de células únicas e dados de microbioma e metaboloma de amostras humanas (13). No entanto, este crescimento exponencial de dados biológicos não levou a um aumento exponencial de nosso conhecimento. O próximo desafio está justamente em como analisar e integrar estes diferentes tipos de dados ômicos.

A biologia de sistemas utiliza a enorme quantidade de dados gerados por estas tecnologias para descrever as complexas interações entre os componentes de um sistema biológico frente a diferentes tipos de perturbações (14). Além disso, modelos computacionais e matemáticos podem ser construídos para predizer o comportamento do sistema (15, 16) e, desta forma, esta estratégia promete revolucionar nosso entendimento dos sistemas biológicos.

A aplicação da biologia de sistemas na vacinologia e doenças infecciosas é definida como Imunologia de Sistemas. Esta nova área de pesquisa pode potencialmente preencher gaps fundamentais na imunologia básica, assim como permitir o desenho racional e auxiliar no teste de novas vacinas (17). O estudo detalhado das redes biológicas que controlam e definem a resposta imune para infecção pode nos ajudar a entender melhor as próprias doenças (18) e identificar alvos terapêuticos importantes (19). A Imunologia de Sistemas é, portanto, a estratégia ideal para a análise de redes biológicas complexas relacionadas ao sistema imune. São essas redes que coordenam os diferentes processos biológicos relevantes às respostas imunológicas induzidas por patógenos, vacinações, inflamação ou doenças auto-imunes.

Existem 2 tipos de Imunologia de Sistemas. Um sistema biológico pode ser modelado usando uma estratégia “bottom-up”, o que significa que o conhecimento mecanicista obtido de experimentos é usado para desenvolver um conjunto de equações que tentam explicar uma parte maior do sistema (Figura 2). Graças às tecnologias modernas de larga-escala, hoje é possível fazer uma outra estratégia chamada “top-down”. Nela, experimentos medem, em paralelo, o comportamento de milhares de componentes biológicos frente a uma perturbação. Modelos computacionais e estatísticos são então aplicados para tentar encontrar mecanismos (de uma forma reducionista) que podem ser depois validados experimentalmente (Figura 2).

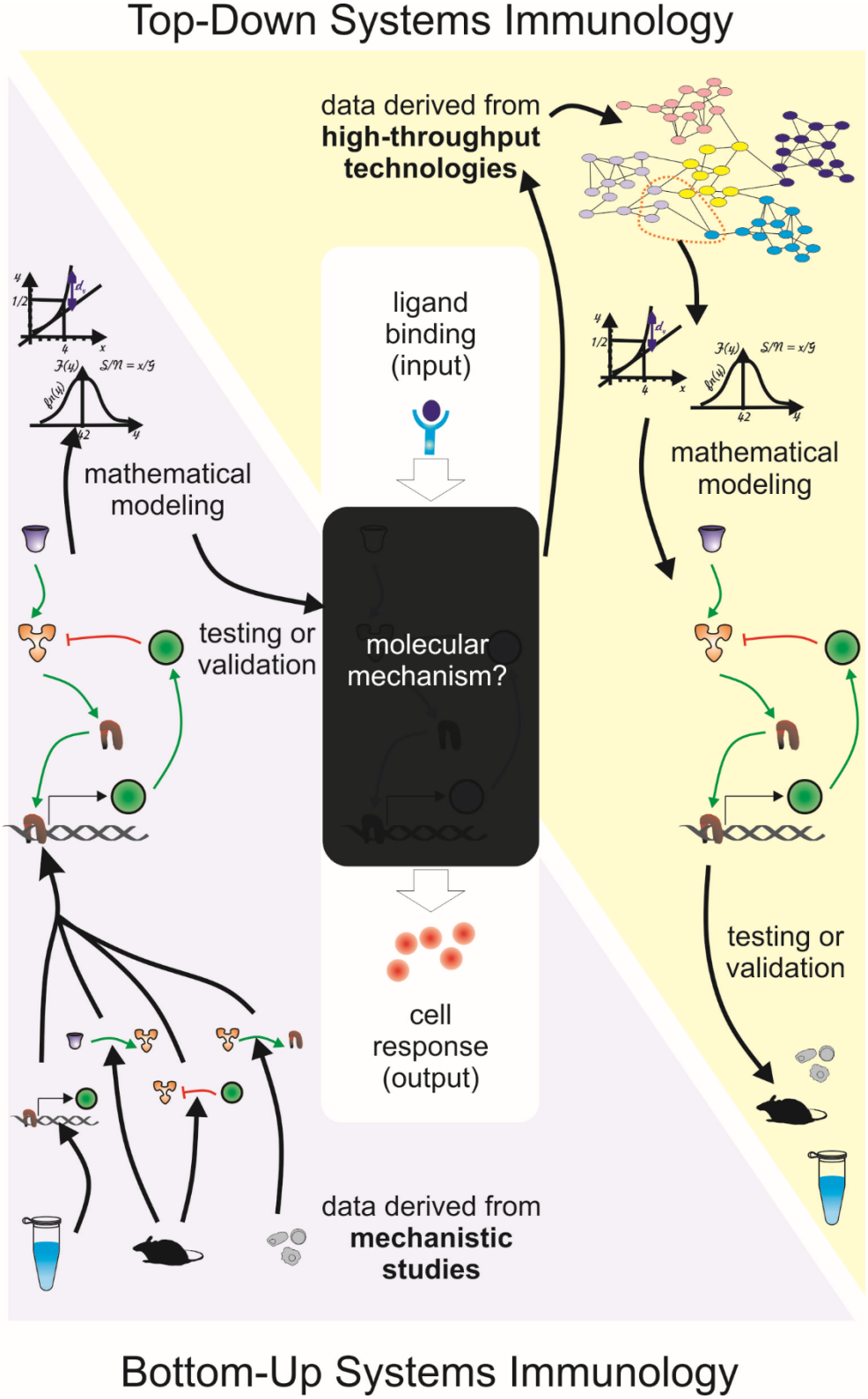
Top-Down Systems Immunology
data derived from
high-throughput
technologies
ligand
binding
(input)
mathematical
modeling
testing or
validation
molecular
mechanism?
cell
response
(output)
mathematical
modeling
testing or
validation
data derived from
mechanistic
studies
Bottom-Up Systems Immunology

Figura 2. Dois tipos de Imunologia de Sistemas. A abordagem "top-down" começa com uma grande quantidade de dados derivados de tecnologias de larga-escala. A modelagem matemática é então usada para reduzir a informação e, assim, obter um mecanismo putativo que pode ser posteriormente testado ou validado em modelos animais ou experimentos *in vitro* e *in vivo*. A abordagem "bottom-up" integra dados mecanicistas derivados de experimentos anteriores a fim de construir uma rede ou circuito que possa explicar o processo biológico estudado. A modelagem matemática é então aplicada para prever o comportamento do sistema, que pode ser posteriormente testado ou validado em um experimento real. Extraído do meu capítulo "Lever M, Hirata TD, Russo P, & Nakaya H (2018) Systems Immunology" (20).

A vacinologia de sistemas emergiu como uma área multidisciplinar cujo objetivo é entender e prever a imunidade induzida por vacinas (17, 21, 22). Apesar de fazer parte da Imunologia de Sistemas já que utiliza as mesmas ferramentas computacionais e bancos de dados (9), as coortes de um estudo de Vacinologia de Sistemas se difere das de estudos de infeção. Na Vacinologia de Sistemas, podemos saber com precisão o momento da perturbação (ou seja, a vacinação) do sistema, a dose utilizada, avaliar a pré-imunidade do indivíduo e trabalhar com pessoas saudáveis. Isso permite estudar a dinâmica do sistema imune e desenvolver modelos computacionais e estatísticos capazes de prever a imunidade. Anteriormente ao trabalho descrito nesta tese, publicamos estudos de Vacinologia de Sistemas para diversas vacinas, como as vacinas contra a Febre Amarela (23), a gripe (24) e a meningite meningocócica (25).

A Figura 3 mostra exemplos de desenhos experimentais e análises aplicados ao estudo de vacinas. Em geral, tecnologias de larga escala como *microarrays*, RNA-seq,

espectrometria de massa, CyTOF, etc são usadas para medir o comportamento de todos os componentes biológicos de células ou tecidos do sistema imunológico durante uma infecção ou vacinação. Estes dados então são submetidos à análise de expressão diferencial para se encontrar genes e vias de sinalização ou metabólicas relevantes para a perturbação. Análise de enriquecimento de vias e de redes biológicas auxiliam na interpretação dos resultados e colocam os achados em um contexto. E finalmente, técnicas avançadas de aprendizado de máquina (26) podem ser aplicadas para se identificar genes cuja expressão pode predizer a imunogenicidade de uma vacina (23).

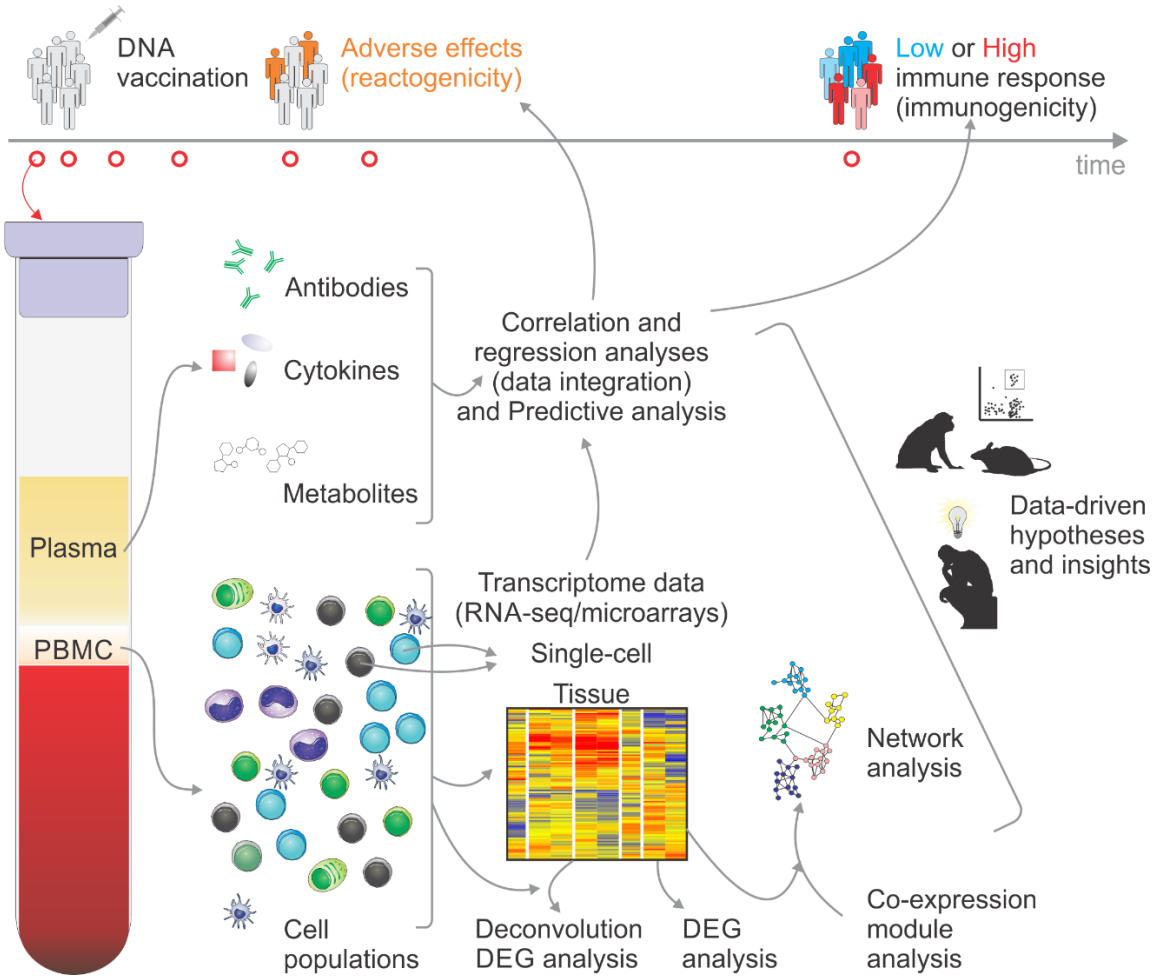

Figura 3. Visão geral da metodologia de um típico estudo de Vaccinologia de Sistemas. Amostras de sangue são coletadas em momentos pré e pós-vacinação e tecnologias de baixo

e alto rendimento são aplicadas para caracterizar a resposta imune. As respostas dos anticorpos, os perfis de citocinas, o metaboloma e o proteoma podem ser avaliados a partir do plasma sanguíneo. As respostas de células T e B podem ser determinadas a partir da fracção de células mononucleares do sangue periférico (PBMC). O perfil de expressão gênica de pessoas vacinadas pode ser obtido a partir de experimentos de RNA-seq e *microarrays* em populações de células ou *single-cells*. Uma ampla gama de métodos computacionais, tais como métodos de deconvolução, análise de genes diferencialmente expressos (DEG) pode ser aplicada para determinar as assinaturas gênicas. Esta informação é então integrada para determinar as medidas de imunogenicidade e segurança da vacina, desvendar os mecanismos de vacinação e conduzir novas hipóteses e *insights*. Figura extraída de meu capítulo "Lever M, Silveira EL, & Nakaya HI (2017) Systems Vaccinology Applied to DNA Vaccines: Perspective and Challenges." (27).

1.3. RNAs não codificadores

Nas últimas décadas, ficou demonstrado que muitos RNAs que não eram traduzidos em proteínas possuíam um papel regulatório em diversos processos celulares. As principais categorias de RNAs não-codificadores com papel regulatório são os RNAs curtos como os miRNAs, siRNAs, e piRNAs e os RNAs longos, incluindo os RNA intergênicos, os RNAs *enhancer*, os RNA antisensos associados a promotores e os RNAs intrônicos (28). Durante meu doutorado, participei da identificação e caracterização funcional de RNAs intrônicos (29, 30), assim como na investigação de seu papel como marcador do câncer de próstata (31).

A classe de RNAs não codificadores regulatórios mais estudada no contexto de infecções é a dos miRNAs. Estas pequenas moléculas de aproximadamente 22 nucleotídeos

podem tanto degradar um RNA mensageiro alvo quanto inibir sua tradução. Já foi mostrado que miRNAs podem moldar a imunidade e a resposta inflamatória durante infecções (32) e inclusive ter uma aplicação clínica servindo de alvos terapêuticos ou de biomarcadores (33).

Longos RNAs não-codificadores (lncRNAs) são definidos como RNAs maiores que 200 pares de bases e que não são traduzidos em proteínas. Estes lncRNAs podem ser basicamente divididos de acordo com sua função regulatória, sua localização no locus gênico ou sua orientação em relação aos genes codificadores de proteínas (34). Enquanto lincRNAs (*large intergenic non-coding RNAs*) estão localizados em regiões intergênicas no genoma, os lncRNAs intrônicos, como o próprio nome sugere, são transcritos em regiões intrônicas de genes codificadores de proteína. NATs (*natural antisense transcripts*) são RNAs que possuem uma orientação de transcrição oposta (antisenso) ao gene codificador de proteína. Em meu doutoramento, nós fizemos uma extensa análise bioinformática, utilizando 6 milhões de sequências parciais de RNAs depositadas em banco de dados público, e revelamos que cerca de 3/4 dos genes humanos possuem lncRNAs intrônicos (29). Outros estudos sugerem que ~40% dos genes codificadores de proteína tenham um ou mais NATs (35) e que no genoma humano haja mais de 8.000 lincRNAs (36).

Uma clara indicação de que a transcrição de lncRNAs não é aleatória ou espúria consiste no fato de que certos lncRNAs possuem sua expressão regulada temporalmente, espacialmente ou em resposta a um estímulo (Mercer, Dinger et al. 2009). Foi mostrado, por exemplo, que, em camundongos, lncRNAs possuíam um perfil de expressão específico para certas regiões do cérebro (37) ou estágios de diferenciação celular (38). Além disso, a biogênese, regulação e processamento de muitos lncRNAs se assemelham aos genes codificadores de proteínas, já que ambos são transcritos pela RNA polimerase II, são regulados por fatores de transcrição e modificação na cromatina, podem sofrer *splicing*,

possuem *cap* 5’ e cauda poli-A (39). Porém, seus níveis de expressão tendem a ser menores que os genes codificadores de proteína (40) e sua expressão mais restrita a um tecido específico (41). Entender como esses lncRNAs são regulados e onde e em que condições eles são expressos é imperativo para o estudo de suas funções.

Embora a vasta maioria dos lncRNAs ainda precise ser caracterizada, um conjunto interessante de funções e mecanismos de regulação podem ser atribuídos a eles (42) (Figura 4). *HOTTIP* (43) é um lncRNA que promove a expressão de vários genes *HOXA* ao manter a cromatina ativa (Figura 4, item 1). Vários outros lncRNAs reprimem a expressão de genes codificadores de proteína através de processos epigenéticos, como *Air* (44), HOTAIR (45) e *Xist*/*Tsix* (46, 47) (Figura 4, item 2), ou por interferência transcricional, como *bxd* (48) (Figura 4, item 3). LncRNAs com orientação antisenso ao gene codificador de proteína como *α-280/250* (49), *E2F4 antisense* (50) e *Emx2OS* (51) podem hibridizar com o mRNA senso e promover sua degradação (Figura 4, item 4). O lncRNA *Saf* (52) pode alterar o padrão de *splicing* do gene *Fas* ao mascarar sítios importantes de reconhecimento do spliceossomo (Figura 4, item 5). O próprio lncRNA, como *Au76* (53), pode também servir como precursor de RNAs pequenos, indiretamente modulando a expressão gênica (Figura 4, item 6). Finalmente, já foi descrito que lncRNAs como *Evf-2* (54), *MALAT1* (55), *centromeric α-satellite RNA* (56), *NEAT1* (57) podem interagir diretamente com proteínas e complexos proteicos e alterar suas atividades ou localizações na célula, assim como possuir um papel estrutural no complexo proteína-RNA (Figura 4, itens 7 a 9). Esses exemplos deixam claro o papel central dos lncRNAs na regulação fina de redes gênicas.

