Research Article

Modeling the dynamics of t helper 17 induction and differentiation

Abstract

Background: CD4+ T cells play important roles in orchestrating immune responses, maintaining homeostasis and offering a functional control of immune responses. Depending on the cytokine milieu, a naïve CD4+ T cell differentiates into different phenotypes. The T helper 17 (Th17) phenotype remains controversial since it has been identified as an inductor of inflammation, but it contributes to the regulatory arm of the immune response. With the increased availability of high-throughput datasets, data-driven computational modeling built upon a robust conceptual base becomes a powerful tool to build predictive, counterintuitive, dynamic gene expression networks.

Results: We used a publicly available microarray dataset generated by Yosef et al.,[1] to construct a dynamic model of Th17 differentiation. Following data treatment to tailor the data to our 2,000+ genes of interest, our data-driven modeling pipeline used the Ingenuity Pathway Analysis (IPA) platform to infer a proof-of-concept modeling network composed of 67 genes involved in induction, function, and maintenance of Th17 differentiation. *In vitro* time-course mRNA expression data was used to calibrate the network using COPASI. Computational simulations highlighted the potential role of IL-24 in shaping IL-17A producing responses, and local sensitivity analysis demonstrated that IL-24 is negatively correlated with FOXP3 expression, further regulating the balance between FOXP3 and RORγt. Moreover, NLRP3 was identified as another potential therapeutic target to treat Th17-driven pathologies. Local sensitivity analyses on NLRP3 highlighted the positive correlation of NLRP3 with IP-10, GM-CSF, IFNγ, and IL-17A.

Conclusions: We established a modeling pipeline for constructing and calibrating data-driven models of Th17 differentiation. Our results highlighted the role of IL-24, SGK1, and NLRP3 as key modulators of Th17 cell differentiation. This modeling pipeline is an example of a modeling environment for building computational models with deterministic and stochastic features to generate new mechanistic immunological knowledge and to identify novel therapeutic targets for human diseases.

Keywords: CD4+ T cell differentiation, Computational modeling, Immunoinformatics, Computational Immunology, Th17 cells

Adria Carbo,[1,2,3] Raquel Hontecillas Hontecillas,[1,2] Monica Viladomiu Viladomiu,[1,2] Andrew Leber Viladomiu,[1,2] Stefan Hoops,[2] Josep Bassaganya Riera[1,2]

[1]Utritional Immunology and Molecular Medicine Laboratory, Virginia Bioinformatics Institute, USA
[2]Center for Modeling Immunity to Enteric Pathogens, Virginia Bioinformatics Institute, USA
[3]Biotherapeutics Inc., Blacksburg, USA

Correspondence: Josep Bassaganya- Riera ,Virginia Tech, Virginia Bioinformatics Institute, Nutritional Immunology and Molecular Medicine Laboratory, 1015 Life Science Circle, Blacksburg, VA, 24061, USA, Email jbassaga@vt.edu

Received: February 20, 2015 | **Published:** March 30, 2015

Abbreviations: Th17, T Helper 17; IL, Interleukin; SGK1, serum and glucocorticoid inducible kinase 1; FOXP3, fork head box P3; SBML, systems-biology markup language; IBD, inflammatory bowel disease; ODE, ordinary differential equation; IPA, ingenuity pathways analysis

Introduction

Immunity represents a highly specialized, complex, hierarchical, and networked system that shields humans from foreign pathogens, injury and disease. The CD4+ T cell subsets are functionally and phenotypically heterogeneous, consisting of different differentiated populations involved in coordinating various aspects of adaptive immunity. Upon antigen recognition and co-stimulatory signaling, naïve CD4+ T cells will differentiate into a specific phenotype depending on the cytokine milieu composition. The current understanding of CD4+ T cell differentiation comprises at least 9 phenotypically different states: T helper type 1 and 2 (Th1 and Th2),[2] T helper type 9 (Th9),[3] T helper type 17 (Th17),[4] T helper type 22 (Th22),[5] Follicular T helper cells (Tfh),[6,7] induced regulatory T cells (Treg),[8,9] and type 1 regulatory cells (Tr1).[10] In this panicle of CD4+ T cell subsets, the Th17 population is a key player in chronic inflammatory diseases such as Inflammatory Bowel Disease (IBD),[11] type 2 diabetes (T2D),[12] rheumatoid arthritis,[13] or in the infectious disease setting such as during Helicobacter pylori[14] or Clostridium difficile infection.[15] Even though Th17 cells have been highly characterized in the context of inflammatory pathologies, they remain elusive and controversial because of their plasticity and many of their disease-causing mechanisms that are not fully understood. For example, intestinal IL-17A+ IL-10+ Th17 cells are known to induce an immunosuppressive environment,[16] however, in the onset of IBD, increased IL-17A expression has been reported[17] and Th17 cells expressing IFNγ or GM-CSF can accumulate during gut inflammatory disorders.[18-20]

Theoretical models have been used to decipher the mechanisms controlling differentiation and function of Th17 cells by studying the relationship of RORγt with FOXP3.[21,22] The relationship of FOXP3 and RORγt is especially important since these two transcription factors are master regulators of agonist cell differentiation programs.[23,24] On one side, RORγt promotes pro-inflammatory secretion of IL-21 and IL-17A. On the other side, FOXP3 promotes anti-inflammatory responses through secretion of IL-10. We published a computational and mathematical model of CD4+ T cell differentiation that predicted and validated the role of peroxisome proliferators activated receptor gamma (PPARγ) in modulating the plasticity between Th17 and Treg cells.[25] Due to the increasing availability of high throughput data, combining theoretical and data-driven approached becomes more feasible.[26,27]

While theory-driven modeling in the context of CD4+ T cell differentiation has been linked to reductionist approaches, combining data-driven and theoretical approaches has emerged as a new and improved strategy for multivariate analysis and systems-level hypothesis generation. In this context, network inference is a key step in integrating information from various datasets in a manner that combines data and theory. An example on how to use high-throughput data to construct a CD4+ T cell comprehensive network is the study published by Yosef et al.[1] where they used transcriptional profiling with microarrays at high temporal resolution to build a Th17 induction system. In this study, 1,291 genes were differentially identified and clustered into 20 groups, depending on their temporal profiles. Another advantage highlighted in this study is the use of modules to explain the processes controlling Th17 differentiation. Four regulatory modules were identified: the positive module that increased IL-17 levels, the negative module that down-regulated IL-17, the signature of Th17 genes and signature of other CD4+ T cell subtypes. This work supported the finding of 3 novel key regulators of Th17 function: Mina, Fas, and Pou2af1.

Our current paper describes the creation of a computational ordinary differential equation (ODE)-based model of Th17 induction and differentiation based on Yosef's data[1] that incorporates both dynamics and sensitivities across19 time points. To build the dynamic model we analyzed the microarray data and inferred a computational modeling network using the Ingenuity Pathways (IPA) platform, we translated the inferred network into a Systems-Biology Markup Language (SBML)-compliant system, imported the network into COPASI, and calibrated the system by using the original microarray datasets used to build the dynamic network model. Our in silico predictions about the function of NLRP3 and IL-24 during Th17 differentiation, revealed a modulatory role over Th17 function. Furthermore, our local sensitivity analysis highlighted the correlations between FOXP3, CEBPB, FOSL1, IRF4, and PPARγ with IL-24, as well as the correlations of IP-10, GM-CSF, IFNγ, and IL-17A with NLRP3.

