

UNIVERSIDADE DE SÃO PAULO
Faculdade de Ciências Farmacêuticas
Programa de Pós-Graduação em Farmácia
Área de Fisiopatologia

**Estudando a Artrite Idiopática Juvenil mediante uma
abordagem de medicina de redes**

Cesar Augusto Prada-Medina

TESE PARA OBTENÇÃO DO TÍTULO DE
DOUTOR

Orientador: Prof. Dr. Helder I. Nakaya

São Paulo
2019

UNIVERSIDADE DE SÃO PAULO
Faculdade de Ciências Farmacêuticas
Programa de Pós-Graduação em Farmácia
Área de Fisiopatologia

**Estudando a Artrite Idiopática Juvenil mediante uma
abordagem de medicina de redes**

Cesar Augusto Prada-Medina

Versão Corregida

TESE PARA OBTENÇÃO DO TÍTULO DE
DOUTOR

Orientador: Prof. Dr. Helder I. Nakaya

São Paulo
2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica elaborada eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

P896e Prada-Medina, Cesar Augusto
Estudando a artrite idiopática juvenil mediante
uma abordagem de medicina de redes / Cesar Augusto
Prada-Medina. - São Paulo, 2019.
68 p.

Tese (doutorado) - Faculdade de Ciências
Farmacêuticas da Universidade de São Paulo.
Departamento de Análises Clínicas e Toxicológicas.
Orientador: Nakaya, Helder Imoto

1. artrite idiopática juvenil. 2. expressão
gênica. 3. medicina de redes. 4. dimorfismo sexual.
5. neutrofilo. I. T. II. Nakaya, Helder Imoto,
orientador.

Cesar Augusto Prada-Medina

**Estudando a Artrite Idiopática Juvenil mediante uma
abordagem de medicina de redes**

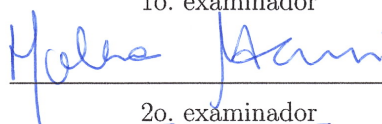
COMISSÃO JULGADORA
DA
TESE PARA OBTENÇÃO DO TÍTULO DE
DOUTOR



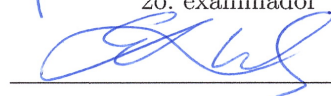
Orientador: Prof. Dr. Helder I. Nakaya



1o. examinador



2o. examinador



3o. examinador

4o. examinador

São Paulo, 12 de Setembro de 2019

“Never is he more active when he does nothing,
never is he less alone than when he is by himself.”

Cato

To the patients, their families, and the scientists around the world that made this work possible.

Acknowledgement

Firstly, I would like to deeply thank my supervisor Prof. Dr. Helder Nakaya. He offered me a great opportunity and permanent scientific advice to develop this project. His confidence and unconditional support on my work was fundamental during this process. Beyond his discipline and scientific rigorosity, I will carry for the rest of my career his teachings on science communication and the socio-political aspects of science.

Without my colleagues of the Computational Systems Biology Laboratory and the Center for Research in Inflammatory Diseases, I would have never realized the findings of this work. Our discussions and *'happy hours'* had substantial contributions to this project. Thank you so much fellows. Similarly, my gratitude to Dr. Prof. Vinicius Maracaja-Coutinho and his Laboratory of Integrative Bioinformatics at *Universdad de Chile* for their contributions to this work.

My time at the European Molecular Biology Laboratory - European Bioinformatics Institute was extremely valuable. I would like to sincerely thank Dr. Irene Papatheodorou, leader of the Expression Atlas. Her thoughtful advice and unconditional support while being part of her team enriched this project. More importantly, it inspired me to keep following a scientific career. I would like to thank the Expression Atlas team as well. Their feedback, help and kindness were responsible for the accomplishment of my time in their team.

I would also like to thank the agencies that financially supported the development of this project. Principally, this study was financed in part by the grant #2017/12487-2 São Paulo Research Foundation (FAPESP) and the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil* (CAPES) – Finance Code 001. Additionally, I would like to thank the project Capacity Building for Bioinformatics in Latin America (CABANA) which granted me a research secondment fellowship to partially develop the project at the EMBL-EBI. Likewise, the *Centro Latinoamericano de Formación Interdisciplinaria (CELFI)* granted me a fellowship to complement my bioinformatics courses at the *Universidad de Buenos Aires*.

Additionally, I wish to convey my special gratitude to Dr. Melissa Lever for her thoroughly English language proofreading of this thesis.

Finally, I would like to thank everyone that accompany me during this process. Many people contributed in different ways. My family has always encouraged and supported me to follow my scientific endeavour. My partner whose grounded company has been extremely supportive during this phase. My friends, colleagues, and acquaintances made my life easier and culturally richer over the last four years. Because all of them, I have successfully accomplished this scientific milestone in my life. My infinite gratitude to all of them.

Resumo

PRADA-MEDINA, C. A. **Estudando a Artrite Idiopática Juvenil mediante uma abordagem de medicina de redes**. 2019. 68f. Tese (Doutorado) - Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2019.

A Artrite Idiopática Juvenil (AIJ) é um grupo de condições inflamatórias de etiologia desconhecida, cuja patofisiologia molecular subjacente ainda não está bem caracterizada. Vários estudos tentaram preencher essa lacuna, caracterizando os perfis de expressão gênica de pacientes com AIJ. No entanto, há uma falta de avaliação sistemática desses resultados transcriptômicos na classificação, prescrição e monitoramento da doença. Além disso, apesar de esta doença ser mais comum em mulheres, nenhum desses estudos tentou avaliar o impacto do sexo na fisiopatologia da doença. Neste projeto, realizamos uma revisão sistemática abrangente e uma metanálise de expressão gênica para revelar a fisiopatologia molecular da AIJ levando em consideração o sexo do paciente. Reunimos e catalogamos mais de 60.000 entradas de características genômicas reportadas como relacionadas à AIJ na literatura. Entre os estudos de transcriptoma, encontramos uma disparidade dramática. Cerca de 15.000 genes foram reportados como perturbados nos leucócitos da AIJ, sendo que menos de um por cento desses genes foram relatados em pelo menos um quarto dos estudos revisados. Em seguida, re-analisamos mais de 700 transcriptomas pediátricos de nove estudos usando uma abordagem analítica comum. Os resultados de expressão diferencial foram combinados usando meta-análise de modelo de efeitos aleatórios. Implementamos esta abordagem de meta-análise de expressão gênica diferencial no pacote MetaVolcanoR R que disponibilizamos no Bioconductor. Usando este pacote, confirmamos várias assinaturas de expressão gênica previamente associadas à AIJ e descobrimos novos genes cuja expressão está perturbada em pacientes com AIJ. Os tamanhos dos efeitos dos genes mais reportados como perturbados coincidem com os resultados da nossa meta-análise. Por meio de uma análise de meta-co-expressão, caracterizamos as assinaturas dos tipos de leucócitos circulantes. Além disso, caracterizamos o dimorfismo sexual da AIJ. Descobrimos que pacientes do sexo feminino com AIJ sistêmica super-ativam genes característicos de mielócitos precoces e neutrófilos bastonetes. Esta assinatura está correlacionada com o estado clínico da doença e à resposta ao tratamento por bloqueio do receptor de IL-1. Isto sugere que a fisiopatologia da AIJs é caracterizada por uma neutrofilia sexualmente dimórfica que afeta a progressão da doença e a resposta aos tratamentos anti-IL-1. Avaliamos ainda esta assinatura neutrofílica em outros contextos. Descobrimos que essa assinatura apresenta uma expressão dependente do sexo ao longo da vida humana, em outras doenças inflamatórias, e sua expressão aumenta durante a gravidez.

Palavras-chave: Artrite Idiopática Juvenil, expressão gênica, medicina de redes, revisão sistemática, meta-análise, dimorfismo sexual, neutrofilo.

Abstract

PRADA-MEDINA, C. A. **Studying Juvenile Idiopathic Arthritis through a Network Medicine Approach**. 68p. Douctoral thesis - Faculty of Pharmaceutical Sciences, University of São Paulo, São Paulo, 2019.

Juvenile Idiopathic Arthritis (JIA) is a group of inflammatory conditions of unknown etiology whose underlying molecular pathophysiology is still not well characterized. Several studies have attempted to fill this gap by characterizing the gene expression profiles of JIA patients. However, there is a lack of systematic assessment of the reliability of these transcriptome results on the disease classification, prescription, and monitoring. In addition, despite this disease is more common in females, none of these studies have tried to assess the impact of sex on disease pathophysiology. In this project, we performed a comprehensive systematic review and a gene expression meta-analysis to reveal the core molecular JIA pathophysiology taking into consideration the patient sex. We gathered and cataloged more than 60,000 entries of genomic features reported as JIA-related in the functional genomics literature, and found a dramatic disparity among the JIA transcriptome studies. Near 15,000 genes have been reported as perturbed in JIA leukocytes. Less than one percent of these genes were reported in at least a quarter of the reviewed studies. We then removed the study-specific analytical bias by re-analyzing more than 700 unique pediatric transcriptome profiles from nine JIA studies using a common analytical framework. The differential expression results from different studies were combined using a random effect model meta-analysis approach. We implemented this differential gene expression meta-analysis methodology in the MetaVolcanoR R package that we made available in Bioconductor. Using this package, we confirmed several gene expression signatures previously associated with JIA and uncover new genes whose expression was perturbed in JIA patients. The effect sizes of the topmost reported perturbed genes coincide with our meta-analysis results. Through a meta-coexpression approach, we characterized the cell type signatures of circulating leukocytes in the JIA affected children. Additionally, we characterized the JIA sexual dimorphism. We found that systemic JIA female patients over-activate a gene expression signature which comprises early myelocytes and band neutrophil expression markers. This signature is correlated with the disease status and response to IL-1 receptor blockade. This suggests that sJIA pathophysiology is characterized by a sexually dimorphic neutrophilia that impacts disease progression and the response to anti-IL-1 treatments. We further assessed this immature neutrophil and female-biased signature in other contexts. We found that this signature presents a sex-dependent expression over human lifetime, in other inflammatory diseases, and its expression increases during pregnancy.

Keywords: Juvenile Idiopathic Arthritis, gene expression, systematic review, meta-analysis, sexual dimorphism, neutrophil.

Contents

List of Abbreviations	xix
List of Figures	xx
1 Juvenile Idiopathic Arthritis gene expression systematic review and meta-analysis: sexual dimorphism of an immature neutrophil-related signature	1
1.1 Introduction	1
1.1.1 Juvenile idiopathic arthritis	1
1.1.2 JIA pathophysiology and treatment	2
1.1.3 Peripheral leukocyte gene expression of JIA patients	3
1.1.4 Gene expression systematic review and meta-analysis	4
1.1.5 Network medicine	5
1.1.6 Justification	5
1.2 Objectives	5
1.2.1 General	5
1.2.2 Specific objectives	5
1.3 Results	6
1.3.1 Systematic review of JIA-functional genomics literature	6
1.3.2 Most genes have been indicated as differentially expressed in JIA patients	6
1.3.3 JIA classes display specific gene expression signatures	9
1.3.4 Novel gene expression features associated with systemic JIA	11
1.3.5 Meta-co-expressed module activity characterizes the leukocyte population composition of sJIA patient clinical statuses	15
1.3.6 Expression of immature neutrophil markers in sJIA is sexually dimorphic	16
1.3.7 JIAexpDB: a JIA gene expression database	20
1.4 Discussion	20
1.5 Methods	24
1.5.1 JIA-functional genomics systematic review	24
1.5.2 Inclusion criteria of the JIA gene expression meta-analysis	24
1.5.3 Gene expression data pre-processing	24
1.5.4 JIA gene expression meta-analysis	25
1.5.5 JIA gene co-expression meta-analysis	25
1.5.6 Single-cell gene expression analysis	26
1.5.7 Interactome reconstruction	26

2	MetaVolcano: making gene expression meta-analysis easily accessible	27
2.1	Introduction	27
2.2	Results	28
2.2.1	Random effect model MetaVolcano	28
2.2.2	Combining <i>p</i> -value MetaVolcano	29
2.2.3	Vote-counting MetaVolcano	29
2.2.4	Discussion	29
2.3	Methods	30
3	Conclusions	32
4	Bibliography	33
5	Appendix	41

List of Abbreviations

ACR	American College of Rheumatology
eJIA	Enthesitis-related juvenile idiopathic arthritis
GEO	Gene Expression Omnibus
GSEA	Gene Set Enrichment Analysis
JIA	Juvenile idiopathic arthritis
JIAexpDB	Juvenile idiopathic arthritis gene expression database
LPS	Lipopolysaccharide
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
NES	Normalized enrichment score
NK	Natural killer
oJIA	Oligoarticular juvenile idiopathic arthritis
PBMC	Peripheral blood mononuclear cells
pJIA	Polyarticular juvenile idiopathic arthritis
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
REM	Random effect model
RF	Rheumatoid factor
RNA	Ribonucleic acid
sJIA	Systemic juvenile idiopathic arthritis

List of Figures

1.1	Juvenile Idiopathic Arthritis clinical classification. The JIA definition comprises at least seven disease classes. This classification is mainly based on the disease severity (e.g. number of affected joints), the co-occurrence with enthesitis and psoriasis, and the presence of the Rheumatoid Factor. Following Petty et al., (2004) definitions, we proposed an update of the Experimental Factor Ontology on the JIA classification (https://github.com/EBISPOT/efo/issues/362). This was permanently incorporated and can be consulted through http://www.ebi.ac.uk/efo/EFO_0002609	2
1.2	Study identification of the JIA-functional genomic systematic review and dataset definition of the JIA transcriptome meta-analysis. A. PRISMA flow diagram of the JIA-functional genomic systematic review. B. PubMed records distribution from 2000 to May 2018 on JIA and JIA-functional genomics. C. Classification of the selected studies for the systematic review. D. Pair-wise (JIA vs healthy control) contrasts defined from the nine selected JIA studies for the meta-analysis. Their available metadata is also presented.	7
1.3	Transcriptome raw-data availability of the selected studies for the JIA systematic review. A. The number of available samples and employed technology of the JIA-related gene expression GEO accessions. The number of JIA and pediatric healthy samples are shown in green. Other non-relevant samples are shown in grey. B and C. Cross-study pattern of sample reusing. Arrow thickness is proportional to the number of samples shared between the GEO accessions. B. Newer GEO accessions comprise previously published samples. Sample GEO accessions were identical among GEO series. C. Some studies comprise samples whose raw data files are identical (<i>md5sum</i> hashes) to samples with different GEO accessions and previously published.	8

1.4	Human sub-interactome core network associated with the JIA pathophysiology. Previously identified JIA-related genes by functional genomics approaches are presented. Genes were classified according to the genomic approach and if they were or not cited in the main written part of the studies. A. The number of genes that we found for each class. In addition, the intersection distribution among the gene classes is presented. B. Physical interaction network among the gene products of the 21 genes reported being associated with the JIA pathophysiology across all the functional genomics approaches. The closest interactome neighbors of these core JIA genes are also included. Node size represents interactome centrality and color stands for the intersection category.	9
1.5	JIA transcriptome systematic review. A. Reverse cumulative distribution of the reported perturbed genes among the studies that published supplementary tables. B. The 85 genes that were identified as perturbed in at least seven studies. Gene size stands for the number of studies where the genes were reported. C. Functional network enrichment of the 85 literature-based most consistently perturbed genes. Node size represents the intersected number of genes between the reference gene-set term and the 85 genes. The thickness of the edges represents the number of intersected genes between gene-set terms.	10
1.6	JIA gene expression meta-analysis results. A. The number of differentially expressed genes per dataset ($p < 0.05$). B. Reverse cumulative distribution of differentially expressed genes across 21 JIA contrasts. C. The 55 genes identified as differentially expressed in at least 15 contrasts. Blue genes are consistently down-regulated, whereas red genes are consistently up-regulated. The gene size represents the number of datasets where the gene was identified as perturbed. D. Mean number of consistently perturbed genes of the main JIA classes. The mean number of consistently differentially expressed genes (REM $p < 0.05$, summary $\log_2(\text{fold-change}) > 0.5$) of the combinatorial pairs of contrasts are presented per JIA class. The standard error of the mean is also presented. E. The intersection among the consistently perturbed genes (REM $p < 0.05$, summary $\log_2(\text{fold-change}) > 0.5$) of the sJIA, pJIA and eJIA conditions. The number of down-regulated, up-regulated, and intersected genes are presented. F. REM summary $\log_2(\text{fold-change})$ and its confidence intervals of the <i>ELOVL7</i> , <i>S100A12</i> , <i>RASSF1</i> , and <i>TIA1</i> genes.	11

1.7 **sJIA gene expression systematic review and meta-analysis.** A. MetaVolcano of the sJIA gene expression. The y -axis depicts the $-\log_{10}(\text{REM } p)$, the x -axis shows the REM summary and their confidence intervals, and color stands for the $\log_2(\text{fold-change})$ sign consistency, which is the number of datasets where the gene presented positive $\log_2(\text{fold-change})$ minus the times it had negative. The number of consistently perturbed genes is presented (REM $p < 0.05$ and an absolute sign consistency score higher than six). B. The intersection between the consistently perturbed genes (meta-analysis) and the previously reported genes in sJIA (systematic review). C. Reverse cumulative distribution of the number of studies where genes have been reported. Top 55 genes most reported in the sJIA transcriptome literature. D. Correlation among the REM summary $\log_2(\text{fold-change})$ and the mean reported $\log_2(\text{fold-change})$ of the top 55 genes most reported in the sJIA transcriptome literature. E. Correlation between manners of reporting genes in the sJIA transcriptome literature. The number of times where a gene was cited in the main written part of the paper is compared with the number of times it was reported in a supplementary table. F. The 978 consistently perturbed genes in our meta-analysis organized by the sign consistency score and the number of times mentioned in the literature. G. Top 20 central genes in the sub-interactome containing the 209 genes whose expression has not been associated with sJIA. 12

1.8 **sJIA co-expression meta-analysis.** A. Reverse cumulative distribution of co-expression module co-membership. Pairs of genes which were part of the same co-expression module in at least 50% of the datasets are included in the consensus meta-co-expression network. Node and edge intersections between the healthy and sJIA consensus networks are presented. B. Both healthy and sJIA consensus network modularity. C. Modularity comparison between the healthy and sJIA consensus networks by intersecting the gene modules and assessing their intersecting significance through the Fisher exact test. D. Differential meta-coexpression. The healthy and sJIA consensus networks are visualized together. Dark nodes and edges represent intersected elements among both consensus networks; and, the red and blue elements stand for the exclusively sJIA and healthy co-expressed genes respectively. E. Expression of the meta-co-expressed genes across the 33 cell clusters of the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). Z-score normalized expression of the meta-co-expressed genes is presented. Red and blue indicate expression levels higher and lower than the mean expression per gene respectively. 13

- 1.9 **Functional enrichment and activity of the consistently co-expressed gene modules.** A. Module M14 gene composition and correlation structure. B. Module M14-functional enrichment. C. Module M36-functional enrichment. D. Module M36 gene composition and correlation structure. E. Functional enrichment and differential expression of the consistently co-expressed gene modules. The sJIA \log_2 (fold-change) of the 13 sJIA datasets was used as ranking criteria for gene set enrichment analysis (GSEA) method (Subramanian et al. 2005) to quantify the module expression activity. The normalized enriched score (NES) for each module in the 13 sJIA datasets is presented. Functional enrichment was performed using the Enrichr method (Kuleshov et al. 2016). F-H. NES of the modules M14, M29, and M36 in the GSE80060, GSE17590, and GSE8650 datasets. 14
- 1.10 **Sex difference in sJIA whole blood gene expression.** A. Genes whose REM $p < 0.01$ are ranked based on the REM summary \log_2 (fold-change) difference between female and male patients. Genes with a REM summary sex difference higher than one are highlighted. B. Sex-biased genes displayed based on their female and male REM summary \log_2 (fold-change); confidence intervals are also shown. Color gradient as in legend of A. C. REM summary \log_2 (fold-change) and confidence intervals of selected sex-biased genes in sJIA patients. D. Gene expression of the female-biased genes represented across the 33 cell clusters of the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). E. tSNE plots highlighting in blue the expression of the *LTF*, *MMP8*, *AZU1*, *ELANE* and *CNRIP1* genes across near 45,000 human hematopoietic cells. 17
- 1.11 **The sJIA female-biased genes represent gene markers for two immature neutrophil populations, early myelocytes and band-neutrophils.** A. Scheme of the five main stages of neutrophil maturation. B and C. Cumulative expression of the sJIA female-biased and neutrophil-related genes for each of the five main stages of neutrophil maturation. Expression values were obtained from Grassi et al., (2018) supplementary tables. D and E. tSNE plots of the neutrophil maturation stages cell clusters from the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). D. Cells are colored by the cluster they come from. E. Cells are colored based on their estimated pseudotime. F and G. Expression of the sJIA female-biased genes along with the estimated pseudotime. F. Counts of the *ELANE*, *LTF*, and *ADM* genes along with the pseudotime. Each dot represents a single cell. G. Z-score normalized expression of the sJIA female-biased and neutrophil-related genes along with the pseudotime. 18

1.12 **Band neutrophil gene expression signature.** A. Top 150 most expressed genes in the band neutrophil cell cluster of the dataset GSE92274 (Lai et al. 2018). Gene size is proportional to the mean raw counts across the cluster cells. B. Band neutrophil signature (*LTF*, *PGLYRP1*, *HP*, *BPI*, *TCN1*, *TFF3*, *CEACAM8*, *CRISP3*, *CAMP*, *MMP8*, *CLC*, *MS4A3*) activation in blood transcriptomes of sJIA patients. Mean expression of the band neutrophil genes is presented per patient. Patients are grouped by their sex and clinical status. C. Band neutrophil signature activity in rheumatoid arthritis patients grouped by their ACR score (275 samples). The ACR score describes the relative improvement of the rheumatoid arthritis symptoms (e.g. ACR0 means no improvement). D. Mean gene expression of the band neutrophil signature along human life span. Over 1,000 samples were considered from 15 to 94 years old individuals. Mean expression values per samples were used to fit a linear regression and its 95% confidence interval. E-F. Band neutrophil gene expression signature in the context of asthma patients (498 samples) and human pregnancy (512 samples). The Pairwise Wilcoxon Rank Sum Tests method implemented in the *pairwise.wilcox.test* R function was utilized to assess the expression differences among groups in B, C, E, and F. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$ 19

2.1 **MetaVolcano implemented visualizations of gene expression meta-analysis results.** Differential expression results from five demo studies are combined. This data and tutorial to reproduce the figures are available in the *MetaVolcanoR* R package vignette (<https://bioconductor.org/packages/MetaVolcanoR>). A. REM MetaVolcano. Top 1% most perturbed genes according to the *Topconfects* ranking are coloured and the 95% confidence intervals are depicted. Color scale represent the sign consistency of each gene. B. Combining p -values MetaVolcano. Top 1% most perturbed genes are coloured in red or blue. C. Number of perturbed genes by study. Users can set the fold-change and p -value thresholds to define the differentially expressed genes. D. Vote-counting meta-volcano. Top 1% of the consistently perturbed genes are coloured. E. Intersection of the 1% top perturbed genes of the three MetaVolcano approaches. The red and blue numbers indicate up-regulated and down-regulated genes respectively. F and G. Forest plot of two genes that were identified as part of the 1% top consistently perturbed genes by the three MetaVolcano approaches. 31

Chapter 1

Juvenile Idiopathic Arthritis gene expression systematic review and meta-analysis: sexual dimorphism of an immature neutrophil-related signature

1.1 Introduction

1.1.1 Juvenile idiopathic arthritis

Juvenile idiopathic arthritis (JIA) is the most common pediatric rheumatological disease worldwide (Manners and Bower 2002). Up to four in one thousand children suffer from the disease, which is more prevalent in females (Cattalini et al. 2019). This disease limits patient's quality of life and reduces their life expectancy (Anon 2015; Klatchoian et al. 2008; Foster et al. 2003; Davies et al. 2014). After 10 years from diagnosis, less than 50% of JIA patients achieve clinical remission under standard treatment (Shoop-Worrall et al. 2017). Therefore, it is fundamental to uncover the immunological mechanisms of JIA to improve the ability of doctors to stratify patients, prescribe effective medications and monitor the disease progression.