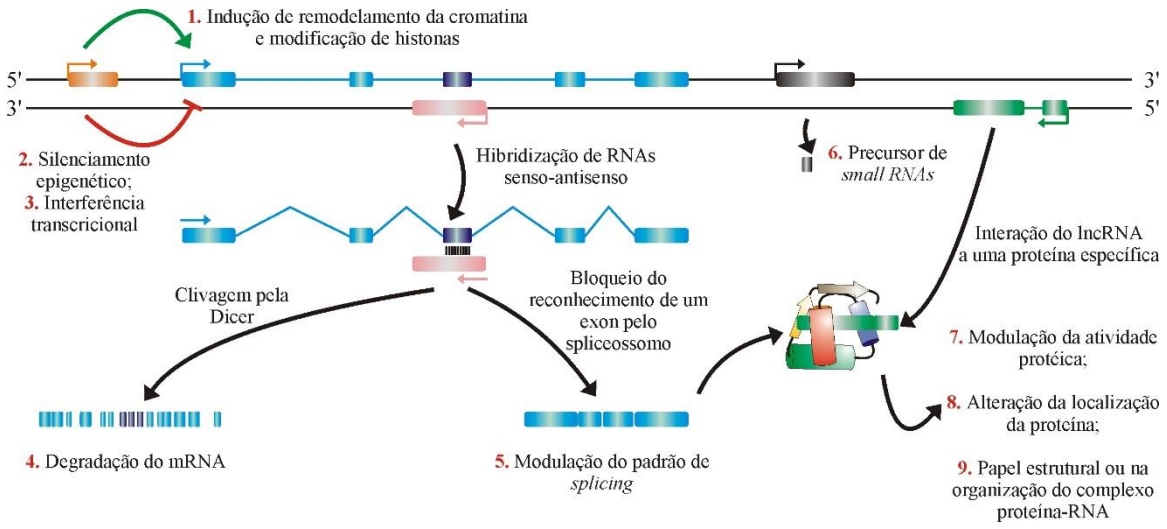

Figura 4. Algumas funções atribuídas aos longos RNAs não-codificadores (lncRNAs). A transcrição de um lncRNA (bloco laranja) pode afetar positivamente (1) ou negativamente (2 e 3) a expressão de um gene codificador de proteína (blocos azuis) através da indução de remodelamento da cromatina ou da inibição do recrutamento da RNA polimerase II e silenciamento epigenético, respectivamente. A hibridização de um lncRNA antisenso (bloco rosa) com um mRNA senso (blocos azuis) pode levar à degradação do mRNA (4) ou ao bloqueio do reconhecimento de um exon (bloco roxo) pelo spliceossomo causando uma alteração no padrão de *splicing* do mRNA (5). LncRNAs (bloco preto) também podem ser processados em RNAs pequenos como microRNAs, piRNAs, dentre outros (6). Finalmente, um lncRNA (bloco verde) ao interagir com uma proteína ou complexo proteico pode afetar a atividade da proteína (7), alterar sua localização dentro da célula (8) ou servir como um componente estrutural do complexo RNA-proteína (9).

Estudos mostram que lncRNAs podem associados à regulação do sistema imunológico (28, 58, 59). Esta regulação vai desde a indução da produção de IL-8 após

infecção viral (60) até da ativação de células CD8+ em tuberculose (61). Vários outros lncRNAs são conhecidos por exercer funções relacionadas ao sistema imune, como controle da diferenciação de células dendríticas (62), controle da expressão de *Foxp3* em Células T (63) e diferenciação de células T CD4+ (64).

2. Aplicação da Biologia de Sistemas no contexto da vacinação

Esta seção descreve os resultados obtidos em projetos de Vacinologia de Sistemas. Apenas o resumo do trabalho e os achados principais serão descritos. A metodologia é brevemente mencionada, mas os detalhes dela se encontram nas publicações citadas.

2.1. Vacina contra gripe

Em 2011, como pós-doutor na *Emory University* no laboratório do prof. Bali Pulendran, publicamos um artigo na *Nature Immunology* (24) descrevendo o primeiro estudo de Vacinologia de Sistemas para a vacina contra Gripe. Neste trabalho identificamos as assinaturas gênicas relacionadas à imunogenicidade induzida pela vacina em adultos jovens e saudáveis.

Porém, já é sabido que a vacina contra Gripe não funciona tão bem em idosos quando comparados à jovens adultos (65). Para esta faixa etária, uma dose maior da vacina geralmente é recomendada (66). Nós fomos então investigar se as assinaturas encontradas em jovens adultos poderiam predizer a imunogenicidade em idosos. Já como professor da USP, em colaboração com o prof. Bali Pulendran, publicamos um artigo na revista *Immunity* em 2015 (67) que comprovava esta hipótese. Investigamos pela primeira vez também o papel

de microRNAs no contexto da vacinação (67). Esta publicação recebeu destaque na *Immunity* e também foi mencionada em diversos jornais e revistas de divulgação científica (Figura 5).

Figura 5. Notícia da revista "The Scientist" sobre o artigo (67).

Da mesma forma que idosos, crianças muito jovens com menos de 2 anos possuem uma resposta de anticorpo menor do que adultos jovens quando recebem a vacina da Gripe inativada (68). Até aquela época, nenhum estudo anterior tentou avaliar o mecanismo molecular induzida pela vacina em crianças menores de 2 anos de idade. Mais importante, em uma parceria com a empresa Novartis (hoje parte da GSK) que desenvolveu um adjuvante chamado MF59, nós pudemos testar como era a resposta da vacina inativada em crianças com ou sem este adjuvante. O artigo foi publicado na revista PNAS em 2016 (69) e mostrou que as respostas imunológicas e as assinaturas gênicas da vacina com o adjuvante eram mais fortes do que a sem adjuvante.

Um achado importante em nossas análises era que a expressão do gene TLR5 (*toll-like receptor 5*) estava diretamente correlacionado com os níveis de anticorpos específicos para o vírus Influenza após a vacinação (24). Isto era algo inesperado já que o ligante natural

do receptor TLR5 era flagelina, um componente bacteriano. Não teria, portanto, nenhuma relação com uma resposta imunológica induzida por uma vacina inativada contra um vírus. Análises adicionais mostraram que essa correlação era robusta e experimentos em camundongo nocaute para TLR5, assim como o uso de antibióticos e administração oral de bactérias com e sem flagelina, demonstrou, pela primeira vez, que a microbiota intestinal era importante na indução de uma boa resposta de anticorpos contra o vírus da Gripe. Este trabalho foi publicado na revista *Immunity* em 2014 (70).

Uma outra comprovação de que as hipóteses e *insights* gerados pelas análises de Imunologia de Sistemas podem fornecer novos conhecimentos sobre os mecanismos de processos imunológicos vieram em dois trabalhos publicados em colaboração com os pesquisadores da *Emory University*. No primeiro publicado na revista *Science* em 2014 (71) mostra que o gene *GCN2* era importante na apresentação de antígeno das células dendríticas aos linfócitos T após a vacinação contra a Febre Amarela. A hipótese sobre esta relação surgiu >5 anos antes, no artigo de Vacinologia de Sistemas da vacina da Febre Amarela (23). Neste artigo, *GCN2* era um dos genes que eram capazes de predizer a resposta de célula T CD8+ citotóxica induzida pela vacina (23). Ao notar que camundongos nocautes para o gene *GCN2* não controlavam a inflamação intestinal, foi proposta a hipótese de que a expressão deste gene também estivesse alterada em pacientes com doenças crônicas inflamatórias (Doença de Crohn e Retocolite Ulcerativa). Os pesquisadores da *Emory University* nos pediram então que investigasse essa hipótese. Realizamos então análise do perfil de expressão desses pacientes usando dados públicos de transcritoma e mostramos que existia alteração da expressão de *GCN2* nestes pacientes comparado com pessoas saudáveis. Esses resultados foram incorporados no artigo publicado na revista Nature em 2016 (72).

2.2. Vacina experimental contra malária

Uma vacina contra malária direcionada para a proteína CSP (*circumsporozoite protein*) do *Plasmodium falciparum* foi desenvolvida. Esta vacina chamada RTS,S consiste em 19 repetições de NANP (R) e o C-terminal de CSP, incluindo epítopos de células T (T) fundidos com o antígeno de superfície do vírus da hepatite B (HBsAg) (S) (73). Uma versão da vacina contendo um adjuvante AS01 mostrou que a mesma é segura e com uma eficácia de 55.8% contra malária clínica num período de 1 ano e em crianças de 5 a 17 meses de idade (74). Para melhorar nossa compreensão dos mecanismos de proteção induzida por esta vacina, nós usamos uma estratégia de Vacinologia de Sistemas para rastrear as variações temporais da resposta transcricional desencadeada pela vacina e, assim, identificar assinaturas gênicas associadas à proteção e imunogenicidade.

Em um trabalho publicado na revista PNAS em 2016 (75) nós comparamos dois regimes de vacinações. Um chamado de RRR onde 21 pessoas receberam a mesma dose de RTS,S por 3 vezes e outro chamado de ARR onde na primeira vacinação 25 pessoas receberam uma dose de *adenovirus 35* (Ad35) seguidas de 2 doses de RTS,S. Cada dose foi administrada com 1 mês de intervalo e amostras de sangue foram coletadas em diversos tempos experimentais (Figura 6). Todos os voluntários foram então desafiados com mosquitos infectados com *Plasmodium falciparum* e a proteção foi determinada. Uma série de experimentos foram utilizados para se medir o perfil transcricional destas pessoas nos diferentes tempos antes e depois da vacinação e do desafio, assim como as respostas imunológicas e os principais tipos celulares (Figura 6).

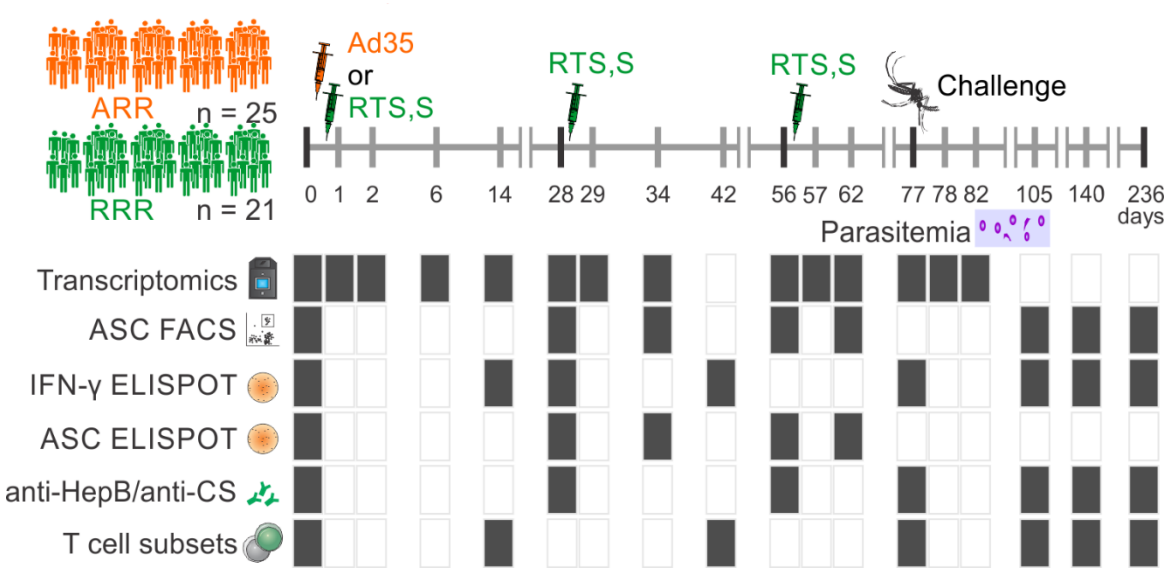

Figura 6. Desenho experimental do estudo. Os retângulos preenchidos indicam os pontos de tempo da coleta de dados. Adaptado de (75).

Dentre os achados principais, conseguimos encontrar módulos gênicos associados com a proteção induzida pela vacina nos 2 regimes (Figura 7). Foi interessante notar que dependendo do regime e dependendo do tempo experimental, módulos associados com diferentes tipos celulares pareciam estar associados com a proteção (Figura 7).

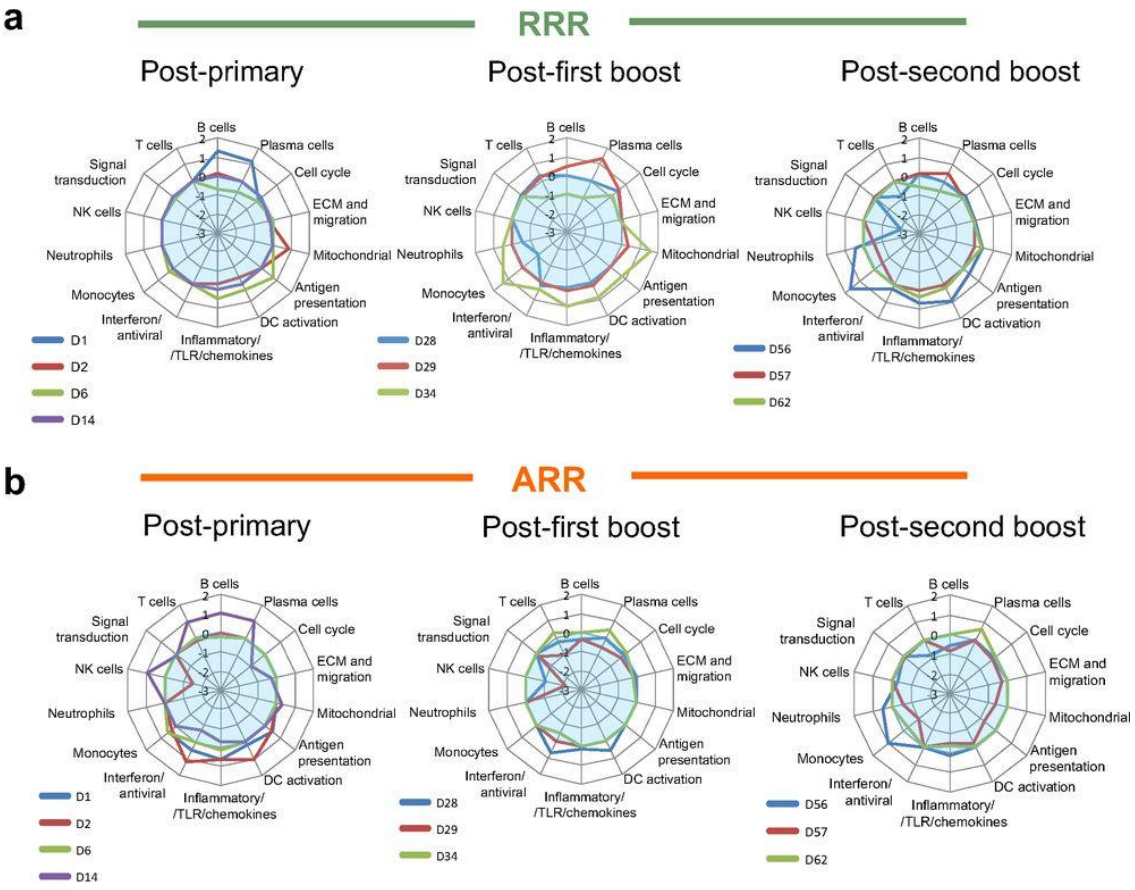

Figura 7. Módulos Gênicos Associados à Proteção. (a e b) O gráfico de teia mostra os valores médios de *Normalizing Enrichment Score* obtidos pela análise de GSEA (*Gene Set Enrichment Analysis*) usando como *ranking* os valores de razão de expressão entre pessoas protegidas ou não de acordo com o desafio e como gene sets os módulos de sangue desenvolvidos em (76). Para indivíduos protegidos, o dia do esfregaço é definido como 28 (o final do período de observação). Apenas módulos com FDR menores que 0,05 são incluídos na contagem. Zonas azuis translúcidas indicam enriquecimento negativo. (A) Correlatos de proteção para a vacina RRR. (B) Correlatos de proteção para a vacina ARR. Adaptado de (75).

2.3. Outras vacinas

Uma estratégia de Vacinologia de Sistemas também foi aplicada para estudar a vacina contra *Neisseria meningitidis*, bactéria que causa mais de 1 milhão de casos de meningite e septicemia no mundo todo por ano (77). As novidades neste trabalho, publicado na revista *Nature Immunology* em 2014 (76), foram que: (1) comparamos as assinaturas gênicas de 5 diferentes vacinas e (2) desenvolvemos um conjunto de módulos gênicos de sangue que foi largamente utilizado pela comunidade científica. Estes módulos chamados BTMs (*Blood Transcription Modules*) foram obtidos a partir de uma meta-análise de mais de 32.000 amostras de *microarrays* de sangue (Figura 8).

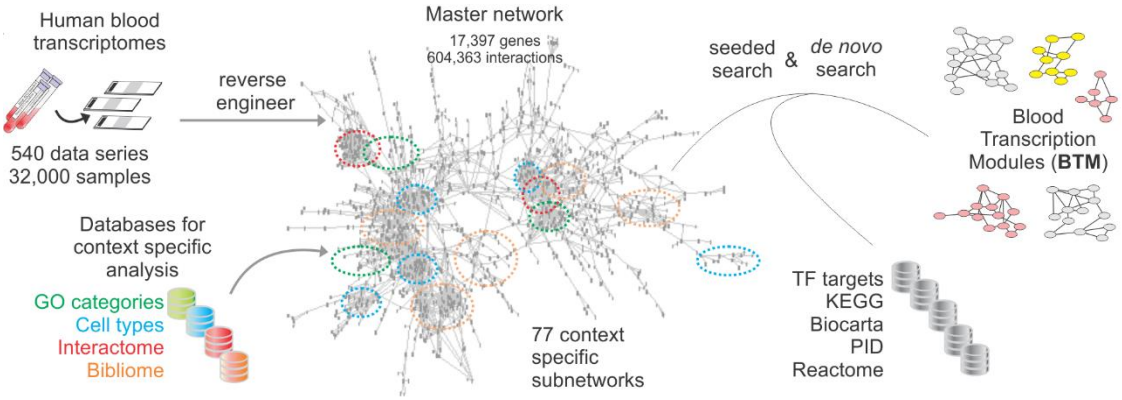

Figura 8. Módulos gênicos de sangue (BTMs). Construção dos módulos BTM através da integração de dados em larga escala. Dados de *microarrays* de amostras de sangue foram obtidas de bancos de dados públicos, integrados e informação mútua entre genes foi usada para se construir a rede principal. Análise de redes e bancos de dados funcionais foram então usados para se identificar os módulos gênicos e determinar sua possível função. Adaptado de (76).