Materials and methods

Algorithms for data treatment

In order to adjust and tailor the input data into Ingenuity Pathways Analysis (IPA) platform we created a python3-based script that trimmed and averaged probe sets. To run, the script needs a file with a list of genes that the user wants to trim. The script runs as follows: first, the script opens all files and deletes any white spaces at the end of the line. The script then reads and stores the name of the first gene in the list and scans the input microarray dataset. Once found, it transfers the information in another file. If different entries are found, the script averages them and only annotates the averaged value. The algorithm also provides the standard deviation to decrease error insertion. We excluded probe sets that highly deviated from the others. The script completes once all the genes in the list have been assessed and it creates an output file that will be used as input for IPA.

IPA Analysis

Ingenuity Pathways Analysis (IPA) (Ingenuity Systems, Redwood City, CA, USA) was used for the identification of key nodes in Th17 cells for network inference purposes. The microarray data was uploaded as log2 changes and a Core Analysis was run to map expression data to IPA's Knowledge database. Upstream Activation analysis was performed and selected genes were moved to Pathway Analysis to create the topology interaction network. To specifically characterize the interactions, the IPA's Knowledge database of each interaction was consulted to discern between activation and inhibition reaction**s.**

Model parameter estimation

The parameter estimation task was run in COPASI using the time-course experimental microarray data provided in.[1] Data was uploaded into COPASI using IL-6 and TGFβ as independent initial concentrations. The rest of the nodes were set up as dependent nodes on transient concentration for mapping purposes. The Particle Algorithm with 2000 iterations was used first. Secondly, using Particle Swarm algorithm results as initial values, the Genetic Algorithm was run. Within each parameter estimation task, objective value, root mean square, and standard deviation were calculated. Quality control was checking on the results of parameter estimation by contra-posing the fitted curves to the experimental data. Sensitivity analyses on the parameters were performed to observe how robust and how affected to change these parameters are.

Sensitivity Analysis

Sensitivity analysis was run with COPASI on our computational model using a delta factor of 0.0001 and a delta minimum of 1e-12. The subtask run for the analysis was a time-series with t=100h and correlation of all the variables of the model against activated SGK1, NLRP3, and IL-24 were assessed, showing positive correlation with key transcription factors that determine phenotype differentiation on Th17.

Creation of an IL-24 and NLRP3 knock-out systems

In order to assess the modulation of NLRP3 and IL-24 over Th17 we created an IL-24 and a NLRP3 knock-out models. In order to do so, we selectively chose the mass action-based reactions that up regulate both nodes and set up their parameters to zero. Therefore, none of these two nodes were able to up regulate its product concentration. Quality control to ensure complete ablation of either NLRP3 concentration in the NLRP3 null model, or IL-24 concentration in the IL-24 null model was performed by executing time-course and scan analyses.

Results and discussion

Modeling approach and objectives

Integrated pipeline for data-driven modeling

To streamline the process of model construction from high-dimensional microarray data, we built a semi-automated analysis pipeline that provides a methodological approach for dynamic model construction. The increase in publicly available microarray and RNA-seq data opens a new avenue for extracting new knowledge by using modeling and simulation approaches. Although the availability of similar proteomics datasets is not there yet, RNA-sequencing datasets can be used to build frameworks to predict novel behaviors. Our pipeline efficiently builds a computational model based on experimental data (Figure 1) generated in-house or downloaded from publicly available data repositories. The strategy proposed here allows us to obtain a list of genes reads which are subjected to trimming algorithms (see next section "Microarray Data Analyses" and methods section) and used as input in the Ingenuity Pathways Analysis (IPA) platform for creating an initial network model by using IPA's network inference algorithms. The resulting static network model was implemented into Cell Designer by through a Cell Designer plug in and it became an SBML-compliant network model. Next, we imported the SBML-compliant network into the Complex Pathway Simulator (COPASI).

COPASI allowed us to perform parameter estimation and local/global sensitivity analysis (see "Modeling implementation") and obtain a fully calibrated dynamic model of Th17 differentiation. Using time-courses and sensitivity analyses, we were able to run in silico experimentation with the implemented Th17 differentiation model and come up with a series of predictions in regards to Interleukin-24 (IL-24), and the Noll-Like Receptor Protein 3 (NLRP3).
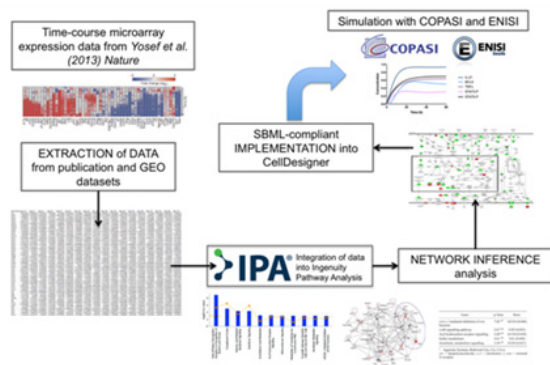
Figure 1 Integrated data-driven modeling pipeline to generate SBML-compliant, comprehensive, and predictive computational models of Th17 differentiation. Briefly, publicly available data from GEO was downloaded and treated to represent a tailor input for Ingenuity Pathways (IPA). IPA was used to infer a network that was imported into an SBML-compliant environment. Using the microarray experimental data in COPASI, the computational model was created and used to generate novel predictions.