The JIA prevalence worldwide is reported as 0.07 to 4.01 per 1000 children and the annual incidence as 0.008 to 0.226 per 1000 children (Manners and Bower 2002). In spite of the absence of official data regarding the prevalence of JIA in Brazil, at least 464 children were diagnosed in 2010 in the city of São Paulo (Teresa Terreri et al. 2013), constituting the most common pediatric rheumatologic condition in the city. Because of the tropical localization and the common occurrence of parasitic infections, Brazilian patients might have different disease courses in comparison with the well documented cases from some countries at increasing latitudes (Braz et al. 2015; de Sousa Studart et al. 2015; Oliveira et al. 2016). However, as in the rest of the world, the underlying pathophysiological differences among different JIA disease courses are not well understood.

The clinical definition of JIA comprises a heterogeneous group of seven immuno-pathological conditions commonly characterized by the chronic inflammation of joints in subjects under the age of 16 (Prakken et al. 2011) (Figure 1.1). Due to its idiopathic nature, JIA is classified mostly based on the disease severity (Petty et al. 2004). The most common form of this condition, the

oligoarticular JIA (oJIA) manifestation, is clinically characterized by the involvement of no more than four joints during the first six months after onset of symptoms (Petty et al. 2004). If more than four joints are involved, patients are classified as polyarticular JIA (pJIA), which is sub-classified based on the presence of rheumatoid factor (RF). These mild manifestations, oJIA and pJIA are mainly driven by adaptive immune mechanisms (Anon 2015; Klatchoian et al. 2008; Foster et al. 2003; Davies et al. 2014; Ombrello et al. 2017). Less prevalent and most severe, the systemic JIA (sJIA) class exhibits fever and skin rashes alongside arthritis. It also resembles the characteristics of an autoinflammatory syndrome (Klatchoian et al. 2008; Foster et al. 2003; Davies et al. 2014; Ombrello et al. 2017). The presence of enthesitis inflammation distinguishes the enthesitis-related JIA (eJIA) class. In addition, the comorbidity of arthritis with psoriasis is classified as a separate condition (Petty et al. 2004).

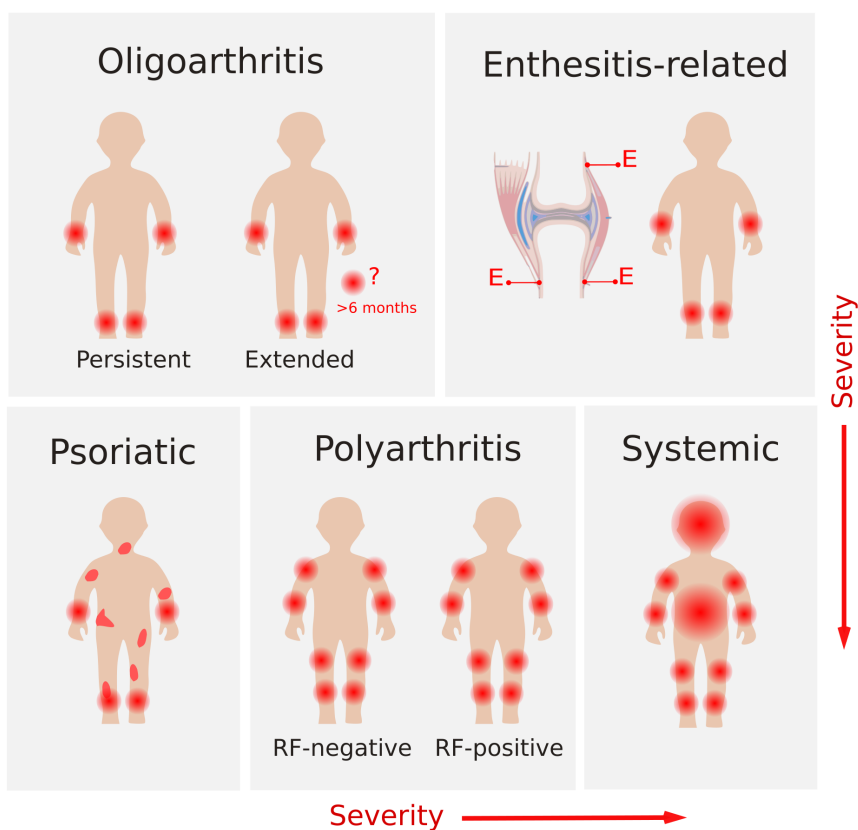


Figure 1.1: Juvenile Idiopathic Arthritis clinical classification. The JIA definition comprises at least seven disease classes. This classification is mainly based on the disease severity (e.g. number of affected joints), the co-occurrence with enthesitis and psoriasis, and the presence of the Rheumatoid Factor. Following Petty et al., (2004) definitions, we proposed an update of the Experimental Factor Ontology on the JIA classification (<https://github.com/EBISPOT/efo/issues/362>). This was permanently incorporated and can be consulted through http://www.ebi.ac.uk/efo/EFO_0002609.

1.1.2 JIA pathophysiology and treatment

The JIA pathophysiology is mainly characterized by the inflammation of the synovial membrane (Prakken et al. 2011). The infiltration of reactive leukocytes and the increase of pro-inflammatory cytokines in the synovial fluid have been associated with the synovium inflammation (Hahn and Kim 2010). Reactive polymorphonuclear neutrophils are the most common infiltrated leukocytes in the inflamed joints of all JIA classes. However, the lymphocyte infiltration is different

among the mild manifestations, oJIA and pJIA, and the systemic class. In contrast to sJIA, oJIA and pJIA patients display an infiltration of activated T and B cells and an unbalanced ratio of Th1 and Th2 cells (Hahn and Kim 2010). On the other hand, sJIA patients experience leukocytosis, neutrophilia, and monocytosis observed in their peripheral blood (Cepika et al. 2017). This sJIA monocytosis can lead to a Macrophage Activation Syndrome which can be fatal. In addition, oJIA and pJIA distinguish from sJIA in their main circulating and synovium-infiltrated pro-inflammatory cytokines (Prakken et al. 2011; Rosenkranz et al. 2010). Although, all JIA classes present an increased concentration of TNF α and IL-6 in synovial fluid and plasma, sJIA patients are less responsive to anti-TNF α treatment (Prakken et al. 2011). Less notorious, IL-1 has also been observed increased in JIA, however, in sJIA patients is possible to observed a clear IL-1-induced gene expression signature (Allantaz et al. 2007). All together these JIA pathophysiology knowledge led to an improvement in the classical immuno-suppressive treatment practices. Several targeted biologics against TNF α , IL-6, IL-1, and other targets are now available. However, it is still necessary effective strategies to stratify patients according to their potential response to the blockade of these targets. For this is also crucial to gain more mechanistic detail on how these drugs operate. For instance, although, anti-IL-1 drugs have been designed to block the IL-1 binding to its receptor by inhibiting either the IL-1 cytokine or its receptor, it is not fully understood how these approaches mechanistically differ.

1.1.3 Peripheral leukocyte gene expression of JIA patients

Over the last two decades, several groups have described the leukocyte gene regulatory programs of several independent JIA cohorts around the world. These efforts have revealed several aspects of the underlying immuno-molecular JIA pathophysiology. Fundamentally, the gene expression profile of circulating leukocytes in JIA patients shows significant perturbation compared to healthy children (Jiang et al. 2016; Jarvis et al. 2009; Lamot et al. 2014; Fall et al. 2007; Knowlton et al. 2009; Griffin et al. 2009; Jiang et al. 2016; Lamot et al. 2014; Gaur et al. 2017). In accordance with the clinical characterization, the gene expression profiles of circulating leukocytes in JIA patients vary among the main JIA classes – oJIA, eJIA, pJIA and sJIA (Barnes et al. 2009; Ishikawa et al. 2009; Hinze et al. 2010). Additionally, the global gene expression of JIA leukocytes changes concomitantly with the clinical status progression (Knowlton et al. 2009; Hinze et al. 2010; Gorelik et al. 2013; Jiang et al. 2013; Shenoi et al. 2015; Jiang et al. 2015; Cepika et al. 2017; Ogilvie et al. 2007; Hunter et al. 2010), and in response to treatment (Moncrieffe et al. 2010; Moncrieffe et al. 2017; Quartier et al. 2011; Jiang et al. 2014; Cui et al. 2016; Omoyinmi et al. 2016; Brachat et al. 2017; Jiang et al. 2013). In addition, peripheral blood mononuclear cells (PBMC) gene expression profiles can be used to distinguish if the age at onset of the oJIA and pJIA symptoms occurred before or after the age of six (Barnes et al. 2010). Furthermore, it has been observed that the expression of natural killer (NK) cells (Gorelik et al. 2013; Barnes et al. 2009; Put et al. 2017; Fall et al. 2007) and mitochondrial-related genes (Omoyinmi et al. 2016; Ishikawa et al. 2009) is lower in the peripheral blood of sJIA patients compared to healthy individuals. Other studies have shown that monocyte-, erythrocyte-, and neutrophil-related genes are overactivated in peripheral blood transcriptomes of JIA patients. In particular, monocyte activation signatures in PBMC transcriptome profiles of sJIA (Macaubas et al. 2012; Gorelik et al. 2013) and eJIA patients (Gaur et al. 2017) have been observed. Immature erythroid- and platelet-related genes

are also over-activated in sJIA (Hinze et al. 2010; Gorelik et al. 2013; Allantaz et al. 2007; Fall et al. 2007; Barnes et al. 2009; Quartier et al. 2011). Likewise, a neutrophil signature has been found to be highly expressed in the sJIA whole blood transcriptome profiles (Allantaz et al. 2007; Fall et al. 2007; Quartier et al. 2011). This neutrophil signature is also present in PBMC transcriptomes of JIA patients (Ramanathan et al. 2018; Fall et al. 2007). It has been found to correlate with an increase in circulating mature and immature neutrophil populations (Ramanathan et al. 2018; Ter Haar et al. 2018). The activity of immature neutrophil genes has been shown to correlate with the counts of ‘low density neutrophils’ (LDNs) in PBMC preparations (Ramanathan et al. 2018; Ter Haar et al. 2018) and to decline when sJIA remission is achieved by interleukin-1 (IL-1) blockade (Ramanathan et al. 2018; Ter Haar et al. 2018). Correspondingly, the transcriptome profile of isolated mature neutrophils is dependent on whether the patient is in remission and relapse state (Jiang et al. 2015; Jarvis et al. 2009; Jiang et al. 2015; Omoyinmi et al. 2016; Hu et al. 2016).

1.1.4 Gene expression systematic review and meta-analysis

Since the sequencing of the human genome in the early 2000s, hundreds of groups around the world have studied how this genetic information is differentially activated in many disease contexts. This has produced enormous amounts of functional genomics data (Banchereau et al. 2017). JIA has not been an exception. As shown in the previous section, during the last couple of decades, several JIA cohorts have been studied through functional genomics approaches. This data availability offers a great opportunity to characterize the common immune molecular features of children who have developed arthritis across several cohorts. However, this is a challenging quest. It requires to be systematic, reproducible, and the results easily accessible. International initiatives like PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al. 2009) have appeared to set a scientifically rigorous workflow to tackle the knowledge-synthesis problem. In the functional genomics context, a systematic review is the qualitative synthesis of the reported genes that has been associated with a certain biological condition in a well defined literature compendium. It means to catalog all the instances where genes were mentioned within a literature corpus. Meta-analysis, on the other hand, is the quantitative synthesis of the gene expression levels. Given that is a common practice to make transcriptome raw-data publicly available, this allows to uniformly pre-process and re-analyze the transcriptome data of independent studies before meta-analyze the gene’s differential expression. After the study identification phase, gene expression meta-analysis implies at least three steps: data curation, data analysis, and results accessibility. Data curation refers to the uniform description of the available clinical and technical information associated to the raw-data. Usually, ontologies like the Experimental Factor Ontology (<https://www.ebi.ac.uk/efo/>) are followed for this purpose. This allows to combine results from independent studies by using the same restricted vocabulary to characterize the clinical and technical variables. For instance, the GEO database (Barrett et al. 2013) (<http://www.ncbi.nlm.nih.gov/geo/>) is one of the main global repositories of gene expression data. It is organized in series accessions (GSE<number>) that are linked to specific publications. Each GEO series comprises a set of samples (GSM<number>). Each sample may have clinical or technical associated information called hereafter as metadata. Each study might use specific jargon to name the metadata variables, that is why data curation is a critical primary step to allow the later meta-analysis. The data analysis phase implies the initial

data pre-processing, normalization, and differential expression analysis for each study independently. Differential expression values ($\log_2(\text{fold-change})$ or differential expression effect-sizes) are then combined using statistical meta-analysis approaches (see Chapter 2). Finally, results must be made accessible for a broad community. It implies results are easily searchable and visualized in a way that allows an easy interpretation. In this work, we made an enormous effort to follow the above mentioned steps to couple a comprehensive JIA gene expression systematic review and meta-analysis.

1.1.5 Network medicine

Since interactions among cellular components (e.g. genes, RNAs, proteins, metabolites, etc.) are responsible for cell physiology, the identification of the specific molecular interactions affected by disease can provide mechanistical insight on its pathophysiology. All these known molecular interactions in a cell can be represented as a network called interactome (Menche et al. 2015). Network medicine is an emergent approach that aims to identify highly interconnected sub-interactome regions associated with diseases (Cho et al. 2012; Vidal et al. 2011; Barabasi et al. 2011; Menche et al. 2015). Proteins which are closer in the human interactome tend to have common biological functions, such as participating in the same metabolic or signaling pathways, and tend to be involved in similar diseases (Menche et al. 2015). Hence, the identification of specific interactome modules associated with JIA can enhance the understanding regarding its molecular basis.

1.1.6 Justification

Although the aforementioned JIA transcriptome studies identified several pathways and cell signatures as perturbed in JIA sufferers, there is not systematic assessment on the the reproducibility of these previously identified signatures across several JIA cohorts. In addition, even though the JIA prevalence is sexually dimorphic, the influence of the patient's sex in the disease progression has not been systematically assessed. We have addressed these limitations by performing a comprehensive functional genomics systematic review and characterizing the JIA immune molecular features of several cohorts of children who have developed arthritis under a consistent framework. Additionally, we characterized the JIA gene expression sexual dimorphism for the first time.

1.2 Objectives

1.2.1 General

To synthesize all publicly available JIA gene expression data to unravel the transcriptional programs underlying the JIA pathophysiology.

1.2.2 Specific objectives

- To systematically catalog the genomic features that have been associated with the JIA pathophysiology.
- To comprehensively meta-analyze the peripheral leukocytes gene expression profiles of JIA.

- To characterize gene expression differences between female and male JIA patients.

1.3 Results

1.3.1 Systematic review of JIA-functional genomics literature

We looked for studies that analyzed functional genomic data obtained from JIA patients' primary samples following the PRISMA guidelines (Moher et al. 2009) (Figure 1.2A). Initially, we identified nearly 6,200 JIA-related PubMed records available in May 2018. From those, we gathered 427 studies specifically related to either JIA genome association, transcriptome, or epigenome studies by adding functional genomics-related query words (Figure 1.2B). During the eligibility phase, we screened their abstracts and discarded 340 studies. We read the resultant 87 studies fully and from them, we kept 67 studies which we considered for the JIA systematic review (Figure 1.2B). Some of these 67 selected studies were integrative analyses. A total of 52 studies included JIA transcriptomes, 17 with JIA genome association and 5 with epigenome analyses (Figure 1.2C).

Available raw data were almost exclusively found in the JIA transcriptome studies. We found 29 GEO series accessions published from 2004 to 2018 ranging from 6 to over 200 samples from either healthy pediatric individuals or JIA patients (Figure 1.3A). Microarray hybridization was the most common technology. Only 5 recent JIA GEO series used sequencing platforms (Figure 1.3A).

Re-using samples from primary GEO series was a common practice among the available JIA transcriptome datasets. In some cases, the reuse of samples was explicitly done by gathering previous samples accessions (Figure 1.3B). In contrast, other authors created GEO series with previously published samples but with new sample accessions (Figure 1.3C). Out of the 24 microarray JIA accessions, nine contained more than three original samples of both JIA patients and healthy pediatric individuals (Figure 1.2D).

Several clinical designs have been used to assess the JIA leukocyte transcriptomes. Most of the older works set a transversal study design while clinical trials are more common in recent studies. Most of the transversal design works allowed the inclusion of patients that were under the standard of care treatment at the date of the sample collection. We found three different JIA diagnostic criteria among the selected nine studies; corresponding to the diagnostic revisions of 1986, 1997, and 2001. Also, in most cases, the authors utilized the disease activity criteria from the year 2004.

Based on the clinical metadata of the selected nine GEO series, such as the presence of fever and the treatment regime, we defined 21 pairwise contrasts. Each contrast contained a uniform group of JIA patients and healthy pediatric controls from the same GEO series (Figure 1.2D).

