Exploring T cell regulatory networks in autoimmune diseases: shared mechanisms and disease-specific pathways

*Proposal by Judith Zaugg (EMBL) and Enrico Ferrero (GSK)*

**Background**

Despite their heterogeneity and the variety of involved organs and systems, autoimmune diseases (ADs) often share similar underlying pathogenic mechanisms and might have a common origin (*1*, *2*). T cells are key players in ADs (*3*–*5*), such as rheumatoid arthritis (*6*), multiple sclerosis (*7*) systemic lupus erythematosus (*8*) and Crohn’s disease (*9*), among others. However, the molecular basis for T cell-mediated autoimmunity and how similar the underlying processes are across different ADs are still unclear.

Genome-wide association studies (GWASs) have identified hundreds of loci significantly associated with ADs, with many of these shared across individual diseases (*1*, *10*, *11*). These genetic variants are often found in regulatory regions such as open chromatin regions and enhancers (*12*–*16*). It is thus likely that disease-associated variants have a deleterious impact on the binding or assembly of transcription factor (TF) complexes at regulatory regions, which ultimately results in dysregulated regulatory networks and disease phenotypes.

Here, we propose to generate a regulatory network of healthy CD4+ T cells and subsequently identify and characterize the common and distinct regulatory pathways that are involved in four different autoimmune diseases: Graves’ disease (GD), juvenile idiopathic arthritis (JIA), systemic lupus erythematosus (SLE) and Psoriasis.

**Project plan and aims**

Our overall goal is to understand the global and disease-specific regulatory cascades that occur in CD4+ T cells leading to autoimmune pathogenic processes in the different diseases. The overall goal is split into four specific aims:

* **Aim 1:** To generate a regulatory network based on publicly available data from the IHEC consortium (*17*). We will use an approach that we have previously used to construct a regulatory network in pulmonary epithelial cells (Reyes-Palomares *in preparation*). Briefly, we will use correlations between H3K27ac (as readout of TF binding activity) and RNA-seq (as proxy for the expression level of a TF) across individuals to determine the regulatory elements that are controlled by a particular TF. The regulatory elements will then be linked to genes by proximity and publicly available promoter capture Hi-C data generated in T cells (*18*, *19*). This will provide us with a general network of TFs and target genes specific to CD4+ T cells.
* **Aim 2:** To identify the TFs that are differentially active between a specific autoimmune disease cohort and healthy controls, we will use our approach for estimating TF activity differences based on histone acetylation signals (Berest et al *in preparation)*. Briefly, we estimate the differential activity of a TF across two conditions based on the overall changes in histone acetylation signal over all its putative binding sites. For this, we will use (i) publicly available data for GD (*20*) and JIA (*21*), and (ii) generate new regulatory genomic and transcriptomic data for patients suffering from SLE and psoriasis. This will provide us with a list of TFs that significantly change their activity in each of the autoimmune diseases under investigation, which will be integrated with the regulatory network generated in Aim 1.
* **Aim 3:** To understand the common pathways of autoimmune diseases as well as the disease-specific parts we will plug the differentially active TFs for each disease obtained in Aim 2 into the regulatory network generated in Aim 1. This might reveal specific regulators driving individual diseases as well as highlighting functional modules shared across ADs. We will then expand the disease regulatory networks by integrating pathway knowledge and protein – protein interactions (e.g. from Reactome, IntAct, KEGG, PANTHER, GO, STRING, etc) to identify potential therapeutic targets or biomarkers.
* **Aim 4:** To overlay genetics data onto the regulatory network to highlight functional modules with genetics evidence. We expect Aim 3 will retrieve the key transcriptional regulators for the diseases under investigation and we plan to validate our approach by overlaying known GWAS hits for these conditions and assessing enrichment over differential peaks. Importantly, this process will also guide us to the identification of disease-causing functional modules with genetic evidence, which could point us to more attractive points of therapeutic intervention (*22*).

