

## **Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis**

### **heritability**

Xiangyu Ge<sup>1\*</sup>, Mojca Frank-Bertoncelj<sup>2\*</sup>, Kerstin Klein<sup>2</sup>, Amanda McGovern<sup>1</sup>, Tadeja Kuret<sup>2,3</sup>, Miranda Houtman<sup>2</sup>, Blaž Burja<sup>2,3</sup>, Raphael Micheroli<sup>2</sup>, Miriam Marks<sup>4</sup>, Andrew Filer<sup>5,6</sup>, Christopher D. Buckley<sup>5,6,7</sup>, Gisela Orozco<sup>1</sup>, Oliver Distler<sup>2</sup>, Andrew P Morris<sup>1</sup>, Paul Martin<sup>1</sup>, Stephen Eyre<sup>1\*</sup> & Caroline Ospelt<sup>2\*,#</sup>

<sup>1</sup>Versus Arthritis Centre for Genetics and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

<sup>2</sup>Department of Rheumatology, Center of Experimental Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

<sup>3</sup>Department of Rheumatology, University Medical Centre, Ljubljana, Slovenia

<sup>4</sup>Schulthess Klinik, Zurich, Switzerland

<sup>5</sup>Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK

<sup>6</sup>NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust, University of Birmingham, Birmingham, UK

<sup>7</sup>Kennedy Institute of Rheumatology, University of Oxford Roosevelt Drive Headington Oxford UK

\*These authors contributed equally

#corresponding author: [caroline.ospelt@usz.ch](mailto:caroline.ospelt@usz.ch)

## Abstract

Genome-wide association studies have reported >100 risk loci for rheumatoid arthritis (RA). These loci have been shown to be enriched in immune cell-specific enhancers, but analysis so far has excluded stromal cells, such as synovial fibroblasts (FLS), despite their crucial involvement in the pathogenesis of RA. Here we integrated DNA architecture (ChIP-seq), 3D chromatin interactions (HiC, capture HiC), DNA accessibility (ATAC-seq) and gene expression (RNA-seq) in FLS, B cells and T cells with genetic fine mapping of RA loci. We identified putative causal variants, enhancers, genes, and cell types for 30 - 60% of RA loci and demonstrated that FLS account for up to 24% of RA heritability. TNF stimulation of FLS altered the organization of topologically associating domains (TADs), chromatin state and the expression of putative causal genes (e.g. *TNFAIP3*, *IFNARI*). Several putative causal genes constituted RA-relevant functional networks in FLS with roles in cellular proliferation and activation. Finally, we demonstrated that risk variants can have joint-specific effects on target gene expression in RA FLS, which may contribute to the development of the characteristic pattern of joint involvement in RA. Overall, our research provides the first direct evidence for a causal role of FLS in the genetic susceptibility for RA accounting for up to a quarter of RA heritability.

## Main

A major challenge of the post-genome-wide association study (GWAS) era is to decipher the functional consequences of genetic risk variants in individual cell types and their contribution to the development of polygenic diseases. The identification of the cell types and conditions in which genetic risk variants are effective is an essential prerequisite for achieving this goal. Rheumatoid arthritis (RA) is a symmetric inflammatory and destructive autoimmune arthritis with a complex genetic basis. RA affects 0.5-1% of the world population and leads to disability, high morbidity burden and premature mortality<sup>1</sup>. GWAS have identified over 100 loci for RA susceptibility<sup>2</sup>. Genetic risk variants at the majority of these loci do not map to the exons of protein coding genes. Potential gene regulatory functions of these noncoding genetic risk variants have been investigated in immune cells based on genome-wide mapping of epigenetic modifications<sup>3</sup>, chromatin interactions<sup>4</sup>, correlation with variation in gene expression (eQTLs)<sup>5</sup> or linear proximity to coding genes in DNA sequence<sup>2</sup>. These studies have demonstrated an enrichment of RA genetic risk variants in immune cell enhancers<sup>3</sup>, but omitted the analysis of synovial fibroblasts or Fibroblast-like Synoviocytes (FLS), the resident stromal cells of the joints, even though they are responsible for the production of many immune related cytokines and chemokines<sup>6,7</sup>.

In addition to immune cells, FLS play a decisive role in the pathogenesis of RA and are essential for the maintenance of normal joint functions. FLS from different joints have different epigenomes, transcriptomes and functions, which may contribute to the characteristic pattern of joint involvement in different types of arthritis<sup>8,9</sup>. FLS substantially contribute to joint inflammation and destruction in RA<sup>10</sup>. RA FLS have an activated phenotype characterized by resistance to apoptosis, increased proliferation, secretion of

matrix-degrading enzymes and production of cytokines and chemokines that promote immune cell differentiation and survival. However, the cause of the activation of FLS in RA is unknown and it is unclear whether this activation leads to or is a consequence of the disease. Defining the contribution of FLS to the heritability of RA will provide essential insights into this question.

For the first time, we have comprehensively mapped RA genetic risk variants to active regulatory DNA elements in FLS. We generated multidimensional epigenetic data in primary FLS, isolated from patients, to create a detailed outline of their chromatin landscape. We conducted genetic fine-mapping of RA loci by computing sets of credible single nucleotide polymorphisms (SNPs) driving GWAS signals. We integrated the credible SNP sets and chromatin datasets to provide evidence that RA risk variants can be functionally relevant in FLS. We used chromatin conformation data to determine enhancer–promoter interactions between risk variants in non-coding DNA regulatory regions of FLS and their target genes. Furthermore, we assessed the influence of the pro-inflammatory cytokine tumour necrosis factor (TNF) on these interactions, chromatin accessibility and gene expression in FLS. We combined FLS data with published data of human tissues and cells<sup>4,11,12</sup> to identify putative causal SNPs, enhancers, genes and cell types for RA risk loci. Finally, we functionally verified enhancer-promoter interactions by CRISPR-Cas technology and showed transcriptional effects of fine-mapped risk variants in FLS samples from RA patients.

## **Results**

### **Integration of epigenetic datasets to define the chromatin landscape of FLS**

As a first step in our analysis, we generated diverse epigenetic and transcriptomic datasets from our primary FLS samples (Supplementary Table 1): chromatin immunoprecipitation

sequencing (ChIP-seq) for six histone marks (H3K4me3, H3K4me1, H3K27me3, H3K36me3, H3K27ac, H3K9me3), Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq), cap analysis gene expression sequencing (CAGE-seq), chromatin conformation analysis (HiC, Capture HiC) and RNA sequencing (RNA-seq) (Supplementary Table 2, quality control metrics in Supplementary Dataset 1, details in Online methods). We used these datasets to annotate open chromatin and to assign 18 pre-trained chromatin states to the genome of FLS using chromHMM<sup>13</sup>. We identified Topologically Associating Domains (TADs) and determined significant chromatin interactions. Finally, we incorporated RNA-seq data from FLS. These analyses provided a comprehensive annotation of the epigenome and transcriptome of FLS (Fig. 1 a, b).

We cross-validated the individual datasets to confirm the quality of the generated FLS data. The open chromatin regions as identified with ATAC-seq showed high enrichment of promoter chromatin states (Transcription Start Sites [TSS]) and active enhancers identified by chromHMM (Fig. 2a). In Capture Hi-C (CHiC), interactions of the selected prey fragments containing previously-reported lead SNPs at RA loci<sup>2</sup> (details in Online Methods) were enriched for promoters (TSS), sites of transcription and enhancers (as defined by chromHMM) (Fig. 2b). At TAD boundaries, transcription and promoter states defined by chromHMM were enriched (Fig. 2c). Basal gene expression as measured by RNA-seq was highest in active TSS (Fig. 2d). Taken together, these analyses validated that we accurately captured chromatin states and chromatin interactions in FLS and that we have generated a comprehensive epigenetic and transcriptomic map of FLS genomes.

## **TNF induces changes in chromatin organization that correspond to altered gene expression in stimulated FLS**

To explore the effect of a pro-inflammatory environment on the chromatin landscape of FLS, we performed Hi-C, CHiC, ATAC-seq and RNA-seq experiments in FLS with and without stimulation with TNF (Supplementary Table 2).

We first computed changes in A/B compartments, which are large, cell-type specific organisational units of the genome, associated with chromatin activity (A = open chromatin, B = closed chromatin)<sup>14</sup>. 94.8% of A and 95.7% of B compartments were consistent between basal and stimulated FLS. Small changes in A/B compartments after stimulation are expected, as A/B compartments infer chromatin activity at DNA segments in low resolution. One of the genomic regions that changed from an inactive (B) to an active (A) compartment on TNF stimulation contains RA associated variants that interact with the *TNFAIP3* gene.

We then explored the influence of TNF on the organization of TADs in FLS. Genes within the same TAD tend to be co-regulated and gene promoters and enhancers often interact within the same TAD<sup>15</sup>. Changes in TAD boundaries, as found in cancer, can induce major changes in gene expression<sup>15</sup>. We compared TADs between basal and stimulated FLS with TADCompare<sup>16</sup>. TADCompare classifies TADs as either non-differential or differential based on changes in position or strength of TAD boundaries. Between our conditions, we identified an average of 4,116 TAD boundaries in FLS samples. Whilst the vast majority (3259, 79.2% ) of TAD boundaries were unchanged between basal and stimulatory conditions, 20.8% of differential TAD boundaries exhibited change in position (9.15 % boundary change complex, split, merge or shifted) or strength (11.66 % differential boundary magnitude) (Fig. 3a), indicating that the FLS genome exhibits changes in 3D structure upon TNF stimulation.

By analysing CHiC data (details in Online Methods), we observed around 800 quantitatively differentially interacting regions between basal and stimulated FLS. The intensity of the differential interactions between the regions correlated with the fold-change of expression of the interacting genes (Fig. 3b). Notably, interaction strength increased after stimulation for genes with differential expression, irrespective of whether expression increased or decreased after stimulation, thereby suggesting that chromatin interactions influence activating and repressive TNF transcriptional responses in FLS (Fig. 3b).

To further explore the regulation of gene transcription after TNF stimulation, we focused on CHiC baits and prey that exhibited increased interaction strength with regulated genes after stimulation. We overlapped these regions with the measurements of open chromatin peaks (ATAC-seq) in stimulated cells. We then used Hypergeometric Optimization of Motif EnRichment (HOMER) to detect known transcription factor binding sites (TFBS) or DNA motifs with high similarity to known TFBS, that were overrepresented at the sites with open chromatin, increased chromatin interactions and differential gene expression compared to random background sequences (chosen by HOMER) (Fig. 3c).