1.3.2 Most genes have been indicated as differentially expressed in JIA patients

We gathered and cataloged 68,348 entries of genomic features reported as JIA-related in the functional genomic literature. Each entry corresponds to an instance where we found a genomic feature within a paper. These 68,348 entries comprise over 16,000 unique genomic features. We classified these features into five categories: genes containing variants associated with JIA, genes

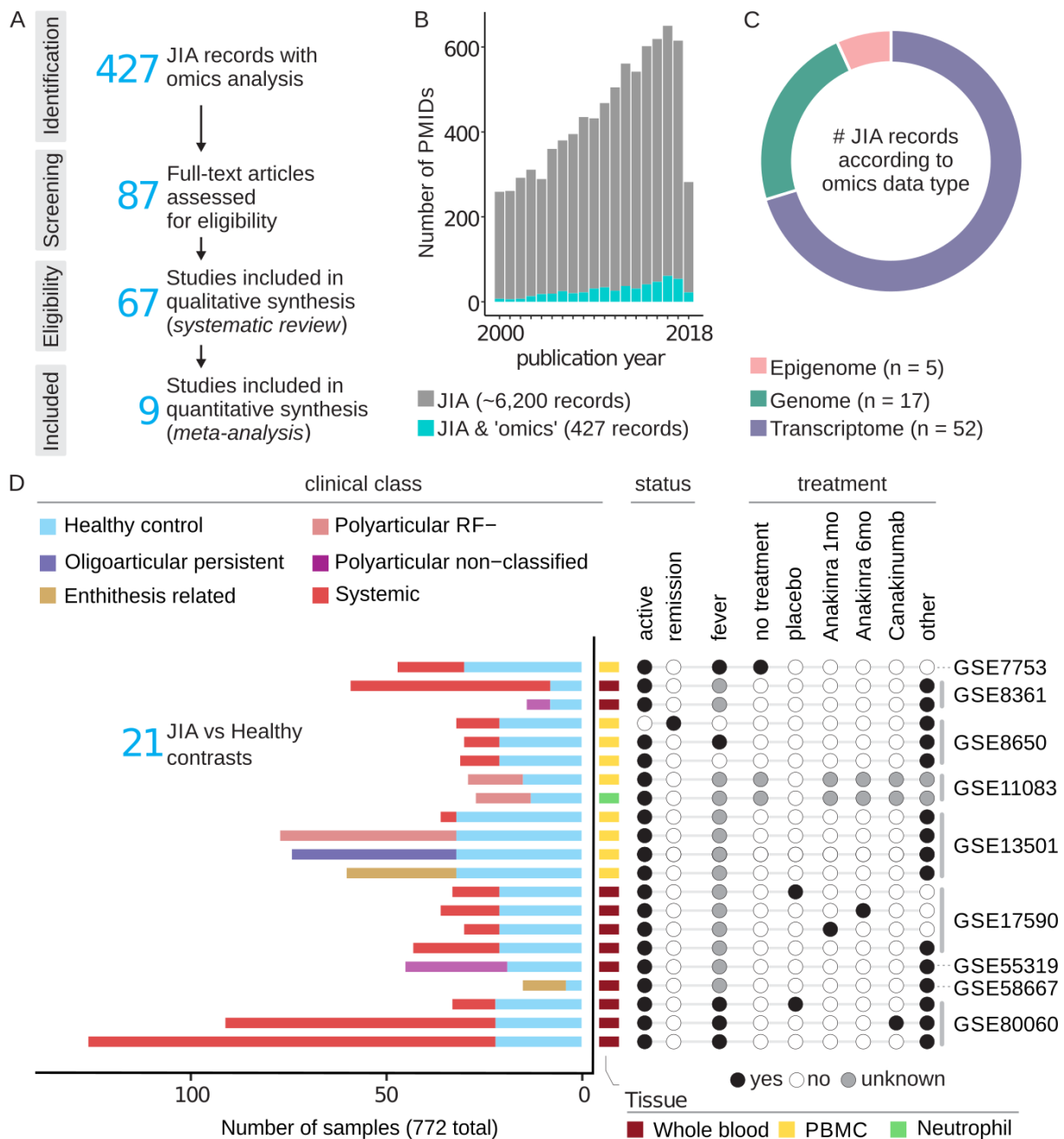


Figure 1.2: Study identification of the JIA-functional genomic systematic review and dataset definition of the JIA transcriptome meta-analysis. A. PRISMA flow diagram of the JIA-functional genomic systematic review. B. PubMed records distribution from 2000 to May 2018 on JIA and JIA-functional genomics. C. Classification of the selected studies for the systematic review. D. Pair-wise (JIA vs healthy control) contrasts defined from the nine selected JIA studies for the meta-analysis. Their available metadata is also presented.

whose expression was perturbed in JIA patients, genes within JIA-associated epigenome marks, genes reported as part of JIA-related gene networks, and the cited genes within the main written part of the 67 included studies (Figure 1.4A). The number of gene expression markers reported drastically outperform other classes of JIA genomic features (Figure 1.4A). More than 15,000 genes have been indicated as perturbed at least once in the transcriptome JIA literature (Figure 1.4A). These gene expression markers mostly intersected with genes reported within JIA-associated epigenetic marks and with the in-paper cited genes. 21 genes have been associated with JIA in all

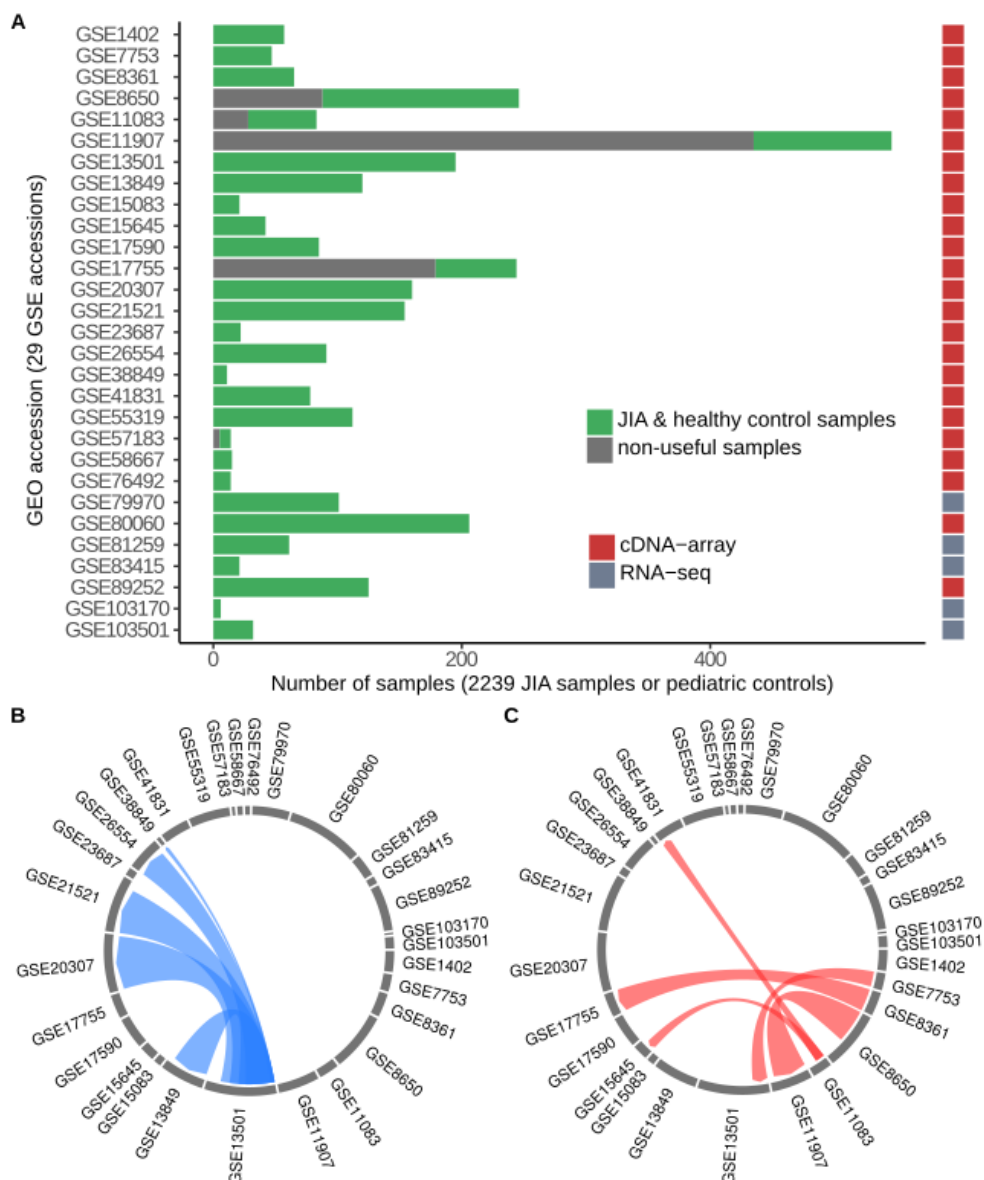


Figure 1.3: Transcriptome raw-data availability of the selected studies for the JIA systematic review. A. The number of available samples and employed technology of the JIA-related gene expression GEO accessions. The number of JIA and pediatric healthy samples are shown in green. Other non-relevant samples are shown in grey. B and C. Cross-study pattern of sample reusing. Arrow thickness is proportional to the number of samples shared between the GEO accessions. B. Newer GEO accessions comprise previously published samples. Sample GEO accessions were identical among GEO series. C. Some studies comprise samples whose raw data files are identical (md5sum hashes) to samples with different GEO accessions and previously published.

of the five contexts studied. To explore their functional relationship, we reconstructed a JIA-associated sub-interactome with the known physical interactions among the proteins coded by these 21 JIA core genes (Figure1.4B). Most of these JIA core proteins physically interact with each other and are constituents in some of the most well-known pathways involved in the JIA pathophysiology. For instance, the JIA core genes include the interferon-induced genes – *IRF1*, *IRF8*, *IRF2*, and *IRF4*; and genes that are involved in the IL6 and tumor necrosis factor (TNF) pathways, as well as macrophage activation genes like *MIF* and *CD74*. The structure of the physical interaction network among the JIA core genes reveals the central role of STAT1 and STAT2 as signal transducers of the proinflammatory signals in the JIA context (Figure1.4B).

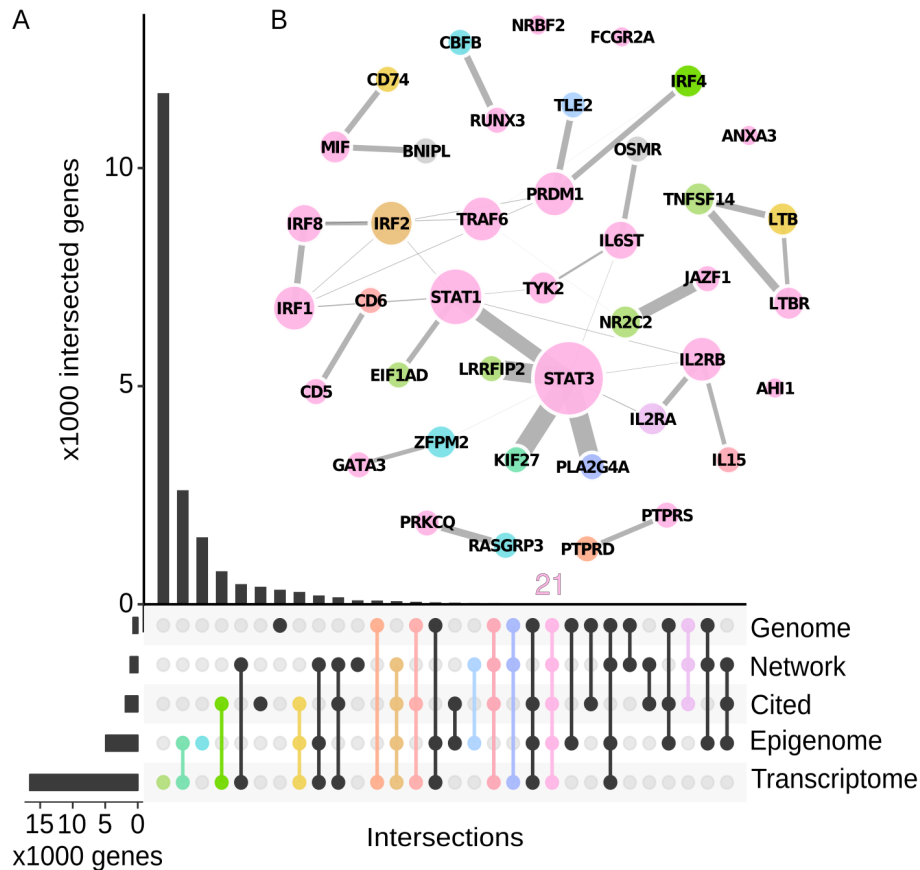


Figure 1.4: Human sub-interactome core network associated with the JIA pathophysiology. Previously identified JIA-related genes by functional genomics approaches are presented. Genes were classified according to the genomic approach and if they were or not cited in the main written part of the studies. A. The number of genes that we found for each class. In addition, the intersection distribution among the gene classes is presented. B. Physical interaction network among the gene products of the 21 genes reported being associated with the JIA pathophysiology across all the functional genomics approaches. The closest interactome neighbors of these core JIA genes are also included. Node size represents interactome centrality and color stands for the intersection category.

The majority of the identified near 15k JIA differentially expressed genes were reported in a few out of the total studies (Figure 1.5A). Out of the 27 transcriptome studies, only 85 genes were reported as perturbed genes in at least seven studies (Figure 1.5B). This core of JIA gene expression markers enriches for neutrophil functions, anti-rheumatic drug signatures, and interestingly, progesterone and retinoic acid signatures (Figure 1.5C).

1.3.3 JIA classes display specific gene expression signatures

We uniformly pre-processed and normalized the selected nine JIA transcriptome datasets, and then performed a differential expression analysis for each of the 21 pairwise contrasts. The number of identified perturbed genes per dataset varied (Figure 1.6A). A simple vote counting approach revealed that 55 genes were differentially expressed in at least 15 contrasts (Figure 1.6B). From those, 26 and 29 genes were consistently up- and down-regulated, respectively (Figure 1.6C).

We assessed if the level of transcriptome perturbation is different among the JIA classes. For this, we compared the number of consistently perturbed genes among the JIA classes (Figure 1.6D). We set the same gene perturbation criteria (REM $p < 0.05$ and summary $\log_2(\text{fold-change}) > 0.5$) to fairly estimate the transcriptome perturbation as the number of consistently perturbed genes in

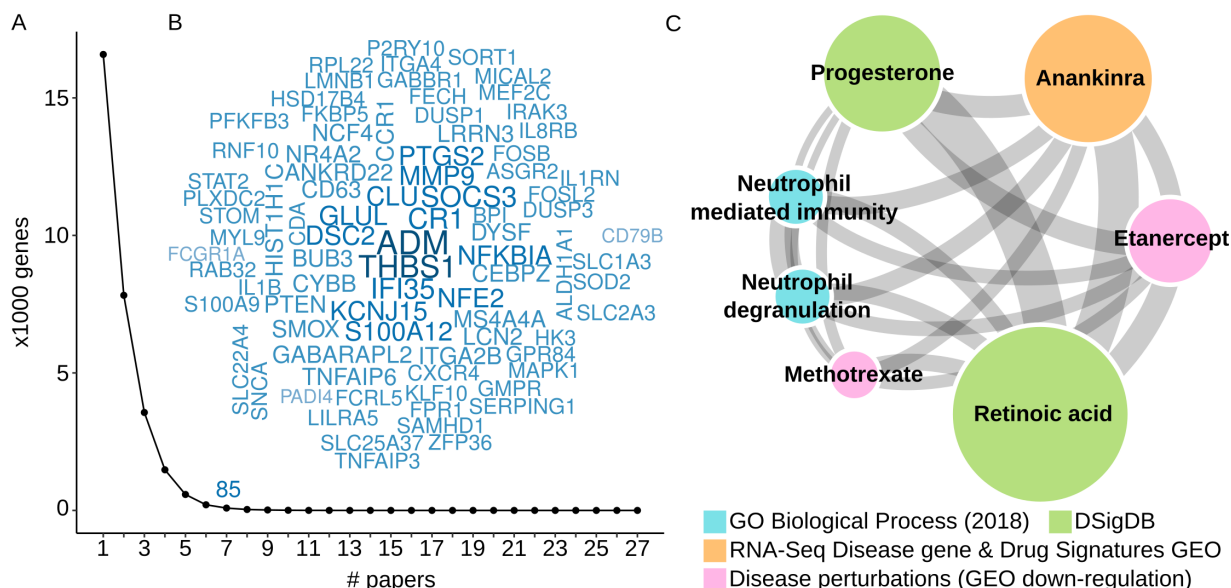


Figure 1.5: JIA transcriptome systematic review. A. Reverse cumulative distribution of the reported perturbed genes among the studies that published supplementary tables. B. The 85 genes that were identified as perturbed in at least seven studies. Gene size stands for the number of studies where the genes were reported. C. Functional network enrichment of the 85 literature-based most consistently perturbed genes. Node size represents the intersected number of genes between the reference gene-set term and the 85 genes. The thickness of the edges represents the number of intersected genes between gene-set terms.

each JIA class. Given that the number of contrasts per JIA class were different, we gathered pairs of contrasts to summarize the gene expression \log_2 (fold-change) in JIA patients using random effect model (REM) meta-analysis. For each JIA class, we obtained the mean number of consistently perturbed genes among all combinations of pair contrasts. We found that the gene expression global perturbation, as shown by the number of consistently perturbed genes, corresponded to the severity of the JIA class, being the highest in sJIA and the lowest in enthesitis-related JIA (Figure 1.6D).

Next we compared the consistently differentially expressed genes among the JIA classes (Figure 1.6E). We summarized the gene expression \log_2 (fold-change) of all contrasts per JIA class using REM meta-analysis, and found few genes that were shared by two or more JIA classes. (Figure 1.6E). The major intersection was between sJIA and pJIA. By contrast, eJIA and pJIA shared 5 perturbed genes. Notably, up-regulated genes tend to be more consistently perturbed in JIA. Our meta-analysis revealed *ELOVL7* as consistently up-regulated among all JIA classes (Figure 1.6F). *ELOVL7* was not part of the most commonly reported perturbed genes in the JIA transcriptome literature (Figure 1.5B). In addition, both the vote-counting approach and the REM meta-analysis revealed the down-regulation of *RASSF1* across all JIA classes (Figure 1.6C and 1.6F). This gene, again, was not part of the most reported expression markers in the JIA transcriptome literature (Figure 1.5B). Moreover, the gene *S100A12* is one of the most commonly reported JIA expression markers (Figure 1.5B) and it was also identified by the vote-counting method (Figure 1.6C); however, *S100A12* up-regulation is more pronounced in sJIA compared to pJIA and eJIA classes (Figure 1.6F).

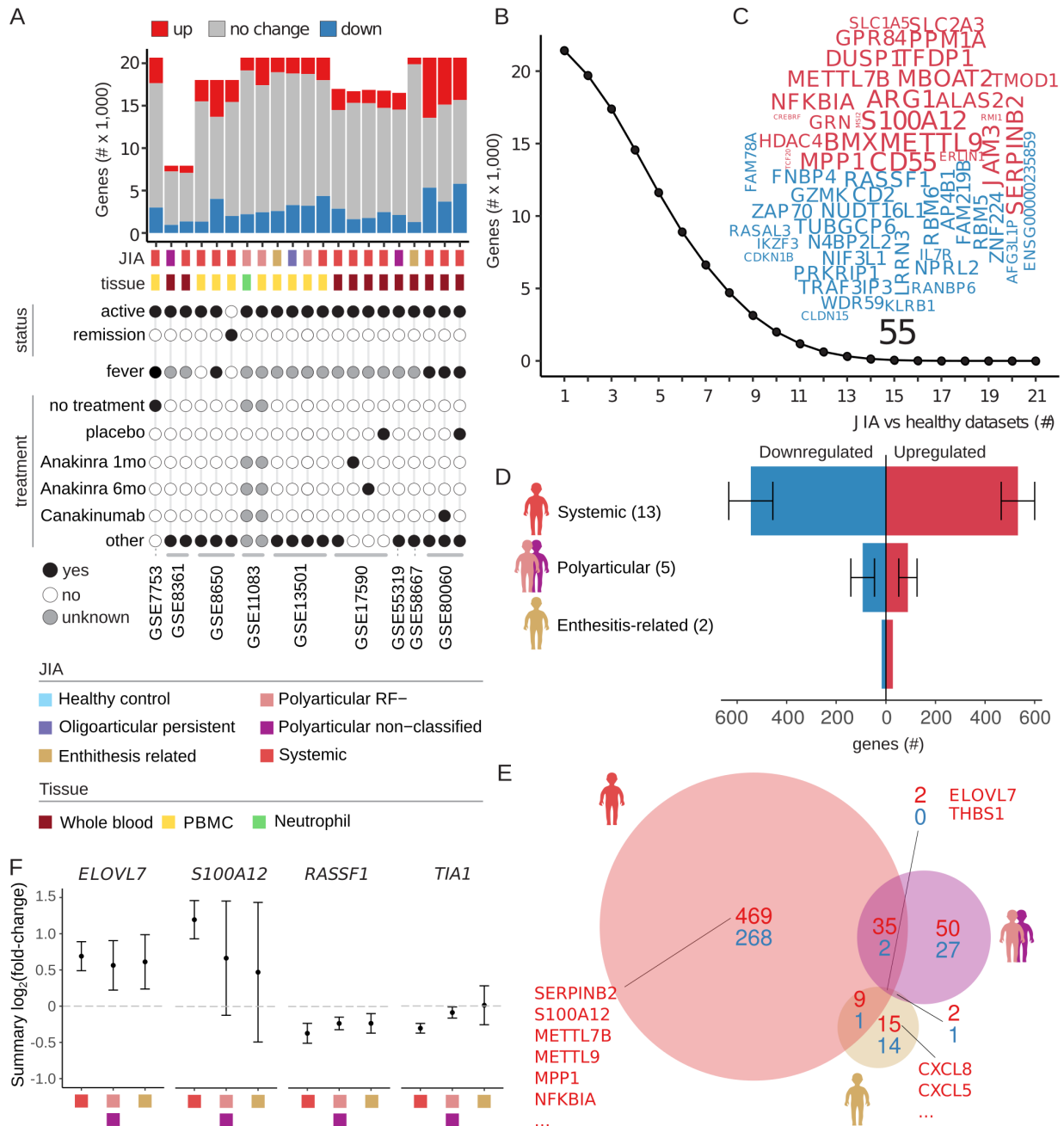


Figure 1.6: JIA gene expression meta-analysis results. A. The number of differentially expressed genes per dataset ($p < 0.05$). B. Reverse cumulative distribution of differentially expressed genes across 21 JIA contrasts. C. The 55 genes identified as differentially expressed in at least 15 contrasts. Blue genes are consistently down-regulated, whereas red genes are consistently up-regulated. The gene size represents the number of datasets where the gene was identified as perturbed. D. Mean number of consistently perturbed genes of the main JIA classes. The mean number of consistently differentially expressed genes ($REM\ p < 0.05$, $\text{summary } \log_2(\text{fold-change}) > 0.5$) of the combinatorial pairs of contrasts are presented per JIA class. The standard error of the mean is also presented. E. The intersection among the consistently perturbed genes ($REM\ p < 0.05$, $\text{summary } \log_2(\text{fold-change}) > 0.5$) of the sJIA, pJIA and eJIA conditions. The number of down-regulated, up-regulated, and intersected genes are presented. F. REM $\text{summary } \log_2(\text{fold-change})$ and its confidence intervals of the ELOVL7, S100A12, RASSF1, and TIA1 genes.