Microarray data analyses

The data used for this study was obtained from the Gene Expression Omnibus (GEO) NIH repository with the accession number GSE43955. This dataset contains a microarray expression-profiling array for Th17 differentiation, including a Th0 control. Briefly, Yosef et al.[1] isolated naïve CD4+ T cells from wild-type mice and treated them with IL-6 and TGFβ1. Samples for mRNA analysis were collected at 19 different time points (0, 0.5, 1, 2, 4, 6, 8, 10, 12, 16, 20, 24, 30, 42, 48, 50, 52, 60, and 72 hours). The Th0 control consisted of time- and culture-matched wild-type naïve T cells stimulated under Th0 conditions. Furthermore, during microarray or sequencing analysis strategies, the expression results are often annotated with more than one entry per gene. For this reason, the GEO dataset extracted from.[1] Contained gene reads that had entries that were annotated with an unspecified gene name (eg. 1424888_at), and also had different entries for one same gene. At the same time, due to the nature of the experimental design, the data obtained had two different profiles: either non-differentiated or Th17-induced CD4+ T cells. For this variety of reasons, we created a list of genes that contained 2,013 genes related to CD4+ T cell differentiation, function, progression, and maintenance, metabolism, inflammatory genetic profiles, and a variety of arrays of transcription factors (Supplementary Information Table 1). We then used the python3-based script trimming algorithm to trim the microarray gene list to only these 2,000+ genes and average probe sets (see Methods for more detail). Furthermore, the algorithm averaged the value of these genes if it found more than one transcript for each one. Next, we calculated the fold change of each of the genes by dividing the Th17 sample by its corresponding Th0 control at each time point. Finally, in order to limit the value of expression and contain a more meaningful list, we calculated the log2 value of the fold-change of expression as it has been described in other publications.[28]
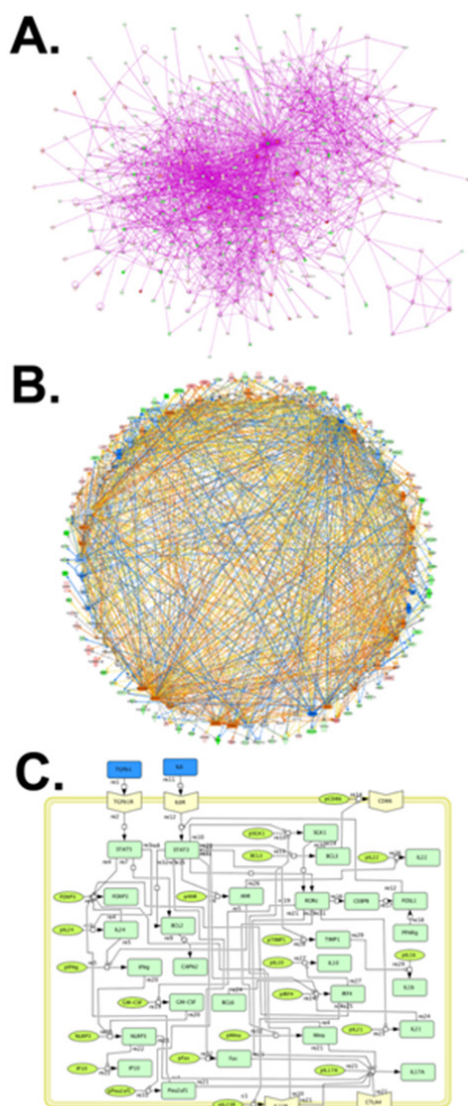
Network Inference and Analysis

In order to analyze the microarray data and infer functionally relevant networks based on data values, we used the Ingenuity Pathway Analysis (IPA) platform. IPA is a system that transforms a list of genes into a set of relevant networks based on extensive records maintained in the IPA Knowledge base. Indeed, reliable network inference methods for gene expression data remain an unsolved problem. Moreover, most inference algorithms fail to accurately predict combinatorial gene regulation (i.e. several molecules regulating one gene).[29] IPA provides an advantage for accurate network predictions since the program contains a broad knowledge base curated from literature. Therefore IPA was used to identify the molecular pathways and functional groupings for significantly differentially expressed mRNAs from the Th17 differentiation data that was uploaded. Briefly, expression data was uploaded into IPA and overlaid onto a global molecular network developed from information contained in the application along with predictions added manually based on in house data. Canonical and predicted gene networks were generated by IPA based on their connectivity, each ranked by a score. This score is based on the hyper geometric distribution, calculated with the right-tailed Fisher's Exact Test, and corresponds to the negative log of this p-value.

Functional analysis in IPA helps identify the published biological functions and canonical pathways that are most significantly associated with the dataset. Genes or gene products are represented as nodes, where shape indicates functional groups, and the biological relationship between two nodes is represented as an edge (line). All lines are supported by at least one reference in literature, textbook, or from canonical information stored in the Ingenuity Pathways knowledge database. Other studies have already used similar strategies[30,31] Furthermore, IPA uses powerful causal analytics that help to build a regulatory picture and a better understanding of the biology underlying a given gene expression dataset. More specifically, IPA's algorithms could retrieve not only the experimental validated interactions, but also predicted ones, as well as networks that were optimized for triangular connections between genes. With this approach, IPA favors denser networks over more sparsely connected ones. Of note, IPA assumes an unbiased, genome-wide, input. Our approach consisted of trimming the dataset so it was CD4+ T cell specific.

For this reason IPA was solely used for network analysis and generation and not for gene enrichment. Our treated and modified microarray data, based solely of the gene list and the log2 values, was uploaded into IPA as a list of target genes with expression values. Next, we computed a Core Analysis so IPA could retrieve all regulatory interactions without restricting the search. Moreover, by running a Core Analysis, we were able to map the microarray data to IPA's Knowledge Base (IPAKB), create molecular networks (algorithmically generated pathways), divide the data into different diseases and biological functions, and determine different canonical pathways. Including both experimental and predicted connections, IPA provided a list of upstream activators, where the most up- and down-regulated genes could be observed and the information of such genes could be accessed. Based on the all the genes uploaded into IPA, we created a first inferred network (Figure 2A). The expression values of this first network were calculated based on single observations extracted from the microarray dataset. In some cases, time-points were repeated and/or different replicates were given. In those cases, we averaged the values so the final expression value was representative of three replicates, when available. The microarray data was used to generate a comprehensive gene regulatory interaction network in IPA.

To ensure experimental data validity, we overlaid the data to canonical pathways already set by IPA. This approach allowed us to add a layer of validity. Given the amount of interactions generated, we decided to select 67 genes based on the levels of expression over time that showed the core of Th17 differentiation (Figure 2B). More specifically, we evaluated the most and least differentially expressed genes and included them in the list of 67 nodes to generate the model for this use-case study. Future work will include a significant higher amount of nodes that could potentially modulate key pathways of the network as well. This inferred Th17 interaction network represents the key and indispensable genes for a naïve CD4+ T cell to differentiate into Th17 when the cytokine milieu is rich in IL-6 and TGFβ1, based on published data.[24,32-42] Following modeling reduction to accommodate more efficient simulation capabilities, we selected 35 genes and their interactions between them and imported them into Cell Designer (Figure 2C).

Figure 2 Network inference and analysis prior to importation into an SBML-compliant environment.

A. Out of all the nodes in IPA, 67 genes (up-regulated or down-regulated) were selected based on differences in expression over the Th0 compartment.

B. SBML-compliant Cell designer network with a subset of genes in (B) to import into COPASI.

This selection was performed based on inclusion of the more critical nodes to mount a Th17 differentiation cascade following IL-6 and TGF β1 induction. To reduce the network by 35 genes, we located IL-6 and TGFβ1 in the IPA network and we built downstream pathways based on the centrality of RORγt and STAT3. Pathways resulting from these interactions were depicted in our SBML-based network in Cell Designer. This SBML-compliant model represents the activating and inhibitory reactions that take place in the cytoplasm and nucleus space and that will ultimately differentiate the cell into a Th17 cell.

Mathematical model implementation and calibration

Model implementation

The transition from a static diagram into a dynamic model of Th17 differentiation helps to not only understand the connections of the network but also observe how these connections changed, in a dynamic manner, over time. Furthermore, the dynamic model allowed us to run simulations and detect novel behaviors that could not be observed by just looking at the static picture. Thus, adding dynamics to the process increased the predictability of the network. Indeed, other comprehensive models of CD4+ T cell differentiation have been constructed.[24]

However, this novel strategy allows using data-driven approaches not only to construct its network, but also to dynamize the static picture that the data offers. Using this rationale, we next sought to import the inferred network, built by IPA and translated in Cell Designer, into the Complex Pathway Simulator (COPASI).[43] After the importation, the name of the species and reactions were annotated, the model, which contains 32 equations and 58 parameters, was adjusted into COPASI by creating Global Quantities and events, and the rate laws were adjusted to create the ordinary differential equations automatically (Figure S1).