Enrichment analysis of known TFBS in open chromatin identified TPA response elements (TREs; TGA(G/C)TCA) as the most enriched motif in the data sets with increased as well as decreased gene expression (Supplementary Dataset 2). TPA response elements serve as canonical binding sites for the subunits of the Activator Protein-1 (AP-1) transcription factor that is closely linked to the pathogenesis of RA<sup>17</sup>. The different subunits of the AP-1 family form homo- and heterodimeric transcription factor complexes with distinct activating and repressing functions<sup>18</sup>. Open chromatin sites with increased CHiC interactions, but decreased gene expression in stimulated FLS were additionally enriched for BACH2 (broad complex-

tramtrack-bric a brac and Cap'n'collar homology 2) binding sites (Supplementary Dataset 2).

Like AP-1, BACH2 belongs to the basic region leucine zipper (bZIP) family, but has a slightly different DNA sequence binding site (TGCTGAGTCA) and has a bric-a-brac-tramtrack-broad-complex (BTB) domain, which specifically interacts with co-repressors to repress transcription<sup>19</sup>. BACH2 is a highly conserved repressor with a central function in terminal differentiation, maturation and activity of B and T cells<sup>20</sup>. Intronic SNPs within the *BACH2* gene have been associated with the risk of different immune-mediated diseases, including RA<sup>21,22</sup>. Our data suggest that in addition to regulating immune cell functions, BACH2 may play a notable role in regulating the TNF response of FLS in RA.

De novo DNA motif discovery in the dataset with decreased levels of gene expression after TNF stimulation was enriched for two DNA motifs (DNA motifs 2 and 3) with high similarity to binding sites for several homeobox and forkhead box proteins (Fig. 3d). Together, they were present in 5.5% of the gene sites (motif 2: 3.7%, background 0.2%; motif 3: 1.8%, background 0%; p-value =  $10^{-12}$ ). These results indicated a potential role for these developmental transcription factors in transcriptional repression of genes in TNF stimulated FLS. Since some of these transcription factors are exclusively expressed in FLS at distal joint locations (HOXA13, HOXD13)<sup>8</sup>, this suggests that TNF responses of FLS may be different at specific joint locations.

In summary, by combining ChIP, ATAC-seq and RNA-seq analyses we confirmed the activating and repressive actions of AP-1 in regulating the TNF response of FLS and we suggest that developmental transcription factors can serve as potential novel repressors of transcriptional response to TNF in FLS.

## **FLS and immune cells are drivers of RA heritability**

We then used the generated knowledge on regulatory DNA elements in FLS to quantify the heritability of RA that can be attributed to active regulatory DNA elements in FLS. We computed the partitioned heritability<sup>23</sup> in FLS and other cell types (HLA regions excluded; details in Methods section). Epigenetic data for non-FLS cell types were acquired from published datasets<sup>11</sup>. We defined active regulatory elements of the genome as the union of H3K4me1, H3K4me3 and H3K27ac peaks, as these histone modifications are associated with transcriptional activity and enhancer/promoter elements. With this approach, we estimated that 12%-24% of the non-HLA RA heritability can be attributed to the active DNA regulatory elements in FLS samples (Fig. 4a).

We then considered RA risk loci attaining genome-wide significance ( $p < 5 \times 10^{-8}$ ) in the European ancestry component of the largest published trans-ethnic RA GWAS meta-analysis (HLA regions excluded)<sup>2</sup>. Where lead SNPs at genomic loci mapped within 1 Mb of each other, the loci were merged (Supplementary Table 3). Using approximate conditional analyses implemented in GCTA<sup>24</sup>, we identified 73 distinct signals of association with RA at locus-wide significance ( $p < 10^{-5}$ ), with each signal being potentially driven by different underlying causal variants (Supplementary Table 4). For each signal, we performed fine-mapping to derive credible SNP sets that together account for  $\geq 99\%$  of the posterior probability of causality for the RA association. Across all 73 signals, the RA credible sets included a total of 8,787 variants, of which 2,654 variants had posterior probability of causality  $> 0.01\%$  (Supplementary Table 5). We then overlapped these 2,654 RA credible SNPs with the FLS epigenome and identified 274 SNPs mapping to active DNA regulatory elements in FLS (Fig. 4b). We also calculated the total posterior probability found within active DNA regulatory elements across the credible SNP sets for 111 primary cell types and

tissues, whose epigenomes were published by the Roadmap Epigenomics Mapping Consortium<sup>11</sup> (Fig 4b). As expected, several credible sets exhibited high posterior probability in active DNA regulatory elements from B and T cells (n=35), of which some (n=14) also overlapped active DNA regulatory regions in FLS (Fig 4b, Supplementary Table 5 column R). Intriguingly, we identified several credible sets that were active in FLS only, but not in B and T cells (n=9; Fig. 4b, Supplementary Table 5 column R). In summary, these analyses showed that both immune cells and FLS mediate the effects of association signal at risk loci and contribute notably to the heritability of RA.

### **TNF-induced alterations in 3D chromatin structure assign additional RA risk loci to FLS**

To assign putative target genes in FLS at RA risk loci, we identified significant CHiC interactions between the region containing a credible SNP set (CHiC baits, see details in Methods) and a gene promoter. We defined gene promoters by downloading all transcripts from Ensembl (version 98) and assigning a 1000 base pair window directly upstream of each transcript as a promoter. In total, we determined 220,000 promoters for 57,602 genes, including non-coding RNA. Across RA risk loci, gene target assignments yielded a total of 228 and 227 interacting, expressed FLS target genes in basal and TNF stimulated conditions, respectively, with 188 gene targets shared between the conditions (Supplementary Table 5 columns W and X).

Since TADs have been shown to define probable limits of gene regulation, we overlapped TADs with RA credible sets. We observed that each credible SNP set is usually found within one or two adjacent TADs in unstimulated and stimulated cells. We then examined the genes within the TADs containing our credible SNPs in basal and stimulated FLS. The alterations

in TAD boundaries between basal and stimulated FLS led to associated TADs overlapping different genes. Genes found within stimulation-specific TADs included *TNFAIP3*, *JAZF1*, *ZFP36L1*, *INFGR1* and *LBH*. Several of these genes, including *TNFAIP3*, *JAZF1*, *IFNGR1* and *LBH* also showed differential gene expression between basal and stimulated states (Table 1, Supplementary Table 5 columns AH and AI).

Furthermore, we observed a change in chromatin interactions as defined by CHiC in stimulated FLS at 17 RA risk loci, which were linked to 35 genes (Table 2). RNA-seq showed that the expression of 17 of the 35 genes was increased upon TNF stimulation in FLS (FDR < 0.05), 13 of which had a log<sub>2</sub> fold change of > 0.5 (e.g. *TRAF1*, *TNFAIP3*, *IFNAR2*) (Table 2, Supplementary Figure 1). Nine of the 35 genes were downregulated after TNF in FLS (FDR < 0.05), two of them with a log<sub>2</sub> fold change of > -0.5 (*RBPJ* and *RNF41*) (Table 2).

This correlated change in chromatin structure, interaction strength of RA implicated regions and gene expression upon stimulation, demonstrates how these loci are dynamic and active in FLS, and suggests that RA associated variants could thereby affect the transcriptional response to TNF in FLS.

### **Epigenetic annotation of the fine-mapped SNPs in immune cells and FLS refines the putative causal credible set SNPs for more than 30% of the RA risk loci**

Based on our genetic fine-mapping analysis, we identified three categories of RA risk loci. First, well characterised loci (“category 1”, n=19), where the credible set included ten or fewer SNPs or ≤ 3 SNPs contributing >80% of the posterior probability of causality. Second, loci with a localised signal (“category 2”, n=26), where the credible set included ≤ 20 SNPs

with similar low posterior probabilities (0.1 - 0.2) or the lead credible SNP accounted for >20% of the posterior probability. Third, poorly characterised loci (“category 3”, n=28), where genetic fine mapping was largely ineffective, resulting in large (>20) credible SNP sets with equally negligible posterior probabilities (<0.05) (Supplementary Table 5 column K, Table 3). Examples of category loci 1-3 are shown in Supplementary Figures 2-4.

By mapping the credible SNP sets to the annotated active promoters and enhancers in T cells, B cells and FLS, we further refined nine of 19 category 1 loci to  $\leq 3$  credible SNPs in active enhancers in either immune cells (n=5 loci), FLS (n=1 loci) or both (n=3 loci) (Table 3, Supplementary Table 5 columns L, M, N). Similarly, we narrowed down the number of putative causal SNPs to  $\leq 3$  for 18 of the 26 category 2 loci, after mapping marks of chromatin activity to the credible set SNPs in immune cells (n=7 loci), FLS (n=3 loci) or both (n=8 loci) (Table 3, Supplementary Table 5 columns L, M, N).

Thus, by integrating genetic fine-mapping with functional chromatin annotation in immune cells and FLS, we identified 27 loci (37%) that harbour  $\leq 3$  putative causal RA risk variants having high posterior probabilities and mapping to cell type-specific active enhancers.

### **Integrative analysis of genetic, expression and epigenetic datasets links putative causal genes and cell types**

We then used our genetic fine-mapping and epigenetic datasets to determine the cell types in which credible SNPs at RA loci are active (FLS, immune cells, both, neither), the effector genes (proximal and interacting in FLS/immune cell types) and their expression in relevant cell types.

In total, 9 of the 73 loci were assigned exclusively to FLS, with 2 further loci assigned to FLS and B cells, and 12 to all three analysed cells types based on SNPs in cell-type specific enhancers (Table 4, Supplementary Table 5 column based on O, P, Q, labelled in column R). Credible SNPs in category 1 and 2 loci found predominantly in FLS active enhancers implicated genes including *GRHL2*, *MYCBP* and *RUNX1*. FLS assigned genes that were associated with category 3 risk loci, which showed negligible posterior probability (< 2%) in immune enhancer SNPs, included *SPRED2*, *RCAN1*, *CDK6* and *RBPJ* (Table 4). Notably, the 24 credible SNPs in the *RBPJ* locus and the 41 credible SNPs in the *CDK6* locus were reduced to just six and three SNPs, respectively, mapping to FLS specific enhancers. The *RBPJ* SNPs were localized in FLS specific enhancers, with none found in T or B cells (Supplementary Table 5, rs11933540). This indicated that the putative causal SNPs in the *RBPJ* locus might specifically affect the function of FLS in RA.