Um outro de estudo de Vacinologia de Sistemas importante foi publicado na revista Cell em 2017 (78). Nele, mostramos pela primeira vez assinaturas de metabólitos relacionados à vacina contra o *Herpes zoster*. Nosso trabalho consistiu em integrar os dados de transcritômica, metabolômica e dados imunológicos das pessoas antes e depois de receberem a vacina. O trabalho recebeu grande atenção pela mídia, inclusive gerando um vídeo de divulgação produzido pela Ciência USP. O vídeo pode ser visualizado pelo link: https://goo.gl/gqtJ7n.

Em colaboração com pesquisadores da *University of Oxford*, realizamos uma análise de Vacinologia de Sistemas para estudar duas vacinas contra a bactéria *Salmonella enterica* serovar Typhi que causa febre entérica e infecta anualmente mais de 9 milhões de pessoas no mundo (79). A abordagem de biologia de sistemas foi capaz de mostrar que a vacina Ty21a é melhor do que a M01ZH09 em induzir uma resposta de células NK (80).

2.4. RNAs não codificadores longos no contexto da vacinação

Dezenas de estudos mostraram que lncRNAs podem modular diferentes processos imunológicos (28, 58, 59). Porém, nenhum estudo mostrou o possível papel de lncRNAs no contexto da vacinação. Nós realizamos uma meta-análise massiva de dados de transcritoma de mais de 2000 amostras de sangue para identificar possíveis lncRNAs envolvidos na vacinação contra Gripe e Febre Amarela (Figura 9).

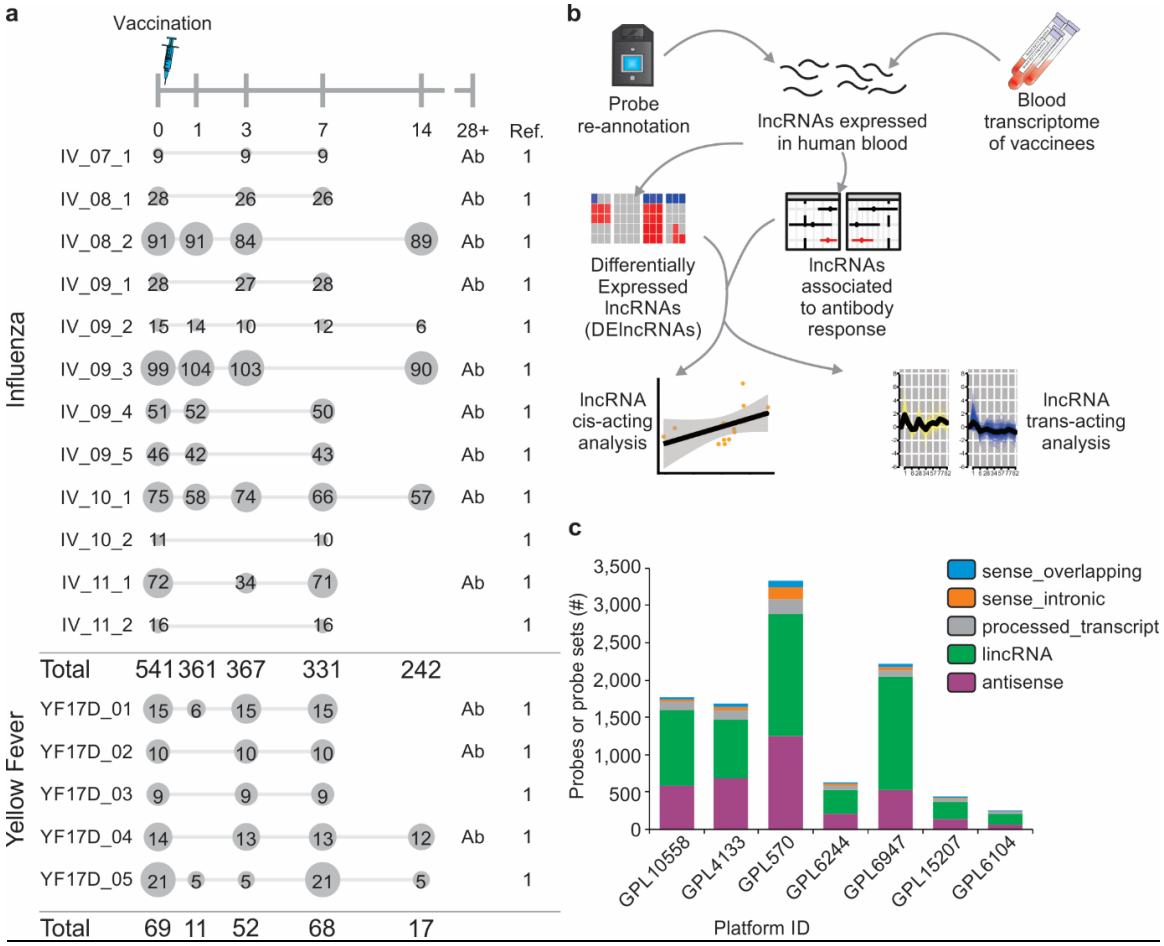

Figura 9. Desenho experimental da meta-análise de 17 coortes de vacinação. (a) Representação gráfica do regime de amostragem. Amostras de sangue foram coletadas nos dias 0, 1, 3, 7, 14 e 28 após a vacinação contra a Gripe ou contra a Febre Amarela. Os números dentro dos círculos representam o número de amostras usadas em cada coorte após o processamento. A temporada de Gripe é mostrada no nome da coorte (07 = 2007/08; 08 = 2008/09; etc). (b) Resumo das análises computacionais realizadas neste trabalho. O transcritoma de sangue das coortes foi avaliado com plataformas de *microarrays* comerciais. Após o remapeamento das sondas, cada coorte foi submetida a análise de expressão diferencial, correlação com títulos de anticorpos, correlação com genes vizinhos e análise de

co-expressão. (c) Caracterização das classes de lncRNA representadas em cada plataforma de *microarray*.

Nossa meta-análise revelou centenas de lncRNAs possivelmente envolvidos na imunidade induzida pela vacina. Os resultados deste trabalho foram disponibilizados em um banco de dados públicos chamado VaccineDB (https://vaccinedb.sysbio.tools/) que foi desenvolvido para este projeto. Os usuários podem buscar por genes codificadores de proteína e lncRNAs e verificar se os mesmos estão diferencialmente expressos ou correlacionados com as respostas imunológicas induzidas pela vacina contra Gripe ou Febre Amarela. No momento, este trabalho está sob revisão na revista PNAS.

3. Aplicação da Biologia de Sistemas em doenças infecciosas

Esta seção descreve os resultados obtidos em projetos de Imunologia de Sistemas em doenças infecciosas. Apenas o resumo do trabalho e os achados principais serão descritos. A metodologia é brevemente mencionada, mas os detalhes dela se encontram nas publicações citadas.

3.1. Arboviroses

Nós analisamos o perfil transcricional do sangue de pacientes infectados com o vírus da Dengue e comparamos com a fase convalescente. Uma característica marcante da infecção pelo vírus da Dengue é a expansão maciça de plasmoblastos (linfócitos B produtores de anticorpos) no sangue poucos dias após a infecção (81). No entanto, os mecanismos específicos que levam à este aumento são pobremente conhecidos. Esta análise apontou para

o papel de uma população de monócitos (CD14+CD16+) que estimulava a diferenciação de plasmoblastos (82). As assinaturas gênicas reveladas no trabalho indicaram os possíveis fatores de transcrição envolvidos e as vias de sinalização ativadas no processo.

Em uma colaboração com pesquisadores da Austrália, utilizamos a Imunologia de Sistema para entender melhor a patogenicidade do vírus Chikungunya (83-86). Nestes trabalhos, as análises foram usadas para revelar as pistas e vias de sinalização importantes que depois foram validadas ou testadas experimentalmente pelos pesquisadores Australianos.

Recentemente, analisamos o perfil transcricional do sangue de pessoas infectadas na fase aguda com o vírus Chikungunya (CHIKV) (Figura 10). Este trabalho está em revisão na revista *PLoS Pathogens* e disponível no BioRxiv (https://goo.gl/9C2qnJ). Um dos achados interessantes deste trabalho foi mostrar que a expressão de vários fatores de iniciação da tradução eucariótica (eIFs) estava negativamente correlacionada com a quantidade de RNA do CHIKV no sangue dos pacientes (Figura 10). Os eIFs são proteínas importantes que controlam o processo de iniciação da tradução, que é um passo importante para a síntese proteica viral (87). Não havia sido mostrado, no entanto, um possível papel destes genes na replicação viral de CHIKV.

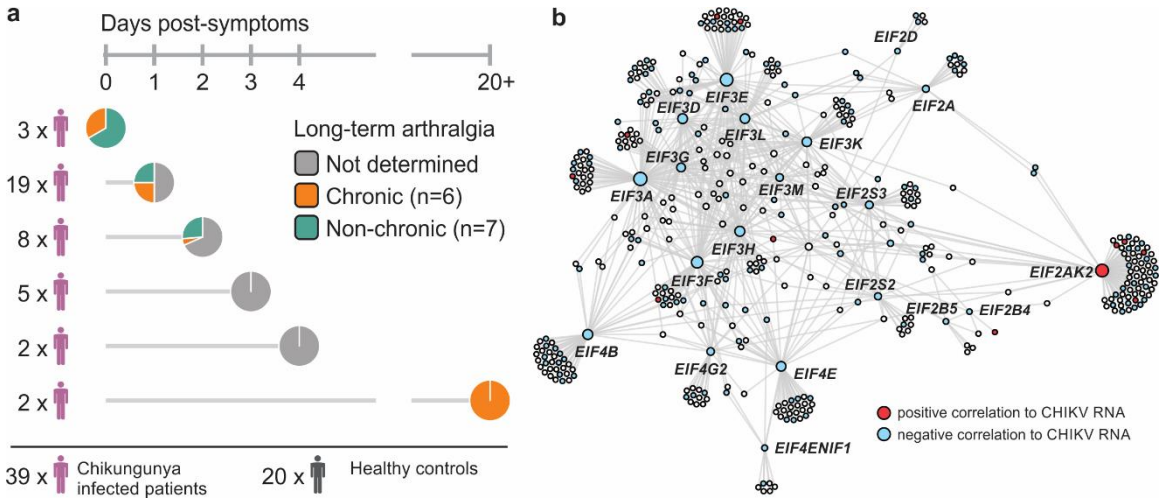

Figura 10. Coleta de amostras de pacientes infectados com o CHIKV e a associação de genes eIFs com os níveis de RNA do CHIKV no sangue. (a) Desenho experimental do estudo que representa o número de indivíduos e os dias após o início dos sintomas que as amostras de sangue foram coletadas. Artralgia de longa duração foi avaliada em alguns pacientes, conforme indicado pelo esquema de cores a seguir: crônico (laranja), não crônico (verde) e indeterminado (cinza). (b) Rede de interação proteína-proteína construída com genes eIF correlacionados com o RNA do CHIKV. A ferramenta NetworkAnalyst.ca foi usada para construir a rede e o programa Gephi para visualização. A cor dos nós representa se a correlação entre a expressão gênica e o RNA do CHIKV é positiva (vermelho) ou negativa (azul).

Em relação à infecção por vírus Zika (ZIKV), nós participamos de 3 trabalhos em colaboração (88-90), sendo 1 deles como co-autor correspondente. A aplicação da Imunologia de Sistemas foi importante para avaliar o papel do glutamato na neurotoxicidade causada pelo vírus (88), para ver a contribuição genética na microcefalia causada pela

infecção do ZIKV durante a gravidez (89) e para ver o perfil de citocinas de indivíduos infectados (90).

Recentemente, em colaboração com diversos pesquisadores do Brasil, realizamos uma análise integrada de dados genômicos, proteômicos e transcritômicos de sangue e cérebro de bebês com a Síndrome de Zika Congênita. Esta síndrome representa um conjunto de anormalidades que pode envolver microcefalia, calcificações cerebrais, ventriculomegalia, malformações corticais decorrentes de distúrbios migratórios, anormalidades oculares, dentre outras (91). Nossas análises integrativas revelaram genes de colágeno e de organização de matriz extracelular, além de genes associados a outras vias de sinalização, como importantes na Síndrome de Zika Congênita (Figura 11). No momento o trabalho está em revisão na revista *Cell Reports* e já foi depositado no BioRxiv (https://goo.gl/WmC8fp).

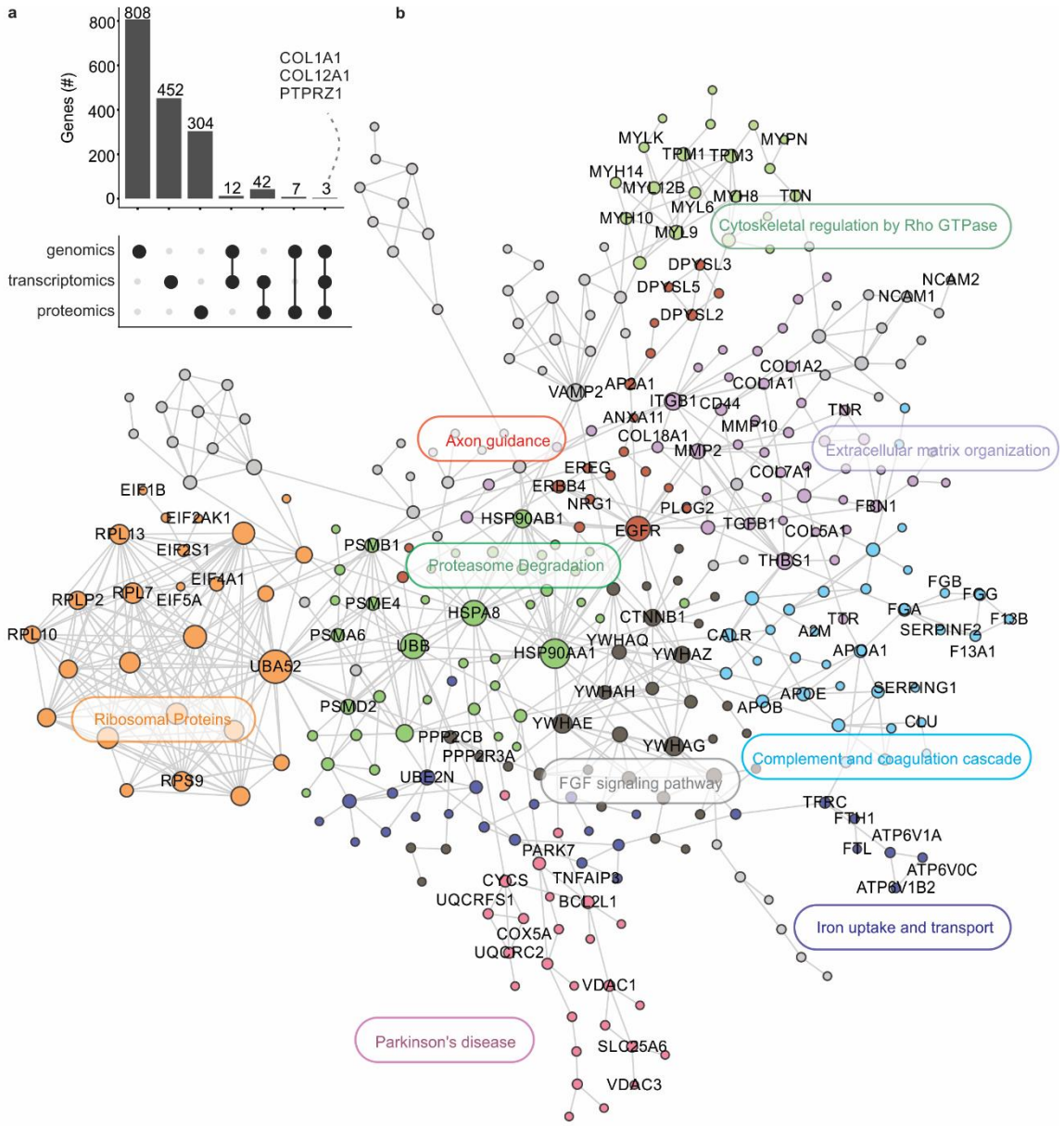

Figura 11. Integração de três camadas de informação biológica em amostras de pacientes com a Síndrome de Zika Congênita. (a) Interseções entre três camadas de informação (genômica, transcritômica e proteômica) envolvidas com CZS. O diagrama de pontos na parte inferior representa as interseções entre as camadas. Gráfico de barras mostra o número de genes em cada intersecção. Uma linha tracejada indica os genes presentes em

todas as camadas. (b) Rede de interação proteína-proteína de genes relacionados com CZS e as suas vias de sinalização associadas.

3.2. Tuberculose

A tuberculose (TB) ainda é um enorme problema de saúde pública em muitos países. Na Índia, há também um grande número de pessoas com diabetes mellitus tipo 2 (DM) e já foi mostrado que a comorbidade entre TB e DM afeta o tratamento de TB e a sua gravidade (92). Nós aplicamos a Imunologia de Sistemas para tentar entender os mecanismos moleculares responsáveis pela susceptibilidade e gravidade da TB na DM. Em um artigo publicado na *Scientific Reports* em 2017 (93), nós mostramos que havia grandes diferenças no perfil de citocinas plasmáticas (Figura 12), no perfil transcritômico e em dados clínicos de pacientes com a comorbidade (TBDM) comparado aos pacientes apenas com TB ou com DM.