Model calibration using microarray experimental data

The data used in this study performed by Yosef et al. in[27] consists of a time-course in vitro study where gene expression analysis was performed at 19 different time points (0, 0.5, 1, 2, 4, 6, 8, 10, 12, 16, 20, 24, 30, 42, 48, 50, 52, 60, and 72 hours). CD4+ T cell differentiation is an extremely dynamic process that can vary in a time-window of 1h ranges. Often times, experimental strategies in CD4+ T cell differentiation assess the levels of cytokines and transcription factors once the cell is fully differentiated around 60h post induction. With this approach, we are missing the specific cell dynamics of the process, as well as the subtle details that may open new strategies for therapeutic development. In this case, the 19 time points offer a clear picture of the exact levels of expression of each gene at each time point. Therefore, this dataset represents a perfect candidate to be used for calibration purposes.

To accomplish this goal, we extracted the experimental data from the raw database for the 35 nodes (except for Mina, Fas, and Pou2af1, which their data was used for validation -see next section-) that our computational model is composed of. Using the parameter estimation function in COPASI, we uploaded these 32 curves and used the Genetic Algorithm followed by a Particle Swarm Optimization (PSO),[44] both embedded in COPASI, to adjust the model parameters. PSO is a global search algorithm and thus depends only minimally on the initial guess of each parameter and therefore avoids the subjective estimation caused by initial guesses in local methods as Leven berg-Marquardt. PSO has been used in other publications for the same purpose.[45] In contrast, the genetic algorithm is an evolutionary

algorithm that mimics the process of natural selection by mutating or altering different parts of the system.

The database used for calibration was a time-point based database with values of the selected nodes per time point. COPASI can run parameter estimation algorithms based on time-course data. Therefore, the result of such estimation is a combination of all time-points executed automatically by COPASI. Furthermore, we found that some of the parameters estimated are comparable to values available in the literature in computational models that have been fully validated with in vivo experiments.[24] Th17 cells produce IL-17. TGFβ1, together with pro-inflammatory cytokines IL-6 or IL-21 and IL-23, orchestrate the differentiation of CD4+ T cells into the Th17 phenotype in a concentration-dependent manner.[32,33] It has been demonstrated that TGFβ1 synergizes with IL-6[34] or IL-21[35,36] to promote the expression of IL-17. This is achieved through stimulation of retinoid-related orphan receptor (ROR)γt by IL-6 through the transcription of STAT-3,[37,38] which in turn induces expression of IL-17.[39,40]

The combination and iterations of the genetic and PSO algorithms resulted in the fitting of these key molecules expression data with the computational model. First, the phosphorylated form of STAT3 controls and plays a central role during Th17 differentiation and function.[46] Our computational model was able to fit the experimental data correctly (Figure 3A), which was characterized by an early rise and a constant degradation over time. Next, the related orphan nuclear receptor C gene (RORc) transcribes the related orphan nuclear receptor gamma t (RORγt), which is central during Th17 differentiation. As our results show, our model was able to fit the bell-shaped expression profile that RORc possesses during differentiation (Figure 3B). As a result of Th17 differentiation, IL-21 and IL-17 are produced and secreted by the cell. Our computational model fit IL-17 (Figure 3C) and IL-21 (Figure 3D).
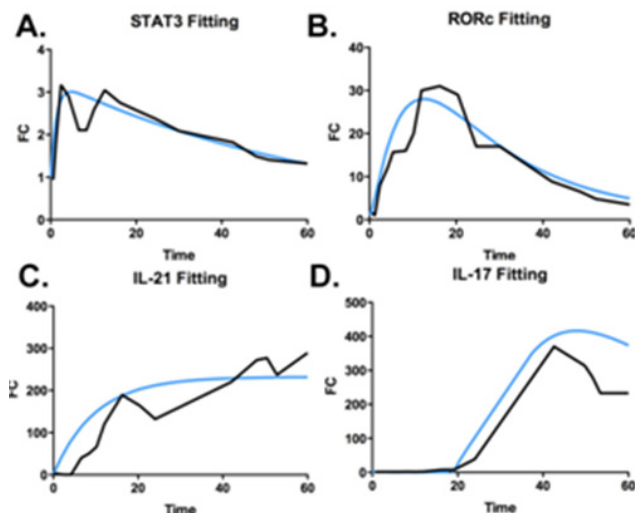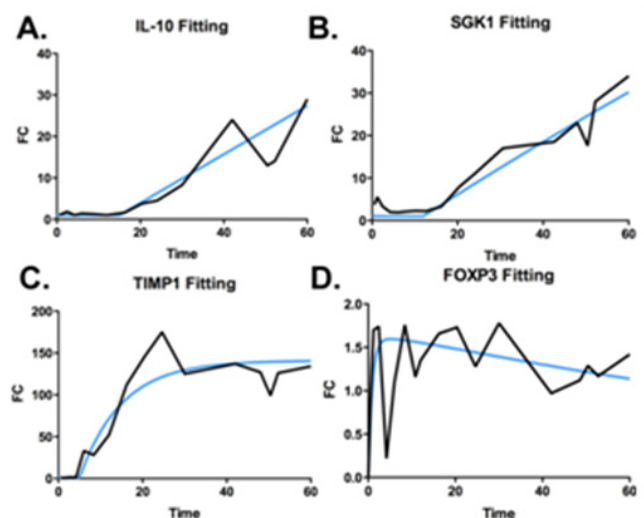

Figure 3 Computational fitting of Th17-related molecules. COPASI was used to run parameter estimation based on the microarray experimental data. Using a Genetic Algorithm and PSO, the model fitted (in blue) the experimental data (in black) of

A. Phosphorylated STAT3

B. RORc

C. Interleukin 21

D. Interleukin 17.

Other fitted molecules of the Th17 model are also depicted in (Figure 4). IL-10 has usually been considered an anti-inflammatory

molecule. Th17 cells produce high levels of IL-10 under certain conditions. In this case, our computational model was able to reproduce the dynamics of this cytokine over time (Figure 4A). As observed, the parameter estimation results fit the experimental data. The set of final parameters allowed us to simulate the Th17 differentiation model. However, robustness and sensitivity analyses studies would be useful to confirm the global and local effects to parameter change. TIMP1 is a transcription factors that, together with the kinase SGK1, were selected due to their high expression during Th17 differentiation. Although their role during Th17 induction has not fully been elucidated, they remain key players for IL-17 production. Our computational model was able to fit SGK1 (Figure 4B) and TIMP1 (Figure 4C).