1.3.4 Novel gene expression features associated with systemic JIA

We meta-analysed the 13 sJIA contrasts results and compared these with the results of our systematic review to identify new sJIA-associated genes (Figure 1.7). From the near 9,000 genes

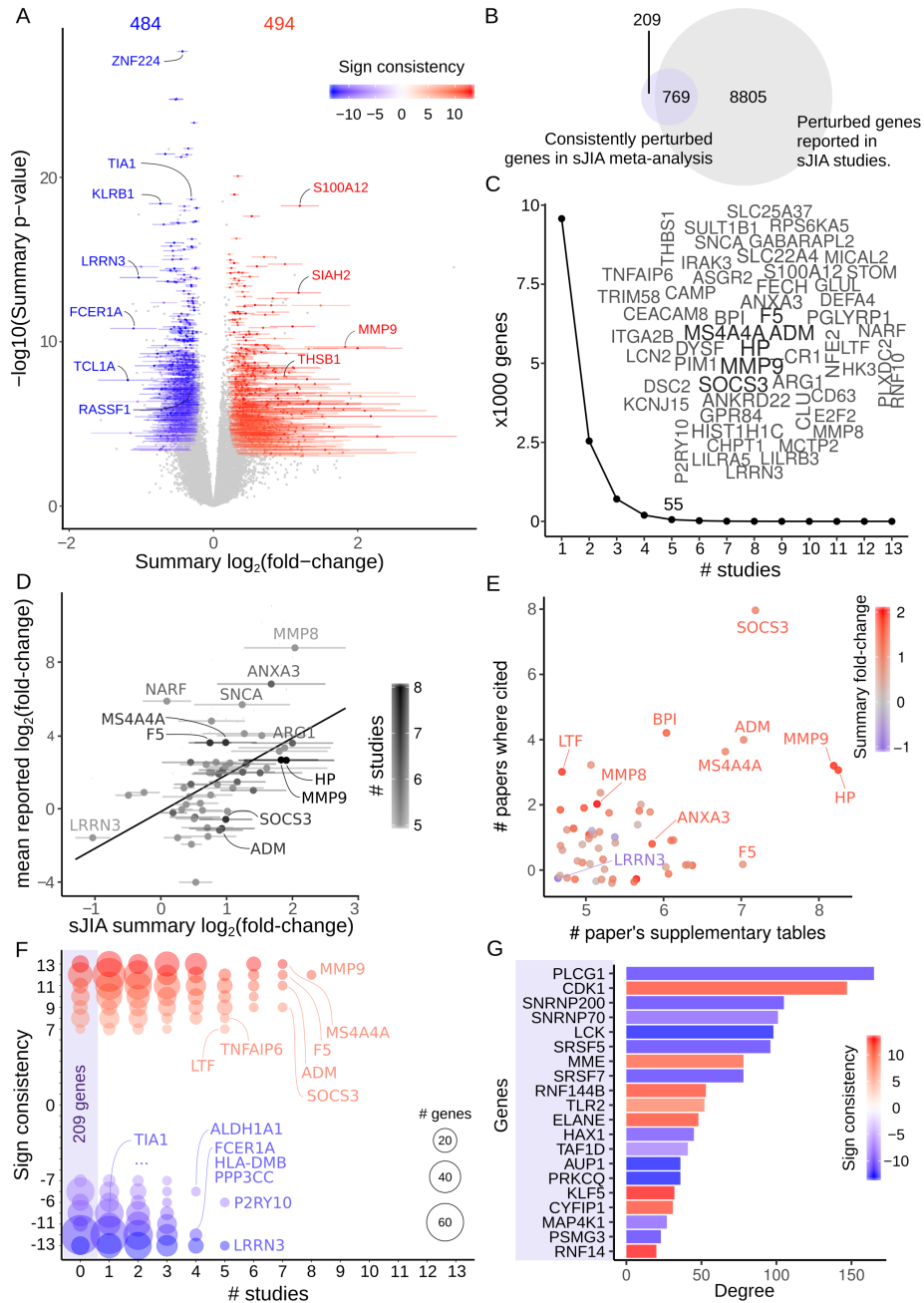


Figure 1.7: sJIA gene expression systematic review and meta-analysis. A. MetaVolcano of the sJIA gene expression. The y-axis depicts the $-\log_{10}(\text{REM } p)$, the x-axis shows the REM summary and their confidence intervals, and color stands for the $\log_2(\text{fold-change})$ sign consistency, which is the number of datasets where the gene presented positive $\log_2(\text{fold-change})$ minus the times it had negative. The number of consistently perturbed genes is presented ($\text{REM } p < 0.05$ and an absolute sign consistency score higher than six). B. The intersection between the consistently perturbed genes (meta-analysis) and the previously reported genes in sJIA (systematic review). C. Reverse cumulative distribution of the number of studies where genes have been reported. Top 55 genes most reported in the sJIA transcriptome literature. D. Correlation among the REM summary $\log_2(\text{fold-change})$ and the mean reported $\log_2(\text{fold-change})$ of the top 55 genes most reported in the sJIA transcriptome literature. E. Correlation between manners of reporting genes in the sJIA transcriptome literature. The number of times where a gene was cited in the main written part of the paper is compared with the number of times it was reported in a supplementary table. F. The 978 consistently perturbed genes in our meta-analysis organized by the sign consistency score and the number of times mentioned in the literature. G. Top 20 central genes in the sub-interactome containing the 209 genes whose expression has not been associated with sJIA.

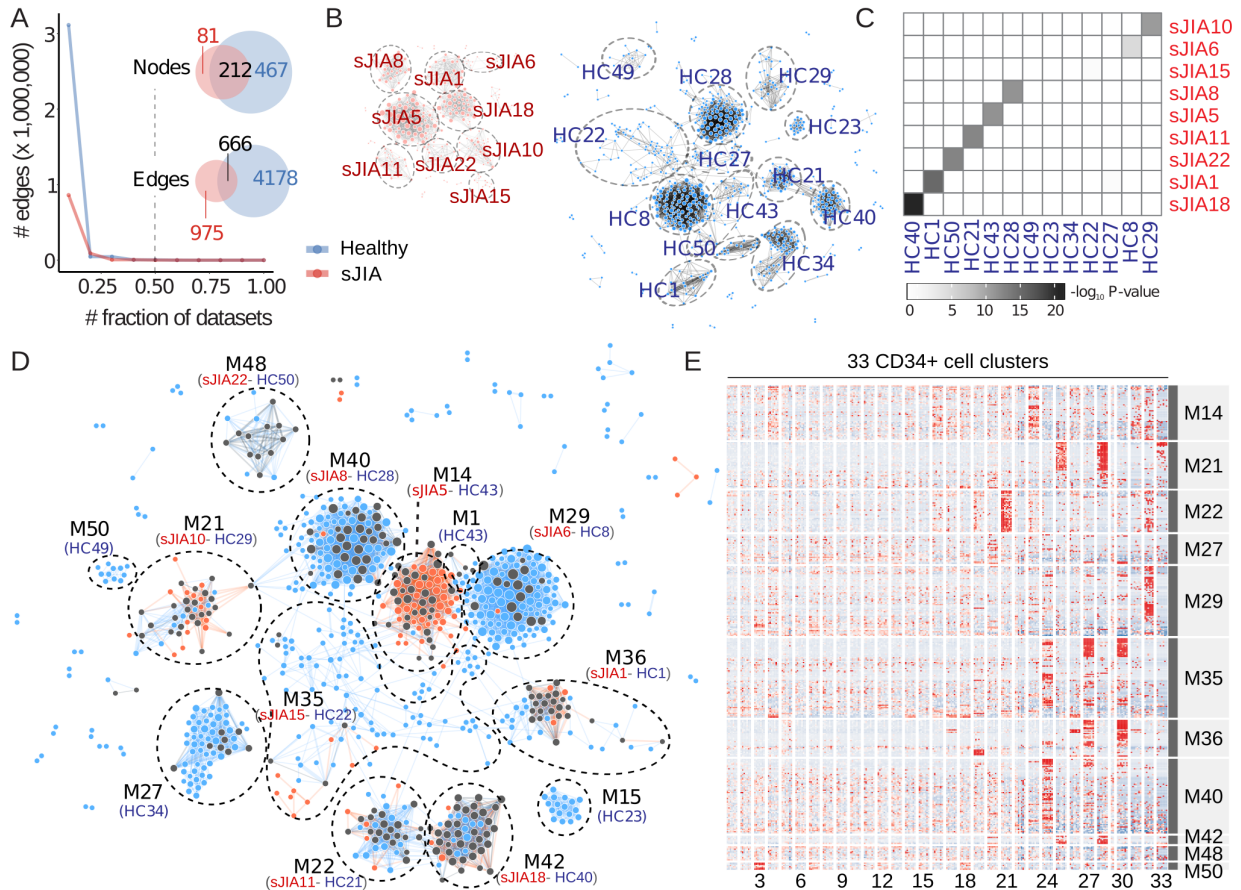


Figure 1.8: sJIA co-expression meta-analysis. A. Reverse cumulative distribution of co-expression module co-membership. Pairs of genes which were part of the same co-expression module in at least 50% of the datasets are included in the consensus meta-co-expression network. Node and edge intersections between the healthy and sJIA consensus networks are presented. B. Both healthy and sJIA consensus network modularity. C. Modularity comparison between the healthy and sJIA consensus networks by intersecting the gene modules and assessing their intersecting significance through the Fisher exact test. D. Differential meta-co-expression. The healthy and sJIA consensus networks are visualized together. Dark nodes and edges represent intersected elements among both consensus networks; and, the red and blue elements stand for the exclusively sJIA and healthy co-expressed genes respectively. E. Expression of the meta-co-expressed genes across the 33 cell clusters of the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). Z-score normalized expression of the meta-co-expressed genes is presented. Red and blue indicate expression levels higher and lower than the mean expression per gene respectively.

that have been reported in the literature as perturbed in sJIA, we found 769 consistently perturbed in our meta-analysis (Figure 1.7A-B). Most of these 9,000 genes were mentioned only in one sJIA transcriptome study (Figure 1.7C). A total of 55 genes have been reported in at least five sJIA studies (Figure 1.7C). The mean reported effect size of these consistently reported sJIA genes positively correlates with the REM summary \log_2 (fold-change) of our meta-analysis (Figure 1.7D). Most of these 55 genes are cited more frequently in the supplementary materials rather than discussed on the main written part of the sJIA transcriptome studies (Figure 1.7E). For instance, the *MMP9* gene was mentioned in the paper main written part half of the times it was identified as perturbed according to the supplementary tables (Figure 1.7E). *MMP9*, however, is one of the top most up-regulated genes in sJIA (Figure 1.7A, F). By contrast, the *TIA1* gene, exclusively and very down-regulated in sJIA patients (Figure 1.6F, 1.7A), was reported as perturbed in only one study (Figure 1.7F). However, in our meta-analysis, it was of the most down-regulated genes

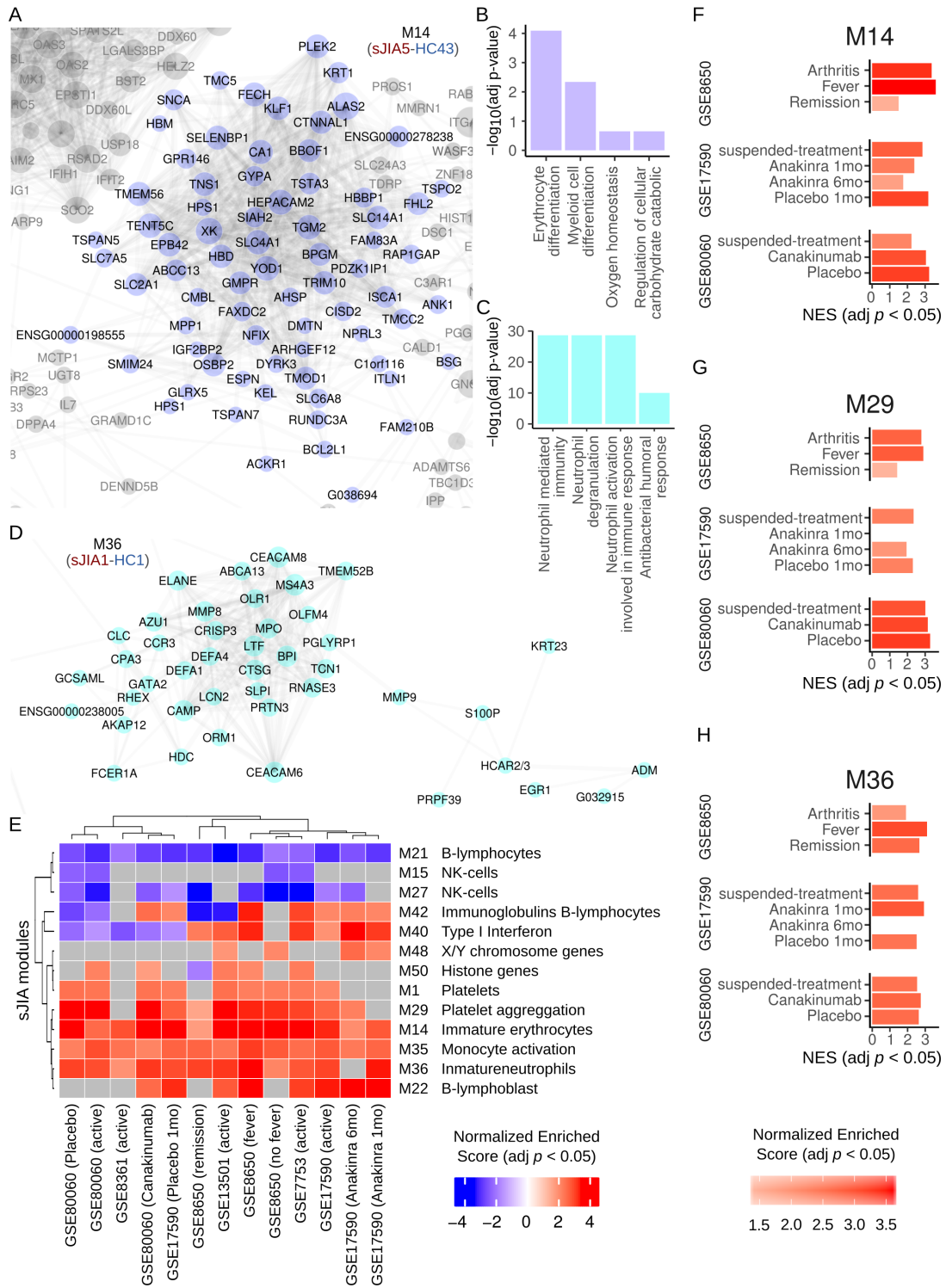


Figure 1.9: Functional enrichment and activity of the consistently co-expressed gene modules. A. Module M14 gene composition and correlation structure. B. Module M14-functional enrichment. C. Module M36-functional enrichment. D. Module M36 gene composition and correlation structure. E. Functional enrichment and differential expression of the consistently co-expressed gene modules. The sJIA $\log_2(\text{fold-change})$ of the 13 sJIA datasets was used as ranking criteria for gene set enrichment analysis (GSEA) method (Subramanian et al. 2005) to quantify the module expression activity. The normalized enriched score (NES) for each module in the 13 sJIA datasets is presented. Functional enrichment was performed using the Enrichr method (Kuleshov et al. 2016). F-H. NES of the modules M14, M29, and M36 in the GSE80060, GSE17590, and GSE8650 datasets.

(Figure 1.7A). In addition, we identified 209 genes consistently perturbed genes that were not previously reported as differentially expressed in sJIA patients (Figure 1.7B, F). We reconstructed a sub-interactome network among these newly sJIA expression features to prioritize them based on its centrality. We considered the known physical interactions among the proteins encoded by these 209 genes. *PLCG1* (consistently down-regulated) and *CDK1* (consistently up-regulated) were the most central genes (Figure 1.7G).

1.3.5 Meta-co-expressed module activity characterizes the leukocyte population composition of sJIA patient clinical statuses

The identification of gene sets whose expression is highly correlated allows the biological characterization of genome-wide gene expression profiles (Russo et al. 2018). These co-expressed gene modules have been shown to represent functionally related groups of genes. Therefore, to investigate the JIA transcriptome profiles in terms of the altered biological functions, we uncovered modules of genes consistently co-expressed in the circulatory leukocytes of pediatric healthy individuals and sJIA patients. This was achieved by integrating co-expression results of independent datasets using a co-membership network approach (see methods). The co-expression consensus network contained nodes and edges representing genes and the number of datasets where a pair of genes are members of the same co-expression module respectively. This co-membership weighted network mostly comprises gene pairs which were members of the same co-expression module only in one dataset (Figure 1.8A). Gene pairs were found to be more consistently correlated with leukocytes from healthy pediatric individuals than sJIA patients (Figure 1.8A). Despite most of the sJIA consistently co-expressed genes overlap with the co-expressed genes in healthy individuals, the correlation structure of the sJIA and healthy networks is different (Figure 1.8A). These consensus networks of both healthy and sJIA were highly modular (Figure 1.8B). We defined 13 and nine consistently co-expressed modules in healthy and sJIA individuals, respectively (Figure 1.8B). Five of the healthy consensus modules were absent in sJIA, and one module were exclusively found in sJIA (Figure 1.8C). To better understand the gene expression regulatory rewiring in sJIA leukocytes, we integrated both healthy and sJIA consensus networks and found 13 co-expressed modules (Figure 1.8D). Modules M14, M21, M36, and M42 presented the most extensive re-wiring (Figure 1.8D). These pediatric co-expression modules were associated with known molecular functions and cell types (Figure 1.9, A-F).

To better understand the biological meaning of the co-expressed modules, we investigated its expression in blood leukocytes at the single-cell level. Given that sJIA patients presents CD34+ leukocytosis, we gathered and clustered a publicly available scRNA-seq dataset of 50,000 CD34+ blood cells from adult healthy donors (GEO dataset GSE92274) (Lai et al. 2018). We defined 31 cell clusters. We found cell-cluster specific expression patterns for the consistently co-expressed gene modules (Figure 1.8E). For instance, the M36 module contains genes whose expression is characteristic of immature neutrophil subpopulations (Figure 1.8E and Figure 1.9). Likewise, modules M21, M22, M29, M35, M40, and M42 comprise gene expression markers for specific leukocyte subpopulations (Figure 1.8E). Notably, the small module M50, which corresponds to the HC49 module that disappears in sJIA and that did not enrich to any known molecular functions (Figure 1.9), comprises the genes *HIST1H2BD*, *HIST1H4H*, *HIST1H2BG*, and *HIST1H2AE* which are cell markers of a specific subpopulation of B cells (Figure 1.8E). Therefore, this co-expression

consensus-co-membership network approach constitutes an unsupervised approach to reveal the cell type composition of tissue transcriptomes. The M21 module, marker of B cells, and the M27, which comprises NK cell markers, are consistently down-regulated in sJIA patients. In addition, M29, M14, M35, M22, and M36 were consistently up-regulated (Figure 1.9F). These modules comprise markers for megakaryocytes, immature erythrocytes, monocytes, B-lymphocytes, and immature neutrophils (Figure 1.9F). From these, M14, M29, and M36 expression levels correlate with the disease status (Figure 1.9 F-H).