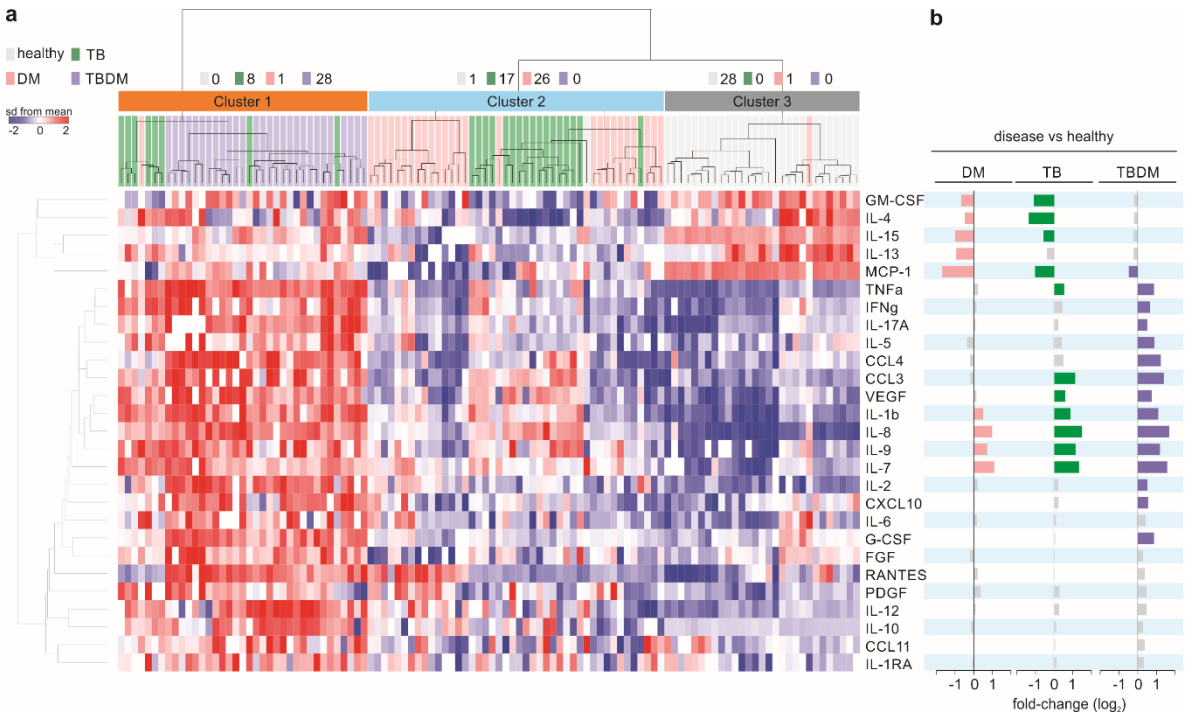

Figura 12. Perfil de citocinas no plasma de pacientes com TB e/ou DM. (a) As concentrações de citocinas foram normalizadas pelo *z-score* em todos os indivíduos. Cada coluna representa um paciente. Os nomes das citocinas são mostrados à direita do mapa de calor. Os perfis de citocinas foram ordenados por agrupamento hierárquico (distância euclidiana e agrupados com o método de Ward). O dendrograma no topo mostra 3 clusters principais. O número de indivíduos de cada fenótipo clínico (classe) em cada cluster é indicado. (b) Diferenças nos níveis de citocinas para cada doença em comparação com indivíduos saudáveis. Diferenças que não alcançaram significância estatística (P Ajustado < 0,05, razão > 1,4) são representadas como barras cinzas. Modificado de nossa publicação (93).

Investigamos também o grau de perturbação molecular (MDP) destes pacientes para avaliar a heterogeneidade das doenças e da comorbidade. Para isso, desenvolvemos uma

ferramenta chamada MDP descrita na seção 6.2 desta tese que realiza esse cálculo do grau de perturbação molecular. A mediana de MDP usando o transcritoma dos indivíduos saudáveis como referência estava aumentada de forma similar nos subgrupos TB e TBDM em comparação com os participantes com DM sozinho (Figura 13a). Fizemos então a correlação do MDP com dados clínicos que revelou uma correlação negativa com IMC, % e total de linfócitos, colesterol total e todas as subfrações de colesterol e correlações positivas foram identificadas com HbA1c, WBC (*Whole Blood Count*) total,% e total de neutrófilos e % de monócitos (Figura 13b). Estes achados sugerem uma associação de baixo colesterol HDL com a resposta hiperinflamatória de pacientes com TBDM, enquanto o alto IMC foi associado com um nível mais baixo de inflamação sistêmica em todos os subgrupos.

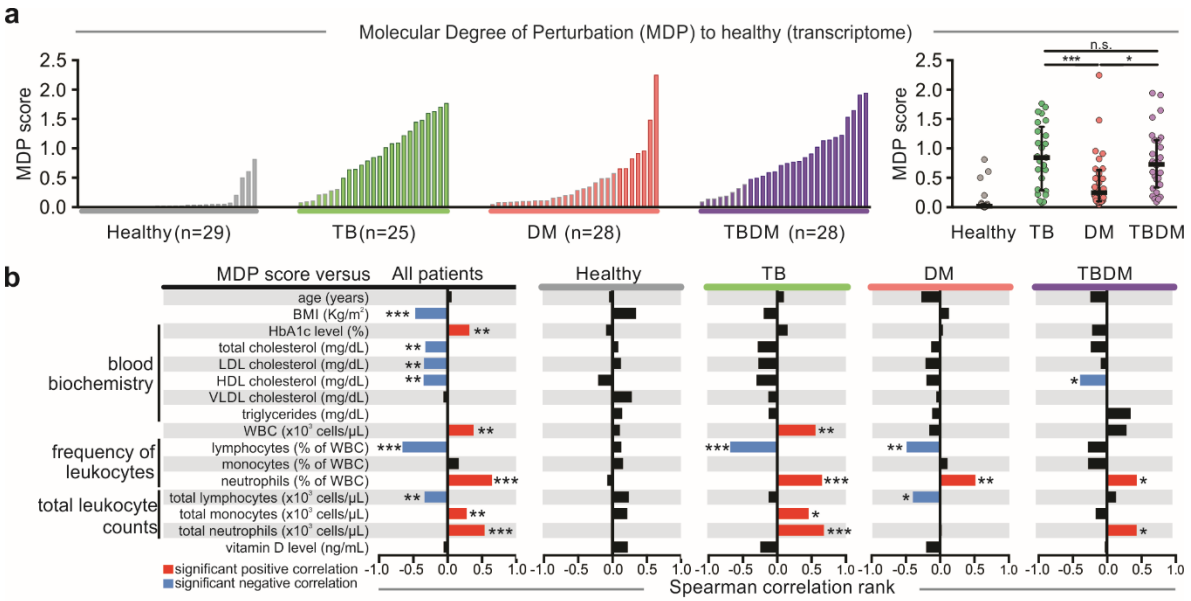

Figura 13. Fatores associados ao aumento do grau molecular de perturbação. (a) O grau de perturbação molecular (MDP) dos pacientes com TB, DM e TBDM. Painéis esquerdos mostram histogramas de indivíduos em cada subgrupo. O painel direito mostra valores individuais com mediana por grupo. Os dados foram comparados usando *one-way ANOVA*

com correção para comparações múltiplas. (b) Correlações de Spearman entre os valores de MDP com os parâmetros indicados em cada subgrupo. Correlações estatisticamente significantes, após o ajuste de Holm-Bonferroni para comparações múltiplas, são destacadas (vermelho, correlação positiva; azul, correlação negativa; preto, correlação não significativa). Modificado de nossa publicação (93).

3.3. Pneumonia

Uma complicação comum da Gripe e que é potencialmente letal é a pneumonia bacteriana secundária (94). É sabido que a colonização por pneumococo no trato respiratório superior é importante, tanto como determinantes da pneumonia quanto da transmissão para a população (95). Apesar de os mecanismos imunológicos da relação entre a infecção pelo vírus da Gripe Influenza e a subsequente colonização por pneumococo ser bem estudados em camundongos (96, 97), pouco se sabe como isso ocorre em humanos. E umas das principais razões é a falta de um modelo humano controlado. Em uma colaboração com pesquisadores da *Liverpool School of Tropical Medicine*, nós obtivemos dados de transcritoma, citocinas plasmáticas e carga viral/bacteriana coletados de um ensaio clínico onde indivíduos eram infectados com o vírus Influenza na cavidade nasal e depois desafiado com a bactéria *Streptococcus pneumoniae*. A infecção pelo vírus Influenza foi feita usando a vacina de vírus atenuada da Gripe (LAIV) que promove a replicação viral na mucosa nasal. De fato, indivíduos que foram primeiro infectados pelo vírus da Gripe possuíam maior aquisição do pneumococo e uma densidade de colonização nos primeiros 14 dias após a inoculação da bactéria maior do que indivíduos que não foram antes infectados pelo vírus da Gripe. A análise do transcritoma das amostras de lavado nasal mostrou que os pacientes primeiramente

infectados pelo vírus da Gripe possuíam um perfil transcricional diferente dos indivíduos controles (Figura 14).

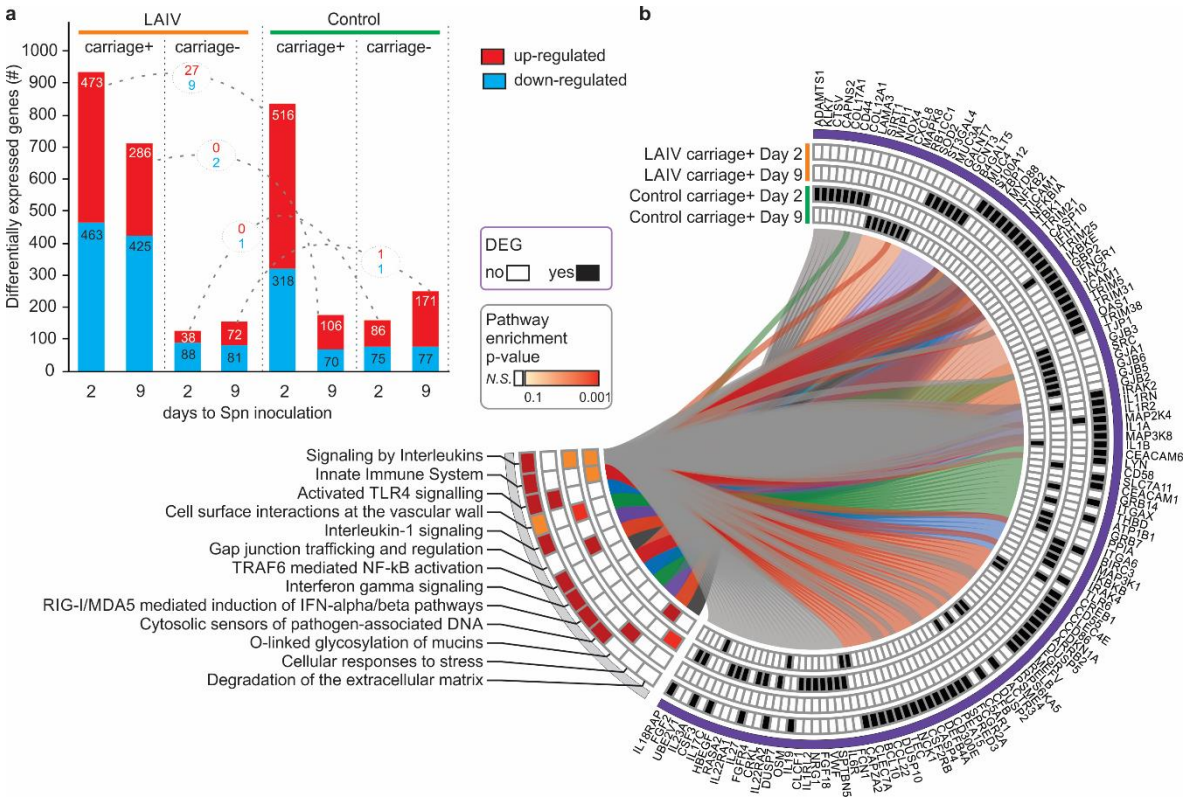

Figura 14. Transcritômica nasal após coinfecção entre vírus da Gripe e a bactéria *Streptococcus pneumoniae*. (a) Número de genes diferencialmente expressos em cada tempo experimental para cada grupo. Genes regulados positivamente e negativamente são representados em vermelho e azul, respectivamente. As conexões entre as barras mostram o número de genes comuns entre a LAIV e controle. (b) Representação circular de genes diferencialmente expressos e análise de enriquecimento em indivíduos LAIV e controle colonizados (+). Genes e vias de sinalização estão conectados por linhas coloridas. Extraído de nosso artigo (98).

Para analisar o padrão de co-expressão dos genes neste contexto de infecção, usamos também uma ferramenta que desenvolvemos chamada CEMiTool que está descrita no item 6.1 desta tese. Ela nos permitiu encontrar módulos gênicos relacionados à inoculação da bactéria, assim como à infecção pelo vírus da Gripe. Por exemplo, um módulo (M1) possui uma maior atividade no dia 9 após a inoculação em indivíduos controles que foram colonizados (Figura 15a). Observamos que a atividade do módulo M1 estava correlacionada com os níveis de monócitos sugerindo que os genes de M1estavam refletindo uma infiltração de monócitos ($r = 0,61$, $p = 0,03$, Figura 15b). Para investigar esses monócitos, realizamos uma análise de enriquecimento de genes entre os genes do Módulo M1 e genes de diferentes subconjuntos de monócitos (82). Esses genes foram enriquecidos para monócitos CD14+ CD16- ($p = 0,00002$) e não para outros subconjuntos de monócitos (Figura 15c). Além disso, o módulo M1 estava também associado a "receptores que ligam quimiocinas" e "sinalização por *Interferon* α / β" (Figura 15d). Como a resposta de interferon tipo I mostrou ser necessária para a limpeza do pneumococos em modelos murinos (99), esses achados sugerem a atividade de monócitos pode ser crítica para essa limpeza em humanos. Integrando estes dados de co-expressão com dados de interação proteína-proteína, vimos que os genes CXCL6 e seu receptor CXCR2 são *hubs* neste módulo M1 (Figura 15e).

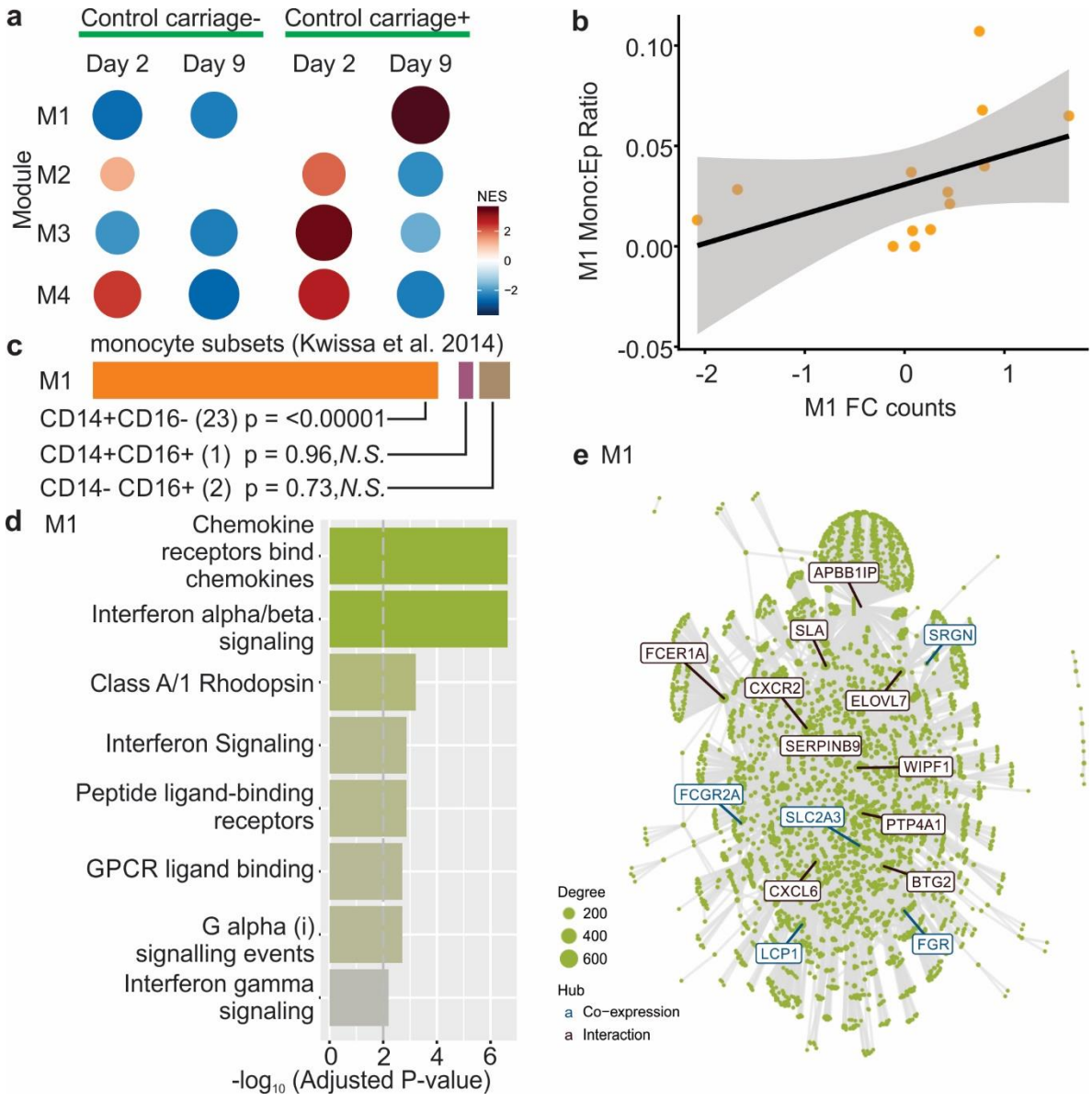

Figura 15. CEMiTool aplicado aos indivíduos controle (não previamente infectados pelo vírus Influenza), Módulo 1. (a) Análise de GSEA mostrando a atividade do módulo em cada tempo experimental para os indivíduos colonizados (+) ou não (-); (b) Correlação entre a expressão do módulo M1 (eixo x) e os níveis de monócitos (eixo y) no dia 9; (c) O módulo M1 contém genes altamente expressos em CD14+ CD16- (578 genes), CD14+ CD16 + (108 genes), CD14- CD16+ (162 genes). O número de genes entre M1 e subconjuntos de monócitos estão mostrados entre parênteses. A significância foi analisada pelo teste qui-

quadrado; d) Análise de enriquecimento do módulo M1 usando conjuntos de genes da base de dados Reactome; e) Gráfico de interação proteína-proteína para M1, com os genes hub destacados. Extraído de nosso artigo (98).

Este trabalho do qual sou co-autor correspondente foi publicado em 2018 na revista Nature Immunology (98). Nele, identificamos mecanismos celulares e moleculares chaves no controle da colonização pneumocócica em humanos, e como este controle é interrompido após infecção pelo vírus da Gripe na mucosa nasal. O estudo foi amplamente divulgado em revistas de divulgação científica e até em rádios e TV.

3.4. Outras doenças infecciosas

Também aplicamos a Imunologia de Sistemas para estudar os aspectos imunológicos (100, 101) e a cardiopatia (102-104) relacionada à Doença de Chagas. Esta foi uma colaboração com pesquisadores da UFRJ, da USP de Ribeirão Preto e da Faculdade de Medicina da USP. Em uma outra colaboração com pesquisadores da UFMG, ajudamos a investigar a relação entre o ciclo celular do parasita da malária com o ritmo circadiano dos hospedeiros vertebrados (105).