Figure 4 Computational fitting of Th17-related molecules. COPASI was used to run parameter estimation based on the microarray experimental data. Using a Genetic Algorithm and PSO, the model fitted (in blue) the experimental data (in black) of

A. Interleukin-10

B. Glucocorticoid Kinase 1 (SGK1)

C. TIMP1

D. FOXP3.

Local sensitivity analysis on SGK1, or also known as salt-sensing glucocorticoid kinase 1, showed a strong correlation with RORc, IL-17A, and most notably, the IL-23R (Figure 5). Indeed, the relationship between SGK1 and the promotion of pathogenic Th17 cells has already been reported.[47] In this case, Wu et al. demonstrated that SGK1 is an essential node downstream of the IL-23R pathway and is critical for regulating its expression by de-activating FOXO1.The last example of calibration is represented by FOXP3. FOXP3 inhibits the RORγt-driven transcription of IL-17 by directly suppressing RORγt.[33,42] Furthermore, the IL-2/STAT-5 axis constrains Th17[48] in part through a FOXP3-dependent mechanism, since STAT-5 activates FOXP3,[49] as well as through the inhibition of the STAT-3/IL-21 pathway.[50] Indeed, double positive FOXP3 RORγt T-helper cells have been identified as an intermediary that displays suppressive function[22] and are being investigated due to the plasticity of Th17. Our computational model also captures the relationship of FOXP3 and RORγt and it fits the expression data of FOXP3 (Figure 4D). These calibration results proved grounds to start in silico experimentation with our mathematical and computational model of Th17 differentiation.

In silico experimentation and computational modeling predictions

In silico validation of microarray experimental knock-out studies

Predictive models are characterized by generating results that can be further validated using subsets of data or in vivo experimentation. In order to initially validate our Th17 differentiation computational model, three genes of interest were selected based on the knock-out studies results published in.[27] These genes were Mina, Fas, and Pou2af1. The data for Mina, Fas, and Pou2af1 was not used during our calibration studies. Therefore, using a subset of data for validation ensures that the model here presented contains predictive validation.
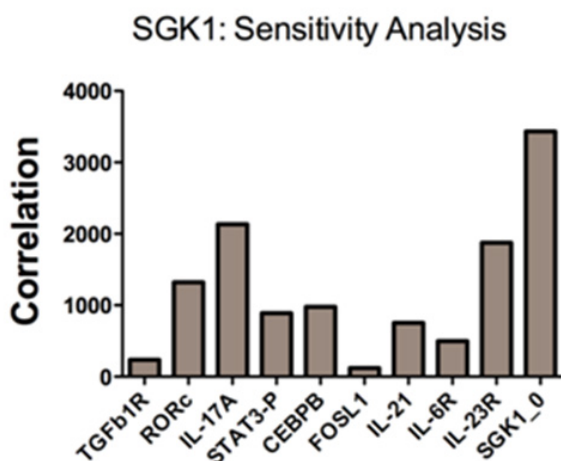

Figure 5 Sensitivity analysis on salt-sensing glucocorticoid kinase 1 (SGK1) by the Th17 differentiation computational model. Sensitivity analysis was run with COPASI on our computational model using a delta factor of 0.0001 and a delta minimum of 1e-12. The subtask run for the analysis was a time-series with t=100h and correlation of all the variables of the model against activated SGK1 was assessed, showing high correlation with key transcription factors that determine phenotype differentiation on Th17.

Mina is a chromatin regulator from the Jumonji C (JmjC) family and it represses Rorc, Batf, and Irf4 while increasing the expression of FOXP3.[27] Our computational Th17 differentiation model was able to fit the microarray experimental data for Mina (Figure 6A). Secondly, Fas pertains to the TNFR family[51] and it plays a key role in many human and murine autoimmune pathologies.[52,53] Based on the microarray results, Fas is early induced and is a target of STAT3. Similarly to Mina, our computational model was able to fit the model Fas concentration to the experimental data (Figure 6B). Of note, it is interesting to see how residual levels of Fas at t=20h can trigger IL-17A expression. These effects could be potentially attributed to stochasticity or inhibition effects. Last, Pou2af1 encodes the POU domain class 2-associating factor 1, a protein that has been linked to Th17 up regulation and that also mediates the inhibition of FOXP3, STAT4, and GATA3.[27]

Fas and Pou2af1 knockout systems that was validated using the knock-out experimental data from Yosef et al.[1]

Consistently to our previous fittings, Pou2af1 was also fitted by our computational model (Figure 6C). Next, in order to validate some of the experimental results extracted from the fifth figure of,[27] where T cells extracted from Mina -/-, Fas -/-, and Pou2af1 -/- mice were differentiated to Th17 with IL-6 and TGFβ, we generated three knock-out systems for Mina, Fas, and Pou2af1. Our computational IL-17A expression results were validated with the experimental data with

knock-out studies (Figure 6D). These results partially validate the Th17 differentiation model and allowed us to generate further novel predictions.
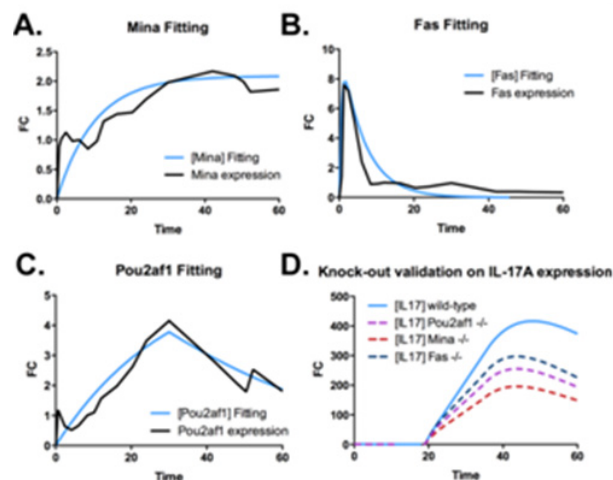
Figure 6 Fitting and validation of Mina, Fas, and Pou2af1 expression. Our computational model was able to fit the expression of

A. Mina

B. Fas

C. Pou2af1

D. In silico experimentation using Mina.

IL-24 modulates the balance between FOXP3 and RORγt during Th17 differentiation

One of the first targets for in silico experimentation was Interleukin-24 (IL-24), a member of the IL-10 family of cytokines. Upon binding to its receptors, IL-24 induces rapid activation of STAT1 and STAT3 transcription factors, both of which activate effector profiles in CD4+ T cell differentiation.[54] The role of IL-24 during Th17 differentiation, however, is very poorly understood and there are no publications explicitly stating how IL-24 fits in the complex Th17 differentiation story.

The microarray analysis of the Th17 differentiation datasets used in this study proved that IL-24 is highly expressed when a naïve CD4+ T cell is induced with TGFβ1 and IL-6. At the same time, our computational model was able to fit such dynamics observed in the experimental data (Figure 7A). In order to computationally shed some light on the role of IL-24 during Th17 differentiation, we performed local sensitivity analysis on the IL-24 node. Results showed how IL-24 negatively correlates with FOXP3 and positively correlates with Th17-related molecules such as STAT3, RORc, CEBPB and FOSL1 (Figure 7B). At first glance, IL-24 seems to promote an effector response by down regulating FOXP3 and minimizing its inhibition towards RORγt.