1.3.6 Expression of immature neutrophil markers in sJIA is sexually dimorphic

The transcriptome perturbation found in the peripheral blood of sJIA patients indicates a sexual dimorphism on its pathophysiology (Figure 1.10). We combined the differential expression results of the five whole-blood sJIA contrasts for male and female individuals separately. We compared the REM summary $\log_2(\text{fold-change})$ of the genes that we identified as perturbed in either female or male sJIA patients, and found 104 differentially expressed genes displaying a summary $\log_2(\text{fold-change})$ difference higher than one among sexes. (Figure 1.10, A-B). A total of 55 genes were highly perturbed in sJIA males. For instance, the up-regulation of the genes *CAPZA1* and *ZNF699* is male-specific (Figure 1.10C). Other 49 genes presented a female-biased perturbation. For instance, the *RNASE3*, *ELANE*, and *LTF* genes are exclusively up-regulated in sJIA female patients (Figure 1.10C). Given that most of these female-biased genes are part of the consistently co-expressed M36 gene module, which enriches for neutrophil activation (Figure 1.9, C-D), we explored the expression of these genes among the different CD34+ cell types (Figure 1.10D). Among the sJIA patients, female-biased genes were expression markers of specific immature sub-populations of neutrophils (Figure 1.10, D-E). We wanted to know which of the hematopoietic neutrophil stage correspond to these sJIA female-related immature neutrophils. Neutrophils maturation is characterized by five stages ranging from metamyelocytes in the bone marrow to polymorphonuclear neutrophils in the peripheral blood (Figure 1.11A). We followed two strategies to determine the neutrophil stages represented in the sJIA female-biased genes. First, we looked at the expression of these genes in a study that isolated and characterized the bulk RNA-seq transcriptome of each of the five main neutrophil stages (Figure 1.11 B-C). We found that the sJIA female-biased and neutrophil-related genes displayed two expression patterns along with the neutrophil maturation. One profile is characteristic of the early immature myelocytes whereas the second pattern is characteristic of band neutrophils (Figure 1.11 B-C). Genes from each of the two patterns corresponded to marker genes of different hematopoietic cell clusters (Figure 1.10, D-E, 1.11B-C, F), which confirm these genes are markers of specific immature neutrophil populations. In the second strategy, we sub-settled the CD34+ hematopoietic cell clusters annotated as neutrophil-like cells (Figure 1.11D) and estimated their developmental order by estimating its pseudotime. This allowed us to reconstruct the expression dynamics of the sJIA female-biased genes along with the neutrophil maturation process (Figure 1.11 E-F). We found that these sJIA female-biased and neutrophil-related genes cluster in two types of expression dynamics along with the neutrophil maturation (Figure 1.11E). This corroborates our findings on the independent bulk RNA-seq data from the isolated neutrophil stages (Figure 1.11 B-C). Altogether, these results indicate that female sJIA patients present a consistent over-expression of genes whose expression is

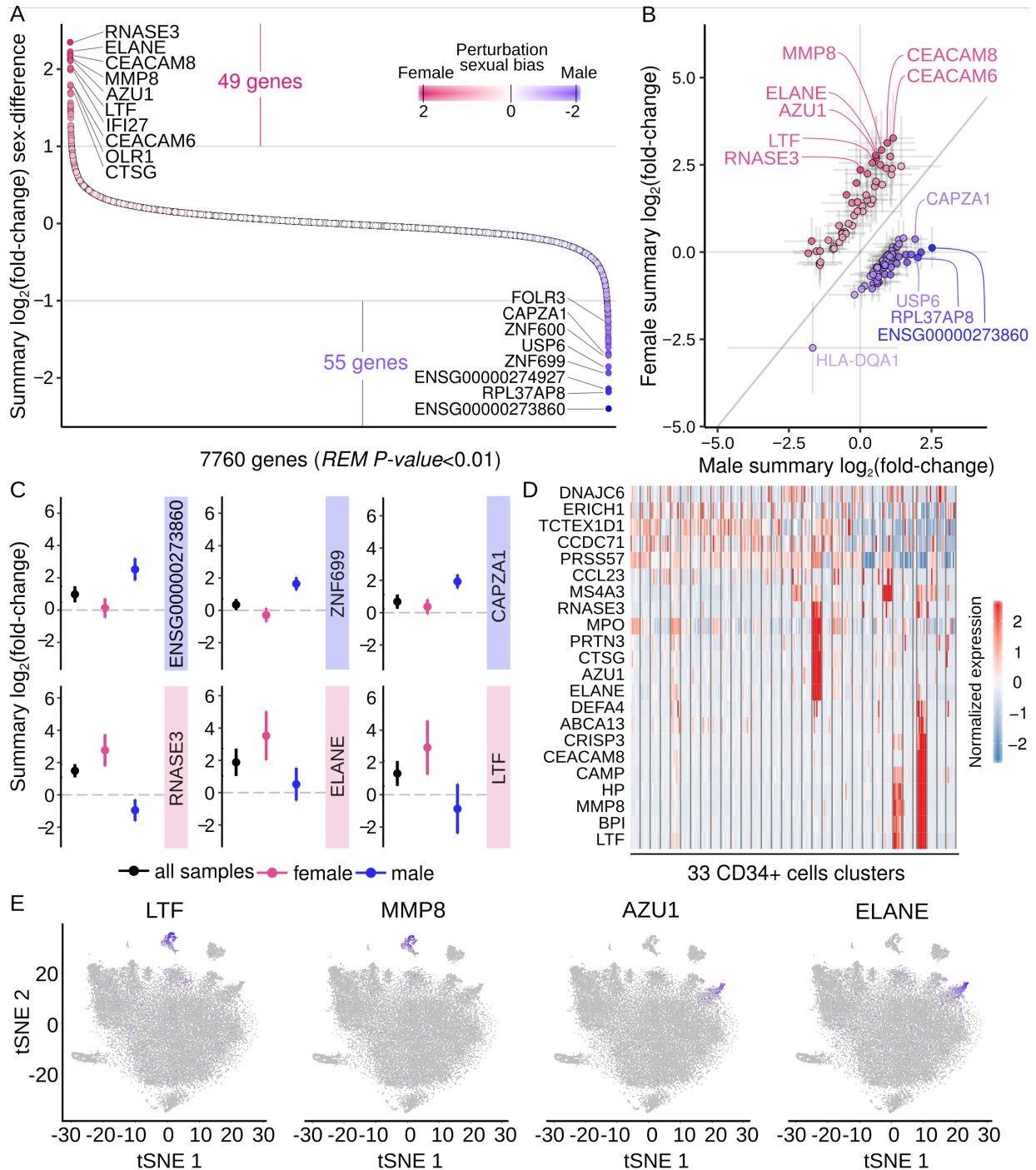


Figure 1.10: Sex difference in sJIA whole blood gene expression. A. Genes whose REM $p < 0.01$ are ranked based on the REM summary $\log_2(\text{fold-change})$ difference between female and male patients. Genes with a REM summary sex difference higher than one are highlighted. B. Sex-biased genes displayed based on their female and male REM summary $\log_2(\text{fold-change})$; confidence intervals are also shown. Color gradient as in legend of A. C. REM summary $\log_2(\text{fold-change})$ and confidence intervals of selected sex-biased genes in sJIA patients. D. Gene expression of the female-biased genes represented across the 33 cell clusters of the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). E. tSNE plots highlighting in blue the expression of the LTF, MMP8, AZU1, ELANE and CNRIP1 genes across near 45,000 human hematopoietic cells.

characteristic of immature neutrophils, specifically, early myelocytes and band-neutrophils.

Given that this female-biased and neutrophil-related expression signatures overlap with the meta-co-expressed gene module, M36, which correlates with sJIA patient clinical status, we as-

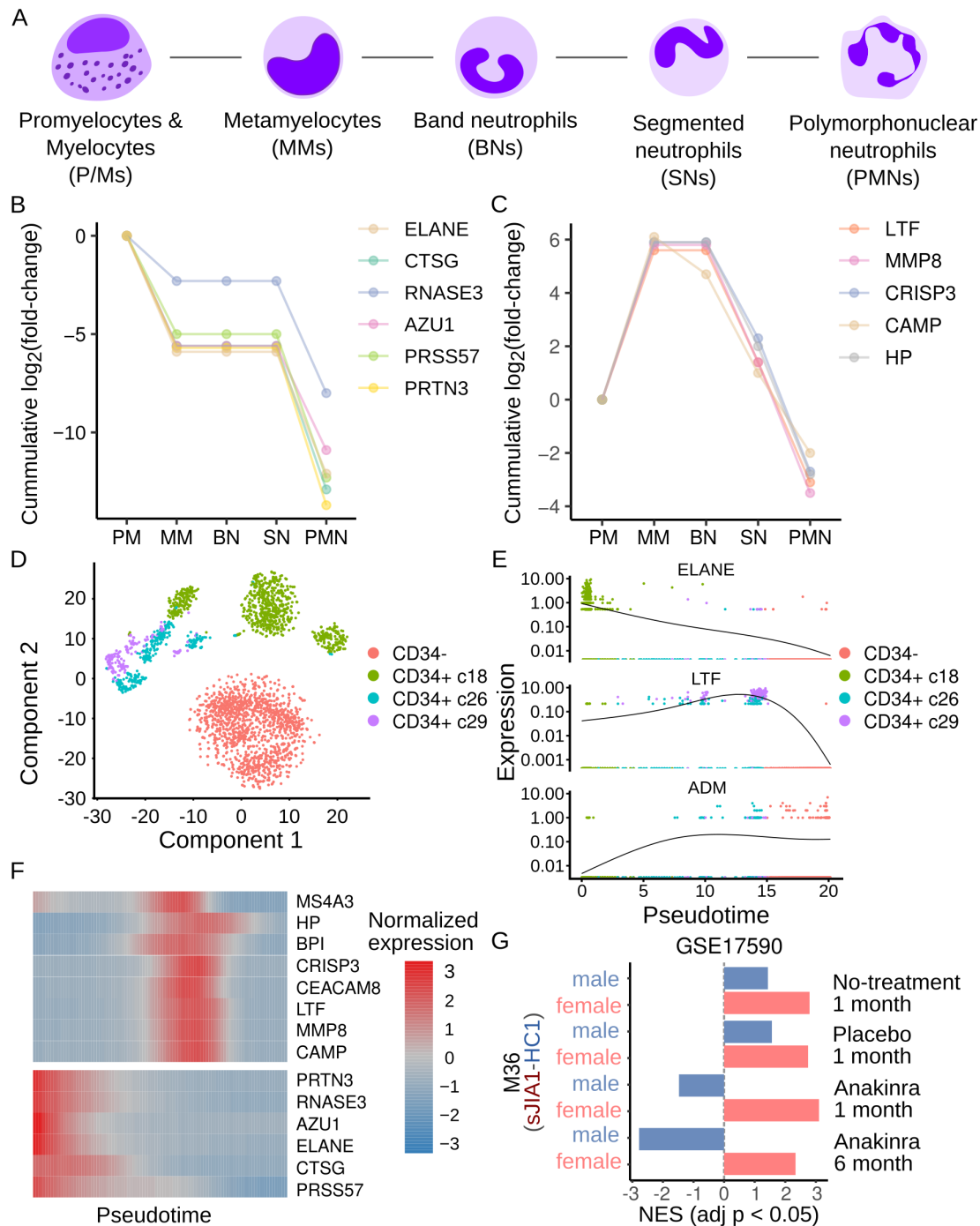


Figure 1.11: The sJIA female-biased genes represent gene markers for two immature neutrophil populations, early myelocytes and band-neutrophils. A. Scheme of the five main stages of neutrophil maturation. B and C. Cumulative expression of the sJIA female-biased and neutrophil-related genes for each of the five main stages of neutrophil maturation. Expression values were obtained from Grassi et al., (2018) supplementary tables. D and E. tSNE plots of the neutrophil maturation stages cell clusters from the scRNA-seq human hematopoietic dataset (GSE92274) (Lai et al. 2018). D. Cells are colored by the cluster they come from. E. Cells are colored based on their estimated pseudotime. F and G. Expression of the sJIA female-biased genes along with the estimated pseudotime. F. Counts of the ELANE, LTF, and ADM genes along with the pseudotime. Each dot represents a single cell. G. Z-score normalized expression of the sJIA female-biased and neutrophil-related genes along with the pseudotime.

sessed the sexual dimorphism of this signature in response to IL-1 receptor blockade. We found that the diminution of this neutrophil signature is sex-dependent (Figure 1.11G, 1.12B). Male

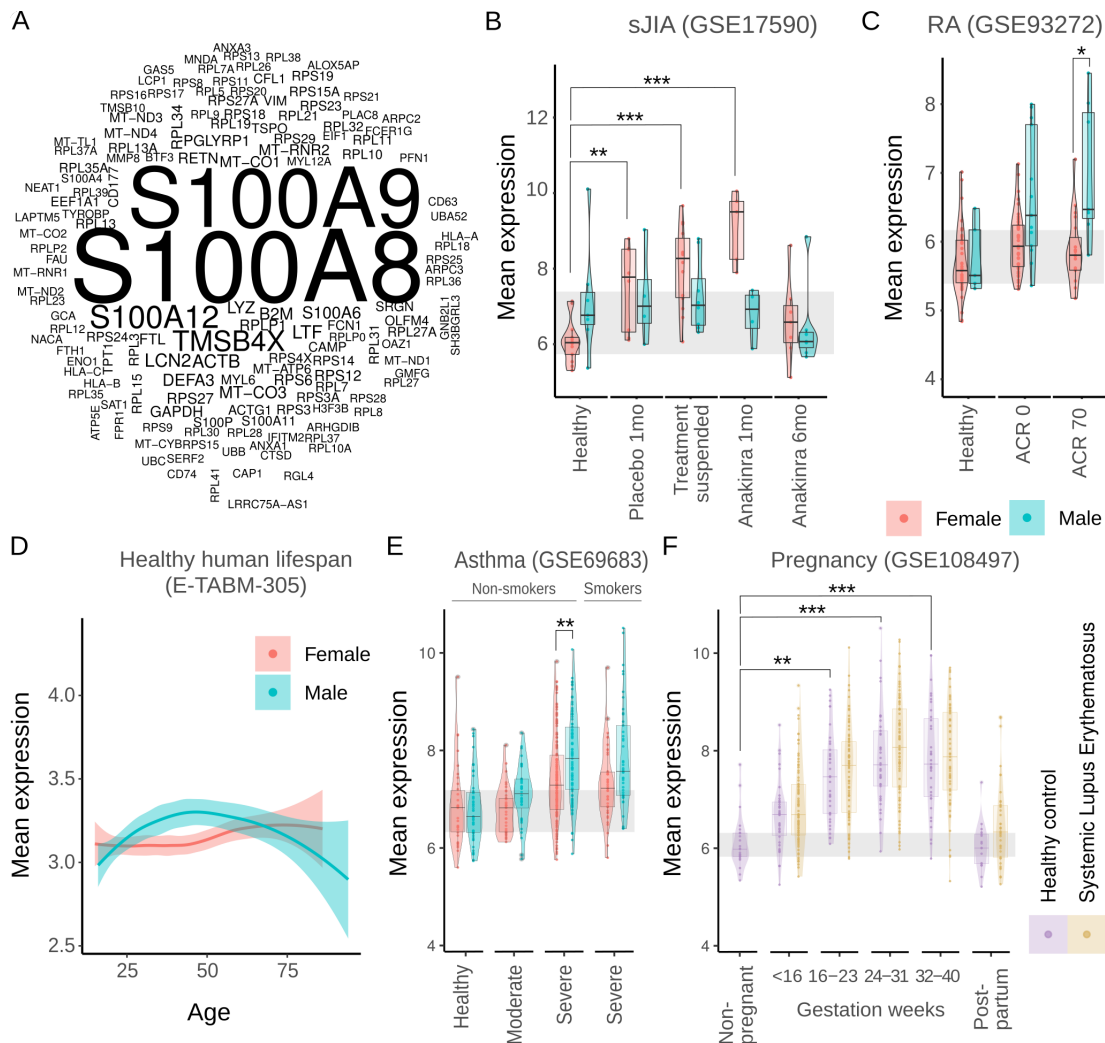


Figure 1.12: Band neutrophil gene expression signature. A. Top 150 most expressed genes in the band neutrophil cell cluster of the dataset GSE92274 (Lai et al. 2018). Gene size is proportional to the mean raw counts across the cluster cells. B. Band neutrophil signature (*LTF*, *PGLYRP1*, *HP*, *BPI*, *TCN1*, *TFF3*, *CEACAM8*, *CRISP3*, *CAMP*, *MMP8*, *CLC*, *MS4A3*) activation in blood transcriptomes of sJIA patients. Mean expression of the band neutrophil genes is presented per patient. Patients are grouped by their sex and clinical status. C. Band neutrophil signature activity in rheumatoid arthritis patients grouped by their ACR score (275 samples). The ACR score describes the relative improvement of the rheumatoid arthritis symptoms (e.g. ACR0 means no improvement). D. Mean gene expression of the band neutrophil signature along human life span. Over 1,000 samples were considered from 15 to 94 years old individuals. Mean expression values per samples were used to fit a linear regression and its 95% confidence interval. E-F. Band neutrophil gene expression signature in the context of asthma patients (498 samples) and human pregnancy (512 samples). The Pairwise Wilcoxon Rank Sum Tests method implemented in the `pairwise.wilcox.test` R function was utilized to assess the expression differences among groups in B, C, E, and F. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

patients drastically diminish this signature after one month of Anakinra treatment. In contrast, even after 6 months of Anakinra, females patients keep on over-expressing these genes (Figure 1.11G, 1.12B).

We investigated the expression of the band neutrophil signature in other biological conditions. This signature displays a sexually dimorphic up-regulation in RA patients (GSE93272) (Tasaki et al. 2018) (Figure 1.12C). Contrary to sJIA affected individuals, in RA patients the up-regulation is higher in males compared to females. To account for the age, we examined the activity of this band neutrophil signature in over 1,000 blood transcriptome samples from healthy individuals

ranging between 15 and 94 years (Array Express dataset E-TABM-305) (Kent et al. 2012). Males manifest a higher expression activity during adulthood compared to women (Figure 1.12D). During childhood and late old age the difference is less evident. In these phases, females tend to have higher expression activity compared to men (Figure 1.12D). This baseline difference may account for the contrasting activation of the band neutrophil signature in sJIA and RA. Given that a similar signature was previously observed up-regulated in blood of adult asthma patients, we investigated its sex differences (GEO dataset GSE69683) (Bigler et al. 2017) (Figure 1.12E). The asthma severity was positively correlated with the expression of the band neutrophil signature. Males diagnosed with severe asthma present higher expression levels of this signature compared to affected women (Figure 1.12E). These sexual differences suggests that the circulating band neutrophil counts might be influenced by the sex hormonal context. We investigated the whole blood expression activity of the band neutrophil signature during pregnancy (GEO dataset GSE108497) (Hong et al. 2019) (Figure 1.12F). It increases concomitantly with the gestation weeks in the blood of healthy and systemic lupus erythematosus affected women (Figure 1.12F).

1.3.7 JIAexpDB: a JIA gene expression database

The meta-analysis results of this study have been made available in the online JIA expression database (JIAexpDB) (<https://jiaexpdb.sysbio.tools>). For all analyzed genes, the summary \log_2 (fold-change) and its 95% confidence interval is presented for the main JIA classes and across studies. In addition, for the sJIA studies, results are also presented for both sexes independently.

1.4 Discussion

In this study, we coupled a systematic review and a meta-analysis to assess the previously reported gene expression signatures and identify new consistently perturbed genes in JIA patients. We cataloged more than 60,000 mentions of more than 16,000 genes throughout the genomic JIA pathophysiology literature. From them, nearly 15,000 were reported as differentially expressed genes in the peripheral leukocytes of JIA patients. This large number of perturbed genes found across studies is indicative of the complex heterogeneity of JIA pathophysiology as well as the diversity of approaches that the scientific community use to determine if a gene expression profile is associated with the JIA pathophysiology. Indeed, most of the literature reported genes have been associated with JIA in only one study. We removed this effect of the study-specific analytical approaches by meta-analyzing the available gene expression profiles of several JIA patient cohorts. The meta-analysis results as well as the systematic review results were made available at <https://jiaexpdb.sysbio.tools>. This constitutes a valuable resource for future studies which will be able to systematically compare their results in the light of previous studies. Moreover, our meta-analysis results are the first line of assessment for new potential JIA gene expression biomarkers.

There have been few attempts to find out JIA consensus gene expression patterns among the available literature (Tuller et al. 2013; Tu et al. 2017; Stevens et al. 2014; Ramanathan et al. 2018; Ter Haar et al. 2018). However, these studies lack of systematicity to gather the available JIA-functional genomics studies as well as a proper meta-analysis approach. We tackled these limitations by following the PRISMA guidelines (Moher et al. 2009) to extensively explore the JIA-functional genomics literature, meta-analyze the available transcriptome data under a common

analytical framework, and made the results accessible through the JIAexpDB online database.

We revealed that the peripheral leukocyte gene expression perturbation of the sJIA, pJIA, and eJIA clinical classes are notably different. This supports the correspondence among the JIA clinical features and the leukocyte molecular pathophysiology. We found that the level of transcriptome perturbation correlates with the clinical severity of the JIA classes. Additionally, it corroborates that the JIA peripheral leukocyte transcriptomes carry on clinical information. The sJIA presented more perturbed genes than the pJIA and eJIA as previously reported (Barnes et al. 2009). This results support the hypothesis that sJIA, pJIA and eJIA represent distinct conditions with different underlying molecular pathophysiology.

We demonstrated here that the consistently meta-co-expressed gene modules of tissue transcriptomes reveal the tissue cellular composition. This approach allowed us to estimate the composition of the peripheral leukocytes in sJIA patients. Gene expression signatures of immature neutrophils and erythrocytes, platelets, monocytes, and a sub-population of B-cell are consistently over-expressed in sJIA patients. Interestingly, the immature neutrophil and early erythroid cells as well as the platelet signatures are correlated with the clinical status of the patients and their response to IL-1 blockade. The monocyte, platelet, and immature erythrocyte signatures have been previously observed as over-expressed in sJIA patients (Macaubas et al. 2012; Hinze et al. 2010; Gorelik et al. 2013; Allantaz et al. 2007; Fall et al. 2007; Barnes et al. 2009; Quartier et al. 2011). However, the B-cell signature has not. Given that this meta-co-expressed modules comprise gene markers for specific leukocyte subpopulations and that sJIA patients experience leukocytosis, neutrophilia, and monocytosis in active untreated sJIA patients (Cepika et al. 2017), we argue that this expression activity represents the cellular composition dynamics of sJIA patients. In addition, we found that an NK cell and another B-cell signature are consistently down-regulated in sJIA patients. This NK-cell dysfunction has been previously observed (Gorelik et al. 2013; Barnes et al. 2009; Put et al. 2017; Fall et al. 2007). Contrary to previous studies, our results did not confirm the down-regulation of a mitochondria signature in sJIA patients. Our meta-co-expression results have uncovered two B-cell gene expression signatures with opposite expression activity. The module M21 that enriched for naive B cell-related genes was observed down-regulated. In contrast, the module M27 that enriched for B-lymphoblasts-related genes, was up-regulated in sJIA patients. This indicates that sJIA patients keep on activating B cells regardless of the time after the disease onset. A previous study identified a gene expression signature, which distinguish the age at onset of sJIA patients, whose genes intersect with the B-cell related meta-co-expressed module M21. The study reported that this B cell-related signature is up-regulated in patients with disease onset before 6 years (Barnes et al. 2010). We hypothesized that previously observed differences on the JIA pathophysiology based on the age at onset are mainly driven by B-cells. This difference might be related to the early hormonal event known as adrenarche.