We then integrated the credible set SNPs with our previously established ChIP dataset from B cell (GM12878) and T cell (Jurkat) lines<sup>12,25</sup>. We found that the loci assigned to immune cell types associated with genes that are vital in T and B cell-specific activities (Table 5, Supplementary Table 5 columns AA to AD). Genes in category 1 and 2 loci, which associated with active immune cells enhancer regions included *CTLA4*, *IL2RA* and *GATA3* for T cells and *BLK* for B cells (Table 5, Supplementary Table 5 columns AA to AD). Of note, the *ANKRD55/IL6ST* locus (rs7731626 in Supplementary Table 5) had a single SNP in the credible list, an eQTL with both *ANKRD55* and *IL6ST*<sup>26</sup> confined to an enhancer exclusive to T cells in our analysis. Immune cell assigned genes from category 3, where credible SNPs in immune enhancers accounted for >30% of the posterior probability, but had negligible posterior probability (< 5%) in FLS enhancers included *STAT4*, *CXCR5*, *CD28* and *MYC*.

These analyses highlighted a number of SNP-enhancer-gene combinations that could be assigned to an immune cell or fibroblast-driven risk of developing RA. We were able to assign >60% of the non-HLA RA loci with a putative causal cell type (FLS, B cells, T cells) and putative causal gene (column R not 'none'). Compared to previous gene assignment results<sup>2</sup>, our method provides empirical evidence for an additional 104 RA associated genes at the 73 European risk loci.

### **TNF stimulation induces major changes in chromatin organization of the TNFAIP3 and IFNGR2 risk loci with concomitant effects in the expression of interacting genes in FLS**

Some of the risk loci emerged as particularly interesting in FLS, exemplifying how stimulation induced changes in chromatin conformation and gene expression can affect RA risk in FLS.

The intergenic region on chromosome 6q23 between *OLIG3* and *TNFAIP3*, which contains eight credible SNPs (rs17264332 in Supplementary Table 5), was dynamically linked to the *TNFAIP3* gene through DNA activity, chromatin interactions and gene expression. The organization of this genomic region changed from a closed, inactive (compartment B) to an open, active chromatin conformation (compartment A) upon TNF-stimulation of FLS (Fig. 5a), and TAD boundary strength increased in TNF-stimulated FLS (Table 3). These substantial alterations to the chromatin organization coincided with a strong increase in the expression of the interacting *TNFAIP3* gene in FLS (Fig. 5a, Table 2, Supplementary Figure 1).

Similarly, we demonstrated stimulation induced changes in chromatin activity in the *IFNGR2* region (rs73194058, Supplementary Table 5) in FLS. Our CHiC analysis showed that the credible set SNPs in this region interacted with several genes relevant to the interferon (IFN) pathway, such as *IFNAR2*, *IL10RB*, *IFNAR1* and *IFNGR2* (Supplementary Table 5 columns W to Z and Fig. 5b). TNF stimulation of FLS induced dynamic changes in chromatin interactions at this locus as assessed by CHiC and increased the expression of *IFNAR2*, *IFNAR1*, and *IFNGR2* (Supplementary Figure 1, Table 2, Fig. 5b). Additionally, chromatin activity in the region of *IFNAR2* changed (from inactive B to active A compartment) in stimulated FLS (Fig. 5b).

IFN pathways are strongly associated with the pathogenesis of RA and IFN-responsive genes are induced in FLS upon stimulation with TNF<sup>27</sup>. The *TNFAIP3/IFNGR1* region on chromosome 6 and the *IFNAR1/IFNGR2* region on chromosome 21 interacted with genes encoding five subunits of the IFN I/III receptors in FLS (Fig. 5 a,b), suggesting a close genetic link between FLS function and IFN response in RA.

### **Genes linked to RA risk SNPs in FLS are functionally interlinked and regulate FLS-relevant RA functions**

To predict biological processes influenced by potential transcriptional effects of risk variants active in FLS, we conducted analyses to predict protein-protein interaction, pathway enrichment and functional annotation clustering. For these analyses, we included all target genes of RA risk loci that were assigned to FLS as a causal cell type (“All” and/or “FLS” in column R of Supplementary Table 5).

We found significantly enriched protein-protein interactions for the genes in the loci active in FLS by using STRING protein-protein interaction networks (PPI enrichment p-value:  $<1.0e-016$ ; Fig. 6a) and identified additional functional connections between the assigned genes by literature search. For instance, the transcription factor ZFP36L (rs1950897, Supplementary Table 5) negatively regulates the expression of CDK6<sup>28</sup> by binding to the 3'UTR region of the *CDK6* gene, which contains the credible set SNPs at this locus (rs4272, Supplementary Table 5). CDK6 in turn interferes with DNA binding of Runx1<sup>29</sup> (rs8133843, Supplementary Table 5). *ZFP36L*, *CDK6* and *RUNX1* were all assigned to FLS-active loci (Supplementary Table 5 column R), are functionally connected and regulate cell proliferation.

*CD40* (rs4239702 in Supplementary Table 5), *RBPJ* (rs11933540 in Supplementary Table 5) and *TRAF1* (rs10985070 in Supplementary Table 5) may constitute another genetically-influenced interlinked functional network in FLS. CD40 activation in FLS increased the expression of several cytokines relevant in RA, including VEGF and RANKL<sup>30,31</sup>. RBPJ (also known as CBF1), a regulatory transcription factor of the Notch signalling pathway, has been shown to repress the activation of CD40<sup>32</sup>. Similarly, TRAF1 can negatively regulate CD40 activity<sup>33</sup>.

Gene Ontology (GO) molecular function analysis (Fig. 6b) and functional annotation clustering of enriched pathways with the genes associated with credible set SNPs in FLS (Supplementary Table 5 column R and columns U-Z) revealed several clusters highly relevant to RA pathogenesis. These clusters included enrichment of genes involved in IFN response and viral defence (*IFNAR1*, *IFNAR2*, *CD40*, *IFNGR*, *C5*, *IL-10RB*) (Database for Annotation, Visualization and Integrated Discovery DAVID<sup>34,35</sup> enrichment score 1.22) as well as lipid metabolism and fatty acid synthesis (*FADS1-3*, *ACOX2*, *LCLAT1*, *JAZF1*,

*DAGLA*) (DAVID enrichment score 1.91). In addition, ‘cilium morphogenesis’ emerged as an enriched term (DAVID enrichment score 0.87) and several genes associated with RA risk SNPs in FLS were connected to the formation of the primary cilium (*C5orf30*, *GSN*, *TMEM138*, *TMEM216*, *CNTRL*, *INCENP*, *ACTR2*).

Overall, by integrating epigenetic and transcriptional data in FLS, we identified several functional relationships among RA risk variants and their target genes active in FLS. The multi-level effects of RA risk variants on key signalling pathways may contribute to the accumulated genetic risk in driving FLS activation and proliferation in RA.

### **RA risk SNPs in the *RBPJ* enhancer region confer joint-specific genetic effects in FLS**

Our epigenetic and functional analyses of the *RBPJ* locus identified *RBPJ* as a candidate causal and functional gene in FLS (rs11933540 in Supplementary Table 5). Mapping of the 24 credible SNPs to FLS enhancers in the *RBPJ* locus reduced the number of likely causal SNPs to six and rs874040 was identified as a strong candidate SNP for causality (Supplementary Table 5 rs11933540, Fig. 7a). To functionally establish that the rs874040-containing enhancer region can regulate the expression of *RBPJ*, we transduced FLS with lentiviral particles containing dCas9-VPR and two guide RNAs (g9 or g12) targeting the rs874040-containing enhancer region (Fig. 7a). FLS transduced with the activating dCas9-VPR and guide RNAs increased the expression of *RBPJ* compared to FLS transduced with the respective guide RNAs without dCas9-VPR (Fig. 7b). Even though the upregulation of *RBPJ* expression was modest (30%), which could be due to enhancer redundancy in this region, this experiment verified the regulation of *RBPJ* expression by the rs874040-containing enhancer region.

FLS homozygous for the risk allele of rs874040 exhibited lower expression of *RBPJ* mRNA compared to FLS with the wild-type variant. This effect was, however, present only in FLS from upper extremity joints, and not from lower extremity joints (Fig. 7c). It is known that *RBPJ* binds to the promoter of *HES1* and represses its transcription<sup>36</sup>. Accordingly, the expression of *HES1* was increased in FLS from patients homozygous for rs874040 in upper extremity joints (Fig. 7d). TNF stimulation significantly downregulated *RBPJ* mRNA expression in FLS (Table 2), and FLS from RA patients expressed less *RBPJ* than FLS from patients with arthralgia (Fig. 7e). These data indicate that genetic predisposition and a pro-inflammatory environment can affect *RBPJ* expression in FLS, which might lead to increased activation of the Notch signalling pathway via *HES1*.

To explain the joint-specific effect of rs874040, we explored the enhancer landscape and the chromatin interactions in different upper and lower extremity joints. CAGE-seq data showed that the enhancer activity within the *RBPJ* locus is higher in knee FLS compared to shoulder and hand FLS (Fig. 7e). CAGE-seq enhancer signals largely overlapped with ATAC-seq peaks, being more abundant in knee FLS than in shoulder or hand FLS (Fig. 7f). Shoulder FLS appeared to mainly use an upstream enhancer that interacted with the *RBPJ* risk locus (green boxes, Fig. 7e). Overlap of ATAC-seq and CAGE-seq analyses was weaker in hand FLS, but CAGE-seq data indicated that hand FLS used an enhancer within the risk locus (red box, Fig. 7e). Additionally, chromatin interactions within the locus were generally weaker in hand FLS (Fig. 7g). Knee FLS activated several enhancers (Figs. 7e, f) and exhibited strong chromatin interactions across the locus (Fig. 7g). We analysed DNA-binding motifs in the enhancer overlapping the RA risk locus spanning chr4:26090045-26090465 (hg19) (red box in Fig. 7e) by using the JASPAR2020 database<sup>37</sup>. This enhancer contained TFBS for different HOX transcription factors (HOXA6, HOXA7, HOXA10, HOXB2, HOXB6, HOXB7,

HOXB13, HOXD3, HOXD9, HOXD13), similar to the DNA motifs identified in open chromatin at repressed genes after TNF stimulation in FLS (Fig. 3d) and expressed in a joint-specific manner in FLS<sup>38</sup>.

Together these data suggest that joint-specific differences in chromatin interactions and enhancer usage could underlie the joint-specific effects of rs874040 on *RBPJ* expression in upper extremity joints. This illustrates that RA genetic risk can be different between the joints, thereby shaping a specific pattern of joint involvement in RA.