We next sought to determine what would occur if the expression of IL-24 was completely ablated in a Th17 cell. In order to accomplish this goal, we created an IL-24 null system, where the ability of IL-24 to exert its functions to other nodes was completely impaired. Results showed how the expression of FOXP3 in the IL-24 null when compared to the wild-type system remains unchanged during the first, approximately, 10 hours. However, after 10 hours, FOXP3 starts degrading over time. In contrast, in the IL-24 null system, FOXP3 reached a steady-state and it did not undergo degradation (Figure 7C). Since FOXP3 and RORγt are regulated with such tight balance, we next sought to determine the effect of this un-degraded FOXP3

towards the expression of RORc. Interestingly, the IL-24 null system showed less expression of RORc when compared to the wild-type Th17 model (Figure 7D).
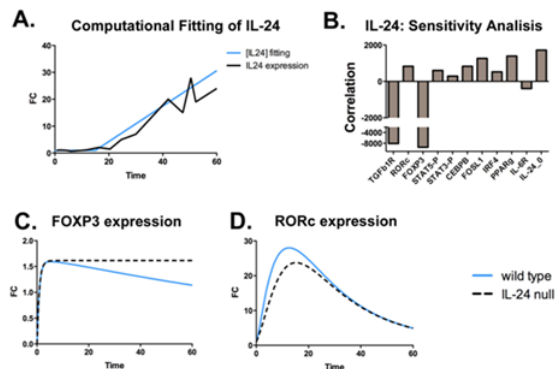
Figure 7 In silico experimentation: IL-24 modulates IL-17A production through a FOXP3-dependent mechanism.

A. IL-24 computational fitting of the microarray experimental data using Genetic Algorithm and PSO in COPASI

B. Sensitivity analysis on Interleukin 24. Time-course analysis of

C. FOXP3

D. RORc in either a wild-type model, or an IL-24 null Th17 differentiation model.

We hypothesize that the relationship of RORc and IL-24 is FOXP3-dependent. Our results show how that, when blocking IL-24, the concentration of FOXP3 increases. The increase of FOXP3 breaks the balance of FOXP3 and RORc, which have been found to antagonize.[22,23] For this reason we observe RORc down-regulated with the absence of IL-24. These counterintuitive results generated in silico by our Th17 differentiation model in regards to the role of IL-24 during Th17 induction should be validated with specific in vitro and in vivo experimental studies. The potential modulating role of IL-24 in the RORc-FOXP3 balance could lead to the development of IL-24 blockers as therapeutic treatment. Therefore, if these predictions were validated, IL-24 would arise as an immune-based, powerful, and potential therapeutic target to modulated inflammatory diseases characterized by a Th17 up regulation.

NLRP3 knockout Th17 cells result in an impaired production of IL-17A

Th17 cells stimulate the tissue environment towards an effector profile and orchestrate the function of cell subsets that are part of the innate immune response, such as macrophages, neutrophils, or epithelial cells. Part of this response, the Nod-like receptors (NLRs) are a subset of pattern recognition receptors (PRRs) found in the cytosol that are essential for detecting invading pathogens and initiating innate immune responses. NLRs are part of a broader group, named the inflammasome, which consists of molecular platforms activated upon cellular infection or stress that triggers the release of pro inflammatory cytokines, such as IL-1β or IL-6.[55] Within the inflammasome, NLRP3 has received particular attention since it interacts with the caspase recruitment domain (ASC) and the cytokine protease caspase-1, forming a cytoplasmic complex (NLRP3).[56]

Although it has been recently demonstrated that activation of NLRP3 in other cell subsets promotes Th17 differentiation in the T cell compartment in the context of allogeneic hematopoietic cell transplantation[57] and allergic lung inflammation,[58] T cell intrinsic mechanisms by which NLRP3 modulates Th17 function remain

highly unexplored. In our study, NLRP3 appeared in our list of the genes up regulated by the induction of Th17 over time. In our inferred network, NLRP3 is linked to the activation of GM-CSF and IFNγ (Figure 2C), both responsible to increase the pathogenic activity of Th17 cells.[59] Our Th17 computational model was able to fit the experimental expression data of NLRP (Figure 8A). Indeed, our sensitivity analysis confirmed that NLRP3 is linked to the activation of the pathogenic machinery in Th17, since NLRP3 is positively correlated with IFNγ, GM-CSF, and IL-17A (Figure 8B). We next sought to determine the role of NLRP3 over the production of IL-17A. We created an in silico NLRP3 null Th17 model.

In this system, NLRP3's ability to up regulate its linked genes was completely ablated. Our results show how the lack of NLRP3 in a Th17 cells also triggers a down regulation of IL-17A (Figure 8C). Given the amount of variability of NLRP3 expression in the first time-points of the Th17 induction process, we next sought to determine if a change in the variability of the expression of NLRP3 at the early stages of differentiation would affect the outcome of IL-17A production. To add stochasticity to the NLRP3 node, we used ENISI SDE, a novel web-based stochastic modeling tool.[60] Our results show how that the addition of 5% variability in the expression of the NLRP3 node slightly down regulated the IL-17A production in the Th17 cell (Figure 8D).
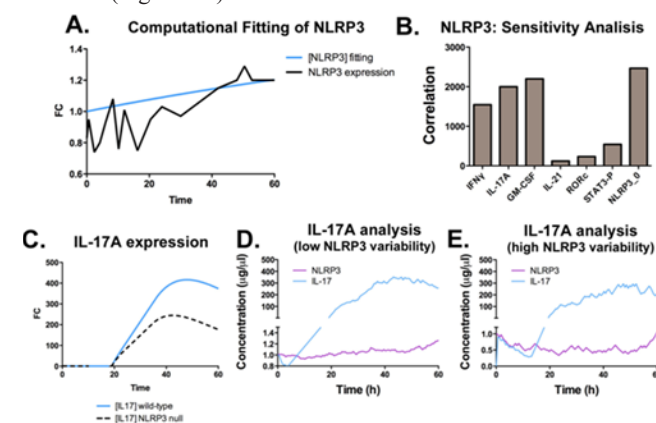
Figure 8 In silico experimentation: NLRP3 helps to modulate IL-17A production in Th17 cells.

A. NLRP3 computational fitting of the microarray experimental data using Genetic Algorithm and PSO in COPASI

B. Sensitivity analysis on NLRP3

C. Time-course analysis of IL-17A in either a wild-type model, or an NLRP3 null Th17 differentiation model

D. Time course analysis of IL-17A and NLRP3 using a Stochastic Differential Equation Simulator with a variability on the NLRP node of 5%

E. Time course analysis of IL-17A and NLRP3 using a Stochastic Differential Equation Simulator with a variability on the NLRP node of 30%.

For this reason, we hypothesized that if the variability was higher, the production of IL-17A would be highly affected. Indeed, our results show how when the expression of the NLRP3 node was set at a 30%, the production of IL-17A was dramatically affected (Figure 8E). These results highlight the role and T cell intrinsic relationship of NLRP3 and IL-17A. Of note, it is very interesting to see how a molecule like NLRP3, which shows a fold-change between 0.8 and 1, computationally exerts great changes in IL-17A. Further studies will evaluate how modest changes in gene expression among several components in a pathway is sometimes more relevant to cell phenotype that the magnitude of a single gene expression change.