4. Aplicação da Biologia de Sistemas em outras doenças humanas

Esta seção descreve os resultados obtidos em projetos de Imunologia de Sistemas em câncer e na artrite idiopática juvenil. Apenas o resumo do trabalho e os achados principais serão descritos. A metodologia é brevemente mencionada, mas os detalhes dela se encontram nas publicações citadas.

4.1. Artrite idiopática juvenil

Nosso grupo também faz parte do CRID (*Center for Research on Inflammatory Diseases*), um CEPID da FAPESP que investiga doenças inflamatórias. Nosso objetivo é encontrar genes e vias de sinalização que possam estar relacionados a estas doenças. Uma delas é a artrite juvenil idiopática e a artrite reumatoide. Para a primeira, estamos preparando um manuscrito para submissão e para a segunda, um artigo já foi publicado em 2018 (106) no qual participamos como colaborador.

A Artrite Idiopática Juvenil (JIA) é um grupo de condições patológicas caracterizadas pela inflamação crônica das articulações em indivíduos com idade inferior a 16 anos (107). JIA é subclassificada baseado na severidade da doença entendida como a quantidade de áreas do corpo afetadas com inflamação crônica (108). Por ser idiopática, diversos estudos de transcritoma de pacientes diagnosticados com JIA foram realizados a fim de encontrar potenciais genes envolvidos na regulação transcricional na doença (109-115). No entanto, devido à enorme heterogeneidade individual dos pacientes (raça, idade, microbiota, status nutricional, severidade da doença, comorbidade, genética, etc) e do ruído e variação intrínseca das técnicas em larga escala e dos métodos analíticos utilizados (9), ainda não há uma assinatura robusta e consistente que caracterize a JIA. Ou seja, apesar de cada um desses estudos ter identificado genes relacionados à JIA, há apenas uma pequena sobreposição de genes e vias de sinalização em comum entre eles. Nós realizamos uma meta-análise de dezenas de estudos que continham dados “ômicos” para doença a fim de encontrar uma assinatura consistente de JIA.

Utilizamos o banco de dados público *Gene Expression Omnibus* (GEO), selecionamos 22 conjuntos de dados de expressão gênica de pacientes diagnosticados com JIA publicados no entre 2007 e 2015 (Figura 16). Em vários destes estudos, os autores re-utilizaram amostras sem deixar explícito este fato. Nós detectamos estes casos para evitar qualquer viés nas análises (Figura 16).

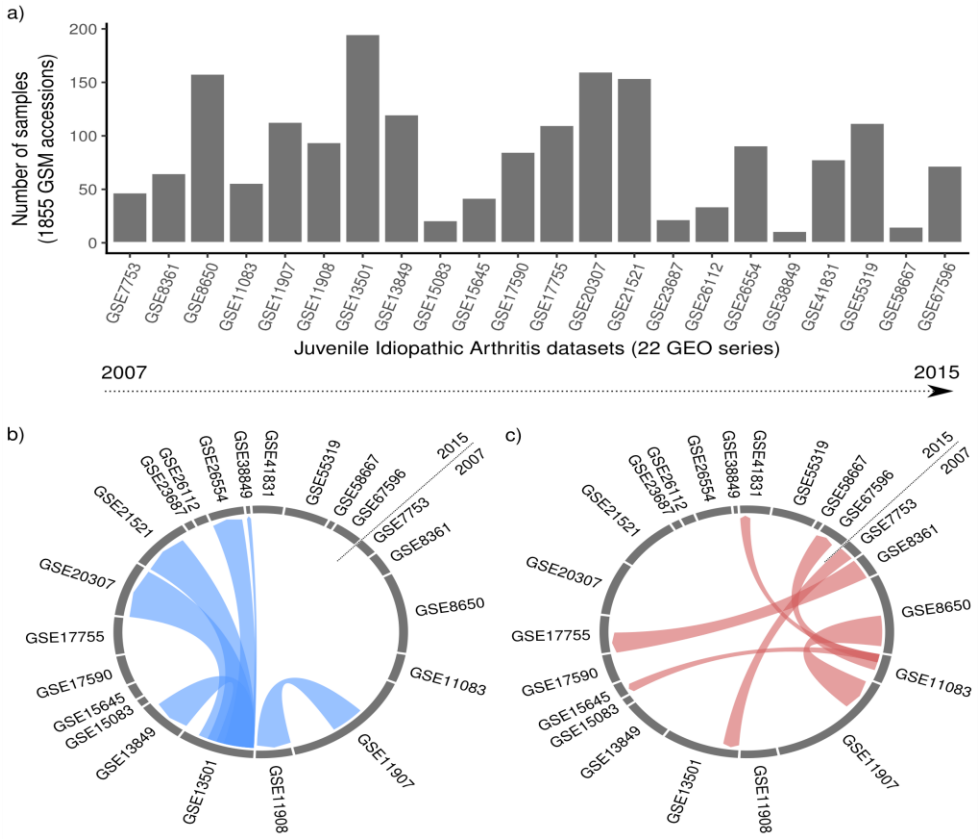

Figura 16. Dados de expressão gênica de microarrays relacionado a Artrite Idiopática Juvenil (JIA) disponível no *Gene Expression Omnibus* (GEO). (a) Número de amostras pelo *GEO series* que incluem amostras com JIA. As séries são organizadas de acordo com a data de publicação. (b) Amostras que foram explicitamente copiados entre *GEO series* (com o mesmo número *GSM accession*) ou (c) não copiados explicitamente (amostra com diferentes números *GSM accession* mas dados brutos com a mesma assinatura *md5sum*). As

setas representam a relação de amostras copiadas entre os *GEO series*. A espessura da seta é proporcional ao número de amostras copiadas.

A análise de expressão diferencial para cada conjuntos de dados revela um número heterogêneo de DEGs (Figura 17a). Apenas 55 desses genes foram diferencialmente expressos em pelo menos 15 conjuntos de dados (Figura 17b). Destes, 26 genes tiveram sua expressão consistentemente aumentada e 29 diminuídos na doença comparado com os controles (Figura 17c). Como havia um número desbalanceado de conjuntos de dados por classe de JIA, assim como uma disparidade do número de amostras entre conjuntos de dados, nós reunimos pares de conjuntos de dados e aplicamos uma meta-análise para todas as combinações de pares para comparar a quantidade de genes diferencialmente expressos entre as classes de JIA. O número de DEGs correlacionou-se com a gravidade da classe de JIA (Figura 17d). Para comparar os DEGs entre as classes JIA, primeiro identificamos os DEGs consistentes entre todos os conjuntos de dados disponíveis por classe. Apenas poucos genes estavam alterados da mesma forma entre as diferentes classes de JIA (Figura 17e). O maior cruzamento ocorreu entre as classes sJIA/pJIA e sJIA/eJIA. Em contraste, eJIA/pJIA têm poucos genes perturbados em comum. Curiosamente, a nossa meta-análise revelou que a expressão do gene ELOVL7 estava aumentada de forma consistente entre diversas classes de JIA (Figura 17f). É importante mencionar que ELOVL7 não foi mencionado como relevante para JIA em nenhum artigo publicado, demonstrando que novos conhecimentos podem ser obtidos a partir da re-análise de dados públicos. Por outro lado, a expressão diminuída e consistente de RASSF1 nas classes de JIA (Figura 17c e 17f), assim como a alteração do gene S100A12, valida nossa meta-análise já que ambos são sabidamente relacionados à doença.

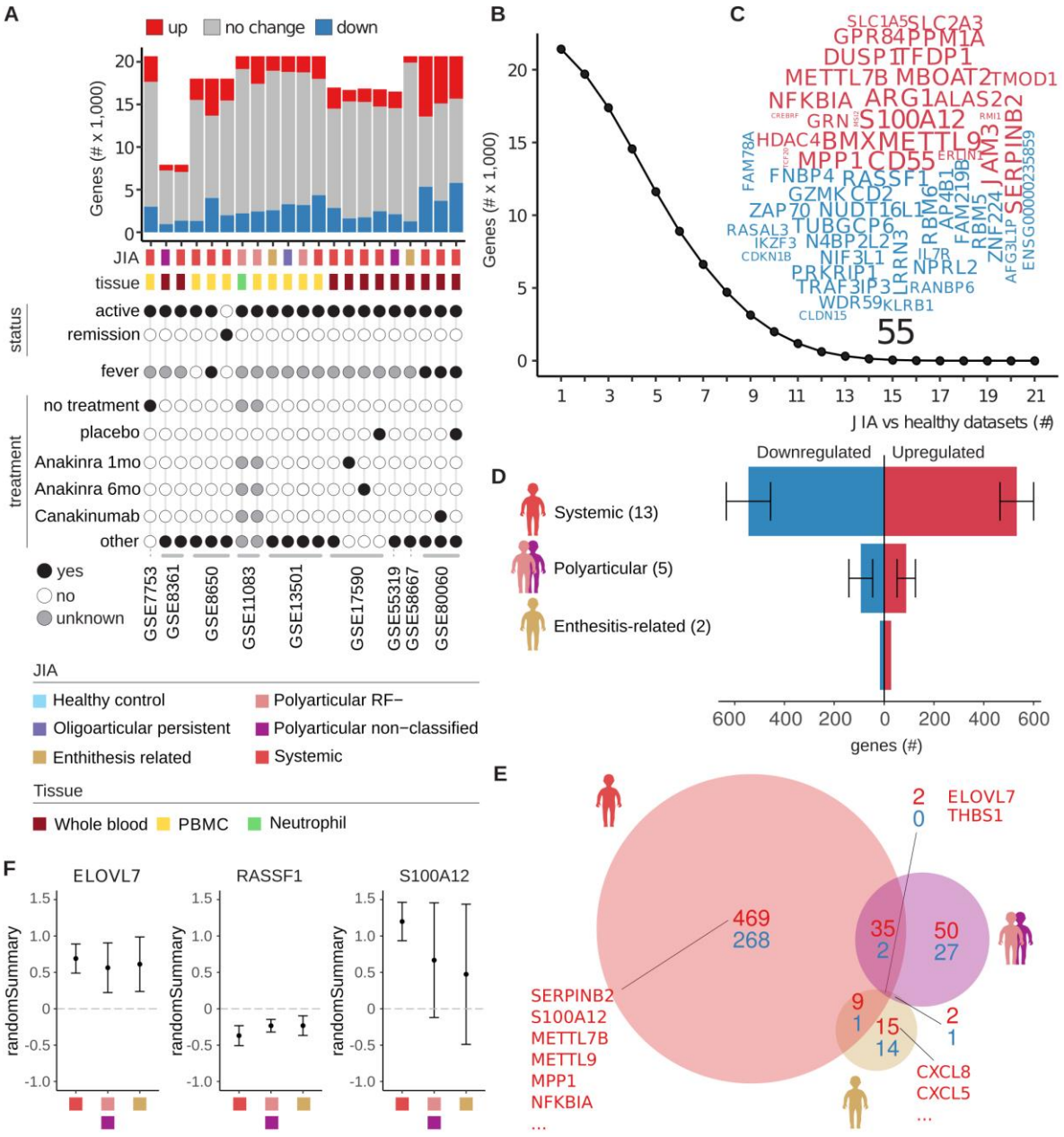

Figura 17. Meta-análise de expressão gênica da Artrite Idiopática Juvenil. (a) Número de genes diferencialmente expressos por conjunto de dados ($p < 0,05$); (b) Distribuição cumulativa reversa dos genes diferencialmente expressos ao longo dos 21 conjuntos de dados da JIA; (c) *Word cloud* dos 55 genes que foram diferencialmente expressos em pelo menos

15 conjuntos de dados. A cor da palavra representa a direção de sua regulação (vermelho: aumentada e azul: diminuída) e o tamanho da letra é proporcional ao número de conjuntos de dados em que o gene estava diferencialmente expresso; (d) Meta-análise para as diferentes classes de JIA utilizando um modelo de efeitos aleatórios (REM) para incorporar a heterogeneidade dos dados. O número médio de genes consistentemente diferencialmente expressos (REM p <0,05) dos pares combinatórios de conjuntos de dados por classe de JIA; (e) Diagrama de Venn dos genes identificados em (d); (f) Valores de REM e intervalos de confiança dos genes ELOVL7, RASSF1 e S100A12.

Realizamos ainda análise de co-expressão e fizemos análises independentes pelo gênero dos pacientes e encontramos assinaturas específicas para homens e mulheres com JIA. Estes resultados irão mostrar mecanismos de regulação inéditos relacionados à Artrite Idiopática Juvenil.

4.2. Câncer

Demonstramos também que a estratégia de análise de transcritoma e de miRNAs pode ser usada para estudar diferentes tipos de câncer, como glioblastoma (116, 117), melanoma (118, 119), tireoide (120) e dezenas de outros (121). Estes trabalhos foram frutos de colaborações com cientistas de diferentes instituições brasileiras e internacionais.

Em um artigo publicado na revista *Nature* em 2016 (122), ajudamos a identificar e caracterizar uma população de células T CD8+ que aumentava a proliferação após o bloqueio do receptor PD-1. Em infecções virais crônicas (123) e no câncer (124), parte das células T CD8+ que são antígeno-específica estão funcionalmente “exaustas”. Como estas células

exaustas aumentam a expressão dos receptores inibitórios PD-1 e CTLA-4 (dentre outros receptores importantes), existem hoje terapias para o câncer que usam anticorpos direcionados a estes receptores (125, 126). Já foi mostrado que a inibição de ambos os receptores CTLA-4 e PD-1 resulta em aumento da sobrevida de pacientes quando comparado com quimioterapias convencionais (127). Usando um modelo de camundongo e infecção crônica por LCMV, nossos colaboradores da *Emory University* isolaram uma população de células T CD8+ que expressavam o receptor CXCR5, um receptor de quimiocina normalmente presente em linfócitos B e T CD4+ *helper* foliculares. Nós então analisamos o perfil transcricional das populações que expressavam ou não o receptor CXCR5 e vimos que a população CXCR5+ quando comparado com a população CXCR5- possuía níveis maiores de várias moléculas co-estimulatórias (*Cd28*, *Icos*, *Tnfsf14*, *Tnfrsf4*) e menores níveis de receptores inibitórios (*Cd244*, *Havcr2*, *Entpd1*, *Lag3*) (Figura 18). Uma análise de GSEA (*Gene Set Enrichment Analysis*) usando assinaturas de células T CD4+ *helper* foliculares, células T CD8+ precursores ou efetores terminais e de células hematopoiéticas progenitoras ou maduras mostrou que as células CXCR5+ pareciam funcionar como células-tronco de memória durante a infecção crônica (Figura 18). Esses achados forneceram um melhor entendimento de como funciona a exaustão de células T e têm implicações na otimização da imunoterapia dirigida por PD-1 em infecções crônicas e câncer (122).

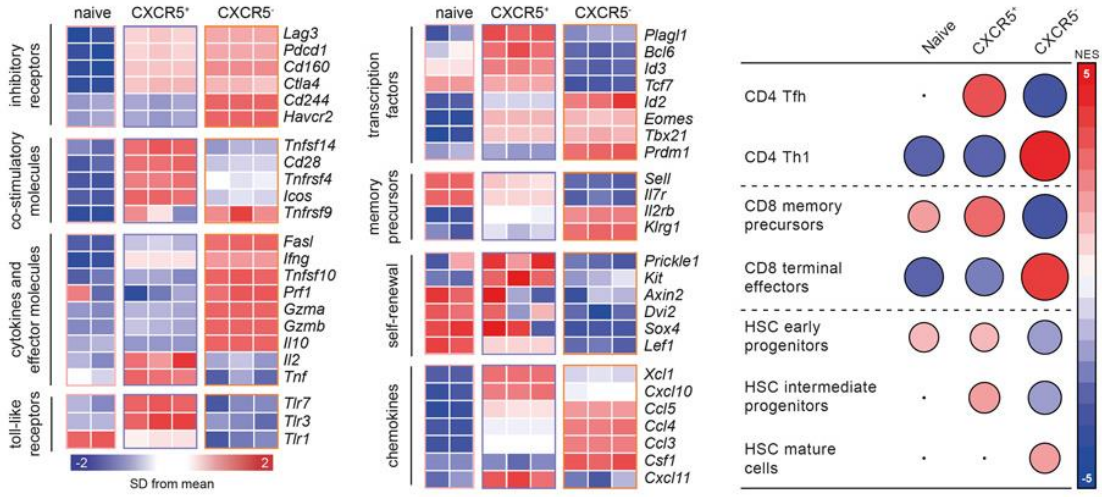

Figura 18. Identificação de uma população de células T PD-1+ CD8+ durante infecção crônica por LCMV que possui uma assinatura gênica semelhante a células T CD4+ Tfh e a células T CD8+ precursoras de memória. O mapa de calor à esquerda ilustra a expressão relativa de genes em células T CD8+ naïve de camundongos não infectados e células T CD8+ de camundongos cronicamente infectados com LCMV (por mais de 45 dias) e CXCR5+ Tim-3- PD-1+ ou CXCR5- Tim-3+ PD-1+. O gráfico à direta mostra os resultados de GSEA dos dois subconjuntos de células T CD8+ em comparação com assinaturas gênicas diferentes. HSC, células-tronco hematopoiéticas; NES, escore de enriquecimento normalizado. Adaptado do nosso artigo (122).

5. Desenvolvimento de ferramentas computacionais de Imunologia de Sistemas

Como demonstrado, a Imunologia de Sistemas é a estratégia ideal para a análise de redes biológicas complexas relacionadas ao sistema imune. No entanto, para realizar análises deste tipo, imunologistas precisam de um profundo conhecimento em bioinformática, dominando pelo menos uma ou duas linguagens de programação. Esse fato limita que muitas análises neste campo sejam realizadas. Nos últimos anos, nosso laboratório tem desenvolvido

diversas ferramentas computacionais que auxiliam as análises dos estudos de imunologia de sistemas. Suas aplicações vão desde a análise de co-expressão em estudos de transcritômica em pacientes infectados por patógenos ou vacinados até quantificar as diferenças individuais entre pacientes saudáveis ou doentes. Duas ferramentas computacionais amigáveis e *online* foram publicadas ou submetidas para publicação: CEMiTool (https://cemitool.sysbio.tools/) e MDP (https://mdp.sysbio.tools/).

5.1. Ferramenta CEMiTool

Um modo de buscar por genes associados a processos imunológicos é através da análise de co-expressão gênica. Essa análise identifica grupos de transcritos (ou módulos gênicos) cuja expressão está correlacionada em uma ou mais condição biológica. Esses transcritos podem ser regulados pelos mesmos fatores, desempenhar a mesma função e/ou atuar como reguladores de expressão gênica de outros transcritos do mesmo módulo. Para facilitar essa análise criamos uma ferramenta chamada *Co-Expression Module identification Tool* ou CEMiTool.