**Deliverables**

1. The main output of this project should be an increased understanding of the function of T cells and the mechanisms by which they can drive autoimmunity in a range of inflammatory diseases. T cells are an important area of focus within the Immuno-Inflammation Therapeutic Area (II TA) at GSK and our results might have an impact in biomarker and/or target selection strategies for Discovery Performance Units (DPUs) interested in this space, such as the Tempero DPU.
2. We also aim to gain novel insights into specific disease areas where the II TA is actively working on (GD, SLE, psoriasis). Similarly, biological mechanisms identified in JIA might be relevant for other areas of interest, such as rheumatoid arthritis and osteoarthritis. Some of the disease-specific functional modules identified might represent novel therapeutic intervention points and we will seek to engage with the relevant partners within the II TA.
3. If successful, the approach described here would represent the development of a capability that could be applied elsewhere (e.g.: a different group of diseases and/or a different cell type) as an integrated workflow to uncover biological pathways that are disease-specific and shared across related diseases.

**Estimated costs and timeline**

The main part of the proposed project is computational data integration and will be carried out by a postdoc jointly supervised by Judith Zaugg and Enrico Ferrero. The experiments will be carried out in collaboration with Pärt Peterson (University of Tartu, already collaborating with Judith Zaugg on an independent project) and the GeneCore sequencing facility at EMBL.

Costs:

The experimental costs are estimated at ~10000 EUR for collecting patient samples, cell sorting and ChIP + Sequencing (45 samples multiplexed into 9 lanes) ~15000 EUR.

Timeline:

Patient sampling and ChIP-Seq experiments will be started immediately and take 7- 8 months

Computational analysis will be carried out by the GSK-funded postdoc, who is planned for 24 months:

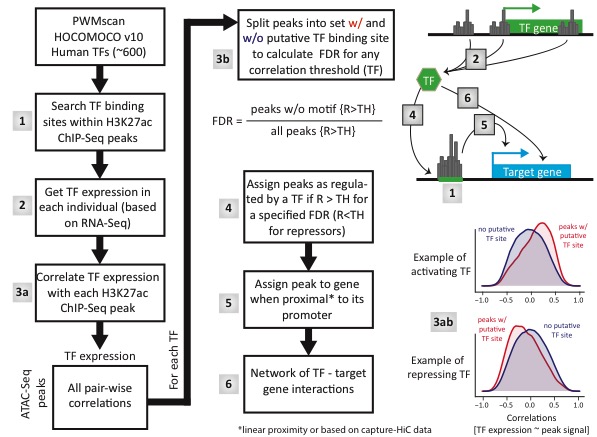
* Aim 1: months 1-6
* Aim 2 (including processing of new data): months 7-12
* Aim 3: months 13-18
* Aim 4: months 15-24

**Detailed description**

Transcription factors (TFs) are low abundant proteins and are thus difficult to measure in a proteomics experiment. However, TFs activities are often good read-outs of the activities of the gene expression pathways they regulate. They are thus important components to study for understanding the changes in biological processes that drive diseases, including autoimmune disorders. Here we propose to use a combination of RNA and active histone marks to build a T-cell specific regulatory network and then use this to further our understanding of common pattern and distinct changes in the regulatory landscape of T cells in patients suffering from different types of autoimmune diseases.

**Aim 1: Gene regulatory network for T cells based of publicly available data of 156 individuals**

To obtain the cell type specific regulatory networks we will follow the approach in **Figure 1**. We will first obtain a set of putative TF binding sites, using PWMscan (*23*), for each TF for which a motif is available (e.g. from the HOCOMOCO data base (*24*)) (1). In parallel we estimate the expression level of every TF based the RNA-seq levels from data generated by the IHEC consortium for CD4+ alpha-beta T cells (available for 155 individuals)(*17*) (2). We then calculate the Pearson correlations between each TF expression level and each H3K27ac-Seq signal across individuals (3a). Then, for each TF, we group the correlations into putative target peaks and other peaks based on whether or not they overlap with a putative binding site of the TF and calculate the False Discovery Rate (FDR) based on the formula given in **Figure 1** (3b) and define the correlation threshold at which we obtain an FDR of 10%. We then assign peaks that correlate more (for activating TFs) or less (for repressive TFs) than said threshold as the target peaks of a given TF. Finally, we assign the target peaks to the genes based on their proximity to a gene promoter (5) and with this obtain a link between TFs and their target genes (6). Importantly, there are several Capture-HiC data sets available for T-cells (*18*, *19*), which will allow us to also link more distal H3K27ac peaks in enhancer regions to their respective promoters.