More importantly, NLRP3 could represent a potential therapeutic target to treat Th17-mediated inflammatory diseases. Indeed, NLRP3 is under investigation to evaluate if it can be a good target to treat type 2 diabetes.[61]

Conclusions

As next generation sequencing technologies for datasets are becoming increasingly available at lower cost, the computational immunology field needs to take advantage of big data to construct more comprehensive, predictive networks. Indeed, the CD4+ T cell field will be soon taking advantage of newer and more novel methods such as single-cell sequencing. These approaches will be key to determine cell-to-cell differences and how heterogeneity affects the population as a whole.[62] Proteomics datasets will also confirm the findings observed using transcriptomics analyses. In parallel, mathematical modeling of immune responses offers an efficient approach to gain a deeper mechanistic systems-level understanding Th17 differentiation and plasticity, but also on how a Th17 cell can be pharmacologically manipulated to treat disease. In this study, we presented a computational pipeline that combines data-driven and theoretical approaches to study mechanisms of Th17 cell differentiation. Indeed, this paper explores a new potential use of microarray and RNA-seq data to produce highly predictive dynamic models. Of note, in order to calibrate the generated Th17 network, we used all 19 time-points available from the datasets.

However, further sensitivity analyses and robustness studies will be needed to evaluate the minimum number of time-points needed to fully and efficiently calibrate this Th17 network. In this study, we have identified two genes that are intimately related to Th17 function and induction: IL-24 and NLRP3. Furthermore, we computationally confirmed the role of SGK1 in modulating Th17 responses via an IL-23R by running local sensitivity analysis on SGK1. We demonstrated how the computational ablation of both of these genes resulted in a modulation of Th17 function. By using stochastic modeling systems, we also found that intrinsic noise in the NLRP3 gene dramatically affected the expression of IL-17A. Furthermore, we identified a molecular mechanism by which IL-24 exerts its modulator effects. By performing sensitivity analysis we found a negative correlation of IL-24 with FOXP3. In the IL-24 null system, we detected that the ablation of IL-24 and the up regulation of FOXP3 a posteriori, triggers an unbalanced equilibrium between FOXP3 and RORγt, since both act in an antagonistic manner. More specifically, the subtle FOXP3 up regulation caused by IL-24 led to a down regulation of RORγt.

Obviously, any prediction generated by the Th17 differentiation model will need to be validated with specific in vitro and in vivo experiments. Despite of this fact, in this paper we demonstrate how our developed data-driven pipeline can generate meaningful and counterintuitive predictions by taking advantage of publicly available datasets and apply a computational modeling framework based on network inference and data-mining to create a fully calibrated computational model, in this case, of Th17 differentiation. Altogether, this work can serve as a template on how to build a data-driven computational model with deterministic and stochastic features to generate new mechanistic knowledge related to immunological mechanisms and to identify novel therapeutic targets for infectious and immune-mediated diseases.

Acknowledgements

Conflicts of interest

None.

Funding

None.

References

1. Yosef N, Shalek AK, Gaublomme JT, et al. Dynamic regulatory network controlling TH17 cell differentiation. *Nature*. 2013;496(7446):461-468.

2. Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu Rev Immunol*. 1989;7:145-173.

3. Ma CS, Tangye SG, Deenick EK. Human Th9 cells: inflammatory cytokines modulate IL-9 production through the induction of IL-21. *Immunol Cell Biol*. 2010;88(6):621-623.

4. Ivanov II, McKenzie BS, Zhou L, et al. The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell*. 2006;126(6):1121-1133.

5. Ramirez JM, Brembilla NC, Sorg O, et al. Activation of the aryl hydrocarbon receptor reveals distinct requirements for IL-22 and IL-17 production by human T helper cells. *Eur J Immunol*. 2010;40(9):2450-2459.

6. Nurieva RI, Chung Y, Hwang D, et al. Generation of T follicular helper cells is mediated by interleukin-21 but independent of T helper 1, 2, or 17 cell lineages. *Immunity*. 2008;29(1):138-149.

7. Vogelzang A, McGuire HM, Yu D, et al. A fundamental role for interleukin-21 in the generation of T follicular helper cells. *Immunity*. 2008;29(1):127-137.

8. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. *Science*. 2003;299(5609):1057-1061.

9. Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol*. 2003;4(4):330-336.

10. Groux H, O'Garra A, Bigler M, et al. A CD4+ T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. *Nature*. 1997;389(6652):737-742.

11. Abraham C, Cho J. Interleukin-23/Th17 pathways and inflammatory bowel disease. *Inflamm Bowel Dis*. 2009;15(7):1090-1100.

12. Wu CC, Sytwu HK, Lu KC, et al. Role of T cells in type 2 diabetic nephropathy. *Exp Diabetes Res*. 2011;2011:514-738.

13. van Hamburg JP, Asmawidjaja PS, Davelaar N, et al. Th17 cells, but not Th1 cells, from patients with early rheumatoid arthritis are potent inducers of matrix metalloproteinases and proinflammatory cytokines upon synovial fibroblast interaction, including autocrine interleukin-17A production. *Arthritis Rheum*. 2011;63(1):73-83.

14. Carbo A, Bassaganya-Riera J, Pedragosa M, et al. Predictive computational modeling of the mucosal immune responses during Helicobacter pylori infection. *PLoS One*. 2013;8(9):e73365.

15. Viladomiu M, Hontecillas R, Pedragosa M, et al. Modeling the role of peroxisome proliferator-activated receptor gamma and microRNA-146 in mucosal immune responses to Clostridium difficile. *PLoS One*. 2012;7(10):e47525.

16. Esplugues E, Huber S, Gagliani N, et al. Control of TH17 cells occurs in the small intestine. *Nature*. 2011;475(7357):514-518.

17. Fujino S, Andoh A, Bamba S, et al. Increased expression of interleukin 17 in inflammatory bowel disease. *Gut*. 2003;52(1):65-70.

18. Ahern PP, Schiering C, Buonocore S, et al. Interleukin-23 drives intestinal inflammation through direct activity on T cells. *Immunity*. 2010;33(2):279-288.

19. El-Behi M, Ciric B, Dai H, et al. The encephalitogenicity of T(H)17 cells is dependent on IL-1- and IL-23-induced production of the cytokine GM-CSF. *Nat Immunol*. 2011;12(6):568-575.

20. Codarri L, Gyulveszi G, Tosevski V, et al. RORgammat drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. *Nat Immunol*. 2011;12(6):560-567.

21. Hong T, Xing J, Li L, et al. A mathematical model for the reciprocal differentiation of T helper 17 cells and induced regulatory T cells. *PLoS Comput Biol*. 2011;7(7):e1002122.

22. Mendoza L. A virtual culture of CD4+ T lymphocytes. *Bull Math Biol*. 2013;75(6):1012-1029.

23. Tartar DM, VanMorlan AM, Wan X, et al. FoxP3+RORgammat+ T helper intermediates display suppressive function against autoimmune diabetes. *J Immunol*. 2010;184(7):3377-3385.

24. Zhou L, Lopes JE, Chong MM, et al. TGF-beta-induced Foxp3 inhibits T(H)17 cell differentiation by antagonizing RORgammat function. *Nature*. 2008;453(7192):236-240.