O CEMiTool é uma ferramenta de identificação de módulos co-expressos a partir de dados de expressão gênica provenientes de *microarrays* ou RNAseq baseada no método WGCNA (128). Inicialmente, o CEMiTool irá traçar uma rede de co-expressão a partir dos dados, onde cada gene será representado por um vértice na rede, e as arestas, o valor de correlação entre o nível de expressão entre cada par de genes i e j (Cij). Tradicionalmente, deve-se então definir um valor de corte para determinar quais arestas da rede são realmente significativas para a estrutura da rede. No entanto, esta dicotomização pode levar a uma perda de informação dependente do valor de *threshold* escolhido. Para evitar isto, os autores do

WGCNA propuseram o uso de um *soft-threshold*, o qual irá pontuar as arestas da rede de acordo com a significância das mesmas. A ponderação das arestas se dá ao utilizar uma função de adjacência, de tal forma que Aij = Cijβ, sendo β um número inteiro positivo.

A escolha do parâmetro β é feita utilizando a propriedade de redes biológicas, em geral, tenderem a seguir o chamado modelo livre de escala. Redes livres de escala são, por definição, aquelas em que a probabilidade de um nó estar conectado com k outros nós (P(k)) decai segundo uma lei de potência, definida por $P(k) \sim k^{-\gamma}$ (128). Consequentemente, uma rede livre de escala possui alguns poucos nós (os chamados hubs) com muito mais conexões que o restante (129). Assim, para cada possível valor de β em um dado intervalo (como padrão, de 1 a 20), a ferramenta traçará a curva de $\log_{10}(p(k))$ e $\log_{10}(k)$. Uma linha reta nesse gráfico (isto é, valores de R^2 próximos de 1) indica topologia livre de escala. De acordo com os autores do WGCNA, escolhe-se como valor de β o menor número que permita com que R^2 seja maior que 0,8, uma vez que valores muito altos de β causarão uma diminuição na conectividade média da rede (128). Apesar de a escolha automática para o valor de β ser disponibilizada pelo pacote em R do WGCNA, nós encontramos que muitas vezes os valores escolhidos não eram satisfatórios de acordo com os próprios critérios estabelecido pelos autores (isto é, possuíam R^2 correspondente abaixo de 0,8). Por este motivo, o CEMiTool implementa um novo método automático de seleção do valor de β, no qual ele seleciona o menor β cujo R^2 seja acima de 0,8, e que a partir dele os seguintes valores de β possuem valores correspondentes de R^2 dentro de um intervalo específico de amplitude.

Escolhido um valor para β, a ferramenta irá realizar a separação dos módulos através do uso de uma matriz de sobreposição topológica, isto é, será determinada uma medida de similaridade entre cada par de genes a partir do número de vizinhos em comum que os genes

possuem entre si, e em seguida esta medida será utilizada como distância em uma clusterização hierárquica, cujo dendrograma resultante definirá os módulos da rede.

A ferramenta CEMiTool é capaz de não apenas identificar módulos gênicos de co-expressão mas também de integrar estes módulos com dados de interação proteína-proteína, além de realizar análises funcionais (Figura 19). Mais importante, esta ferramenta faz a análise de forma automatizada. Isso permitiu realizar análises modulares de co-expressão em mais de 1.000 estudos de transcritoma, algo nunca antes realizado (130).

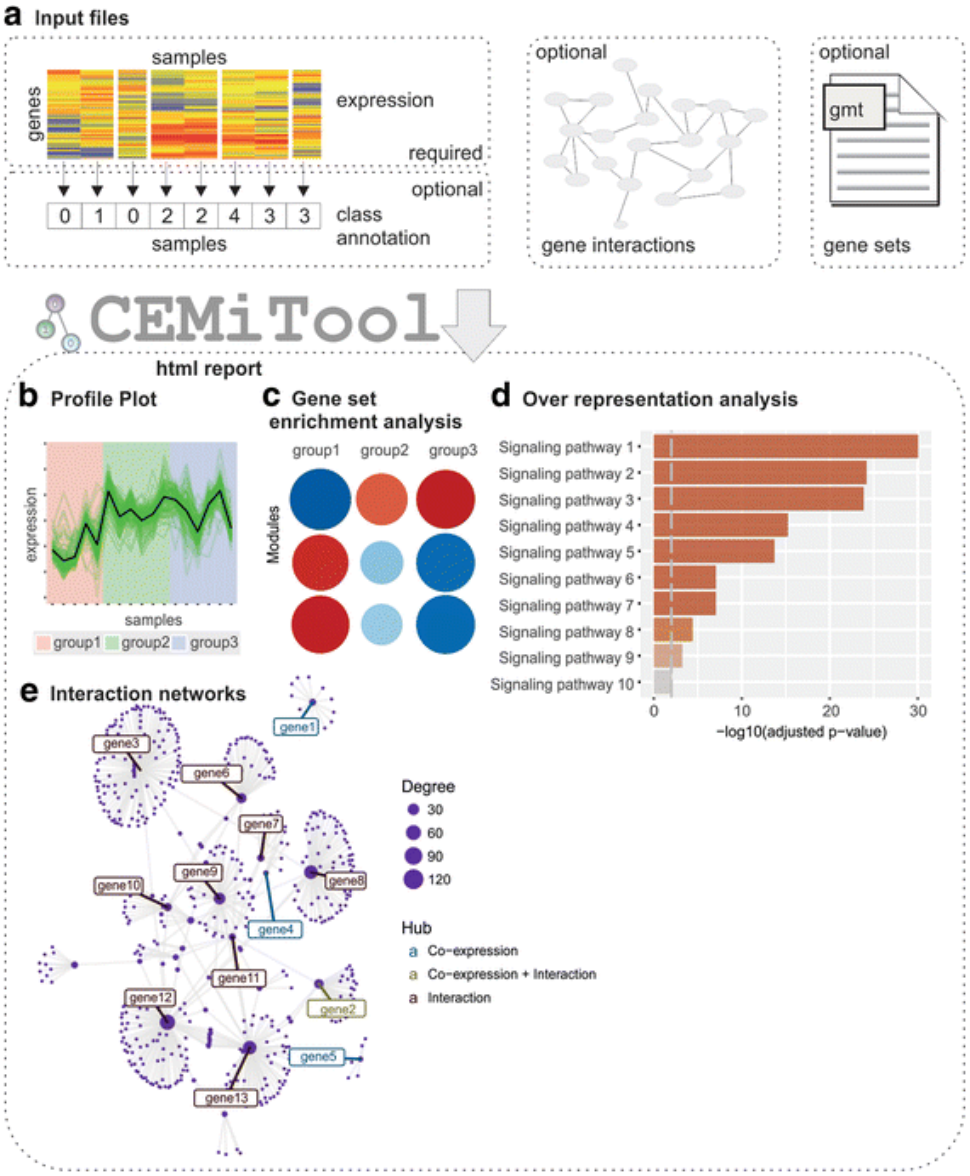

Figura 19. Visão geral da ferramenta CEMiTool. (a) Os arquivos de entrada do CEMiTool consistem em um arquivo de expressão gênica (obrigatório) e arquivos opcionais. (b) Caso um arquivo de anotação das amostras (que determine quais grupos elas pertencem) seja usado, o CEMiTool irá mostrar o perfil de expressão de genes individuais nas amostras nos diferentes grupos; (c) Com este arquivo de anotação das amostras, o CEMiTool também realiza uma Análises de Enriquecimento GSEA que mostre a atividade do módulo em cada grupo de amostras; (d) Se o usuário fornecer um arquivo contendo genes associados a vias de sinalização, o CEMiTool irá executar uma análise de enriquecimento de vias para definir as funções do módulo; (e) Fornecendo um arquivo contendo interações proteína-proteína, a ferramenta irá exibir a rede e os dez principais genes mais conectados (*hubs*). Extraída de nossa publicação (130).

Após o desenvolvimento do pacote em R do CEMiTool disponível no Bioconductor (https://bioconductor.org/packages/release/bioc/html/CEMiTool.html), nós criamos uma ferramenta *web* que permite que usuários realizem as análises sem mesmo conhecer a linguagem de programação R. O artigo que descreve esta versão amigável da ferramenta chamada webCEMiTool foi recentemente publicado na revista *Frontiers in Genetics* (131) em uma sessão especial chamada *User-Friendly Tools Applied to Genetics or Systems Biology* da qual sou um dos editores responsáveis (https://goo.gl/r874ya). Esta nova versão *web* amigável permite que todas as análises sejam realizadas apenas com o uso do *mouse* do computador (Figura 20). Para comprovar seu poder, nós identificamos módulos de co-expressão em dados de RNA-seq de células únicas (*single-cell transcriptomics*).

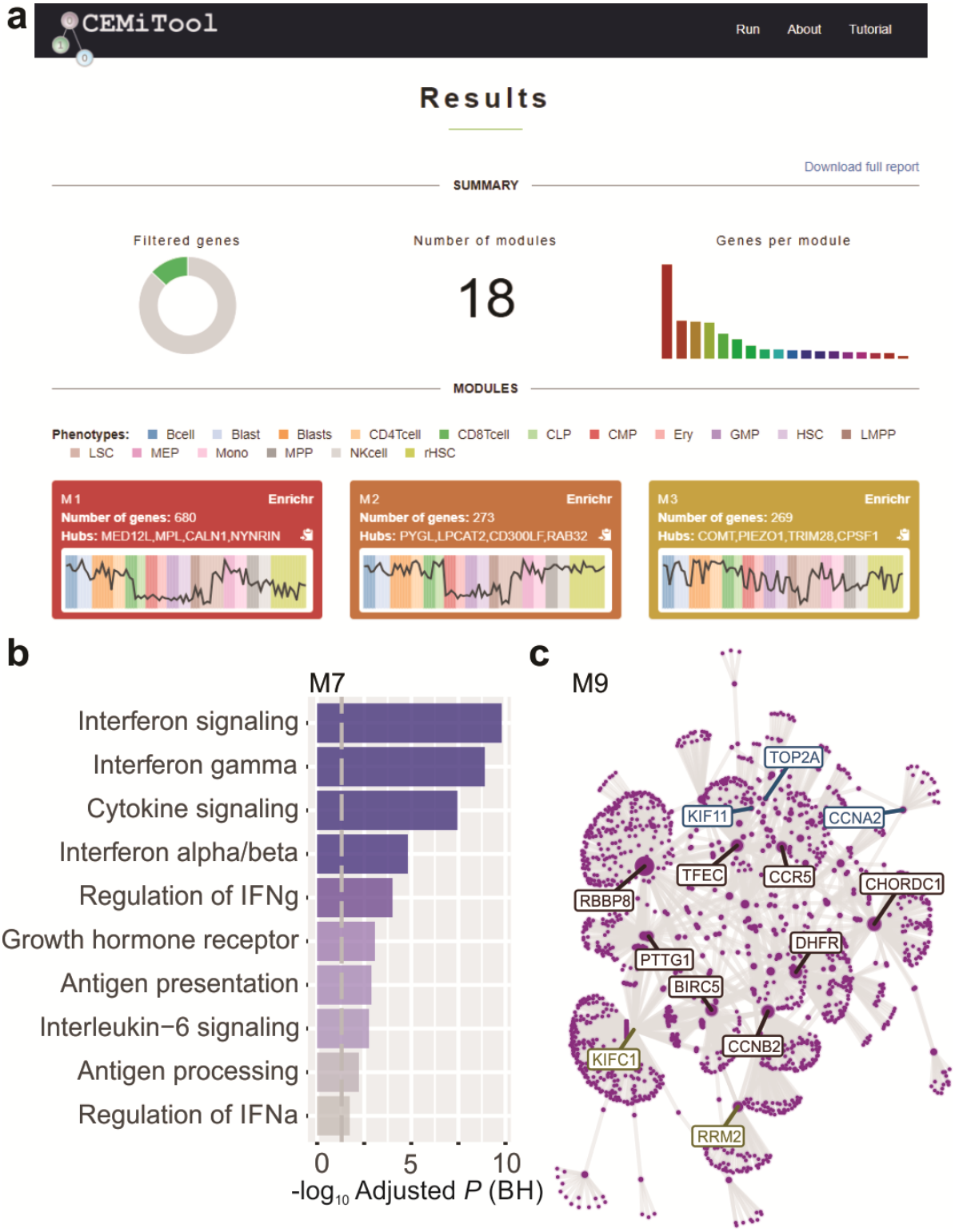
a
CEMiTool
Run
About
Tutorial
Results
Download full report
SUMMARY
Filtered genes
Number of modules
18
Genes per module
MODULES
Phenotypes: Bcell Blast Blasts CD4Tcell CD8Tcell CLP CMP Ery GMP HSC LMPP LSC MEP Mono MPP NKcell rHSC
M1 Enrichr
Number of genes: 680
Hubs: MED12L,MPL,CALN1,NYNRIN
M2 Enrichr
Number of genes: 273
Hubs: PYGL,LPCAT2,CD300LF,RAB32
M3 Enrichr
Number of genes: 269
Hubs: COMT,PIEZO1,TRIM28,CPSF1
b
M7
Interferon signaling
Interferon gamma
Cytokine signaling
Interferon alpha/beta
Regulation of IFNg
Growth hormone receptor
Antigen presentation
Interleukin−6 signaling
Antigen processing
Regulation of IFNa
0
5
10
-log10 Adjusted P (BH)
c
M9
TOP2A
KIF11
CCNA2
TFEC
CCR5
CHORDC1
RBBP8
DHFR
PTTG1
BIRC5
CCNB2
KIFC1
RRM2

Figura 20. Visão geral do webCEMiTool. (a) Resumo dos resultados do webCEMiTool. O gráfico de rosca representa a proporção de genes selecionados pelo nosso filtro de forma não supervisionada. A primeira página também exibe o número de módulos obtidos, bem como um gráfico de barras que mostra o número de genes em cada módulo. Os gráficos de linha ilustram a atividade de expressão mediana de genes dos módulos em cada amostra. As cores representam as diferentes classes ou grupos de amostra. (b) Análise de enriquecimento. O gráfico de barra representa o valor de p ajustado por $-\log_{10}$ (Benjamini-Hochberg) das vias enriquecidas em um módulo (vias de sinalização definidas pelo arquivo .gmt inserido pelo usuário). (c) Rede gênica um módulo. Os principais genes mais conectados (*hubs*) são destacados e coloridos com base na cor do módulo. Extraída de nossa publicação (131).

5.2. Ferramenta MDP

Dado o alto custo dos experimentos com técnicas de larga-escala, estudos de transcritoma normalmente utilizam poucas amostras. Isso faz com que a grande heterogeneidade entre indivíduos causada por fatores ambientais ou doenças não detectadas afete a análise da expressão gênica. Para estudar doenças humanas, a análise inicial geralmente irá buscar genes cuja expressão esteja significativamente alterada na condição perturbada (isto é, pacientes com a doença) em comparação com os indivíduos não perturbados (isto é, os indivíduos saudáveis). No entanto, a definição de saúde e doença é ampla e a variação inerente entre os indivíduos pode tornar qualquer grupo de amostras humanas altamente heterogêneo. Esta variação pode ser devida a fatores genéticos e ambientais, bem como a problemas de saúde não detectados (132, 133). Além disso,

pacientes com a mesma doença podem apresentar grande variação em termos de sintomas ou gravidade (134, 135).

Nós criamos um programa que calcula o grau de perturbação molecular (*Molecular Degree of Perturbation* ou MDP) das amostras de um estudo. Este *score* é baseado no método *molecular distance to health* (11) que classifica as amostras com base em quão distante sua expressão total está em relação às pessoas saudáveis (12-14). A Figura 21a mostra como é calculado o MDP. Mais importante, nós criamos um pacote R disponível no Bioconductor (https://bioconductor.org/packages/release/bioc/html/mdp.html) e uma versão *online* amigável da ferramenta (https://mdp.sysbio.tools/) que permite que qualquer usuário, sem nenhuma base em linguagem de programação, possa rodar as análises (Figura 21).

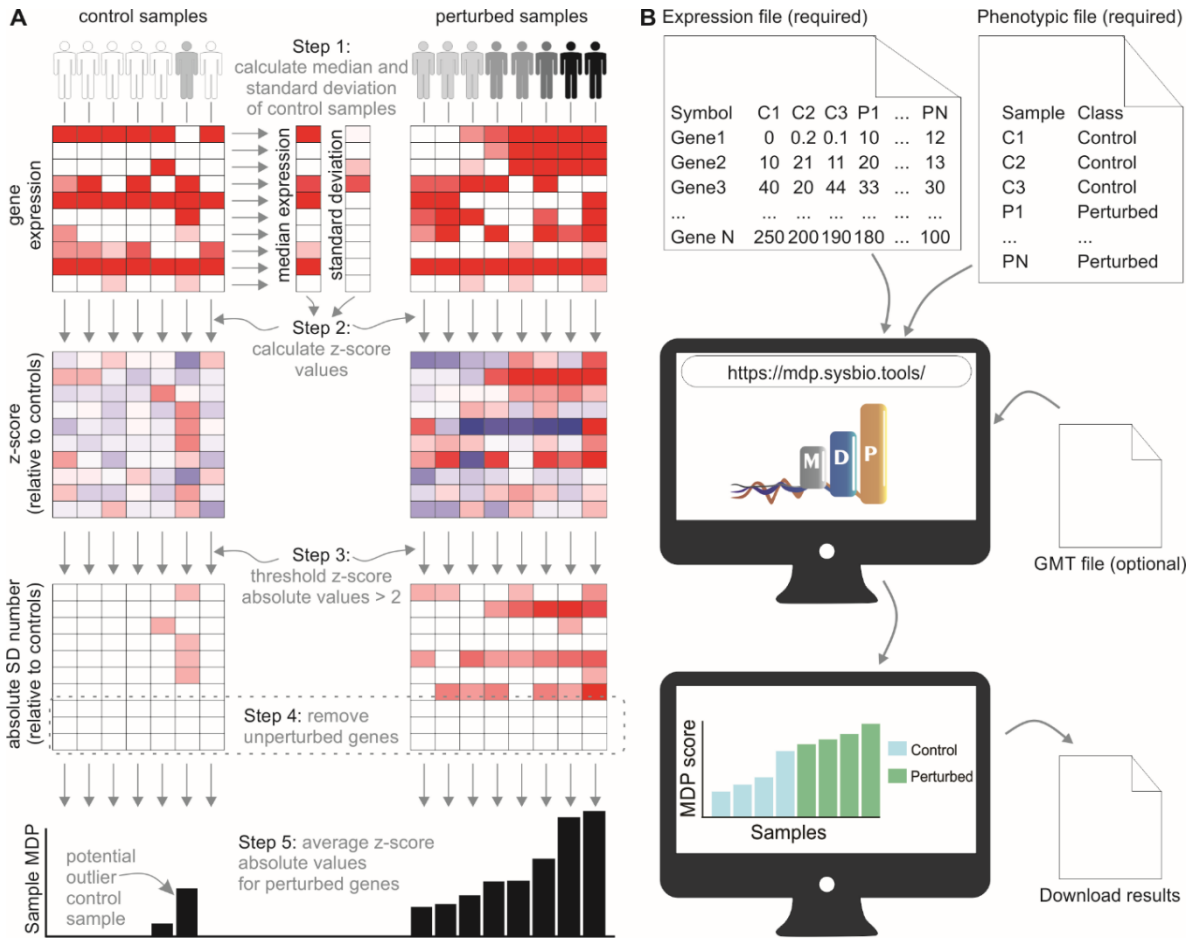

Figura 21. Grau de perturbação molecular (MDP). (a) O algoritmo MDP dá um *score* para as amostras com base em sua perturbação em relação a amostras controle definidas pelo usuário (geralmente indivíduos saudáveis). Uma normalização por *Z-score* é realizada usando as amostras controle como referência. O valor absoluto dos *Z-scores* é calculado e os valores abaixo de 2 são definidos como 0. O *score* da amostra será a média desses valores de *Z-score* para cada amostra. (b) A ferramenta *online* MDP. Arquivos de expressão e fenotípicos são necessários para executar o MDP; os resultados são um simples gráfico de barras e *boxplot* mostrando o MDP para cada amostra submetida. Um recurso opcional permite que os usuários executem o MDP usando um conjunto específico de genes, fornecido como um arquivo .gmt.