## Discussion

Deciphering the role of causal genetic variants underlying GWAS loci in RA, albeit challenging, provides an unbiased strategy to understand the core disease pathways and guide drug discovery<sup>2,39</sup>. Here we demonstrate that a significant proportion of the 73 European ancestry non-HLA RA risk loci contain disease-associated variants that are located within active regulatory DNA elements in FLS. Linking these DNA regions with target genes indicates genes and biological pathways that trigger RA susceptibility by stromal cell activation in the joint. Thus, we provide for the first time substantial evidence for an independent, causal role of FLS in RA genetic susceptibility and pathogenesis.

With our approach, we were able to assign RA risk loci to immune and/or stromal cells. Genes implicated in T cells, but not in FLS, showed a pattern of involvement in ‘canonical’ T-cell immunity, including *CTLA4*, *CD28*, *IL2RA*, and *GATA3*. Similarly, genes enriched in B cell specific enhancers were involved in B cell biology, including *IRF8*, *BLK*, and *TAB1*. The stromal activation observed in RA joints was clearly reflected in the predicted function of the identified FLS-specific regulatory variants, many of which were previously associated

with RA pathogenesis. For example, several genes linked to RA credible SNPs in FLS were implicated in cell proliferation and tumour development (e.g. *SPRED2*<sup>40</sup>, *GRHL2*<sup>41</sup>, *CDK6*<sup>42</sup>, and *ZFP36L*<sup>43</sup>).

Most notably, the risk loci on chromosomes 6 (*TNFAIP3/IFNGR1*) and 2 (*IFNAR/IFNGR2*) could critically impact the contribution of FLS to the development of RA. Chromatin interactions in these regions connected RA risk variants with several genes encoding the subunits of type I (*IFNAR1*, *IFNAR2*), type II (*IFNGR1*, *IFNGR2*) and type III (*IL-10RB*) interferon receptors in FLS. Furthermore, they tightly linked the IFN response to TNF stimulation by interconnection with the *TNFAIP3* gene, encoding the TNF signalling repressor A20, and by their strong reaction to TNF stimulation. All three types of interferons signal via the JAK-STAT signalling pathway, which, along with TNF, is one of the central therapeutic targets in RA. A type I interferon gene signature is detectable in up to two thirds of patients with RA<sup>44</sup> and it associates with an increased risk of developing RA as well as with therapeutic response to biological DMARDs like TNF inhibitors<sup>45,46</sup>. In FLS, TNF induces an extensive interferon gene response via secondary autocrine production of IFN $\beta$  and the activation of the IRF1-IFN $\beta$ -IFNAR-JAK-STAT1 axis<sup>47,48</sup>. Down syndrome (trisomy 21) leads to increased dosage of the IFN receptors encoded on chromosome 21, which results in a type I interferon gene signature with constant activation of interferon pathways in fibroblasts<sup>49</sup>. Notably, people with Down syndrome are at increased risk of developing erosive, inflammatory seronegative arthritis of their hands and wrists<sup>50</sup>. Together this strongly suggests a causal role for stromal activation of IFN pathways in the development of RA.

We showed that RA risk allele rs874040 is associated with reduced expression of *RBPJ* in FLS in a location specific manner. RBPJ, also called CBF1 or CSL, is a key transcriptional

regulator of the Notch signalling pathway<sup>51</sup>. In the absence of Notch signalling, RBPJ represses Notch target genes (e.g. HES1). Upon activation of Notch signalling, RBPJ binds to the intracellular domain of the Notch receptor and enhances Notch-dependent gene expression. Loss of RBPJ leads to activation of dermal fibroblasts and promotes their transformation into cancer associated fibroblasts (CAFs), which play a crucial role in tumour development and growth<sup>52,53</sup>. Activation of Notch signalling was shown in RA FLS and induced FLS proliferation<sup>54</sup>. Furthermore, Notch signalling is critical for shaping the synovial environment by guiding the development of THY1+ sublining FLS, a subset of FLS that is expanded in RA synovial tissues<sup>55</sup>. Constitutive lower levels of RBPJ in FLS from individuals carrying the RBPJ risk variant could favour synovial enrichment of THY1+ sublining FLS, which are considered critical for the development of RA. Joint-specific differences in the chromatin landscape in this locus exemplify how genetic risk could result in the specific patterns of joint involvement that typically occur in chronic inflammatory joint diseases. Additionally, joint-specific expression of HOX transcription factors<sup>8</sup>, for which we suggest a role in gene repression after TNF stimulation in FLS, could contribute to joint-specific differences in the susceptibility to RA.

Further pathways that we found enriched in genes targeted by RA credible SNPs were connected to lipid metabolism and the primary cilium. FADS1 and 2 have been implicated in the production of anti-inflammatory unsaturated fatty acids in LPS-treated macrophages, contributing to the resolution phase of LPS-driven inflammatory response in macrophages<sup>56</sup>. Changes in the lipid metabolism have been suggested in RA FLS<sup>57</sup>, but specific functional data does not exist so far. The primary cilium serves as a hub for several cell signalling pathways, e.g. Notch<sup>58</sup> and wnt signalling<sup>59</sup>. In FLS, it was shown that TNFR1 and TNFR2 localise to the cilium-pit<sup>60</sup>. The cilium connected proteins C5orf30 and GSN, that we found

interacting with RA risk variants in FLS, were previously shown to be negative regulators of arthritis in mice<sup>61,62</sup>. Another ciliary protein, SPAG16 was found to be a genetic risk factor for joint damage progression in RA patients, increasing the production of matrix-metalloproteinases in FLS<sup>63</sup>. Future studies are required to demonstrate how changes in lipid metabolism and primary cilium affect the function of FLS and influence RA pathogenesis.

Overall, our research significantly advances the knowledge about putative causal SNPs, enhancers, genes and cell types affected by genetic risk loci in RA. Our analysis can direct future studies to investigate pathways that are genetically affected in a cell type specific way. This will ultimately enable the connection of an individual's genetic risk with the causal pathways and cell types that drive disease, paving the way to stratified treatment decisions and precision medicine.

## References

1. Smolen, J.S., Aletaha, D. & McInnes, I.B. Rheumatoid arthritis. *Lancet* **388**, 2023-2038 (2016).
2. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-81 (2014).
3. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).
4. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun* **6**, 10069 (2015).
5. Thalayasingam, N. *et al.* CD4+ and B Lymphocyte Expression Quantitative Traits at Rheumatoid Arthritis Risk Loci in Patients With Untreated Early Arthritis: Implications for Causal Gene Identification. *Arthritis Rheumatol* **70**, 361-370 (2018).
6. Croft, A.P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246-251 (2019).
7. Mizoguchi, F. *et al.* Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat Commun* **9**, 789 (2018).
8. Frank-Bertoncelj, M. *et al.* Epigenetically-driven anatomical diversity of synovial fibroblasts guides joint-specific fibroblast functions. *Nat Commun* **8**, 14852 (2017).
9. Ai, R. *et al.* Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. *Nat Commun* **7**, 11849 (2016).
10. Ospelt, C. Synovial fibroblasts in 2017. *RMD Open* **3**, e000471 (2017).
11. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
12. McGovern, A. *et al.* Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. *Genome Biol* **17**, 212 (2016).
13. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
14. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
15. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci Adv* **5**, eaaw1668 (2019).
16. Cresswell, K.G. & Dozmorov, M.G. TADCompare: An R Package for Differential and Temporal Analysis of Topologically Associated Domains. *Front Genet* **11**, 158 (2020).
17. Hannemann, N. *et al.* Transcription factor Fra-1 targets arginase-1 to enhance macrophage-mediated inflammation in arthritis. *J Clin Invest* **129**, 2669-2684 (2019).
18. Shaulian, E. & Karin, M. AP-1 in cell proliferation and survival. *Oncogene* **20**, 2390-400 (2001).
19. Tanaka, H. *et al.* Epigenetic Regulation of the Blimp-1 Gene (Prdm1) in B Cells Involves Bach2 and Histone Deacetylase 3. *J Biol Chem* **291**, 6316-30 (2016).
20. Igarashi, K., Kurosaki, T. & Roychoudhuri, R. BACH transcription factors in innate and adaptive immunity. *Nat Rev Immunol* **17**, 437-450 (2017).
21. McAllister, K. *et al.* Identification of BACH2 and RAD51B as rheumatoid arthritis susceptibility loci in a meta-analysis of genome-wide data. *Arthritis Rheum* **65**, 3058-62 (2013).

22. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-25 (2010).
23. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
24. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
25. Martin, P. *et al.* Chromatin interactions reveal novel gene targets for drug repositioning in rheumatic diseases. *Ann Rheum Dis* **78**, 1127-1134 (2019).
26. Kundu, K. *et al.* Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for twelve immune-mediated diseases. 2020.01.15.907436 (2020).
27. Burja, B., Mertelj, T. & Frank-Bertoncelj, M. Hi-JAKi-ng Synovial Fibroblasts in Inflammatory Arthritis With JAK Inhibitors. *Front Med (Lausanne)* **7**, 124 (2020).
28. Chen, M.T. *et al.* ZFP36L1 promotes monocyte/macrophage differentiation by repressing CDK6. *Sci Rep* **5**, 16229 (2015).
29. Fujimoto, T., Anderson, K., Jacobsen, S.E., Nishikawa, S.I. & Nerlov, C. Cdk6 blocks myeloid differentiation by interfering with Runx1 DNA binding and Runx1-C/EBPalpha interaction. *EMBO J* **26**, 2361-70 (2007).
30. Cho, C.S. *et al.* CD40 engagement on synovial fibroblast up-regulates production of vascular endothelial growth factor. *J Immunol* **164**, 5055-61 (2000).
31. Lee, H.Y. *et al.* CD40 ligation of rheumatoid synovial fibroblasts regulates RANKL-mediated osteoclastogenesis: evidence of NF-kappaB-dependent, CD40-mediated bone destruction in rheumatoid arthritis. *Arthritis Rheum* **54**, 1747-58 (2006).
32. Mann, J., Oakley, F., Johnson, P.W. & Mann, D.A. CD40 induces interleukin-6 gene transcription in dendritic cells: regulation by TRAF2, AP-1, NF-kappa B, AND CBF1. *J Biol Chem* **277**, 17125-38 (2002).
33. Fotin-Mleczek, M. *et al.* Tumor necrosis factor receptor-associated factor (TRAF) 1 regulates CD40-induced TRAF2-mediated NF-kappaB activation. *J Biol Chem* **279**, 677-85 (2004).
34. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
35. Huang da, W. *et al.* Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 11 (2009).
36. Lake, R.J., Tsai, P.F., Choi, I., Won, K.J. & Fan, H.Y. RBPJ, the major transcriptional effector of Notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. *PLoS Genet* **10**, e1004204 (2014).
37. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).
38. Frank-Bertoncelj M *et al.* The Long Noncoding RNA HOTTIP Regulates Cell Cycle and Inflammatory Response in Hand Synovial Fibroblasts. *Arthritis & Rheumatology* **71**(2019).
39. Yarwood, A., Eyre, S. & Worthington, J. Genetic susceptibility to rheumatoid arthritis and its implications for novel drug discovery. *Expert Opin Drug Discov* **11**, 805-13 (2016).
40. Kachroo, N., Valencia, T., Warren, A.Y. & Gnanapragasam, V.J. Evidence for downregulation of the negative regulator SPRED2 in clinical prostate cancer. *Br J Cancer* **108**, 597-601 (2013).
41. Frisch, S.M., Farris, J.C. & Pifer, P.M. Roles of Grainyhead-like transcription factors in cancer. *Oncogene* **36**, 6067-6073 (2017).