25. Carbo A, Hontecillas R, Kronsteiner B, et al. Systems modeling of molecular mechanisms controlling cytokine-driven CD4+ T cell differentiation and phenotype plasticity. *PLoS Comput Biol*. 2013;9(4):e1003027.

26. Carbo A, Hontecillas R, Andrew T, et al. Computational modeling of heterogeneity and function of CD4+ T cells. *Front Cell Dev Biol*. 2014;2:31.

27. Carbo A, Hontecillas R, Andrew T, et al. Computational modeling of heterogeneity and function of CD4+ T cells. *Front Cell Dev Biol*. 2014;2:31.

28. Quackenbush J. Microarray data normalization and transformation. *Nat Genet*. 2002;32 Suppl: 496-501.

29. Marbach D, Prill RJ, Schaffter T, et al. Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci U S A*. 2010;107(14):6286-6291.

30. Wang J, Zhao YC, Lu YD, et al. Integrated bioinformatics analyses identify dysregulated miRNAs in lung cancer. *Eur Rev Med Pharmacol Sci*. 2014;18(16):2270-2274.

31. Wray CJ, Ko TC, Tan FK. Secondary use of existing public microarray data to predict outcome for hepatocellular carcinoma. *J Surg Res*. 2014;188(1):137-142.

32. Bettelli E, Korn T, Kuchroo VK. Th17: the third member of the effector T cell trilogy. *Curr Opin Immunol*. 2007;19(6):652-657.

33. Zhou L, Lopes JE, Chong MMW, et al. TGF-[bgr]-induced Foxp3 inhibits TH17 cell differentiation by antagonizing ROR[ggr]t function. *Nature*. 2008;453:236-240.

34. Bettelli E, Carrier Y, Gao W, et al. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature*. 2006;441(7090):235-238.

35. Korn T, Bettelli E, Gao W, et al. IL-21 initiates an alternative pathway to induce proinflammatory TH17 cells. *Nature*. 2007;448:484-487.

36. Zhou L, Ivanov II, Spolski R, et al. IL-6 Programs TH-17 Cell Differentiation by Promoting the Sequential Engagement of the IL-21 and IL-23 Pathways. *Nat Immunol*. 2007;8(9):967-974.

37. Foley JF. STAT3 Regulates the Generation of Th17 Cells. *Sci STKE*. 2007;2007(380:tw113.

38. Harris TJ, Grosso JF, Yen H-R, et al. Cutting Edge: An In Vivo Requirement for STAT3 Signaling in TH17 Development and TH17-Dependent Autoimmunity. *J Immunol*. 2007;179(7):4313-4317.

39. Mathur AN, Chang HC, Zisoulis DG, et al. Stat3 and Stat4 Direct Development of IL-17-Secreting Th Cells. *J Immunol*. 2007;178(8):4901-4907.

40. Yang XO, Panopoulos AD, Nurieva R, et al. STAT3 Regulates Cytokine-mediated Generation of Inflammatory Helper T Cells. *J Biol Chem*. 2007;282(13):9358-9363.

41. Ivanov II, McKenzie BS, Zhou L, et al. The Orphan Nuclear Receptor ROR³t Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. *Cell*. 2006;126(6):1121-1133.

42. Ichiyama K, Yoshida H, Wakabayashi Y, et al. Foxp3 Inhibits ROR{gamma}t-mediated IL-17A mRNA Transcription through Direct Interaction with ROR{gamma}t. *J Biol Chem*. 2008;283(25):17003-17008.

43. Hoops S, Sahle S, Gauges R, et al. COPASI--a COmplex PAthway SImulator. *Bioinformatics*. 2006; 22(24):3067-3074.

44. Kennedy JE, Eberhart R. Particle Swarm Optimization. Neural Networks, IEEE. *International Conference*. 1995;4:1942-1948.

45. Xu R, Venayagamoorthy GK, Wunsch DC 2nd. Modeling of gene regulatory networks with hybrid differential evolution and particle swarm optimization. *Neural Netw*. 2007;20(8):917-927.

46. Harris TJ, Grosso JF, Yen HR, et al. Cutting edge: An in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity. *J Immunol*. 2007;179(7):4313-4317.

47. Wu C, Yosef N, Thalhamer T, et al. Induction of pathogenic TH17 cells by inducible salt-sensing kinase SGK1. *Nature*. 2013;496(7446):513-517.

48. Laurence A, O'Shea JJ. TH-17 differentiation: of mice and men. *Nature Immunology*. 2007;8:903-905.

49. Passerini L, Allan SE, Battaglia M, et al. STAT5-signaling cytokines regulate the expression of FOXP3 in CD4+CD25+ regulatory T cells and CD4+CD25- effector T cells. *Int Immunol*. 2008; 20(3):421-431.

50. Wei L, Laurence A, Elias KM, et al. IL-21 is produced by Th17 cells and drives IL-17 production in a STAT3-dependent manner. *J Biol Chem*. 2007;282(48):34605-34610.

51. Griffith TS, Brunner T, Fletcher SM, et al. Fas ligand-induced apoptosis as a mechanism of immune privilege. *Science*. 2005;270(5239):1189-1192.

52. Stassi G, De Maria R. Autoimmune thyroid disease: new models of cell death in autoimmunity. *Nat Rev Immunol*. 2002;2(3):195-204.

53. Suvannavejh GC, Dal Canto MC, Matis LA, et al. Fas-mediated apoptosis in clinical remissions of relapsing experimental autoimmune encephalomyelitis. *J Clin Invest*. 2000;105(2):223-231.

54. Wang M, Liang P. Interleukin-24 and its receptors. *Immunology*. 2005;114(2):166-170.

55. Schroder K, Tschopp J. The inflammasomes. *Cell*. 2010;140(6):821-832.

56. Westwell-Roper C, Nackiewicz D, Dan M, et al. Toll-like receptors and NLRP3 as central regulators of pancreatic islet inflammation in type 2 diabetes. *Immunol Cell Biol*. 2014;92(4):314-323.

57. Jankovic D, Ganesan J, Bscheider M, et al. The Nlrp3 inflammasome regulates acute graft-versus-host disease. *J Exp Med*. 2013;210(10):1899-1910.

58. Besnard AG, Togbe D, Couillin I, et al. Inflammasome-IL-1-Th17 response in allergic lung inflammation. *J Mol Cell Biol*. 2012;4(1):3-10.

59. Duhen R, Glatigny S, Arbelaez CA, et al. Cutting edge: the pathogenicity of IFN-gamma-producing Th17 cells is independent of T-bet. *J Immunol*. 2013;190(9):4478-4482.

60. Mei Y, Carbo, A, Hoops S, et al. ENISI SDE: A novel web-based stochastic modeling tool for computational biology. *Transactions of Computational Biology and Bioinformatics In Press.* 2013.

61. Liu SS, Ding Y, Lou JQ. NLRP3, a Potential Therapeutic Target for Type 2 Diabetes? *Cardiovasc Drugs Ther*. 2014;28(4):391-392.

62. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell*. 2010; 141(4):559-563.