Nós aplicamos a ferramenta MDP em dezenas de estudos de transcritoma (11 *microarrays* e 9 RNA-seq) obtidos de bancos de dados públicos. O objetivo era investigar como a heterogeneidade das amostras pode impactar a análise de expressão diferencial. Os estudos são relacionados à tuberculose, câncer, artrite idiopática juvenil, sepse e outras doenças autoimunes e infecciosas. Inicialmente, nós mostramos que o MDP das amostras varia amplamente dentro da mesma doença e entre as diferentes doenças ou tratamentos (Figura 22). A infecção pela bactéria *Staphylococcus aureus*, por exemplo, parece ser uma perturbação mais forte do que a infecção pelo vírus da influenza (Figura 22) (136). Da mesma forma, diferentes tipos de câncer podem apresentar escores de perturbação menores ou maiores, independentemente de seus valores prognósticos conhecidos (Figura 22) (137). Nossa abordagem também diferencia entre vários subtipos de doenças inflamatórias, como artrite idiopática juvenil (JIA), doença de Crohn e retocolite ulcerativa (Figura 22) (138).

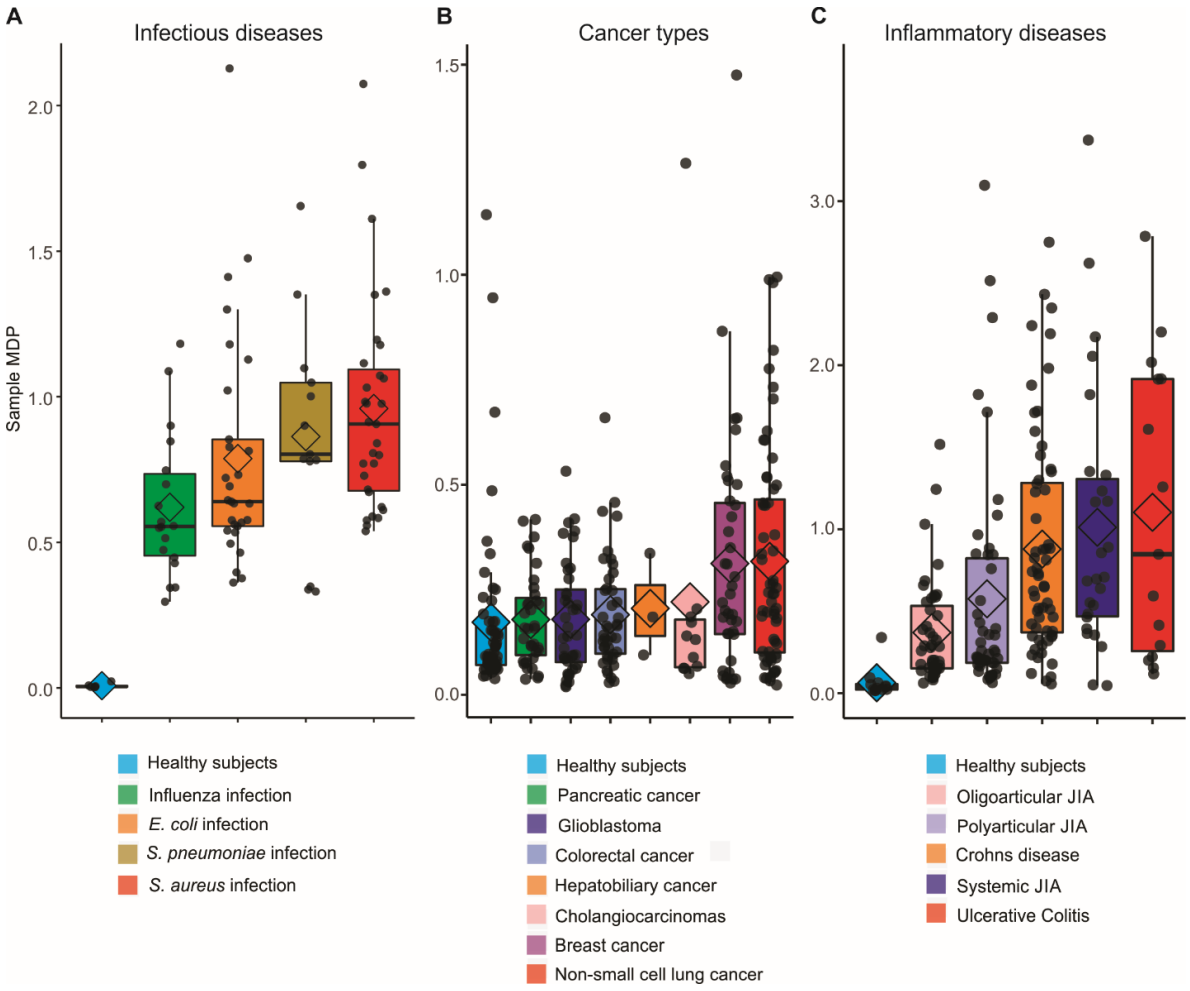

Figura 22. MDP de amostra de diferentes doenças humanas. (a) MDP de pacientes com infecção aguda por vírus ou bactérias. Os dados foram obtidos a partir de leucócitos do sangue e estão disponíveis sob o acesso GEO6269 do GEO. (b) MDP de diferentes tipos de câncer. Os dados foram obtidos a partir de plaquetas e estão disponíveis no âmbito do GEO acesso GSE68086. (c) MDP de pacientes com doenças inflamatórias. Os dados foram obtidos a partir de sangue total e estão disponíveis sob o acesso GEO112057 do GEO. Sujeitos saudáveis (azul) foram usados como grupo de referência.

Ao avaliar o MDP de cada amostra, pudemos identificar possíveis amostras *outliers* para cada um dos 20 estudos de *microarray* e RNA-seq. Um *boxplot* representativo (Figura

23a) mostra que um dos indivíduos saudáveis pode estar de fato “perturbado” quando comparado ao resto do grupo saudável. Da mesma forma, 12 dos pacientes com doença de Crohn não parecem muito perturbados ao nível molecular (Figura 23a). Nós então tratamos essas amostras como *outliers* e as removemos da análise de expressão diferencial. Isso fez com que o número de DEGs aumentasse (Figura 23b). Para a comparação do GSE112057 entre indivíduos saudáveis e pacientes com doença de Crohn, identificamos 188 DEGs antes da remoção de *outliers* (Figura 23b). Após a remoção, o número de DEGs para essa comparação foi de 3.477 (aumento de mais de 18 vezes). Também removemos aleatoriamente o mesmo número de amostras consideradas *outliers* e contamos o número de DEGs para cada comparação. Este processo foi repetido 1.000 vezes. Isso mostrou que o aumento no número DEG não se deve à chance aleatória (Figura 23b). Realizamos esta análise também para as 19 outras comparações. Em todos eles, o número de DEGs aumentou após a remoção dos potenciais outliers (Figura 23b).

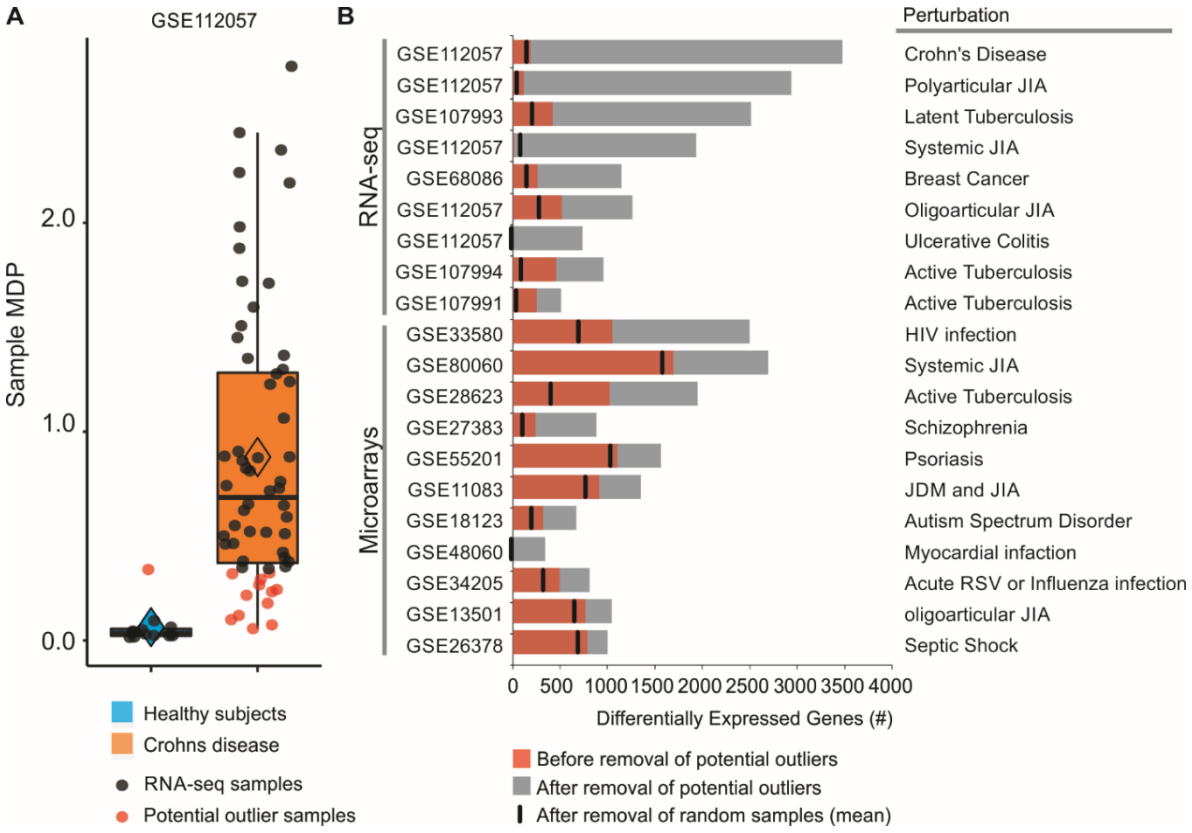

Figura 23. Remoção de possíveis amostras *outlier* impacta as análises de expressão diferencial. (a) MDP de 60 pacientes com doença de Crohn, utilizando como grupo de referência 12 indivíduos saudáveis. Os dados foram obtidos a partir de sangue total e estão disponíveis sob o acesso GEO112057 do GEO. Sujeitos saudáveis (azul) foram usados como grupo de referência. Amostras *outlier* potenciais são mostradas como pontos vermelhos. (b) Análises de expressão diferencial entre pacientes com uma doença e controles saudáveis. O número de DEGs antes e depois da remoção de possíveis amostras *outlier* é mostrado como barras vermelha e preta, respectivamente. A remoção aleatória de amostras seguida de análise de expressão diferencial foi realizada 1.000 vezes para cada comparação e o número de DEGs foi calculado pela média (linha vertical preta).

Nós também aplicamos a ferramenta MDP para dados de *single-cell RNA-seq* e mostramos que há uma grande heterogeneidade molecular ao nível de célula individual (todas da mesma linhagem). Este trabalho apresentou uma análise sistemática da heterogeneidade das amostras em uma variedade de estudos de expressão gênica relacionados a doenças inflamatórias e infecciosas. O aumento do número de DEGs serviu para encontrar novos *insights* imunológicos. O trabalho está em revisão na revista *Frontiers in Genetics*.

5.3. Outras ferramentas computacionais

Também desenvolvemos outras ferramentas computacionais que podem ser aplicadas para diversas funções (Figura 24). Uma delas é a ferramenta MetaVolcano (https://metavolcano.sysbio.tools) que resume e gera gráficos de análise de meta-análise. Outras podem servir para análise de geolocalização espacial de *hotspots* de transmissão de doenças infecciosas (SiPoS), reconhecimento de imagens microscópicas usando técnicas de aprendizado de máquina (MIRe), análise da estrutura secundária de sequências de nucletotídeos (StructRNA finder) (139), dentre outras.

Figura 24. Logotipo das ferramentas computacionais desenvolvidas pelo laboratório.

6. Discussão

O barateamento dos experimentos e a modernização das técnicas atuais, assim como o surgimento de outras tecnologias de larga escala, irão fazer com que a quantidade de dados biológicos ainda aumente muito nos próximos anos. A tecnologia de *single-cell transcriptomics*, por exemplo, permite medir a expressão de todos os genes ao nível de uma única célula (140). A tecnologia de citometria de massa permite caracterizar o fenótipo das células usando dezenas de marcadores e, assim, encontrar inúmeras populações de células envolvidas em processos biológicos (141). A espectrometria de massa está continuamente evoluindo facilitando a medição em paralelo de milhares de proteínas, lipídios e metabólitos dentros e fora das células (142). Bibliotecas de CRISPR permitem fazer um *screening*

funcional de todos os genes numa determinada condição biológica (143). E métodos como Hi-C e derivados de 3C (*chromosome conformation capture*) revelam a organização tridimensional da cromatina e como isso contribui com os processos celulares (144). Mais do que nunca precisaremos de técnicas e conhecimento computacional para analisar este verdadeiro "mar de dados".

Mesmo com o enorme avanço da inteligência artificial e das técnicas de aprendizado de máquina, ainda não há métodos automáticos de análise de dados que permitam encontrar as hipóteses e *insigths* que devem ser testadas experimentalmente. A ciência ainda precisa da intuição, expertise e o modo de pensar não linear de pesquisadores humanos. Fazer as perguntas certas é mais importante do que métodos computacionais que ajudam em respondê-las. A Imunologia de Sistemas é o caminho ideal para integrar os dados e fazer com que surjam as perguntas certas (*data-driven hypothesis*).

7. Conclusão

Esta tese mostrou como a Imunologia de Sistemas pode ser aplicada para estudar diversas doenças humanas e entender melhor os mecanismos moleculares por trás das respostas imunes. Porém, ainda há um longo caminho pela frente.

8. Referências

1. Takeuchi O & Akira S (2009) Innate immunity to virus infection. *Immunol Rev* 227(1):75-86.
2. Kieser KJ & Kagan JC (2017) Multi-receptor detection of individual bacterial products by the innate immune system. *Nat Rev Immunol* 17(6):376-390.
3. Baccala R, *et al.* (2009) Sensors of the innate immune system: their mode of action. *Nature Reviews Rheumatology* 5(8):448-456.
4. Kouwaki T, *et al.* (2016) Extracellular Vesicles Including Exosomes Regulate Innate Immune Responses to Hepatitis B Virus Infection. *Front Immunol* 7:335.
5. Silveira ELV, Dominguez MR, & Soares IS (2018) To B or Not to B: Understanding B Cell Responses in the Development of Malaria Infection. *Front Immunol* 9.
6. Annunziato F, Romagnani C, & Romagnani S (2015) The 3 major types of innate and adaptive cell-mediated effector immunity. *J Allergy Clin Immun* 135(3):626-635.
7. Laiosa CV, Stadtfeld M, & Graf T (2006) Determinants of lymphoid-myeloid lineage diversification. *Annual review of immunology* 24:705-738.
8. Furman D & Davis MM (2015) New approaches to understanding the immune response to vaccination and infection. *Vaccine* 33(40):5271-5281.
9. Nakaya HI, Li S, & Pulendran B (2012) Systems vaccinology: learning to compute the behavior of vaccine induced immunity. *Wiley Interdiscip Rev Syst Biol Med* 4(2):193-205.
10. Nakaya HI & Pulendran B (2012) Systems vaccinology: its promise and challenge for HIV vaccine development. *Curr Opin HIV AIDS* 7(1):24-31.
11. Nakaya H (2013) Systems Biology of Infectious Diseases and Vaccines. *Computational Systems Biology*, eds Kriete A & Eils R (Elsevier), Vol 2, p 548.
12. Hayden EC (2014) Technology: The $1,000 genome. *Nature* 507(7492):294-295.
13. Delhalle S, Bode SFN, Balling R, Ollert M, & He FQ (2018) A roadmap towards personalized immunology. *NPJ Syst Biol Appl* 4:9.
14. Ideker T, Galitski T, & Hood L (2001) A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics* 2:343-372.
15. Kitano H (2002) Computational systems biology. *Nature* 420(6912):206-210.
16. Kitano H (2002) Systems biology: a brief overview. *Science* 295(5560):1662-1664.
17. Pulendran B, Li S, & Nakaya HI (2010) Systems vaccinology. *Immunity* 33(4):516-529.
18. Casadevall A & Pirofski LA (2003) The damage-response framework of microbial pathogenesis. *Nature Reviews Microbiology* 1(1):17-24.
19. Aderem A, *et al.* (2011) A Systems Biology Approach to Infectious Disease Research: Innovating the Pathogen-Host Research Paradigm. *Mbio* 2(1).
20. Lever M, Hirata TD, Russo P, & Nakaya H (2018) Systems Immunology. *Theoretical and Applied Aspects of Systems Biology*, eds da Silva FAB, Carels N, & Junior FPS (Springer), Vol 1.
21. Pulendran B (2009) Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. *Nat Rev Immunol* 9(10):741-747.
22. Pulendran B & Ahmed R (2011) Immunological mechanisms of vaccination. *Nat Immunol* 12(6):509-517.
23. Querec TD, *et al.* (2009) Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* 10(1):116-125.