42. Kollmann, K. *et al.* A kinase-independent function of CDK6 links the cell cycle to tumor angiogenesis. *Cancer Cell* **24**, 167-81 (2013).
43. Suk, F.M. *et al.* ZFP36L1 and ZFP36L2 inhibit cell proliferation in a cyclin D-dependent and p53-independent manner. *Sci Rep* **8**, 2742 (2018).
44. Rodriguez-Carrio, J., Lopez, P. & Suarez, A. Type I IFNs as biomarkers in rheumatoid arthritis: towards disease profiling and personalized medicine. *Clin Sci (Lond)* **128**, 449-64 (2015).
45. Wright, H.L., Thomas, H.B., Moots, R.J. & Edwards, S.W. Interferon gene expression signature in rheumatoid arthritis neutrophils correlates with a good response to TNFi therapy. *Rheumatology (Oxford)* **54**, 188-93 (2015).
46. Mavragani, C.P., La, D.T., Stohl, W. & Crow, M.K. Association of the response to tumor necrosis factor antagonists with plasma type I interferon activity and interferon-beta/alpha ratios in rheumatoid arthritis patients: a post hoc analysis of a predominantly Hispanic cohort. *Arthritis Rheum* **62**, 392-401 (2010).
47. Rosengren, S., Corr, M., Firestein, G.S. & Boyle, D.L. The JAK inhibitor CP-690,550 (tofacitinib) inhibits TNF-induced chemokine expression in fibroblast-like synoviocytes: autocrine role of type I interferon. *Ann Rheum Dis* **71**, 440-7 (2012).
48. Karonitsch, T. *et al.* mTOR Senses Environmental Cues to Shape the Fibroblast-like Synoviocyte Response to Inflammation. *Cell Rep* **23**, 2157-2167 (2018).
49. Sullivan, K.D. *et al.* Trisomy 21 consistently activates the interferon response. *Elife* **5**(2016).
50. Foley, C.M., Deely, D.A., MacDermott, E.J. & Killeen, O.G. Arthropathy of Down syndrome: an under-diagnosed inflammatory joint disease that warrants a name change. *RMD Open* **5**, e000890 (2019).
51. Giaimo, B.D., Oswald, F. & Borggreffe, T. Dynamic chromatin regulation at Notch target genes. *Transcription* **8**, 61-66 (2017).
52. Goruppi, S. *et al.* The ULK3 Kinase Is Critical for Convergent Control of Cancer-Associated Fibroblast Activation by CSL and GLI. *Cell Rep* **20**, 2468-2479 (2017).
53. Procopio, M.G. *et al.* Combined CSL and p53 downregulation promotes cancer-associated fibroblast activation. *Nat Cell Biol* **17**, 1193-204 (2015).
54. Nakazawa, M. *et al.* Role of Notch-1 intracellular domain in activation of rheumatoid synoviocytes. *Arthritis Rheum* **44**, 1545-54 (2001).
55. Wei, K. *et al.* Notch signalling drives synovial fibroblast identity and arthritis pathology. *Nature* (2020).
56. Oishi, Y. *et al.* SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism. *Cell Metab* **25**, 412-427 (2017).
57. Ahn, J.K. *et al.* GC/TOF-MS-based metabolomic profiling in cultured fibroblast-like synoviocytes from rheumatoid arthritis. *Joint Bone Spine* **83**, 707-713 (2016).
58. Ezratty, E.J. *et al.* A role for the primary cilium in Notch signaling and epidermal differentiation during skin development. *Cell* **145**, 1129-41 (2011).
59. Pala, R., Alomari, N. & Nauli, S.M. Primary Cilium-Dependent Signaling Mechanisms. *Int J Mol Sci* **18**(2017).
60. Rattner, J.B., Sciore, P., Ou, Y., van der Hoorn, F.A. & Lo, I.K. Primary cilia in fibroblast-like type B synoviocytes lie within a cilium pit: a site of endocytosis. *Histol Histopathol* **25**, 865-75 (2010).
61. Aidinis, V. *et al.* Cytoskeletal rearrangements in synovial fibroblasts as a novel pathophysiological determinant of modeled rheumatoid arthritis. *PLoS Genet* **1**, e48 (2005).

62. Muthana, M. *et al.* C5orf30 is a negative regulator of tissue damage in rheumatoid arthritis. *Proc Natl Acad Sci U S A* **112**, 11618-23 (2015).
63. de Rooy, D.P. *et al.* Genetic Factors for the Severity of ACPA-negative Rheumatoid Arthritis in 2 Cohorts of Early Disease: A Genome-wide Study. *J Rheumatol* **42**, 1383-91 (2015).

### **Acknowledgements**

This work was supported by the Wellcome Trust (award reference 207491/Z/17/Z), Versus Arthritis (award references 21754 and 21348) and by the NIHR Manchester Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. CO and MFB were supported by the Swiss National Science Foundation (project 320030\_176061) and the Georg and Bertha Schwyzer-Winiker Foundation.

### **Contributions**

X.G., M.F.B., A.P.M., P.M., S.E. and C.O. conceived and designed the study. M.F.B., K.K., A.M., T.K., B.B., G.O. and C.O. performed the experiments. X.G., M.H., R.M., A.P.M., P.M. and S.E. conducted fine mapping and bioinformatics analyses. M.M., A.F., C.D.B and O.D. collected patient samples and clinical data. X.G., M.F.B., A.P.M., P.M., S.E. and C.O. interpreted the data. X.G., M.F.B., A.P.M., S.E. and C.O. wrote the manuscript. All the authors critically discussed the manuscript and approved the submitted version.

## **Online Methods**

### **Patients and cell culture**

Synovial tissues were obtained from OA patients (n=9, 2male/ 8 female, mean age 68, range 54-88) and RA patients (n=30, 8 male/22 female, mean age 66, range 44-78) undergoing joint replacement surgery at the Schulthess Clinic Zurich, Switzerland. Patient's characteristics are described in Supplementary Table 1. RA patients fulfilled the 2010 ACR/EULAR (American College of Rheumatology/European League Against Rheumatism) criteria for the classification of RA<sup>1</sup>. Samples from patients with joint pain without inflammation or cartilage destruction (healthy, 3 male/3 female, mean age 39, range 23-49) were obtained from the Queen Elizabeth Hospital in Birmingham, UK. The studies were approved by the local ethic committees of the University Hospital Zurich, Switzerland and the University of Birmingham, UK. Informed consent was obtained from all patients. Synovial tissues were digested with dispase (37 °C, 1 h) and FLS were cultured in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) supplemented with 10% fetal calf serum (FCS), 50 U ml<sup>-1</sup> penicillin/streptomycin, 2 mM L-glutamine, 10 mM HEPES and 0.2% amphotericin B (all from Life Technologies). Purity of FLS cultures was confirmed by flow cytometry showing the presence of the fibroblast surface marker CD90 (Thy-1) and the absence of leukocytes (CD45), macrophages (CD14; CD68), T lymphocytes (CD3), B lymphocytes (CD19) and endothelial cells (CD31). Cell cultures were negative for mycoplasma contamination as assessed by MycoAlert mycoplasma detection kit (Lonza). FLS were used between passages 4-8. Information on the assays performed on each sample is given in Supplementary Table 2.

### **RNA sequencing**

RNA sequencing data from unstimulated samples (Supplementary Table 2) was retrieved from the European Nucleotide Archive (ENA) with the primary accession code PRJEB14422. A detailed description of sample preparation and sequencing procedures is given in Frank-Bertoncelj et al.<sup>2</sup>. For RNA sequencing of TNF stimulated FLS, cultured FLS were treated with 10 ng/ml human recombinant TNF (Roche) for 24 h or were left untreated. Total RNA was isolated using the RNeasy Mini kit (Qiagen) including on-column DNAase I digestion. Part of the libraries (n=12) were prepared using the NEB Next Ultra Directional RNAseq protocol with ribosomal depletion and were sequenced using Illumina HiSeq4000 with 75bp paired end reads. The additional libraries (n=20) were generated using the Illumina TruSeq Stranded total RNA protocol with the TruSeq Stranded total RNA Sample Preparation Kit and were sequenced using Illumina Novaseq 6000. All Fastq-files were mapped to hg19 and sequence reads assigned to genomic features using STAR<sup>3</sup> and featureCounts<sup>4</sup>, respectively. We used svaseq R<sup>5</sup> package (version 3.36.0) to find and remove hidden batch effects. Differential gene expression analysis was performed with DESeq2<sup>6</sup> R package (version 1.28.1) according to standard protocol.

### **ChIP Sequencing**

ChIP DNaseq was performed on the Illumina HiSeq 2500 (50 bp, single end) as described in Frank-Bertoncelj et al.<sup>2</sup>. Briefly, ChIP assays were performed using 1 million FLS (Supplementary Table 2) per IP and the following antibodies (all Diagenode): H3K4me3 (0.5 µg, C15411003), H3K27me3 (1 µg, C15410195), H3K27ac (1 µg, C15410196), H3K4me1 (1 µg, C15410194), H3K36me3 (1 µg, C15410192) and H3K9me3 (1 µg, C15410193). The reads were mapped to the GRCh38 human genome reference using Bowtie2<sup>7</sup> with default settings. The mapped alignment files were further QC'ed with Picard Tools (Broad-Institute, available at: <http://broadinstitute.github.io/picard/>) to check for

duplication rates, unique mapping reads and library complexity. The duplicated reads and non-unique mapping reads were then removed prior to analysis with Picard Tools.