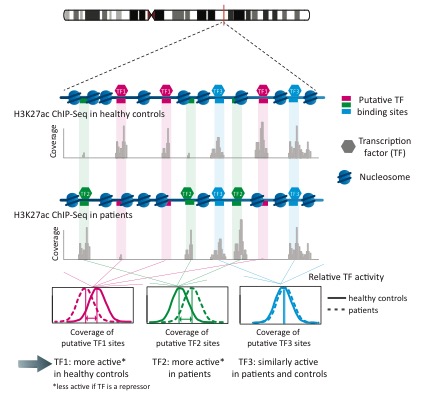
**Figure 1: Schematic of the approach for building a gene regulatory network from H3K27ac ChIP-Seq data.** Approach and data are from Reyes-Palomares et al (*in pre*paration).

This regulatory network will serve as a framework for predicting the effects of changes in TF activity across the different autoimmune diseases. By plugging in the TF activity differences obtained in **Aim 2** we will be able to identify potential master regulators and intervention points that might be causally associated with autoimmune diseases in general or with a specific autoimmune disease in particular.

**Aim 2: Assessing transcription factor binding activity based on histone modification data**

We have developed a method to globally assess TF binding activity for all TF for which a motif (PWM) is known. Briefly, we use open chromatin (e.g. H3K27ac marks) in the vicinity of all putative binding sites of a given TF as a readout its binding activity. By then comparing the binding activity of all TFs between two conditions we can identify the factors that are important for cell-type identity.

The method is based on the results of our hQTL study where we found that genotype-dependent variation in H3K27ac signal can be explained by disruptions of TF motifs whenever the hQTL-SNP overlapped with a TF binding site (*14*). And without knowing the exact mechanism we can assume that TF binding plays an important role in mediating the effect of the DNA variant onto chromatin marks. We can thus turn this relationship around and use H3K27ac, or open chromatin markers in general, at putative binding sites of a given TF as a read-out of its binding activity. Briefly, we obtained the putative binding sites for 644 TF by scanning the genome with the respective position weight matrices (PWMs) using PWMscan (*23*). The PWMs for each TF were obtained from the HOCOMOCO database (*24*), which collects this information from numerous ChIP-Seq experiments. For each TF we then calculate the average fold-change in chromatin accessibility across all its putative binding sites, and normalize this value to the average fold-change across the entire genome to estimate its relative binding activity between two conditions – in this case healthy controls and patients (see schematic in **Figure 2A**). It is important to note that the signal differences at each of the individual putative binding site is very low and rarely significant, however, by averaging across the thousands of putative binding sites per TF we are able to detect the global differences in binding activity of the TF.



**Figure 2: TF activity: approach and validation.** A: Schematic description of our approach to estimate relative TF binding activity. Relative TF activity is calculated based on the fold-change of open chromatin (ATAC-Seq reads) across all its putative binding sites. B: Correlation of the RNA expression level of a TF and its binding activity for all TFs (B) and only TFs that have exceed a certain difference in binding activity (C). Approach and data from Berest et al (*in preparation)*

B)

C)

A)

We have validated the concept of TF binding activity by comparing the change in RNA expression levels and binding activity for a set of TFs between two conditions (mouse GMP cells wildtype vs knock out of TET2, Berest et al (*in preparation, data shown in Fig. 2B-C)*). The specific comparison does not matter, however what is important to see is that the overall Pearson correlation between TF activity and TF expression levels is 0.22 and when only considering TFs that show a significant difference in binding activity, this correlation increases to 0.59 (**Figure 2**).

*Subaim 2.1:* *Estimate TF activities for GD and JIA cohorts from publicly available data.* CD4+ and CD8+ T cells from 38 GD patients and 31 healthy individuals were epigenetically profiled using H3K27ac and H3K4me3 ChIP-seq (*20*) with data accessible from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71957>. Similarly, H3K27ac ChIP-seq data on CD4+ T cells was generated from 14 JIA patients and healthy controls (*21*) with data available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71597>. We will use the H3K27ac ChIP-seq data from these cohorts to calculate relative TF activities in CD4+ T cells by comparing patients and healthy controls. RNA-seq data on the same samples will be used to filter out TFs that are not expressed in the cells.