We demonstrated in this study that sJIA transcriptome perturbation displays sexual dimorphism. A circulating immature neutrophil gene expression signature is more positively activated in female than in male patients. This sJIA female-biased signature comprises 49 genes. These genes enriched for neutrophil-related genes. Given that different populations of neutrophils have been observed in higher frequencies in sJIA patients, we wanted to know if this female-biased signature represents specific subsets of neutrophils. We investigated the expression of these female-biased genes in two hematopoietic gene expression datasets; one of near 50,000 CD34+ single cells and

another containing the transcription profiles of the main five neutrophil maturation stages. The female-biased genes presented similar expression patterns in these two independent datasets. This signature comprises gene expression markers for the early neutrophil myelocytes and the band neutrophils. Interestingly, as our systematic review reveals, these genes have been previously observed up-regulated in sJIA patients; however, previous studies did not assess the sex contribution to this perturbation. For instance, Ramanatham et al. (2018) found neutrophil-related genes (*MMP8*, *MMP9*, *CRISP3*, *CEACAM1*, *CEACAM8*, *CRISP3*, *BPI*, *LCN2*, *DEFA4*, *LTF*, *RNASE3*, *CAMP*, *MPO*, *ELANE*, *CTSG*, *RNASE2*, *CLU*, *CTSA*, and *AZU1*) being upregulated in sJIA patients by re-analyzing the GSE13501 and GSE21521 GEO data series. They also isolated the neutrophils from PBMC samples and demonstrated that they are more abundant in sJIA patients. Because of neutrophils are supposed to precipitate during density-based PBMC isolation protocols, these neutrophils in the PBMC fraction are called low density neutrophils. These low density granulocytes have been also observed circulating in higher numbers in systemic lupus erythematosus patients. For instance, Villanueva et al., (2011) compared the expression profiles of low density neutrophils obtained from lupus patients and healthy individuals with normal density granulocytes. They identified differentially expressed genes that overlaps with our female-biased neutrophilic signature. In addition, Bennett et al. (2003) and Denny et al. (2010) showed that the low density neutrophils isolated from PBMC preparations obtained from lupus patients comprises segmented and band neutrophils as well as myelocyte-like cells by microscopy inspection, confirming that low density neutrophils comprise at least two neutrophil subpopulations. Based on the foregoing, we hypothesized that the sJIA pathophysiology presents a sexual dimorphism mostly driven by band neutrophils characterized by a low density and higher occurrence in female patients presenting systemic inflammation. Beyond the PBMC preparations, the sJIA female-biased signature has also been found in normal density granulocytes but with less intensity. For instance, using qPCR, Ramanatham et al., (2018) observed the upregulation of the *MPO* and *MMP8* genes in sJIA normal density neutrophils compared to healthy controls. However, they did not confirm up-regulation of the *ELANE* gene, which is highly expressed in PBMC transcriptomes. In addition, Brown et al. (2018) characterized the transcriptome of normal density neutrophils in sJIA patients. We observed in their reported dataset (GSE122552) that the genes *MMP8*, *LTF*, and *CEACAM8*, which were lowly expressed compared to gene markers of mature neutrophils, presented higher expression in active sJIA patients but they do not in inactive sJIA patients and healthy controls. They also counted two populations of normal density neutrophil cells using the CD16 and CD62 surface markers, and found that the CD16+CD62Llo population have higher counts in active and lower numbers in inactive sJIA patients compared to healthy pediatric individuals. Moreover, Jarvis et al. (2006) characterized two sub-populations of normal density neutrophils based on the presence of the protein encoded by the *MPO* gene. They found that MPO-positive normal density neutrophils isolated from pJIA samples displays a dysfunction on the NAD(P)H fluctuations after LPS stimulation. Notably, our meta-analysis results indicated the expression of the *MPO* as sexually dimorphic in sJIA patients. We also found it to have a distinctive up-regulation in the early stages of neutrophil maturation. Therefore, we hypothesized that the sJIA female-biased signature represents a non-polymorphonuclear neutrophil population with broad density spectrum, which can get differentially activated while retaining transcriptome features of early myelocytes and band neutrophil cells.

The neutrophil-related gene expression signature that we identified through our meta-co-expression approach correlates with the disease status and the response to IL-1 receptor blockade. This meta-co-expressed module M36 also intersects the identified sJIA female-biased signature and comprises marker genes of band neutrophils. We found that the genes *S100A8*, *S100A9*, and *S100A12* are the most highly expressed genes in the band neutrophil cell cluster (Figure 1.12A). The protein products of these genes are clinical markers to monitor the inflammatory status of sJIA patients. Serum concentrations of the S100A8, S100A9, and S100A12 proteins are higher in relapse and decrease in remission (Ter Haar et al. 2018; Swart et al. 2016; Rothmund et al. 2014). This suggests that band-neutrophil signature correlates with symptom severity, relapse and remission status, and the IL-1 receptor blockade progression. We hypothesize that the reduction of band neutrophils contributes to the observed reduction of these gene products in the peripheral blood. In addition, we showed that this M36 neutrophil signature is correlated with the response to Anakinra in a sex-dependent manner. Gattorno et al. (2008) found that neutrophil counts in sJIA patients are associated with the chances of responding to IL-1 blockade. We propose that this neutrophil-related meta-co-expressed module M36 can be utilized as a tool to stratify and monitor sJIA patient progression.

The transcriptome sexual dimorphism in peripheral leukocytes of sJIA patients indicates that in this disease the neutrophil maturation and activation trajectories are sex-dependent. Similarly, we showed that rheumatoid arthritis and adult asthma patients display a sexually dimorphic activation of the band neutrophil signature. However, in adult arthritis, this signature is more highly expressed in males compared to women. We showed this dimorphism is age dependant and more evident during adulthood in the blood of healthy individuals. This sexual difference suggests that sex hormones may modulate the release of immature neutrophils to the peripheral blood. Blazkova et al., (2017) identified a gene expression signature in whole blood transcriptomes that overlaps with the meta-co-expressed module M36 and displays sexual dimorphism in healthy samples. They also showed that this signature increases during healthy pregnancy. In accordance, Bigler et al. (2017) observed an expansion of immature neutrophil populations during pregnancy. This supports the idea that sexual hormones may play an important role in the sJIA pathophysiology through the modulation of the neutrophil maturation process. Similar immature neutrophil populations have been characterized in cancer patients (Sagiv et al. 2015). These cells are related to poor prognosis and have immunosuppressive profile, principally affecting lymphocytes (Sagiv et al. 2015; Veglia et al. 2019). Given that this neutrophils population expansion have also been observed in several inflammatory contexts (Carmona-Rivera and Kaplan 2013) and, as we showed, it is positively correlated with the sJIA, asthma, and RA severity as well as during gestation. Together, these suggest that instead of a pro-inflammatory profile, this circulating band neutrophils might represent a homeostatic feedback to resolve inflammation through lymphocyte cytotoxicity. Indeed, we showed that the NK cell and a B-lymphocyte expression signatures were consistently down-regulated in sJIA. Additionally, this regulatory feedback might be influenced by the hormonal status. Therefore, we recomend future studies on JIA pathophysiology should consider the diversity of neutrophil sub-populations and its sex-dependent differential activation during the disease relapse.

1.5 Methods

1.5.1 JIA-functional genomics systematic review

To systematically and rigorously review the functional genomics discoveries related to JIA, we performed a systematic review following the PRISMA guidelines (Moher et al. 2009) (Figure 1.2A). We systematically looked for PubMed records made available between 2000 and 2018 (May) related to genomic, transcriptomic, and epigenomic studies on JIA using the *rentrez* R package (Winter 2017) . We identified 427 functional genomics-related records from the near 6.200 JIA- related records (Figure 1.2B). We screened those records by reading their abstract in order to discard the studies whose design does not include the analysis of JIA primary tissue-sourced data. We further inspected the 87 filtered articles by reading articles completely, and found 67 articles analyzing genomic, transcriptomic, or epigenomic data obtained from primary JIA patient samples (Figure 1.2A). We thoroughly explored the 67 selected studies to gather the reported JIA-related genes and raw data when available. We classified the JIA reported genes into five categories, in-paper cited, which includes the mentioned genes in the main written part of the studies; differentially expressed genes; SNP-nearest genes; epigenetically marked genes; and network- related, when the genes were reported in JIA-associated molecular networks. In addition, we collected information regarding the JIA class, tissue, probe or SNP name, genomic region, effect size, and the JIA association probability associated with each gene when such information was provided by the authors.

1.5.2 Inclusion criteria of the JIA gene expression meta-analysis

We gathered the GEO (Barrett et al. 2013) available raw data of the 67 reviewed studies. We compared the sample GEO accessions and the sample raw-data files 128-bit *MD5* hashes among the studies to filter out duplicate samples. When duplicated among studies, we retained the foremost published sample. We included only the GEO datasets containing JIA patients and healthy pediatric controls with more than three original samples per class. We curated the available metadata of the included datasets in order to define clinically uniform groups of JIA patients. We gathered the metadata from GEO and a review of the clinical inclusion criteria reported in the original manuscripts. Based on the curated metadata, we defined pairwise contrasts. Each contrast comprises a clinically homogeneous set of JIA patients and healthy pediatric samples from the same GEO dataset (Figure 1.2E).

1.5.3 Gene expression data pre-processing

We performed background correction, data normalization, probe to gene collapsing, and quality control independently for each GEO dataset with the R software environment (R Core Team 2015). We similarly preprocessed all datasets, however, the background correction was different among the Illumina and Affymetrix platforms. For the Illumina platforms we first discarded the oligos whose fluorescence intensity was not significantly different from the background array fluorescence in all samples p (detection) > 0.05 ; after this, the oligos whose intensity values are less than 10 were replaced by a randomly sampled value from a distribution of mean 10 and standard deviation equal to one. On the other hand, for the Affymetrix datasets, we applied the robust multi-array approach

(RMA) implemented in the *affy* R package (Gautier et al. 2004). We performed normalization by averaging the sample intensities distributions in each dataset through the quantile method implemented in the *affy* R package. In order to allow cross-comparisons between platforms and to consider non-coding regions, we assigned ENCODE (Davis et al. 2018), NONCODE (Zhao et al. 2016), and LNCipedia (Volders et al. 2013) identifiers to the probe/oligos by aligning them on the reference Hg19 genome and intersecting their coordinates with the mentioned reference identifiers. For the further downstream analysis, we retained the probe/oligo with the highest mean expression as representative of each gene symbol.

Patient sex metadata was corroborated by clustering samples based on the expression of *XIST*, *RPS4Y1*, and *KDM5D* genes.

After data pre-processing, we used the principal variation component analysis (PVCA) method implemented in the *pvca* R package (George and Chen 2009) in order to quantify how much of the gene expression variation is explained by the available clinical and technical metadata in each dataset. In the cases where the scanning date explained more than 30% of the observed variance, we removed the batch effect using the ComBat method implemented in the *sva* R package (Johnson et al. 2007).

1.5.4 JIA gene expression meta-analysis

We quantified the circulating leukocyte gene expression perturbation of JIA patients from healthy pediatric individuals through the *LIMMA* method implemented in R (Ritchie et al. 2015). In this way, we calculated the gene $\log_2(\text{fold-change})$ and its confidence interval for each contrast. We estimated the gene summary $\log_2(\text{fold-change})$ of equivalent contrasts by fitting their effect size and variation using the REM implemented in the *metafor* R package (Viechtbauer 2010).

1.5.5 JIA gene co-expression meta-analysis

To identify the set of genes consistently co-expressed in healthy and JIA blood leukocytes, we integrated the co-expression module memberships of independent datasets. First, we defined gene co-expression modules for each dataset with the *CEMiTool* R package (Russo et al. 2018). We set *CEMiTool* to only consider the genes whose variation is out of a null inverse gamma distribution ($p < 0.1$). Based on the gene module membership, we built weighted co-membership networks to combine co-expression results from different datasets. Such consensus networks comprise nodes, which stand for genes and are connected if they are part of the same module in any dataset. The edge weight represents the number of datasets where the gene pair appears in the same module. We analyzed the co-membership network structural modularity to define the consistently co-expressed gene modules. For this purpose, we used the community detection algorithm implemented in the Gephi 9-2.1 software (Bastian et al. 2009), which we also used for network visualization. We assigned either a biological function or a cell type to the consistently co-expressed gene modules through the enriching analysis method implemented in the Enrichr web tool (Kuleshov et al. 2016). We quantified the expression activity of the consistently co-expressed modules using the GSEA method implemented in the *fgsea* R package (Sergushichev 2016). We used the expression $\log_2(\text{fold-change})$ of the JIA patients in relation to the healthy controls as the reference ranking in the *fgsea* analysis.

1.5.6 Single-cell gene expression analysis

We re-clustered the human leukocyte single-cell gene expression dataset available in the GEO accession GSE92274 (Lai et al. 2018) with the *Seurat* R package suite of functions (Butler et al. 2018). Initially, we merged the available raw count batches into a single matrix. We filtered out the cells with no more than 200 and more than 2500 detected genes. In addition, we removed the cells whose mitochondrial-related gene counts represents more than 5% of the total raw counts. We normalized the filtered expression matrix using the *LogNormalize* method. We scaled and centered the genes in the datasets regressing the number of detected genes, the percentage of mitochondrial raw-counts, and the experimental batch. Finally, clustered the hematopoietic cells by using the 27 most exploratory principal components and setting the resolution equal to 3.5. For visualization, we used the t-Distributed Stochastic Neighbor Embedding dimensionality reduction using the same 27 principal components. For pseudotime estimation, we utilized the *Monocle* R package (Trapnell et al. 2014).

1.5.7 Interactome reconstruction

In all sub-interactome reconstruction cases, we utilized the NetworkAnalyst resources (<https://www.networkanalyst.ca>) (Zhou et al. 2019). We used the interest gene list as query and the InnateDB as reference interactome. In all cases, we presented the zero-order network obtained.

Chapter 2

MetaVolcano: making gene expression meta-analysis easily accessible

During the last two decades, the amount of available transcriptome profiles of a great diversity of biological conditions has increased exponentially. In parallel, several analytical methods have been proposed to combine gene expression results from independent studies. However, the lack of open source code, proper documentation and intuitive result visualizations are recurrent issues among the available tools for gene expression meta-analysis. These have been limiting the accessibility and exploitation of the great availability of gene expression data. In this work, we propose MetaVolcano, a tool to perform and visualize gene expression meta-analysis in an easy manner. MetaVolcano utilizes the reasoning of the volcano plot and implements three approaches to produce summary statistics, a random effects model, Fisher's method of combining p -values, and a vote-counting procedure. We have designed a MetaVolcano web tool which allows users to easily meta-analyse more than 100,000 profiles from more than 3,000 studies and 60 species available in the Expression Atlas database (<https://metavolcano.sysbio.tools>). The MetaVolcano implemented methods have also been made available in the Bioconductor R package MetaVolcanoR (<https://bioconductor.org/packages/MetaVolcanoR>). MetaVolcano is an accessible and easy to interpret tool that allows the research community to easily obtain robust expression markers by exploiting the publicly available transcriptome data.

2.1 Introduction

The measurement of how gene expression changes under different biological conditions is necessary to reveal the gene regulatory programs that orchestrate the cellular phenotype. Comparing the expression of genes under a given condition against a reference biological state is a common practice to identify sets of differentially expressed genes. These perturbed genes point out the genomic regions functionally relevant under the biological condition of interest. Although individual genome-wide expression studies have small signal to noise ratios, the huge volume of genomic data available allows the combination of differential gene expression results from independent studies to overcome this limitation.

Transcriptome data repositories such as the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>), ArrayExpress,

(<https://www.ebi.ac.uk/arrayexpress/>), and the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) offer programmatic access to vast amounts of transcriptome data. Despite this great availability, these resources do not offer uniformly curated and preprocessed data. This imposes data curation and analytical efforts to exploit these resources. On the contrary, the Expression Atlas added value database (<http://www.ebi.ac.uk/gxa>) has gathered, consistently curated, pre-processed, and made available hundreds of thousands of transcriptome profiles (Papatheodorou et al. 2018). These features therefore make Expression Atlas appropriate to source gene expression meta-analysis on-demand.

There exist three main approaches to combine differential gene expression results from independent studies: combining effect sizes, p -values, and rank statistics (Tseng et al. 2012). The random effects model (REM) combines effect sizes by a linear model with an underlying true effect size plus a random error in each study and random effects for the inter-study heterogeneity (Chang et al. 2013). Among the methods that combine effect sizes, REM is the most stable and robust at identifying perturbed genes regardless of intra-study partition and study selection (Chang et al. 2013). Despite this, REM has a limited detection capability in comparison to combining p -values methods. Harrison et al. (2019) proposed the Topconfects methodology to improve the detection capability and biological association of effect size based methods. The Fisher’s approach is one of the most used methods to combine p -values in the gene expression meta-analysis literature (Tseng et al. 2012). This method sums the natural logarithm of the K independent p values as a X^2 statistics of $2K$ degrees of freedom (Tseng et al. 2012). Rank-based approaches assess if the position of a certain gene in a effect-size or p -value based rank is consistent across several independent studies (Chang et al. 2013). The underlying assumption of these methods is that genes that consistently appear at the top or bottom of a rank are associated with the biological condition. Finally, vote-counting meta-analysis approaches consist of counting the number of studies where a given gene had a p value and fold-change under certain arbitrary threshold (Tseng et al. 2012). Although this strategy lacks statistical assessment, it is useful to account for the available evidence of a given gene in a group of studies. This is especially relevant given that transcriptome technological platforms do not measure the same genes. Consequently, in most expression meta-analysis, there are several genes partially represented across the combined studies.

In this work, we adjusted the widely used volcano plot, which displays genes in a 2D space comprising the effect size and p -value, to explore three meta-analysis approaches. This tool has been made available to easily exploit the enormous Expression Atlas resource. MetaVolcano is also available as a Bioconductor R package.

2.2 Results

2.2.1 Random effect model MetaVolcano

The REM MetaVolcano allows users to explore and identify consistently perturbed genes in a volcano plot whose x -axis encodes the summarized fold change and the y -axis the probability that the summary fold-change is not different than zero (Figure 2.1A). We utilized the maximum likelihood REM algorithm implemented in the *metafor* R package (Viechtbauer 2010). To improve the detection capabilities of this method, REM MetaVolcano ranks genes based on the *Topconfects* method which takes into account the gene summary fold change and its variance (Harrison et al.

2019). Additionally, to account for the gene representation in the combined studies, we included the gene sign consistency variable as a color-based axis (Figure 2.1A). The sign consistency represent the number of studies where a given gene has the same fold change sign. For instance, a gene detected in five studies, from which in three of them it was down-regulated, will get a sign consistency score of $2 + (-3) = -1$. This score tend to zero when either a gene is poorly represented across the combined studies or it presents contradictory perturbation. Therefore users can combine the summary effect size and the content of consistent evidence to define the consistently perturbed genes. Furthermore, MetaVolcano allows users to explore the forest plot of any given gene based on the summary fold change and the gene effect size across the combined studies (Figure 2.1F-G).

2.2.2 Combining p -value MetaVolcano

The combining p -value MetaVolcano applies Fisher’s method to combine differential gene expression p -values from independent studies. Such combination is possible due to the implementation of the *metap* R package (Viechtbauer 2010). Besides this summary p value, this MetaVolcano visualizes the mean or median fold change across studies (Figure 2.1B). Additionally, the expression sign consistency axis is also visualized (Figure 2.1B).

2.2.3 Vote-counting MetaVolcano

The vote-counting MetaVolcano ranks genes based on the number of studies where genes were defined as differentially expressed based on the user-defined p -value and fold change thresholds. It displays the number of differentially expressed and unperturbed genes per study (Figure 2.1C). It visualizes genes based on the number of studies where genes were identified as differentially expressed and the gene fold change sign consistency (Figure 2.1D).

2.2.4 Discussion

Despite the great availability of gene expression data repositories and gene expression meta-analysis methods, there are few available tools to easily get robust expression markers by exploiting the publicly available transcriptome data. In this work, we introduced MetaVolcano, an open source and well documented tool to easily conduct gene expression meta-analysis. The web tool allows users to meta-analyse the vast Expression Atlas resource. Therefore, the entire meta-analysis workflow, data gathering, curation, differential expression analysis and the combination of independent study results, are all included in the MetaVolcano web tool. Users only need to query a given biological condition and pick up the contrasts they want to meta-analyse.

There are few available web tools that cover the whole gene expression meta-analysis workflow. Recently, Toro-Domínguez et al. (2019) deployed the ImaGEO web tool, which offer gene expression meta-analysis on demand of GEO series. It combines independent study results using the meta-analysis methods implemented in the metaDE package (Toro-Domínguez et al. 2019). *metaDE* is an R package from the metaOmics suit that implements several methods for the three main meta-analysis strategies (Ma et al. 2019). This ImaGEO on-demand meta-analysis tool is study centric and besides the study GEO accession, it requires that the user carries out the curation and study design definition. It utilizes the heatmap visualization to present the fold change

across studies for only the identified perturbed genes. This heatmap approach to explore meta-analysis results make it difficult to prioritize genes based on their effect size since they display normalized values by gene. It also limits users to explore their gene of interest under a given condition if it is not considered as perturbed. Similarly, the NetowrkAnalyst web tool, which utilizes the INMEX meta-analysis approach (Xia et al. 2013), visualizes gene expression meta-analysis results as interactive heatmaps (Zhou et al. 2019). By contrast, the MetaVolcano uses an intuitive volcano plot visualization to present three different meta-analysis approaches which account for the effect sizes and the available gene evidence across the combined studies.

Another recent attempt to make gene expression meta-analysis more accessible involved the implementation of the *metaMA* and *metaRNAseq* R packages into a Galaxy platform (Blanck and Marot 2019). Users are required to gather and curate the data as well as define the study design. The implemented meta-analysis methods of this tool rely on the combination of p -values. Therefore, they cannot account for the effect size and do not consider the consistency of the gene perturbation.

Although the MetaVolcano tool includes the whole gene expression meta-analysis workflow by utilising the curated resources of the Expression Atlas database and applying three meta-analysis approaches, the tool is limited in that it does not include a broader collection of meta-analysis methodologies and it is restricted to the scope of the Expression Atlas resources. Nevertheless, the MetaVolcano presents a powerful tool to perform meta-analysis. The REM approach combines effect sizes, which represent fold changes, considering the unaccountable sources of variability among independent studies. Therefore, its summary results quantitatively encodes gene perturbation, which have a clearer biological meaning than p -values and rank statistics. To complement the sensitivity of this approach, we implemented the *Topconfects* approach to rank the perturbed genes and included the combining p -values Fisher’s method so that results from the REM approach can be compared to this traditional method. Additionally, the vote-counting approach complements the above approaches by summarising the consistency in gene regulation among the available studies. Moreover, the Expression Atlas resources are continually growing which guarantees an ever increasing gene expression curated resource which is explorable through the querying of biological conditions.