### **ChromHMM chromatin state inference**

The de-duplicated, uniquely mapping reads of the ChIP sequencing were binarized with the BinarizeBam script provided by the chromHMM software<sup>8</sup>. This script splits the genome into 200bp bins, and compares the coverage of the alignment file at each bin with the input sequence file to determine if any histone modification is present in the bin (1=yes, 0=no). The pre-trained 18 state chromHMM model based on the six histone marks was applied to the binarized bed files, using the MakeSegmentation script provided and the model parameters downloaded from the Roadmap Epigenomics web portal. The methods employed by Ernst et al.<sup>8</sup> were replicated where possible from the data processing stages to the chromatin state inference.

### **ATAC sequencing**

Cultured RA FLS were stimulated with 10 ng/ml TNF for 24 h or were left untreated (Supplementary Table 2). 50,000 cells were prepared according to the protocol by Buenrostro et al<sup>9</sup>. ATAC-seq libraries were sequenced on Illumina HiSeq 4000 with 75 bp paired end reads. The reads were QC'd with FastQC for read quality, and the Nextera-transposase adaptors were trimmed with cutadapt<sup>10</sup>. The reads were aligned with Bowtie 2 to the GRCh38 human reference. PCR duplicates were identified and removed by Picard Tools prior to peak calling using MACS2<sup>11</sup>. Both broad and narrow peaks were called as ATAC-seq can have properties of both.

### **HiC and capture HiC**

Cultured human FLS from RA patients were treated with 10 ng/ml TNF for 24 h or were left untreated (Supplementary Table 2). Cells ( $1-3 \times 10^7$  per condition), were scratched in 10ml DMEM, spun down, suspended in 35ml DMEM and fixed (2% formaldehyde in DMEM, 10min, RT, with mixing on a rocker). The reaction was quenched with cold 0.125M glycine. Cells were incubated at RT for 5 min, followed by 15 min incubation on ice and centrifugation (1500 rpm, 10 min, 4 °C). Pellets were suspended and washed in cold PBS (1500 rpm, 10 min, 4 °C). Washed pellets were snap frozen and stored at -80 °C. Hi-C libraries from RA FLS samples, were generated as previously described<sup>12</sup>. They were sequenced on Illumina HiSeq 4000 with 75 bp paired end reads. The reads were processed using the HiC Pro pipeline<sup>13</sup>, the correlation between samples were calculated with HiCrep<sup>14</sup>. TADs were called with TADCompare<sup>15</sup>. The regions targeted by the capture HiC (CHiC) were generated based on the LD regions of the lead disease associated SNPs for RA, Juvenile Idiopathic Arthritis (JIA), Psoriatic arthritis (PsA) and Psoriasis (Ps). This resulted in a total of 242 distinct risk variants. 120bp capture baits were designed for all HindIII digestion fragments overlapping these regions as previously described in Martin et al.<sup>16</sup>. Significant CHiC interactions were identified through the CHiCAGO pipeline<sup>17</sup>, where the suggested threshold of CHiCAGO score >5 was used. Differential interactions were identified with DESeq2, where the read counts of each interaction were treated similar to the gene count of RNA-seq.

### **Transcription factor binding site prediction**

We extracted differentially interacting regions from our CHiC data, where the strength in chromatin interaction (log-fold change of read counts between basal and stimulated) correlated with nearby genes. We overlapped these interaction regions (bait and prey fragments) with our ATAC-seq peaks. These ATAC seq peaks were standardised and re-centered to 200bp each.

We then used the findMotifsGenome.pl software from the HOMER suite<sup>18</sup> to identify significantly enriched motifs in these ATAC-peaks.

### **Partitioned Heritability**

We defined active chromatin regions of the genome for each FLS sample and publicly available Roadmap samples, based on the union of H3K27ac, H3K4me1, and H3K4me3 histone peaks. We used the partitioned heritability software from the LDSC<sup>19</sup> package to quantify the non-HLA RA heritability attributed to these active regions in each sample, based on the summary statistics from the Okada et al trans-ethnic meta-analysis<sup>20</sup>.

### **Derivation of RA credible set SNPs**

For each locus, we dissected distinct RA association signals using approximate conditioning implemented in GCTA<sup>21</sup>, based on: (i) European ancestry summary statistics from the Okada et al. trans-ethnic meta-analysis<sup>20</sup>; and (ii) a reference panel of European ancestry haplotypes from the 1000 Genomes Project to approximate linkage disequilibrium between SNPs. We identified index SNPs for each distinct signal, at a locus-wide significance threshold of  $p < 10^{-5}$ , using the --cojo-slc option. For each locus with multiple distinct signals, we derived the conditional association summary statistics for each distinct signal by conditioning out the effects of all other index SNPs at the locus using the --cojo-cond option.

For each distinct signal, we first calculated the posterior probability,  $\pi_j$ , that the  $j$ th variant is driving the association, given by

$$\pi_j = \frac{\Lambda_j}{\sum_k \Lambda_k},$$

where the summation is over all variants at the locus. In this expression,  $\Lambda_j$  is the approximate Bayes' factor<sup>22</sup> for the  $j$ th variant, given by

$$\Lambda_j = \sqrt{\frac{V_j}{V_j + \omega}} \exp\left[\frac{\omega \beta_j^2}{2V_j(V_j + \omega)}\right],$$

where  $\beta_j$  and  $V_j$  denote the estimated allelic effect and corresponding variance from the European ancestry component of Okada et al.<sup>20</sup>. In loci with multiple distinct signals of association, summary statistics were obtained from the approximate conditional analysis. In loci with a single association signal, summary statistics were obtained from unconditional analysis. The parameter  $\omega$  denotes the prior variance in allelic effects, taken here to be 0.04. The 99% credible set for each signal was then constructed by: (i) ranking all variants according to their Bayes' factor,  $\Lambda_j$ ; and (ii) including ranked variants until their cumulative posterior probability of driving the association attained or exceeded 0.99.

### **Pathway analysis and protein-protein interaction network**

The genes assigned to FLS (Supplementary Table 5, column R) and listed in Supplementary Table 5, columns U-Z were analysed by STRINGv11 (interactions settings to medium confidence levels)<sup>23</sup>, ToppFun on ToppGene Suite<sup>24</sup> and DAVID v6.8<sup>25,26</sup> with default settings.

### **Guide RNA design and cloning**

Guide (g)RNAs, targeting the putative upstream *RBPJ* enhancer (locus 23, Supplementary Table 5), were designed using the CRISPOR tool<sup>27</sup> in the DNA region chr4:26106475-26106675 (hg38) comprising 100bp upstream and 100bp downstream of the RA risk SNP rs874040. Complementary gRNA oligo pairs with 5' CACC (fwd) and 5'CAAA (rev) overhangs (Microsynth, 100 mM) were phosphorylated and annealed in a thermocycler (37°C, 30 min; 95°C 5 min, ramp down to 25°C at 5°C/min using T4PNK (NEB) and 10x T4 ligation buffer (NEB). LentiGuide-puro plasmid, a gift from Feng Zhang (Addgene  $\mu$ plasmid # 52963; <http://n2t.net/addgene:52963>; RRID:Addgene\_52963)<sup>28</sup>, was digested with FastDigest BbsI, Fast AP, 10x Fast Digest Buffer at 37°C, 30 min (Fermentas) followed by the ligation of the

annealed gRNA duplex o/n using Quick Ligase (NEB) and 2x Quick Ligation buffer (NEB). One Shot™ Stbl3™ Chemically Competent E. coli (ThermoFisher, C7373-03) were transformed with gRNA containing lentiGuide-Puro plasmids by heat-shocking (45 sec, 42°C) according to manufacturer's instructions. Plasmid DNA from selected colonies was isolated using QIAprep Spin Miniprep Kit (Qiagen) and Sanger sequenced to confirm the insertion and sequences of cloned gRNAs. To prepare gRNA-containing lentiviral particles, HEK293T cells were transfected with psPAX2, pMD2.G gRNA-containing plasmids, (total 10 µg plasmid DNA, mass ratio 2:1:4, respectively). psPAX2 (Addgene plasmid # 12260; <http://n2t.net/addgene:12260>; RRID: Addgene\_12260) and pMD2.G (Addgene plasmid # 12259; <http://n2t.net/addgene:12259>; RRID: Addgene\_12259) were a gift from Didier Trono. Viral particles were precipitated from the supernatants of transfected HEK293T (24h and 48h) using PEG-it™ Virus Precipitation Solution (5x) according to manufacturer's protocol (System Biosciences), resuspended in PBS, and stored at -70°C.

gRNAs	Sequence	5' overhang	5' overhang Complementary reverse strand	PAM
gRNA_9	5' GCCTTATCATGGCATATCACC '3	5'CACC	5'CAAA	TGG
gRNA_12	5' GCTAGAGCACGCAGCTTTTGC '3	5'CACC	5'CAAA	AGG

### Activation of enhancer regions with dCas9-VPR

FLS were transduced with Edit-R Lentiviral dCas9-VPR lentiviral particles (hEF1α promoter, Dharmacon). Edit-R Lentiviral dCas9-VPR is a CRISPR activation system, in which a nuclease-deactivated *S. pyogenes* Cas9 (dCas9) is fused to VP64, p65 and Rta transcriptional activators. Stable populations of dCas9-VPR FLS were blasticidin selected (7.5 µg/ml,

Horizon) and subsequently transduced with gRNA-containing lentiviral particles. Stable gRNA dCas9-VPR FLS were puromycin selected (5 µg/ml Sigma), lysed, RNA was isolated, followed by reverse transcription and SYBR Green real time PCR as described above. Gene expression was normalized to the average expression of *B2M* (Primer sequence Fwd 5' AAGCAGCATCATGGAGGTTTG'3, Rev 5' AAGCAAGCAAGCAGAATTTGGA'3) and *RPLP0* housekeeper genes. Transduction of dCas9-VPR without guide RNA had no effect on *RBPJ* expression.

### **Pyrosequencing**

DNA from FLS was isolated using the QIAamp DNA Blood kit (Qiagen). DNA regions containing rs874040 (*RBPJ*) were amplified by PCR (Primers: FWD: AGT GTG GAT TGT AGC AGA TAT GTC; REV: biotin-ACC AAG GCA GCC ACA GAA TC; GCT CGG ATG GGG TAT TTC TAG). SNPs were genotyped by pyrosequencing using PyroMark Q48 Advanced Reagents and the PyroMark Q48 Autoprep (both Qiagen) according to manufacturer's instructions.