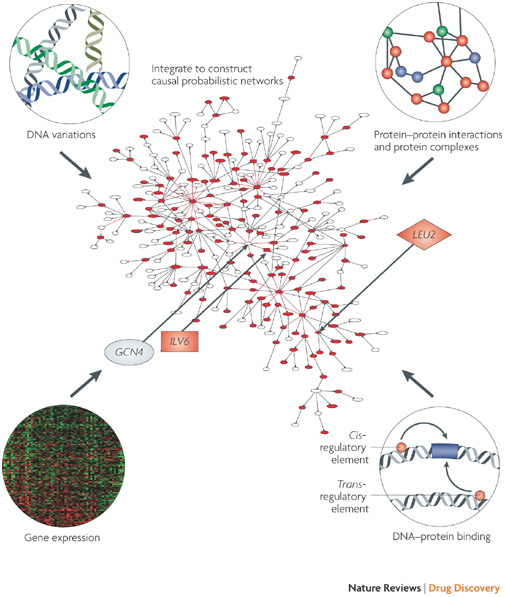
*Subaim 2.2: Generate H3K27ac data and estimate TF activities for SLE and psoriasis cohorts.* We will obtain CD4+ T cells from SLE and psoriasis patients and perform H3K27ac ChIP-Seq as well as RNA-seq on them along with a set of healthy controls. We will then employ the same approach as in Subaim 2.1 to estimate TF activities. The RNA will be used to determine which TFs are expressed in our samples. We are aiming for 10-15 patients for each disease and 10-15 controls, which can be shared between the two diseases. These experiments will be done in collaboration with the group of Pärt Peterson, who have a longstanding expertise in profiling chromatin marks in T cells of autoimmune patients (*20*, *25*, *26*). The sequencing will be performed at the EMBL core facility (GeneCore).

**Aim 3: Identification of regulatory networks and pathways shared across ADs and disease-specific ones.** Overall, we plan to obtain H3K27ac ChIP-seq data on T cells from four patient populations with ADs, with the twofold objective of getting insights into the individual conditions and leveraging the diversity and heterogeneity across ADs to identify shared mechanisms and regulatory networks. After generating disease-specific regulatory networks as described in Aim 2, we will integrate our results across the different diseases to identify functional modules underpinning common mechanisms in multiple ADs and highlight areas of the network that are specific to individual diseases. These findings will help us better understand the molecular details of T cell involvement in ADs and what differences, if any, exist between populations of T cells from patients with different autoimmune conditions.

We will then expand these regulatory network using public domain databases of pathways (e.g.: Reactome, KEGG) and protein – protein interactions (e.g.: IntAct, BioGRID, STRING) upstream and downstream of the key TFs identified. This will allow us to get a better overview of the underlying biology and to identify functional modules that might be attractive from a therapeutic point of view.

**Aim 4: Overlaying GWAS SNPs for autoimmune disease with disease-specific peaks and expanded regulatory network.** We will obtain the list of significant GWAS SNPs for the ADs under investigation from the GWAS catalog (<https://www.ebi.ac.uk/gwas/>) or GRASP (<https://grasp.nhlbi.nih.gov/Overview.aspx>) and expand our collection of variants using linkage disequilibrium data from the 100 Genomes project. We will then test whether the associated SNPs are enriched for regions that show differential acetylation signal, and in particular, for disease-specific TF binding sites. By utilising the same enhancer to gene links described above (i.e.: promoter-capture Hi-C data in T cells and proximity), we will also map GWAS hits onto the network. This will add another layer to our understanding of T cell-mediated autoimmunity by flagging disease-specific or disease-shared modules that carry genetic evidence and are more likely to be causal. These modules will be explored in more detail to assess whether any opportunities from a target discovery perspective exist. A network framework similar to the one described here is depicted in **Figure 3.** Finally, we will use a supervised learning approach to test if the identified regulatory networks carries enough information to predict the key regulators that have associated genetic evidence. If this hypothesis can be demonstrated, we will then use the classifier to identify potentially novel targets based on their disease association profiles, network topological properties and neighbourhood.