24. Nakaya HI, *et al.* (2011) Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 12(8):786-795.
25. Li SZ, Nakaya HI, Kazmin DA, Oh JZ, & Pulendran B (2013) Systems biological approaches to measure and understand vaccine immunity in humans. *Seminars in immunology* 25(3):209-218.
26. Lee EK (2007) Large-scale optimization-based classification models in medicine and biology. *Annals of Biomedical Engineering* 35(6):1095-1109.
27. Lever M, Silveira EL, & Nakaya HI (2017) Systems Vaccinology Applied to DNA Vaccines: Perspective and Challenges. *Curr Issues Mol Biol* 22:1-16.
28. Laffleur B & Basu U (2019) Biology of RNA Surveillance in Development and Disease. *Trends Cell Biol*.
29. Nakaya HI, *et al.* (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol* 8(3):R43.
30. Louro R, *et al.* (2007) Androgen responsive intronic non-coding RNAs. *Bmc Biology* 5.
31. Reis EM, *et al.* (2004) Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene* 23(39):6684-6692.
32. Zhou XK, Li XF, & Wu M (2018) miRNAs reshape immunity and inflammatory responses in bacterial infection. *Signal Transduct Tar* 3.
33. Drury RE, O'Connor D, & Pollard AJ (2017) The Clinical Application of MicroRNAs in Infectious Disease. *Front Immunol* 8:1182.
34. Fernandes JCR, Acuna SM, Aoki JI, Floeter-Winter LM, & Muxel SM (2019) Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA* 5(1).
35. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, & Kinzler KW (2008) The antisense transcriptomes of human cells. *Science* 322(5909):1855-1857.
36. Cabili MN, *et al.* (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25(18):1915-1927.
37. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, & Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A* 105(2):716-721.
38. Dinger ME, *et al.* (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 18(9):1433-1445.
39. van Leeuwen S & Mikkers H (2010) Long non-coding RNAs: Guardians of development. *Differentiation* 80(4-5):175-183.
40. Wilhelm BT, *et al.* (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453(7199):1239-U1239.
41. Birney E, *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799-816.
42. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, & Mattick JS (2011) lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* 39(Database issue):D146-151.
43. Wang KC, *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472(7341):120-124.

44. Sleutels F, Zwart R, & Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415(6873):810-813.
45. Rinn JL, *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129(7):1311-1323.
46. Lee JT, Davidow LS, & Warshawsky D (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 21(4):400-404.
47. Zhang LF, Huynh KD, & Lee JT (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* 129(4):693-706.
48. Petruk S, *et al.* (2006) Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* 127(6):1209-1221.
49. Tasheva ES & Roufa DJ (1995) Regulation of human RPS14 transcription by intronic antisense RNAs and ribosomal protein S14. *Genes Dev* 9(3):304-316.
50. Yochum GS, Cleland R, McWeeney S, & Goodman RH (2007) An antisense transcript induced by Wnt/beta-catenin signaling decreases E2F4. *J Biol Chem* 282(2):871-878.
51. Spigoni G, Gedressi C, & Mallamaci A (2010) Regulation of Emx2 expression by antisense transcripts in murine cortico-cerebral precursors. *Plos One* 5(1):e8658.
52. Yan MD, *et al.* (2005) Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. *Human molecular genetics* 14(11):1465-1474.
53. Watanabe T, *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453(7194):539-543.
54. Feng JC, *et al.* (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes & Development* 20(11):1470-1484.
55. Tripathi V, *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39(6):925-938.
56. Wong LH, *et al.* (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Research* 17(8):1146-1160.
57. Sunwoo H, *et al.* (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Research* 19(3):347-359.
58. Li ZH & Rana TM (2014) Decoding the noncoding: Prospective of lncRNA-mediated innate immune regulation. *Rna Biology* 11(8):979-985.
59. Heward JA & Lindsay MA (2014) Long non-coding RNAs in the regulation of the immune response. *Trends in Immunology* 35(9):408-419.
60. Imamura K, *et al.* (2014) Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli (vol 53, pg 393, 2014). *Mol Cell* 54(6):1055-1055.
61. Wang Y, *et al.* (2015) Long noncoding RNA derived from CD244 signaling epigenetically controls CD8(+) T-cell immune responses in tuberculosis infection. *Proceedings of the National Academy of Sciences of the United States of America* 112(29):E3883-E3892.
62. Wang P, *et al.* (2014) The STAT3-Binding Long Noncoding RNA lnc-DC Controls Human Dendritic Cell Differentiation. *Science* 344(6181):310-313.

63. Zemmour D, Pratama A, Loughhead SM, Mathis D, & Benoist C (2017) Flicr, a long noncoding RNA, modulates Foxp3 expression and autoimmunity. *Proceedings of the National Academy of Sciences of the United States of America* 114(17):E3472-E3480.
64. Zhang H, *et al.* (2013) Profiling of human CD4(+) T-cell subsets identifies the T(H)2-specific noncoding RNA GATA3-AS1. *J Allergy Clin Immun* 132(4):1005-+.
65. Sasaki S, *et al.* (2011) Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies. *Journal of Clinical Investigation* 121(8):3109-3119.
66. Dunning AJ, *et al.* (2016) Correlates of Protection against Influenza in the Elderly: Results from an Influenza Vaccine Efficacy Trial. *Clinical and Vaccine Immunology* 23(3):228-235.
67. Nakaya HI, *et al.* (2015) Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* 43(6):1186-1198.
68. Jefferson T, Rivetti A, Harnden A, Di Pietrantonj C, & Demicheli V (2008) Vaccines for preventing influenza in healthy children. *Cochrane Database Syst Rev* (2):CD004879.
69. Nakaya HI, *et al.* (2016) Systems biology of immunity to MF59-adjuvanted versus nonadjuvanted trivalent seasonal influenza vaccines in early childhood. *Proc Natl Acad Sci U S A* 113(7):1853-1858.
70. Oh JZ, *et al.* (2014) TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. *Immunity* 41(3):478-492.
71. Ravindran R, *et al.* (2014) Vaccine activation of the nutrient sensor GCN2 in dendritic cells enhances antigen presentation. *Science* 343(6168):313-317.
72. Ravindran R, *et al.* (2016) The amino acid sensor GCN2 controls gut inflammation by inhibiting inflammasome activation. *Nature* 531(7595):523-527.
73. Cohen J, Nussenzweig V, Nussenzweig R, Vekemans J, & Leach A (2010) From the circumsporozoite protein to the RTS, S/AS candidate vaccine. *Hum Vaccines* 6(1):90-96.
74. Agnandji ST, *et al.* (2011) First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children. *New England Journal of Medicine* 365(20):1863-1875.
75. Kazmin D, *et al.* (2017) Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. *Proc Natl Acad Sci U S A* 114(9):2425-2430.
76. Li S, *et al.* (2014) Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 15(2):195-204.
77. Rouphael NG & Stephens DS (2012) Neisseria meningitidis: biology, microbiology, and epidemiology. *Methods Mol Biol* 799:1-20.
78. Li S, *et al.* (2017) Metabolic Phenotypes of Response to Vaccination in Humans. *Cell* 169(5):862-877 e817.
79. Antillon M, *et al.* (2017) The burden of typhoid fever in low- and middle-income countries: A meta-regression approach. *PLoS Negl Trop Dis* 11(2):e0005376.
80. Blohmke CJ, *et al.* (2017) Induction of Cell Cycle and NK Cell Responses by Live-Attenuated Oral Vaccines against Typhoid Fever. *Front Immunol* 8:1276.
81. Wrammert J, *et al.* (2012) Rapid and massive virus-specific plasmablast responses during acute dengue virus infection in humans. *J Virol* 86(6):2911-2918.

82. Kwissa M, *et al.* (2014) Dengue virus infection induces expansion of a CD14(+)CD16(+) monocyte population that stimulates plasmablast differentiation. *Cell Host Microbe* 16(1):115-127.
83. Poo YS, *et al.* (2014) CCR2 Deficiency Promotes Exacerbated Chronic Erosive Neutrophil-Dominated Chikungunya Virus Arthritis. *Journal of Virology* 88(12):6862-6872.
84. Poo YS, *et al.* (2014) Multiple immune factors are involved in controlling acute and chronic chikungunya virus infection. *PLoS Negl Trop Dis* 8(12):e3354.
85. Chen W, *et al.* (2017) Specific inhibition of NLRP3 in chikungunya disease reveals a role for inflammasomes in alphavirus-induced inflammation. *Nat Microbiol* 2(10):1435-1445.
86. Prow NA, *et al.* (2017) Lower temperatures reduce type I interferon activity and promote alphaviral arthritis. *PLoS Pathog* 13(12):e1006788.
87. Walsh D & Mohr I (2011) Viral subversion of the host protein synthesis machinery. *Nature Reviews Microbiology* 9(12):860-875.
88. Costa VV, *et al.* (2017) N-Methyl-d-Aspartate (NMDA) Receptor Blockade Prevents Neuronal Death Induced by Zika Virus Infection. *MBio* 8(2).
89. Caires LC, *et al.* (2018) Discordant congenital Zika syndrome twins show differential in vitro viral susceptibility of neural progenitor cells (vol 9, 2018). *Nat Commun* 9.
90. Barros JBS, *et al.* (2018) Acute Zika Virus Infection in an Endemic Area Shows Modest Proinflammatory Systemic Immunoactivation and Cytokine-Symptom Associations. *Front Immunol* 9:821.
91. Melo ASD, *et al.* (2016) Congenital Zika Virus Infection Beyond Neonatal Microcephaly. *Jama Neurol* 73(12):1407-1416.
92. Harries AD, *et al.* (2011) The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. *Int J Tuberc Lung Dis* 15(11):1436-1444, i.
93. Prada-Medina CA, *et al.* (2017) Systems Immunology of Diabetes-Tuberculosis Comorbidity Reveals Signatures of Disease Complications. *Sci Rep* 7(1):1999.
94. Morris DE, Cleary DW, & Clarke SC (2017) Secondary Bacterial Infections Associated with Influenza Pandemics. *Frontiers in Microbiology* 8.
95. Simell B, *et al.* (2012) The fundamental link between pneumococcal carriage and disease. *Expert Review of Vaccines* 11(7):841-855.
96. Lu YJ, *et al.* (2008) Interleukin-17A mediates acquired immunity to pneumococcal colonization. *Plos Pathogens* 4(9).
97. Lu YJ, *et al.* (2010) GMP-grade pneumococcal whole-cell vaccine injected subcutaneously protects mice from nasopharyngeal colonization and fatal aspiration-sepsis. *Vaccine* 28(47):7468-7475.
98. Jochems SP, *et al.* (2018) Inflammation induced by influenza virus impairs human innate immune control of pneumococcus. *Nat Immunol* 19(12):1299-1308.
99. Parker D, *et al.* (2011) Streptococcus pneumoniae DNA Initiates Type I Interferon Signaling in the Respiratory Tract. *Mbio* 2(3).
100. Oliveira AC, *et al.* (2017) Crucial role for T cell-intrinsic IL-18R-MyD88 signaling in cognate immune response to intracellular parasite infection. *Elife* 6.
101. Silva MC, *et al.* (2018) Canonical PI3Kgamma signaling in myeloid cells restricts Trypanosoma cruzi infection and dampens chagasic myocarditis. *Nat Commun* 9(1):1513.

102. Navarro IC, *et al.* (2015) MicroRNA Transcriptome Profiling in Heart of Trypanosoma cruzi-Infected Mice: Parasitological and Cardiological Outcomes. *PLoS Negl Trop Dis* 9(6):e0003828.
103. Ferreira LR, *et al.* (2017) Blood Gene Signatures of Chagas Cardiomyopathy With or Without Ventricular Dysfunction. *J Infect Dis* 215(3):387-395.
104. Ferreira LRP, *et al.* (2017) Integration of miRNA and gene expression profiles suggest a role for miRNAs in the pathobiological processes of acute Trypanosoma cruzi infection. *Sci Rep-Uk* 7.
105. Hirako IC, *et al.* (2018) Daily Rhythms of TNFα Expression and Food Intake Regulate Synchrony of Plasmodium Stages with the Host Circadian Cycle. *Cell Host Microbe* 23(6):796-808.
106. Peres RS, *et al.* (2018) TGF-beta signalling defect is linked to low CD39 expression on regulatory T cells and methotrexate resistance in rheumatoid arthritis. *J Autoimmun* 90:49-58.
107. Prakken B, Albani S, & Martini A (2011) Juvenile idiopathic arthritis. *Lancet* 377(9783):2138-2149.
108. Petty RE, *et al.* (2004) International league of associations for rheumatology classification of juvenile idiopathic arthritis: Second revision, Edmonton, 2001. *J Rheumatol* 31(2):390-392.
109. Jarvis JN, *et al.* (2009) Gene Expression Profiling in Neutrophils From Children With Polyarticular Juvenile Idiopathic Arthritis. *Arthritis and Rheumatism* 60(5):1488-1495.
110. Lamot L, *et al.* (2014) Aberrant Expression of Shared Master-Key Genes Contributes to the Immunopathogenesis in Patients with Juvenile Spondyloarthritis. *Plos One* 9(12).
111. Allantaz F, *et al.* (2007) Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. *Journal of Experimental Medicine* 204(9):2131-2144.
112. Barnes MG, *et al.* (2010) Biologic Similarities Based on Age at Onset in Oligoarticular and Polyarticular Subtypes of Juvenile Idiopathic Arthritis. *Arthritis and Rheumatism* 62(11):3249-3258.
113. Fall N, *et al.* (2007) Gene expression profiling of peripheral blood from patients with untreated new-onset systemic juvenile idiopathic arthritis reveals molecular heterogeneity that may predict macrophage activation syndrome. *Arthritis and Rheumatism* 56(11):3793-3804.
114. Knowlton N, *et al.* (2009) The Meaning of Clinical Remission in Polyarticular Juvenile Idiopathic Arthritis. *Arthritis and Rheumatism* 60(3):892-900.
115. Moncrieffe H, *et al.* (2010) Generation of novel pharmacogenomic candidates in response to methotrexate in juvenile idiopathic arthritis: correlation between gene expression and genotype. *Pharmacogenet Genom* 20(11):665-676.
116. Tamim S, *et al.* (2014) Genomic analyses reveal broad impact of miR-137 on genes associated with malignant transformation and neuronal differentiation in glioblastoma cells. *Plos One* 9(1):e85591.
117. Uren PJ, *et al.* (2015) RNA-Binding Protein Musashi1 Is a Central Regulator of Adhesion Pathways in Glioblastoma. *Mol Cell Biol* 35(17):2965-2978.
118. Munoz M, *et al.* (2016) Antimicrobial peptide LL-37 participates in the transcriptional regulation of melanoma cells. *J Cancer* 7(15):2341-2345.

119. Oliveira EA, *et al.* (2017) Toxicogenomic and bioinformatics platforms to identify key molecular mechanisms of a curcumin-analogue DM-1 toxicity in melanoma cells. *Pharmacol Res* 125(Pt B):178-187.
120. Geraldo MV, Nakaya HI, & Kimura ET (2017) Down-regulation of 14q32-encoded miRNAs and tumor suppressor role for miR-654-3p in papillary thyroid cancer. *Oncotarget* 8(6):9597-9607.
121. Amaral PP, *et al.* (2018) Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol* 19(1):32.
122. Im SJ, *et al.* (2016) Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* 537(7620):417-421.
123. Zajac AJ, *et al.* (1998) Viral immune evasion due to persistence of activated T cells without effector function. *Journal of Experimental Medicine* 188(12):2205-2213.
124. Sharma P & Allison JP (2015) The future of immune checkpoint therapy. *Science* 348(6230):56-61.
125. Hodi FS, *et al.* (2003) Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. *Proc Natl Acad Sci U S A* 100(8):4712-4717.
126. Larkin J, *et al.* (2015) Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* 373(1):23-34.
127. Seidel JA, Otsuka A, & Kabashima K (2018) Anti-PD-1 and Anti-CTLA-4 Therapies in Cancer: Mechanisms of Action, efficacy, and Limitations. *Front Oncol* 8.
128. Zhang B & Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mo B* 4.
129. Barabasi AL & Oltvai ZN (2004) Network biology: Understanding the cell's functional organization. *Nat Rev Genet* 5(2):101-U115.
130. Russo PST, *et al.* (2018) CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *Bmc Bioinformatics* 19(1):56.
131. Cardozo LE, *et al.* (2019) webCEMiTool: Co-expression Modular Analysis Made Easy. *Front Genet* 10(146).
132. Whitney AR, *et al.* (2003) Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences of the United States of America* 100(4):1896-1901.
133. Albert FW & Kruglyak L (2015) The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 16(4):197-212.
134. Hersh AO & Prahalad S (2015) Immunogenetics of juvenile idiopathic arthritis: A comprehensive review. *Journal of Autoimmunity* 64:113-124.
135. Garg N & Smith TW (2015) An update on immunopathogenesis, diagnosis, and treatment of multiple sclerosis. *Brain Behav* 5(9).
136. Ramilo O, *et al.* (2007) Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* 109(5):2066-2077.
137. Best MG, *et al.* (2015) RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* 28(5):666-676.
138. Mo A, *et al.* (2018) Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. *Genome Med* 10(1):48.

139. Arias-Carrasco R, Vasquez-Moran Y, Nakaya HI, & Maracaja-Coutinho V (2018) StructRNAfinder: an automated pipeline and web server for RNA families prediction. *Bmc Bioinformatics* 19(1):55.
140. Schaum N, *et al.* (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562(7727):367-+.
141. Spitzer MH & Nolan GP (2016) Mass Cytometry: Single Cells, Many Features. *Cell* 165(4):780-791.
142. Yang K & Han X (2016) Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends Biochem Sci* 41(11):954-969.
143. Shalem O, *et al.* (2014) Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 343(6166):84-87.
144. Dogan ES & Liu C (2018) Three-dimensional chromatin packing and positioning of plant genomes. *Nat Plants* 4(8):521-529.