The implemented MetaVolcano methods for gene expression meta-analysis are complementary. The REM offer the robustness and stability on the summary calculation while the *Topconfects* and Fisher’s methods improve the detection power. Additionally, the vote-counting approach permits to account for the available evidence. Collectively, this offers flexibility for users defining the genes whose expression is consistently perturbed in a given biological condition.

2.3 Methods

The gene expression meta-analysis methods included in MetaVolcano were implemented in the *MetaVolcanoR* R package. The source code, documentation and tutorials are publically available at <https://github.com/csbl-usp/MetaVolcanoR> and <https://bioconductor.org/packages/MetaVolcanoR>. The suit of *MetaVolcanoR* functions allow users to combine differential gene expression results from independent studies and visualizes the results on interactive plots.

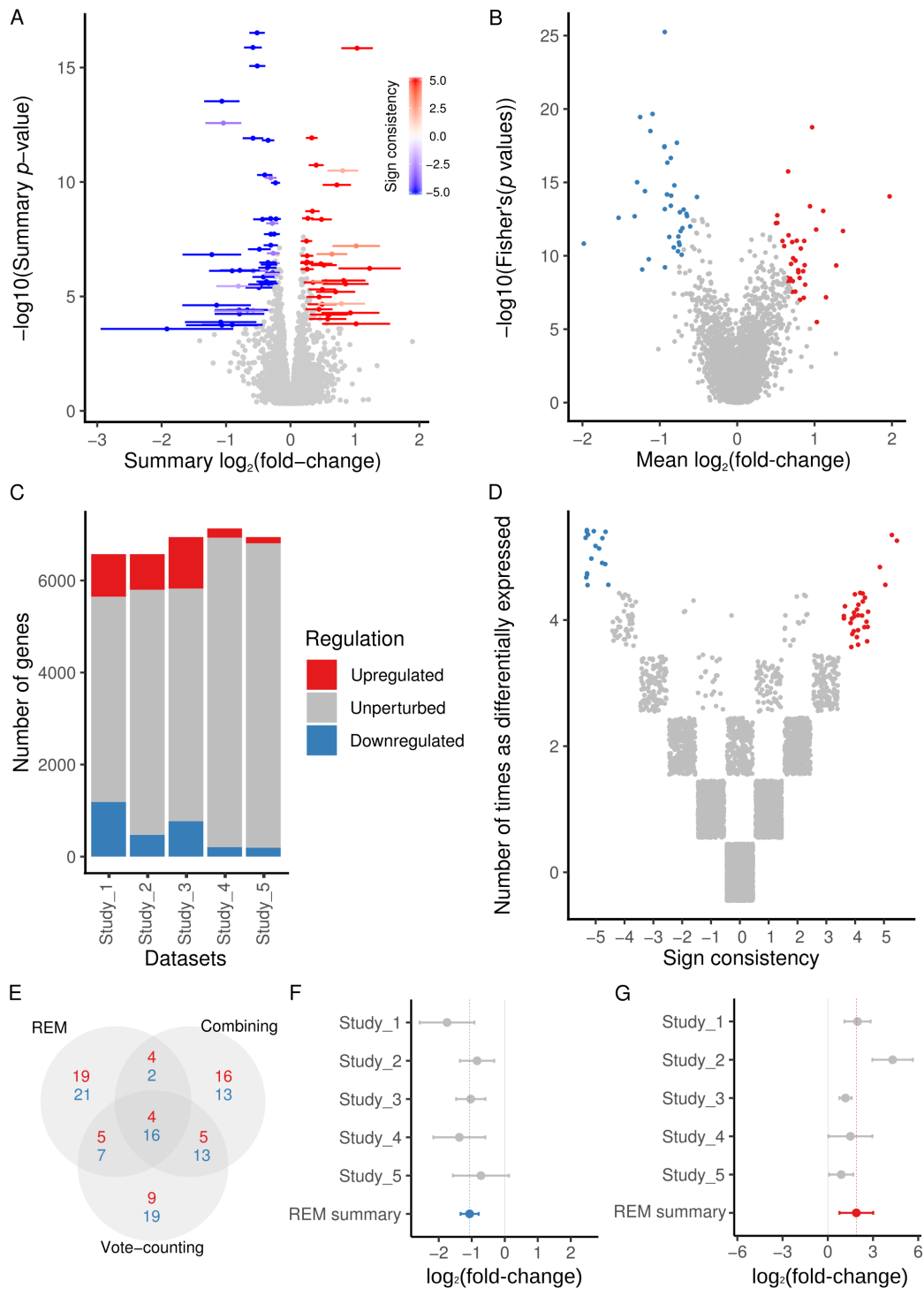


Figure 2.1: MetaVolcano implemented visualizations of gene expression meta-analysis results. Differential expression results from five demo studies are combined. This data and tutorial to reproduce the figures are available in the MetaVolcanoR R package vignette (<https://bioconductor.org/packages/MetaVolcanoR>). **A.** REM MetaVolcano. Top 1% most perturbed genes according to the Topconfects ranking are coloured and the 95% confidence intervals are depicted. Color scale represent the sign consistency of each gene. **B.** Combining p-values MetaVolcano. Top 1% most perturbed genes are coloured in red or blue. **C.** Number of perturbed genes by study. Users can set the fold-change and p-value thresholds to define the differentially expressed genes. **D.** Vote-counting meta-volcano. Top 1% of the consistently perturbed genes are coloured. **E.** Intersection of the 1% top perturbed genes of the three MetaVolcano approaches. The red and blue numbers indicate up-regulated and down-regulated genes respectively. **F** and **G.** Forest plot of two genes that were identified as part of the 1% top consistently perturbed genes by the three MetaVolcano approaches.

Chapter 3

Conclusions

- Approximately 15,000 genes have been associated with juvenile idiopathic arthritis (JIA) in the transcriptome literature. Most of which were reported in only one study.
- The peripheral leukocyte gene expression perturbation in JIA is consistent with its clinical classification.
- The systemic JIA pathophysiology presents sexual dimorphism, which is characterized by an increase in the circulating immature neutrophil gene expression signature in female patients.
- sJIA pathophysiology is partially driven by a diversity of neutrophil sub-populations that display a sex-dependent differential activation during the disease relapse and in response to anti-IL-1 treatment.
- The circulating band neutrophil signature manifest a sexual dimorphic activation in inflammatory diseases like sJIA, RA, and asthma. Likewise along human lifespan and increases during pregnancy.
- The JIAexpDB (<https://jiaexpdb.sysbio.tools>) is a comprehensive resource to consult the expression perturbation of any gene in the peripheral leukocytes of JIA patients.
- MetaVolcano constitutes an easily accessible tool to exploit the enormous publicly available gene expression data through statistical meta-analysis (<https://metavolcano.sysbio.tools>, <https://github.com/csbl-usp/MetaVolcanoR>).

Chapter 4

Bibliography

- Allantaz, F., Chaussabel, D., Stichweh, D., et al. 2007. Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. *The Journal of Experimental Medicine* 204(9), pp. 2131–2144.
- Anon 2015. 266. Standardized Mortality Rates are Increased in Patients with Severe Juvenile Idiopathic Arthritis. *Rheumatology*.
- Banchereau, R., Cepika, A.-M., Banchereau, J. and Pascual, V. 2017. Understanding human autoimmunity and autoinflammation through transcriptomics. *Annual Review of Immunology* 35, pp. 337–370.
- Barabási, A.-L., Gulbahce, N. and Loscalzo, J. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics* 12(1), pp. 56–68.
- Barnes, M.G., Grom, A.A., Thompson, S.D., et al. 2010. Biologic similarities based on age at onset in oligoarticular and polyarticular subtypes of juvenile idiopathic arthritis. *Arthritis and Rheumatism* 62(11), pp. 3249–3258.
- Barnes, M.G., Grom, A.A., Thompson, S.D., et al. 2009. Subtype-specific peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis. *Arthritis and Rheumatism* 60(7), pp. 2102–2112.
- Barrett, T., Wilhite, S.E., Ledoux, P., et al. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* 41(Database issue), pp. D991-5.
- Bastian, M., Heymann, S. and Jacomy, M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks.
- Bennett, L., Palucka, A.K., Arce, E., et al. 2003. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *The Journal of Experimental Medicine* 197(6), pp. 711–723.
- Bigler, J., Boedigheimer, M., Schofield, J.P.R., et al. 2017. A Severe Asthma Disease Signature from Gene Expression Profiling of Peripheral Blood from U-BIOPRED Cohorts. *American Journal of Respiratory and Critical Care Medicine* 195(10), pp. 1311–1320.

- Blanck, S. and Marot, G. 2019. SMAGEXP: a galaxy tool suite for transcriptomics data meta-analysis. *GigaScience* 8(2).
- Blazkova, J., Gupta, S., Liu, Y., et al. 2017. Multicenter systems analysis of human blood reveals immature neutrophils in males and during pregnancy. *Journal of Immunology* 198(6), pp. 2479–2488.
- Brachat, A.H., Grom, A.A., Wulffraat, N., et al. 2017. Early changes in gene expression and inflammatory proteins in systemic juvenile idiopathic arthritis patients on canakinumab therapy. *Arthritis Research & Therapy* 19(1), p. 13.
- Braz, A.S., de Andrade, C.A.F., da Mota, L.M.H. and Lima, C.M.B.L. 2015. Recommendations from the Brazilian Society of Rheumatology on the diagnosis and treatment of intestinal parasitic infections in patients with autoimmune rheumatic disorders. *Revista brasileira de reumatologia* 55(4), pp. 368–380.
- Brown, R.A., Henderlight, M., Do, T., et al. 2018. Neutrophils From Children With Systemic Juvenile Idiopathic Arthritis Exhibit Persistent Proinflammatory Activation Despite Long-Standing Clinically Inactive Disease. *Frontiers in immunology* 9, p. 2995.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 36(5), pp. 411–420.
- Carmona-Rivera, C. and Kaplan, M.J. 2013. Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Seminars in immunopathology* 35(4), pp. 455–463.
- Cattalini, M., Soliani, M., Caparello, M.C. and Cimaz, R. 2019. Sex differences in pediatric rheumatology. *Clinical Reviews in Allergy & Immunology* 56(3), pp. 293–307.
- Cepika, A.-M., Banchereau, R., Segura, E., et al. 2017. A multidimensional blood stimulation assay reveals immune alterations underlying systemic juvenile idiopathic arthritis. *The Journal of Experimental Medicine* 214(11), pp. 3449–3466.
- Chang, L.-C., Lin, H.-M., Sibille, E. and Tseng, G.C. 2013. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. *BMC Bioinformatics* 14, p. 368.
- Cho, D.-Y., Kim, Y.-A. and Przytycka, T.M. 2012. Chapter 5: Network biology approach to complex diseases. *PLoS Computational Biology* 8(12), p. e1002820.
- Cui, A., Quon, G., Rosenberg, A.M., Yeung, R.S.M., Morris, Q. and BBOP Study Consortium 2016. Gene expression deconvolution for uncovering molecular signatures in response to therapy in juvenile idiopathic arthritis. *Plos One* 11(5), p. e0156055.
- Davies, R., Southwood, T.R., Kearsley-Fleet, L., Lunt, M. and Hyrich, K.L. 2014. FRI0557 Standardised Mortality Rates Are Increased in Patients with Severe Juvenile Idiopathic Arthritis. *Annals of the Rheumatic Diseases* 73(Suppl 2), pp. 588.1-588.

- Davis, C.A., Hitz, B.C., Sloan, C.A., et al. 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research* 46(D1), pp. D794–D801.
- Denny, M.F., Yalavarthi, S., Zhao, W., et al. 2010. A distinct subset of proinflammatory neutrophils isolated from patients with systemic lupus erythematosus induces vascular damage and synthesizes type I IFNs. *Journal of Immunology* 184(6), pp. 3284–3297.
- Fall, N., Barnes, M., Thornton, S., 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), pp. 3793–3804.
- Foster, H.E., Marshall, N., Myers, A., Dunkley, P. and Griffiths, I.D. 2003. Outcome in adults with juvenile idiopathic arthritis: a quality of life study. *Arthritis and Rheumatism* 48(3), pp. 767–775.
- Gattorno, M., Piccini, A., Lasigliè, D., et al. 2008. The pattern of response to anti-interleukin-1 treatment distinguishes two subsets of patients with systemic-onset juvenile idiopathic arthritis. *Arthritis and Rheumatism* 58(5), pp. 1505–1515.
- Gaur, P., Myles, A., Misra, R. and Aggarwal, A. 2017. Intermediate monocytes are increased in enthesitis-related arthritis, a category of juvenile idiopathic arthritis. *Clinical and Experimental Immunology* 187(2), pp. 234–241.
- Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. 2004. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20(3), pp. 307–315.
- George, N.I. and Chen, J.J. 2009. Batch Effect Estimation of Microarray Platforms with Analysis of Variance. In: Scherer, A. ed. *Batch effects and noise in microarray experiments*. Chichester, UK: John Wiley & Sons, Ltd, pp. 75–85.
- Gorelik, M., Fall, N., Altaye, M., et al. 2013. Follistatin-like protein 1 and the ferritin/erythrocyte sedimentation rate ratio are potential biomarkers for dysregulated gene expression and macrophage activation syndrome in systemic juvenile idiopathic arthritis. *The Journal of Rheumatology* 40(7), pp. 1191–1199.
- Grassi, L., Pourfarzad, F., Ullrich, S., et al. 2018. Dynamics of transcription regulation in human bone marrow myeloid differentiation to mature blood neutrophils. *Cell reports* 24(10), pp. 2784–2794.
- Griffin, T.A., Barnes, M.G., Ilowite, N.T., et al. 2009. Gene expression signatures in polyarticular juvenile idiopathic arthritis demonstrate disease heterogeneity and offer a molecular classification of disease subsets. *Arthritis and Rheumatism* 60(7), pp. 2113–2123.
- Hahn, Y.-S. and Kim, J.-G. 2010. Pathogenesis and clinical manifestations of juvenile rheumatoid arthritis. *Korean journal of pediatrics* 53(11), pp. 921–930.
- Harrison, P.F., Pattison, A.D., Powell, D.R. and Beilharz, T.H. 2019. Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biology* 20(1), p. 67.

- Hinze, C.H., Fall, N., Thornton, S., et al. 2010. Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis. *Arthritis Research & Therapy* 12(3), p. R123.
- Hong, S., Banchereau, R., Maslow, B.-S.L., et al. 2019. Longitudinal profiling of human blood transcriptome in healthy and lupus pregnancy. *The Journal of Experimental Medicine* 216(5), pp. 1154–1169.
- Hunter, P.J., Nistala, K., Jina, N., et al. 2010. Biologic predictors of extension of oligoarticular juvenile idiopathic arthritis as determined from synovial fluid cellular composition and gene expression. *Arthritis and Rheumatism* 62(3), pp. 896–907.
- Hu, Z., Jiang, K., Frank, M.B., Chen, Y. and Jarvis, J.N. 2016. Complexity and specificity of the neutrophil transcriptomes in juvenile idiopathic arthritis. *Scientific reports* 6, p. 27453.
- Ishikawa, S., Mima, T., Aoki, C., et al. 2009. Abnormal expression of the genes involved in cytokine networks and mitochondrial function in systemic juvenile idiopathic arthritis identified by DNA microarray analysis. *Annals of the Rheumatic Diseases* 68(2), pp. 264–272.
- Jarvis, J.N., Jiang, K., Frank, M.B., et al. 2009. Gene expression profiling in neutrophils from children with polyarticular juvenile idiopathic arthritis. *Arthritis and Rheumatism* 60(5), pp. 1488–1495.
- Jarvis, J.N., Petty, H.R., Tang, Y., et al. 2006. Evidence for chronic, peripheral activation of neutrophils in polyarticular juvenile rheumatoid arthritis. *Arthritis Research & Therapy* 8(5), p. R154.
- Jiang, K., Frank, M., Chen, Y., Osban, J. and Jarvis, J.N. 2013. Genomic characterization of remission in juvenile idiopathic arthritis. *Arthritis Research & Therapy* 15(4), p. R100.
- Jiang, K., Sawle, A.D., Frank, M.B., Chen, Y., Wallace, C.A. and Jarvis, J.N. 2014. Whole blood gene expression profiling predicts therapeutic response at six months in patients with polyarticular juvenile idiopathic arthritis. *Arthritis & rheumatology (Hoboken, N.J.)* 66(5), pp. 1363–1371.
- Jiang, K., Sun, X., Chen, Y., Shen, Y. and Jarvis, J.N. 2015. RNA sequencing from human neutrophils reveals distinct transcriptional differences associated with chronic inflammatory states. *BMC Medical Genomics* 8, p. 55.
- Jiang, K., Wong, L., Sawle, A.D., et al. 2016. Whole blood expression profiling from the TREAT trial: insights for the pathogenesis of polyarticular juvenile idiopathic arthritis. *Arthritis Research & Therapy* 18(1), p. 157.
- Johnson, W.E., Li, C. and Rabinovic, A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1), pp. 118–127.
- Kent, J.W., Göring, H.H.H., Charlesworth, J.C., et al. 2012. Genotype \times age interaction in human transcriptional ageing. *Mechanisms of Ageing and Development* 133(9–10), pp. 581–590.

- Klatchoian, D.A., Len, C.A., Terreri, M.T.R.A., et al. 2008. Quality of life of children and adolescents from São Paulo: reliability and validity of the Brazilian version of the Pediatric Quality of Life Inventory version 4.0 Generic Core Scales. *Jornal de Pediatria* 84(4), pp. 308–315.
- Knowlton, N., Jiang, K., Frank, M.B., et al. 2009. The meaning of clinical remission in polyarticular juvenile idiopathic arthritis: gene expression profiling in peripheral blood mononuclear cells identifies distinct disease states. *Arthritis and Rheumatism* 60(3), pp. 892–900.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* 44(W1), pp. W90-7.
- Lai, S., Huang, W., Xu, Y., et al. 2018. Comparative transcriptomic analysis of hematopoietic system between human and mouse by Microwell-seq. *Cell discovery* 4, p. 34.
- Lamot, L., Borovecki, F., Tambic Bukovac, L., et al. 2014. Aberrant expression of shared master-key genes contributes to the immunopathogenesis in patients with juvenile spondyloarthritis. *Plos One* 9(12), p. e115416.
- Macaubas, C., Nguyen, K.D., Peck, A., et al. 2012. Alternative activation in systemic juvenile idiopathic arthritis monocytes. *Clinical Immunology* 142(3), pp. 362–372.
- Manners, P.J. and Bower, C. 2002. Worldwide prevalence of juvenile arthritis why does it vary so much? *The Journal of Rheumatology* 29(7), pp. 1520–1530.
- Ma, T., Huo, Z., Kuo, A., et al. 2019. MetaOmics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. *Bioinformatics* 35(9), pp. 1597–1599.
- Menche, J., Sharma, A., Kitsak, M., et al. 2015. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347(6224), p. 1257601.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and PRISMA Group 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 6(7), p. e1000097.
- Moncrieffe, H., Bennett, M.F., Tsoras, M., et al. 2017. Transcriptional profiles of JIA patient blood with subsequent poor response to methotrexate. *Rheumatology (Oxford, England)* 56(9), pp. 1542–1551.
- Moncrieffe, H., Hinks, A., Ursu, S., et al. 2010. Generation of novel pharmacogenomic candidates in response to methotrexate in juvenile idiopathic arthritis: correlation between gene expression and genotype. *Pharmacogenetics and Genomics* 20(11), pp. 665–676.
- Ogilvie, E.M., Khan, A., Hubank, M., Kellam, P. and Woo, P. 2007. Specific gene expression profiles in systemic juvenile idiopathic arthritis. *Arthritis and Rheumatism* 56(6), pp. 1954–1965.
- Oliveira, S.M. de, Gomides, A.P.M., Mota, L.M.H. da, Lima, C.M.B.L. and Rocha, F.A.C. 2016. Intestinal parasites infection: protective effect in rheumatoid arthritis? *Revista Brasileira de Reumatologia* (57)5, pp 461-465.

- Ombrello, M.J., Arthur, V.L., Remmers, E.F., et al. 2017. Genetic architecture distinguishes systemic juvenile idiopathic arthritis from other forms of juvenile idiopathic arthritis: clinical and therapeutic implications. *Annals of the Rheumatic Diseases* 76(5), pp. 906–913.
- Omoyinmi, E., Hamaoui, R., Bryant, A., et al. 2016. Mitochondrial and oxidative stress genes are differentially expressed in neutrophils of sJIA patients treated with tocilizumab: a pilot microarray study. *Pediatric rheumatology online journal* 14(1), p. 7.
- Papatheodorou, I., Fonseca, N.A., Keays, M., et al. 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Research* 46(D1), pp. D246–D251.
- Petty, R.E., Southwood, T.R., Manners, P., et al. 2004. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *The Journal of Rheumatology* 31(2), pp. 390–392.
- Prakken, B., Albani, S. and Martini, A. 2011. Juvenile idiopathic arthritis. *The Lancet* 377(9783), pp. 2138–2149.
- Put, K., Vandenhaute, J., Avau, A., et al. 2017. Inflammatory Gene Expression Profile and Defective Interferon-gamma and Granzyme K in Natural Killer Cells From Systemic Juvenile Idiopathic Arthritis Patients. *Arthritis & rheumatology (Hoboken, N.J.)* 69(1), pp. 213–224.
- Quartier, P., Allantaz, F., Cimaz, R., et al. 2011. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). *Annals of the Rheumatic Diseases* 70(5), pp. 747–754.
- Ramanathan, K., Glaser, A., Lythgoe, H., et al. 2018. Neutrophil activation signature in juvenile idiopathic arthritis indicates the presence of low-density granulocytes. *Rheumatology (Oxford, England)* 57(3), pp. 488–498.
- Ritchie, M.E., Phipson, B., Wu, D., et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7), p. e47.
- Rosenkranz, M.E., Wilson, D.C., Marinov, A.D., et al. 2010. Synovial fluid proteins differentiate between the subtypes of juvenile idiopathic arthritis. *Arthritis and Rheumatism* 62(6), pp. 1813–1823.
- Rothmund, F., Gerss, J., Ruperto, N., et al. 2014. Validation of relapse risk biomarkers for routine use in patients with juvenile idiopathic arthritis. *Arthritis Care & Research* 66(6), pp. 949–955.
- Russo, P.S.T., Ferreira, G.R., Cardozo, L.E., et al. 2018. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* 19(1), p. 56.
- R Core Team 2015. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.