**Figure 3: Example of an interaction/regulatory network where several layers of data can be overlaid.** In this proposal, the initial network will be generated from healthy T cells (Aim 1). Transcription factors displaying differential activity in the diseases under investigation will be added following the analysis of the H3K27ac data (Aim 2). Data from the different diseases will then be integrated to reveal common modules of autoimmunity and disease-specific ones. Then, the network will be expanded using protein-protein and causal interactions from pathways (Aim 3). Finally, GWAS hits will be overlaid to identify possible points of therapeutic intervention. (Figure taken from (27))



**Bibliography**

1. M. Parkes, A. Cortes, D. A. van Heel, M. A. Brown, Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–73 (2013).

2. J. M. Anaya, Common mechanisms of autoimmune diseases (the autoimmune tautology). *Autoimmun. Rev.* **11** (2012), pp. 781–784.

3. U. Walter, P. Santamaria, CD8+ T cells in autoimmunity. *Curr. Opin. Immunol.* **17** (2005), pp. 624–631.

4. A. Jäger, V. K. Kuchroo, Effector and regulatory T-cell subsets in autoimmunity and tissue inflammation. *Scand. J. Immunol.* **72** (2010), pp. 173–184.

5. M. Kronenberg, A. Rudensky, Regulation of immunity by self-reactive T cells. *Nature*. **435**, 598–604 (2005).

6. A. P. Cope, T cells in rheumatoid arthritis. *Arthritis Res. Ther.* **10 Suppl 1**, S1 (2008).

7. J. M. Fletcher, S. J. Lalor, C. M. Sweeney, N. Tubridy, K. H. G. Mills, T cells in multiple sclerosis and experimental autoimmune encephalomyelitis. *Clin. Exp. Immunol.* **162** (2010), pp. 1–11.

8. A. Suárez-Fueyo, S. J. Bradley, G. C. Tsokos, T cells in Systemic Lupus Erythematosus. *Curr. Opin. Immunol.* **43** (2016), pp. 32–38.

9. A. Salas, J. Panés, IBD. Regulatory T cells for treatment of Crohn’s disease. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 315–6 (2015).

10. P. S. Ramos, A. M. Shedlock, C. D. Langefeld, Genetics of autoimmune diseases: insights from population genetics. *J. Hum. Genet.* **60**, 657–664 (2015).

11. Y. Kochi, Genetics of autoimmune diseases: Perspectives from genome-wide association studies. *Int. Immunol.* **28** (2016), pp. 155–161.

12. M. T. Maurano *et al.*, Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science (80-. ).* **337**, 1190–1195 (2012).

13. M. Kasowski *et al.*, Extensive variation in chromatin states across humans. *Science (80-. ).* **342**, 750–752 (2013).

14. F. Grubert *et al.*, Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell*. **162**, 1051–1065 (2015).

15. D. Hnisz *et al.*, XSuper-enhancers in the control of cell identity and disease. *Cell*. **155**, 934–947 (2013).

16. K. K.-H. Farh *et al.*, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. **518**, 337–43 (2015).

17. D. Bujold *et al.*, The International Human Epigenome Consortium Data Portal. *Cell Syst.* **3**, 496–499.e2 (2016).

18. I. Chepelev, G. Wei, D. Wangsa, Q. Tang, K. Zhao, Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* **22**, 490–503 (2012).

19. P. Martin *et al.*, Chromosome interaction analysis of risk loci in related autoimmune diseases reveals complex , long-range promoter interactions implicating novel candidate genes. *Nat. Commun.* **6**, 1–17 (2015).

20. M. Limbach *et al.*, Epigenetic profiling in CD4+ and CD8+ T cells from Graves’ disease patients reveals changes in genes associated with T cell receptor signaling. *J. Autoimmun.* **67**, 46–56 (2016).

21. J. G. C. Peeters *et al.*, Inhibition of Super-Enhancer Activity in Autoinflammatory Site-Derived T Cells Reduces Disease-Associated Gene Expression. *Cell Rep.* **12**, 1986–1996 (2015).

22. M. R. Nelson *et al.*, The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

23. G. Ambrosini, PWMTools. *http://ccg.vital-it.ch/pwmtools*.

24. I. V. Kulakovskiy *et al.*, HOCOMOCO: A comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* **41**, D195–D202 (2013).

25. L. Tserel *et al.*, Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. *Sci. Rep.* **5**, 13107 (2015).

26. S. Kasela *et al.*, Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLOS Genet.* **13**, e1006643 (2017).

27. E. E. Schadt, S. H. Friend, D. A. Shaywitz, A network view of disease and compound screening. *Nat. Rev. Drug Discov.* **8**, 286–95 (2009).