### **Quantitative Real-time PCR**

Total RNA was isolated using the RNeasy Mini Kit (Qiagen) and the Quick RNA MicroPrep Kit (Zymo Research) including on column DNaseI digest and was reverse transcribed. SYBRgreen Real-time PCR was performed (primers: *RBPJ* FWD: GGC TGC AGT CTC CAC GTA CGT C, REV: CTC ACC AAA TTT CCC AGG CGA TGG; *HES1* FWD: ATG GAG AAA AGA CGA AGA GCA AG; REV: TGC CGC GAG CTA TCT TTC TT), including controls (samples containing the untranscribed RNA, dissociation curves). Data were analysed with the comparative CT methods and presented as  $\Delta CT$  or  $2^{-\Delta\Delta CT}$  as described elsewhere<sup>29</sup>

using *RPLP0* as a housekeeping gene for sample normalization (FWD: GCG TCC TCG TGG AAG TGA CAT CG, REV: TCA GGG ATT GCC ACG CAG GG).

### Cap Analysis Gene Expression (CAGE)

Cultured human FLS from RA patients were treated with 10 ng/ml TNF, 24h or were left untreated (Supplementary table 2). RNA was isolated using the Quick RNA MicroPrep Kit (Zymo Research). CAGE libraries were prepared and sequenced as previously described in detail<sup>30</sup>. Mapping and identification of CAGE transcription start sites (CTSSs) were performed by DNAFORM (Yokohama, Kanagawa, Japan). In brief, the sequenced CAGE tags were mapped to hg19 using BWA software and HISAT2 after discarding ribosomal RNAs. Identification of CTSSs was performed with the Bioconductor package CAGEr (version 1.16.0)<sup>31</sup>. TSS and enhancer candidate identification and quantification were performed with the Bioconductor package CAGEfightR (version 1.6.0)<sup>32</sup> with default settings.

### Data availability

The following publicly available datasets were used in this study:

RNA sequencing data (Figs. 2d, 3b) : ENA PRJEB14422

Capture HiC data from GM12878 and Jurkat cell lines: NCBI Gene Expression Omnibus (GEO) GSE69600

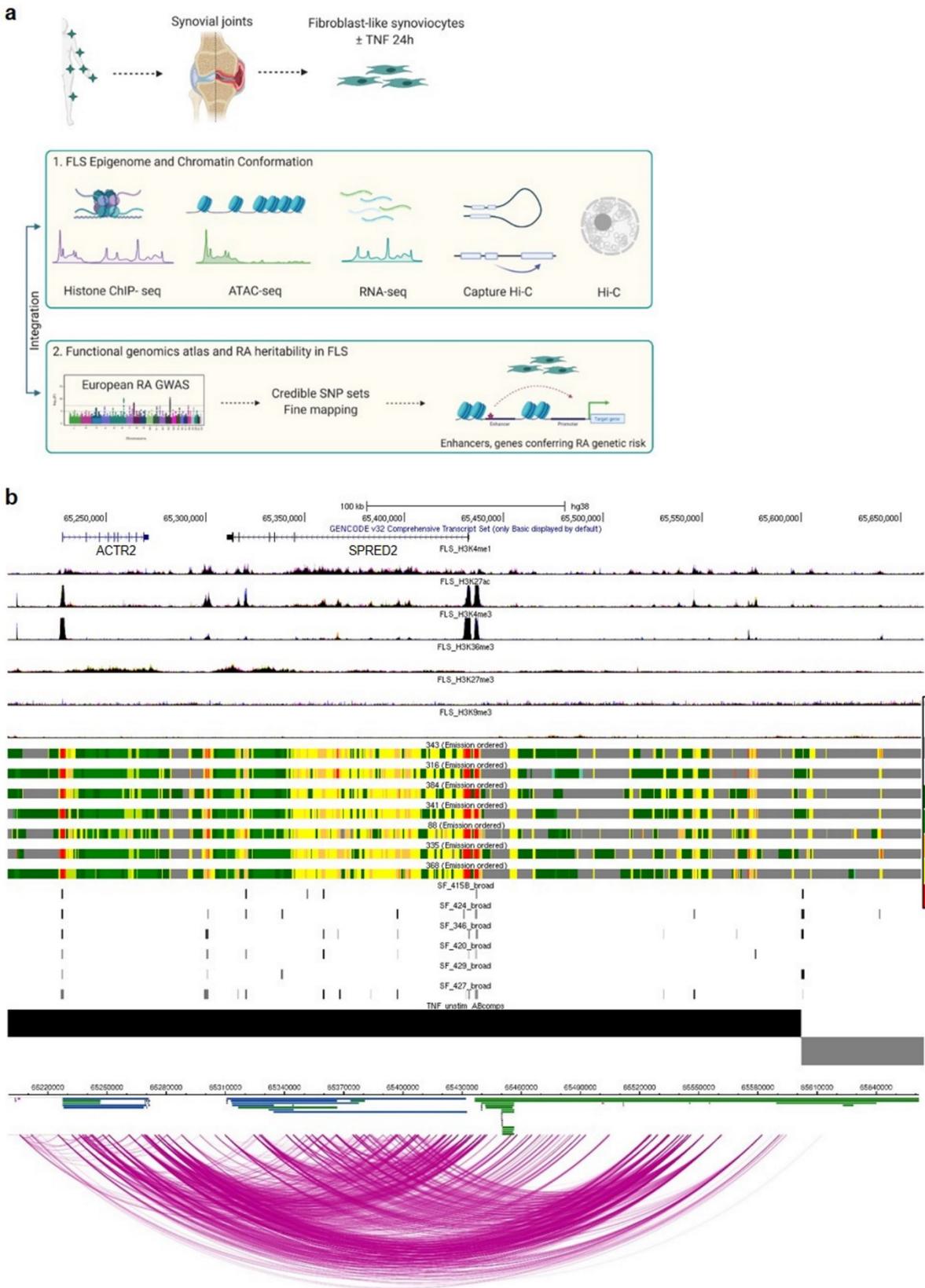
All data are available from the authors upon request.

### References

1. Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* **62**, 2569-81 (2010).
2. Frank-Bertoncelj, M. *et al.* Epigenetically-driven anatomical diversity of synovial fibroblasts guides joint-specific fibroblast functions. *Nat Commun* **8**, 14852 (2017).

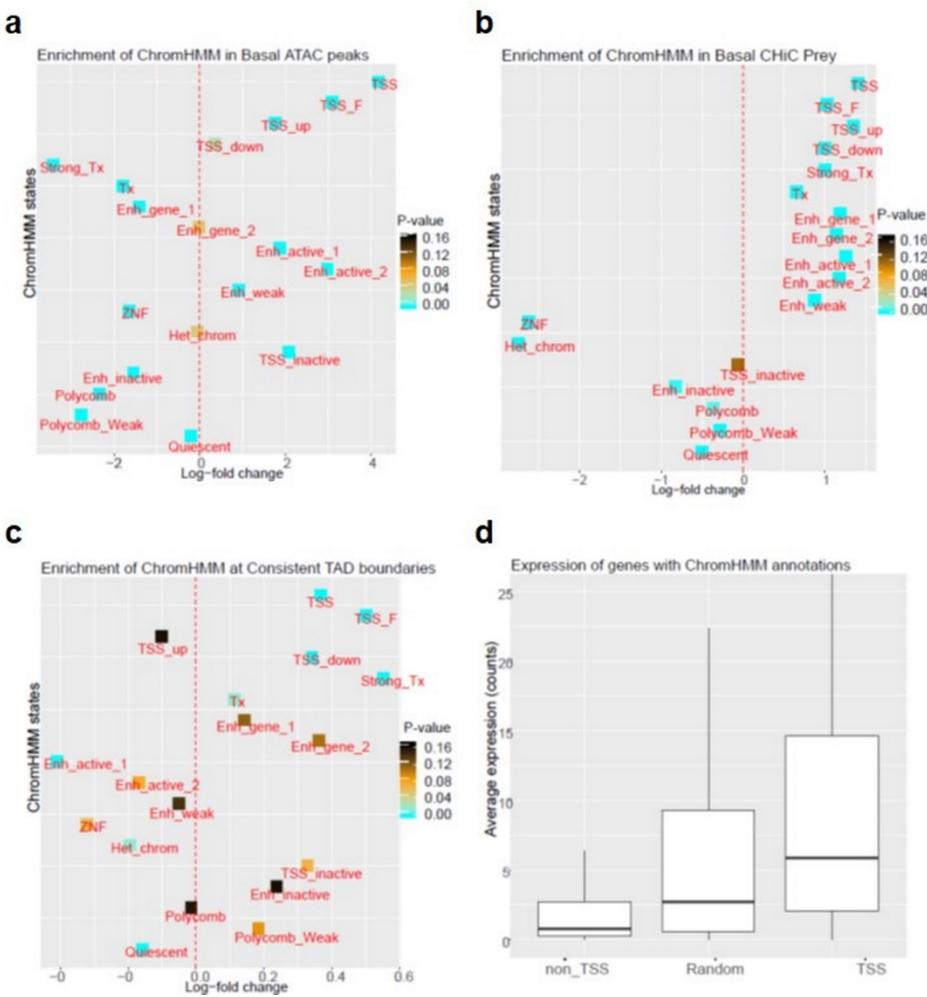
3. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
4. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
5. Leek, J.T. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res* **42**(2014).
6. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
7. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
8. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
9. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 1-9 (2015).
10. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
11. Zhao, S., Zhang, Y., Gamini, R., Zhang, B. & von Schack, D. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA<sup>+</sup> selection versus rRNA depletion. *Sci Rep* **8**, 4781 (2018).
12. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
13. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015).
14. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**, 1939-1949 (2017).
15. Cresswell, K.G. & Dozmorov, M.G. TADCompare: An R Package for Differential and Temporal Analysis of Topologically Associated Domains. *Front Genet* **11**, 158 (2020).
16. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun* **6**, 10069 (2015).
17. Ignatiadis, N., Klaus, B., Zaugg, J.B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* **13**, 577-80 (2016).
18. Boeva, V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Front Genet* **7**, 24 (2016).
19. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
20. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-81 (2014).
21. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
22. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
23. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
24. Chen, J., Aronow, B.J. & Jegga, A.G. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* **10**, 73 (2009).

25. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
26. Huang da, W. *et al.* Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 11 (2009).
27. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).
28. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784 (2014).
29. Schmittgen, T.D. *et al.* Real-time PCR quantification of precursor and mature microRNA. *Methods* **44**, 31-8 (2008).
30. Kodzius, R. *et al.* CAGE: cap analysis of gene expression. *Nat Methods* **3**, 211-22 (2006).
31. Haberle, V., Forrest, A.R., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* **43**, e51 (2015).
32. Thodberg, M., Thieffry, A., Vitting-Seerup, K., Andersson, R. & Sandelin, A. CAGEfightR: analysis of 5'-end data using R/Bioconductor. *BMC Bioinformatics* **20**, 487 (2019).