- Sagiv, J.Y., Michaeli, J., Assi, S., et al. 2015. Phenotypic diversity and plasticity in circulating neutrophil subpopulations in cancer. *Cell reports* 10(4), pp. 562–573.
- Sergushichev, A. 2016. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*.
- Shenoi, S., Ou, J.-N., Ni, C., et al. 2015. Comparison of biomarkers for systemic juvenile idiopathic arthritis. *Pediatric Research* 78(5), pp. 554–559.
- Shoop-Worrall, S.J.W., Kearsley-Fleet, L., Thomson, W., Verstappen, S.M.M. and Hyrich, K.L. 2017. How common is remission in juvenile idiopathic arthritis: A systematic review. *Seminars in arthritis and rheumatism* 47(3), pp. 331–337.
- de Sousa Studart, S.A., Leite, A.C.R.M., Marinho, A.L.L.F., et al. 2015. Vitamin D levels in juvenile idiopathic arthritis from an equatorial region. *Rheumatology International* 35(10), pp. 1717–1723.
- Stevens, A., Meyer, S., Hanson, D., Clayton, P. and Donn, R.P. 2014. Network analysis identifies protein clusters of functional importance in juvenile idiopathic arthritis. *Arthritis Research & Therapy* 16(3), p. R109.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S. and Ebert, B.L. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide.
- Swart, J.F., de Roock, S. and Prakken, B.J. 2016. Understanding inflammation in juvenile idiopathic arthritis: How immune biomarkers guide clinical strategies in the systemic onset subtype. *European Journal of Immunology* 46(9), pp. 2068–2077.
- Tasaki, S., Suzuki, K., Kassai, Y., et al. 2018. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nature Communications* 9(1), p. 2755.
- Teresa Terreri, M., Campos, L.M.A., Okuda, E.M., et al. 2013. Perfil de especialistas e de serviços em reumatologia pediátrica no estado de São Paulo. *Revista brasileira de reumatologia* 53(4), pp. 346–351.
- Ter Haar, N.M., Tak, T., Mokry, M., et al. 2018. Reversal of Sepsis-Like Features of Neutrophils by Interleukin-1 Blockade in Patients With Systemic-Onset Juvenile Idiopathic Arthritis. *Arthritis & rheumatology (Hoboken, N.J.)* 70(6), pp. 943–956.
- Toro-Domínguez, D., Martorell-Marugán, J., López-Domínguez, R., et al. 2019. ImaGEO: integrative gene expression meta-analysis from GEO database. *Bioinformatics* 35(5), pp. 880–882.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., et al. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* 32(4), pp. 381–386.
- Tseng, G.C., Ghosh, D. and Feingold, E. 2012. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Research* 40(9), pp. 3785–3799.

- Tuller, T., Atar, S., Ruppin, E., Gurevich, M. and Achiron, A. 2013. Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes and Immunity* 14(2), pp. 67–82.
- Tu, Z.-Q., Xue, H.-Y., Chen, W., Cao, L.-F. and Zhang, W.-Q. 2017. Identification of potential peripheral blood diagnostic biomarkers for patients with juvenile idiopathic arthritis by bioinformatics analysis. *Rheumatology International* 37(3), pp. 423–434.
- Veglia, F., Tyurin, V.A., Blasi, M., et al. 2019. Fatty acid transport protein 2 reprograms neutrophils in cancer. *Nature* 569(7754), pp. 73–78.
- Vidal, M., Cusick, M.E. and Barabási, A.-L. 2011. Interactome networks and human disease. *Cell* 144(6), pp. 986–998.
- Viechtbauer, W. 2010. Conducting Meta-Analyses in R with the metafor Package. *Journal of statistical software* 36(3).
- Villanueva, E., Yalavarthi, S., Berthier, C.C., et al. 2011. Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. *Journal of Immunology* 187(1), pp. 538–552.
- Volders, P.-J., Helsens, K., Wang, X., et al. 2013. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Research* 41(Database issue), pp. D246–51.
- Winter, D., J. 2017. rentrez: An R package for the NCBI eUtils API. *The R journal* 9(2), p. 520.
- Xia, J., Fjell, C.D., Mayer, M.L., Pena, O.M., Wishart, D.S. and Hancock, R.E.W. 2013. INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Research* 41(Web Server issue), pp. W63–70.
- Zhao, Y., Li, H., Fang, S., et al. 2016. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Research* 44(D1), pp. D203–8.
- Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N. and Xia, J. 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research* 47(W1), pp. W234–W241.

Chapter 5

Appendix



César Augusto Prada Medina

Endereço para acessar este CV: <http://lattes.cnpq.br/9892775067158105>

Última atualização do currículo em 31/07/2019

Resumo informado pelo autor

Possui graduação em Biología - Universidad Industrial de Santander (2014). Atualmente é candidato a Doutor pela Faculdade de Ciências Farmacêuticas da Universidade de São Paulo.

(Texto informado pelo autor)

Nome civil

Nome César Augusto Prada Medina

Dados pessoais

Filiação

Nascimento 1991 - Bucaramanga/ - Colômbia

Carteira de
Identidade

CPF

Passaporte

Formação acadêmica/titulação

2016 Doutorado em Doutorado.
Universidade de São Paulo, USP, Sao Paulo, Brasil
com **período sanduíche** em European Bioinformatics Institute (Orientador: Irene Papatheodorou)
Título: Studying Juvenile Idiopathic Arthritis through a Network Medicine Approach

Orientador: Helder Takashi Imoto Nakaya
Bolsista do(a): Fundação de Amparo à Pesquisa do Estado de São Paulo

2009 - 2014 Graduação em Biología.
Universidad Industrial de Santander, UIS, Bucaramanga, Colômbia
Título: Radiosensibilidad e inducción de la respuesta SOS por tratamiento con radiación ultravioleta B en células de Escherichia coli deficientes para la reparación del daño genético, Año de obtención: 2014
Orientador: Jorge Luis Fuentes Lorenzo

Formação complementar

- 2016 - 2016** Biología Computacional Orientada al Diseño de Fármacos. (Carga horária: 100h).
Universidad de Buenos Aires, UBA, Argentina
- 2015 - 2015** Curso de curta duração em Integrated Analysis in Systems Biology. (Carga horária: 40h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2014 - 2015** Systems Biology Specialization. (Carga horária: 64h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2015 - 2015** Curso de curta duração em Network Analysis in Systems Biology. (Carga horária: 64h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2015 - 2015** Curso de curta duração em Dynamical Modelling in Systems Biology. (Carga horária: 56h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2014 - 2014** Curso de curta duração em II Curso Int. de intervención integral del DENGUE. (Carga horária: 34h).
Instituto Nacional de Salud, INS, Colômbia
- 2014 - 2014** Curso de curta duração em Introduction to Systems Biology. (Carga horária: 100h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2014 - 2014** Curso de curta duração em Experimental Methods in Systems Biology. (Carga horária: 64h).
Icahn School of Medicine at Mount Sinai (Coursera), ISMMS, Estados Unidos
- 2013 - 2013** Curso de curta duração em Reconstrucción y análisis de redes biológicas. (Carga horária: 20h).
Centro de Bioinformática y Biología Computacional de Colombia, BIOS, Colômbia

Atuação profissional

- Universidad Industrial de Santander - UIS

Vínculo institucional

2014 - 2015 Vínculo: Bolsista , Enquadramento funcional: Young researcher , Carga horária: 40, Regime: Dedicación exclusiva
Outras informações:
Grupo de Investigación en Enfermedades Infecciosas y Metabólicas:
<http://scienti1.colciencias.gov.co:8080/gruplac/jsp/visualiza/visualizagr.jsp?nro=0000000012382>

2013 - 2014 Vínculo: Bolsista , Enquadramento funcional: Microbiology teaching assistant , Carga horária: 8, Regime: Parcial

- Centro Nacional de Investigaciones para la Agroindustrialización de Especie - CENIVAM

Vínculo institucional

2015 - 2015 Vínculo: Celetista , Enquadramento funcional: Young researcher , Carga horária: 40, Regime: Dedicção exclusiva
Outras informações:
Grupo de Investigación en Microbiología y Genética:
<http://scienti.colciencias.gov.co:8080/grupiac/jsp/visualiza/visualizagr.jsp?nro=0000000007670>

3. Corpoeducación - CE

Vínculo institucional

2014 - 2014 Vínculo: Celetista , Enquadramento funcional: Open question assessor , Carga horária: 20, Regime: Parcial

4. Grupo Educativo Marín & Rosales - GE-M&R

Vínculo institucional

2011 - 2014 Vínculo: Celetista , Enquadramento funcional: Biology teacher , Carga horária: 8, Regime: Parcial
Outras informações:
Pre-university education.

5. European Bioinformatics Institute - EMBL

Vínculo institucional

2018 - 2019 Vínculo: Bolsista , Enquadramento funcional: Research secondee , Carga horária: 40, Regime: Integral

6. Universidade de São Paulo - USP

Vínculo institucional

2016 - Atual Vínculo: Bolsista , Enquadramento funcional: PhD candidate , Carga horária: 40, Regime: Dedicção exclusiva

Projetos

Projetos de pesquisa

2014 - 2015 Estudio de la actividad biológica de plantas medicinales: enfoque hacia el descubrimiento de productos con uso potencial para control del dengue y su vector *Aedes aegypti*

Descrição: The project was focused on searching plant extracts and plant compounds with antiviral (anti-DENV) activity. Also, the mechanisms of action of the identified bioactive substances were studied. For those purposes, mammalian cell cultures were infected with dengue virus and cultured with the mentioned substances.

Situação: Concluído Natureza: Projetos de pesquisa
Integrantes: César Augusto Prada Medina; Raquel Elvira Ocazonez Jimenez (Responsável); María Camila Flechas Alarcón; Sindi Alejandra Velandia Cruz

2012 - 2016 Estudio del potencial antígeno tóxico frente a la radiación ultravioleta de extractos SFE y aceites esenciales de especies vegetales de la biodiversidad colombiana.

Descrição: The potential protective activity of several plant extracts against UV-induced DNA-damage has been assessed through bacterial and mammalian in-vitro models.

Situação: Concluído Natureza: Projetos de pesquisa
Integrantes: César Augusto Prada Medina; Jorge Luis Fuentes Lorenzo (Responsável); Nathalia Quintero Ruiz; Elena Stashenko; Nathalia Rey Castellanos; Adriana García Forero

Áreas de atuação

1. Biologia de sistemas
2. Systems Immunology
3. Farmacologia Bioquímica e Molecular
4. Bioprospecção
5. Genética Molecular e de Microorganismos
6. Mutagenese

Idiomas

Inglês Compreende Bem , Fala Bem , Escreve Bem , Lê Bem
Espanhol Compreende Bem , Fala Bem , Escreve Bem , Lê Bem
Português Compreende Bem , Fala Bem , Escreve Bem , Lê Bem










Prêmios e títulos

- 2018** Capacity building for bioinformatics in Latin America (CABANA) research fellowship, European Molecular Biology Laboratory - European Bioinformatics Institute
- 2016** Bioinformatics training fellowship, Centro Latinoamericano de Formación Interdisciplinaria (CELFI)
- 2014** Laurate award for the Bachelor degree project., Universidad Industrial de Santander
- 2014** Summa Cum Laude degree, Universidad Industrial de Santander
- 2013** Best Colombian students of natural sciences, Instituto Colombiano para la Evaluación de la Educación - ICFES



Produção

Produção bibliográfica


Artigos completos publicados em periódicos

1.  SANTAMARÍA ACEVEDO, LILIANA; PRADA-MEDINA, CESAR AUGUSTO; RONDÓN GONZÁLEZ, FERNANDO; STASHENKO, ELENA E; MARTÍNEZ PÉREZ, FRANCISCO JOSÉ; LEVY, MORRIS; LEVY, MARÍA MERCEDES; FUENTES, JORGE LUIS
Interspecific variation and genetic relationship among Colombian . based on small ribosomal subunit gene sequence analysis. *Journal of Herbs, Spices & Medicinal Plants*. , v.24, p.99 - 108, 2017.
2.   PRADA-MEDINA, CESAR A.; FUKUTANI, KIYOSHI F.; PAVAN KUMAR, NATHHELLA; GIL-SANTANA, LEONARDO; BABU, SUBASH; LICHTENSTEIN, FLÁVIO; WEST, KIM; SIVAKUMAR, SHANMUGAM; MENON, PRADEEP A.; VISWANATHAN, VIJAY; ANDRADE, BRUNO B.; NAKAYA, HELDER I.; KORNFIELD, HARDY
Systems Immunology of Diabetes-Tuberculosis Comorbidity Reveals Signatures of Disease Complications. *Scientific Reports*.  v.7, p.1999 - , 2017.
3.  FUENTES, J. L.; GARCÍA FORERO, A.; QUINTERO RUIZ, N.; PRADA MEDINA, C. A.; REY CASTELLANOS, N.; FRANCO NIÑO, D. A.; CONTRERAS GARCÍA, D. A.; CÓRDOBA CAMPO, Y.; STASHENKO, E. E.
The SOS Chromotest applied for screening plant antigenotoxic agents against ultraviolet radiation. *PHOTOCHEMICAL & PHOTOBIOLOGICAL SCIENCES*.  v.16, p.1424 - 1434, 2017.
4.   PRADA MEDINA, CESAR AUGUSTO; ARISTIZABAL TESSMER, ELKE TATJANA; QUINTERO RUIZ, NATHALIA; SERMENT-GUERRERO, JORGE; FUENTES, JORGE LUIS
Survival and SOS response induction in ultraviolet B irradiated *Escherichia coli* cells with defective repair mechanisms. *International Journal of Radiation Biology*.  v.92, p.1 - 8, 2016.

Trabalhos publicados em anais de eventos (resumo)

1. PRADA-MEDINA, C. A.; NAKAYA, H. I
Inferring transcriptional regulators in JIA by coupling coexpression and phi-mixing coefficient networks In: XI International Conference on Systems Biology of Human Diseases, 2018, Los Angeles.
XI International Conference on Systems Biology of Human Diseases. , 2018.
2.  PRADA-MEDINA, C. A.; NAKAYA, H. I
Regulatory molecular circuits in leukocytes of Juvenile Idiopathic Arthritis patients In: The fourth International Society for Computational Biology Latin America Bioinformatics Conference (ISCB-LA), 2016, Buenos Aires.
The fourth International Society for Computational Biology Latin America Bioinformatics Conference (ISCB-LA). , 2016.
3. SANTAMARIA, L.; PRADA-MEDINA, C. A.; RONDON, F.; MARTINEZ, F. J.; TORES, L.; LEVY, M. M.; LEVY, M.; STASHENKO, E.; FUENTES, JORGE LUIS
USEFULNESS OF SMALL RIBOSOMAL SUBUNIT AND MICROSATELLITE DNA SEQUENCE ANALYSES IN TAXONOMY AND PHYLOGENY OF *Lippia SPECIES* In: V Congreso Iberoamericano de Productos Naturales, 2016, Bogotá DC.
V Congreso Iberoamericano de Productos Naturales. , 2016.
4.  PRADA-MEDINA, C. A.; Aristizabal-Tessmer, E. T.; Serment-Guerrero, J. H.; Fuentes-Lorenzo J. L.
RADIOSENSITIVITY AND SOS RESPONSE INDUCTION IN ULTRAVIOLET B IRRADIATED *Escherichia coli* CELLS WITH DEFECTIVE DNA-REPAIR MECHANISMS In: L Congreso Colombiano de Ciencias Biológicas, 2015, Bucaramanga.
Resúmenes Congreso Colombiano de Ciencias Biológicas. , 2015.
5. PRADA-MEDINA, C. A.; Acosta, A.; Daza-Espinosa, M. C.
ACETOGENINAS DE ANNONACEAE: RELACIÓN ESTRUCTURA - ACTIVIDAD ANTICANCERÍGENA In: VI Congreso Colombiano de Botánica, 2011, Cali.
Resúmenes VI Congreso Colombiano de Botánica. Cali: Asociación Colombiana de Botánica, 2011. p.7 - 620

Demais produções bibliográficas

1.  PRADA-MEDINA, C. A.
Systems pharmacology approach allows classification of cell cycle regulatory protein inhibitors according to their anticancer potential determined in silico. Capstone project. , 2015. (Outra produção bibliográfica)

Eventos

Eventos

Participação em eventos

1. Latin American Strategy Forum for Research Infrastructure, 2019. (Outra)
2. Apresentação de Poster / Painel no(a) XI International Conference on Systems Biology of Human Diseases, 2018. (Congresso)
Inferring transcriptional regulators in JIA by coupling coexpression and phi-mixing coefficient networks.
3. Jamboré Brasuca, 2017. (Encontro)
4. Second International Symposium on Inflammatory Diseases - INFLAMMA II, 2016. (Simpósio)
5. Apresentação de Poster / Painel no(a) The fourth International Society for Computational Biology Latin America Bioinformatics Conference (ISCB-LA), 2016. (Congresso)
Regulatory molecular circuits in leukocytes of Juvenile Idiopathic Arthritis patients.
6. Workshop Beyond Zika - A Tripartite Initiative - USP/ Institut Pasteur/ FioCruz, 2016. (Encontro)
7. III Congreso Colombiano de Bioinformática y Biología Computacional, 2015. (Congresso)
8. XXII Congreso Latinoamericano de Microbiología, 2014. (Congresso)
9. Apresentação de Poster / Painel no(a) VI Congreso Colombiano de Botánica, 2011. (Congresso)
Acetogeninas de Annonaceae: relación estructura-actividad anticancerígena.



UNIVERSIDADE DE SÃO PAULO

Faculdade de Ciências Farmacêuticas

HISTÓRICO ESCOLAR DE PÓS-GRADUAÇÃO

Nome: Cesar Augusto Prada Medina

Data de Nascimento: 24/03/1991

Cédula de Identidade: RNE: G224587-R - DF

Local de Nascimento: Colômbia

Nacionalidade: Colombiana

Graduação: Biólogo - Universidad Industrial de Santander - Colômbia - 2014

Título: Doutor em Ciências

Obtido no Programa: Farmácia (Fisiopatologia e Toxicologia)

Área: Fisiopatologia

Data da Matrícula: 05/02/2016

Orientador: Prof(a) Dr(a) Helder Takashi Imoto Nakaya

Proficiência em Língua(s): Inglês

Data de aprovação no exame de qualificação: 20/02/2018

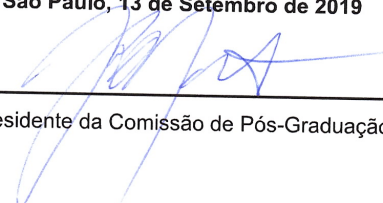
Estágio no exterior: European Molecular Biology Laboratory - European Bioinformatics Institute, Inglaterra - Período de 17/09/2018 até 15/03/2019

Título do Trabalho: "Estudando a Artrite Idiopática Juvenil mediante uma abordagem de medicina de redes"

Data da Defesa: 12/09/2019

Resultado da Defesa: Aprovado

São Paulo, 13 de Setembro de 2019



Presidente da Comissão de Pós-Graduação

13/09/19 11:37:05



UNIVERSIDADE DE SÃO PAULO

Faculdade de Ciências Farmacêuticas

HISTÓRICO ESCOLAR DE PÓS-GRADUAÇÃO

Nome: Cesar Augusto Prada Medina

Sigla	Nome da Disciplina	Início	Término	Créditos	Frequência	Conceito
IBI5011-3	Introdução à Computação para Bioinformática (Curso Interunidades: Bioinformática - Universidade de São Paulo)	22/02/2016	30/05/2016	8	100	A
FBC5792-3	Tópicos em Análises Clínicas III	08/03/2016	20/06/2016	1	87	A
BMI5862-9	Seminário Didático-Científico em Imunologia I (Instituto de Ciências Biomédicas - Universidade de São Paulo)	10/03/2016	30/06/2016	4	100	A
FBC5766-4	Tópicos em Análises Clínicas IV	02/08/2016	14/11/2016	1	100	A
MAE5755-7	Métodos Estatísticos Aplicados às Ciências Biológicas (Instituto de Matemática e Estatística - Universidade de São Paulo)	08/08/2016	25/11/2016	8	91	A
PCS5031-1	Introdução à Ciência dos Dados (Escola Politécnica - Universidade de São Paulo)	12/09/2016	04/12/2016	8	100	B

Créditos atribuídos à Tese:

167

São Paulo, 13 de Setembro de 2019



 Presidente da Comissão de Pós-Graduação

13/09/19 11:37:05

Conceito a partir de 02/01/1997: A - Excelente, com direito a crédito; B - Bom, com direito a crédito; C - Regular, com direito a crédito; R - Reprovado; T - Transferência;
Um (1) crédito equivale a quinze (15) horas de atividade programada.