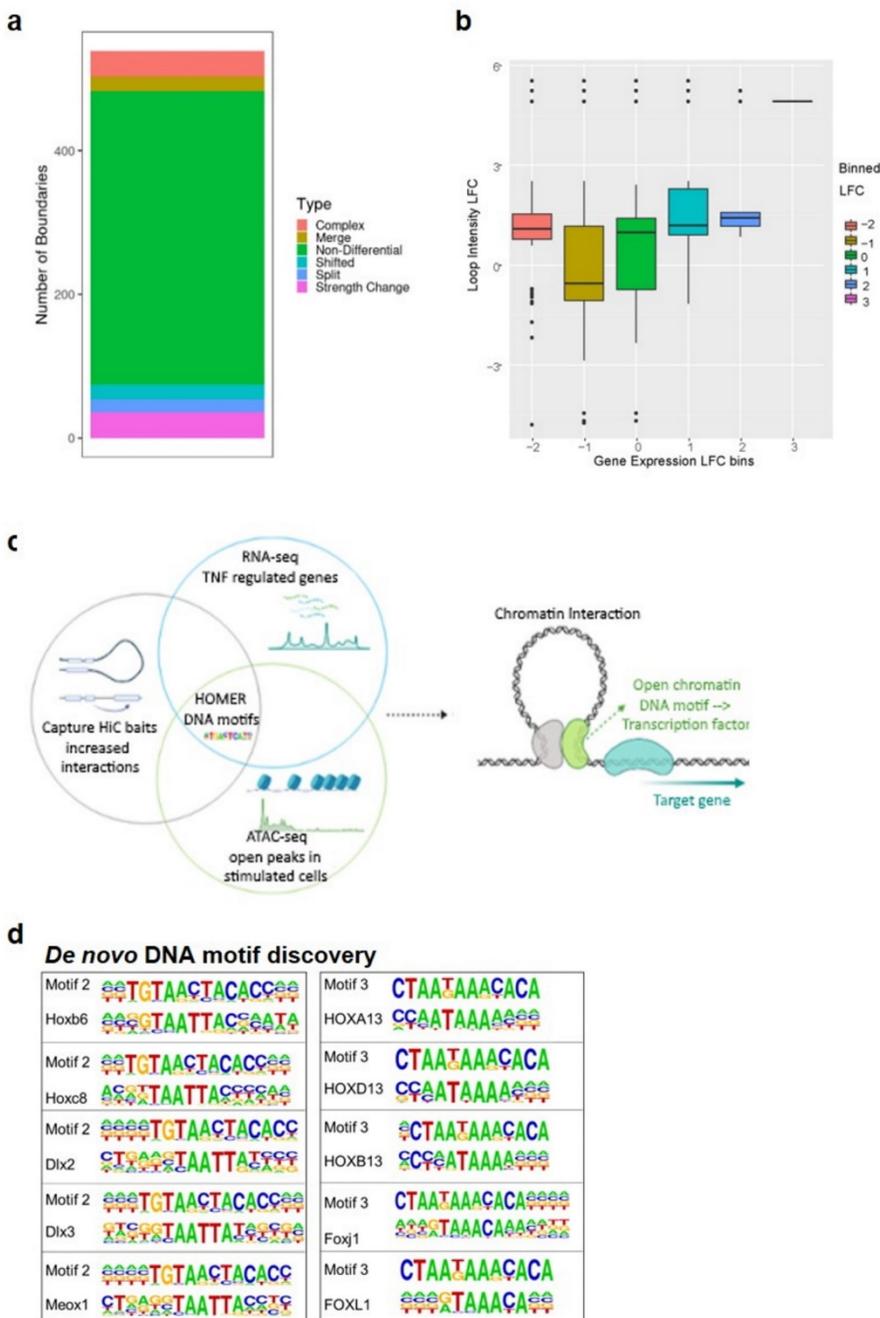
**Fig. 1: Epigenomic and 3D chromatin atlas of human FLS.**

**a)** Schematic representation of the workflow to comprehensively annotate the transcriptome, epigenome and chromatin structure of FLS and define their contribution to RA heritability. This figure was created using BioRender. **b)** The SPRED2 locus as an example genomic region demonstrating the annotation of epigenetic states and chromatin architecture in unstimulated FLS. Shown are from top to bottom, genes, ChIP-seq peaks (H3K4me1, H3K27ac, H3K4me3, H3K36me3, H3K27me3, H3K9me3), ChromHMM annotation in 7 different FLS lines, ATAC-seq peaks in 6 different FLS lines, A/B compartments (black bar open chromatin, grey bar closed chromatin), chromatin interactions.



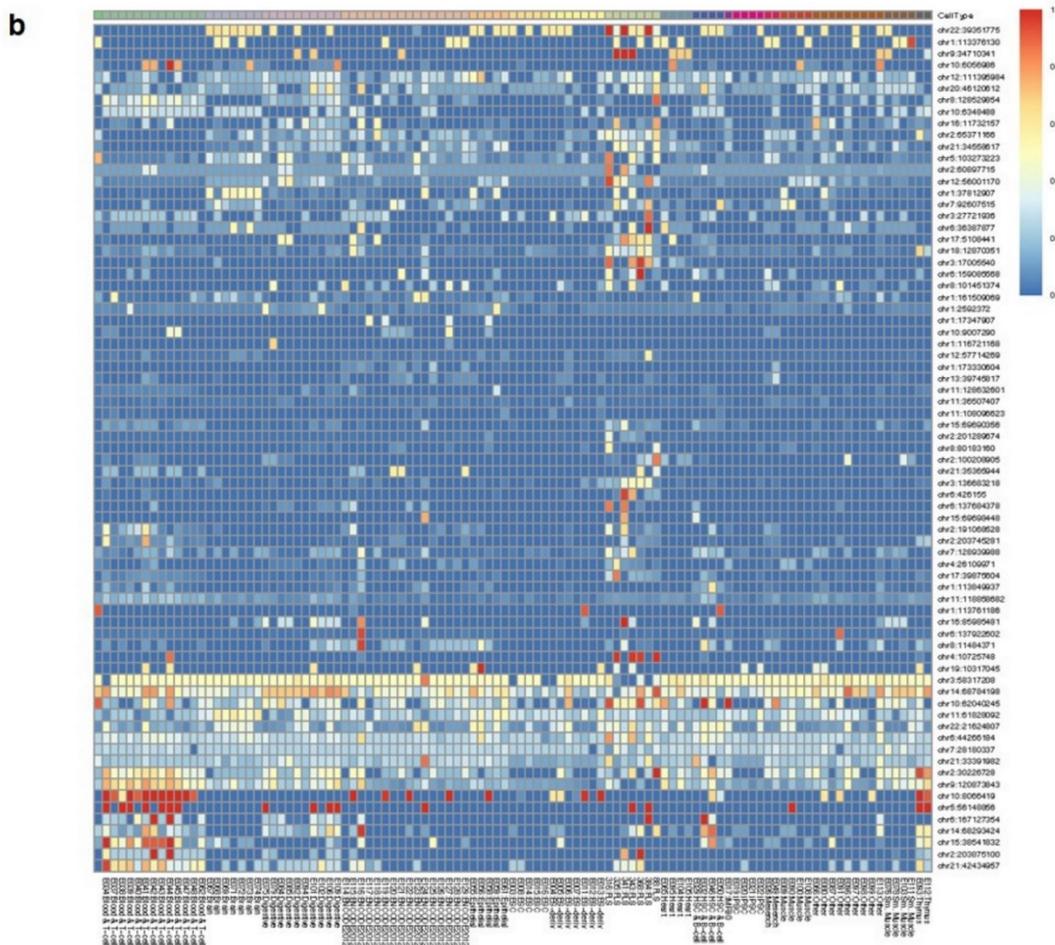
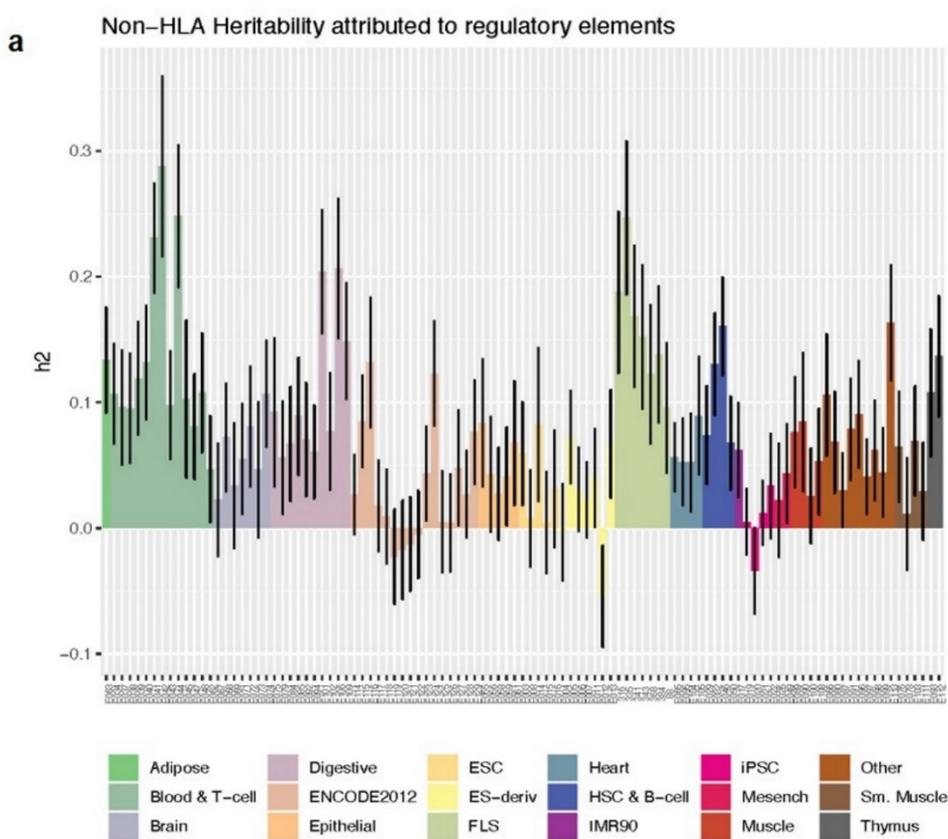
**Fig. 2: Cross validation of generated datasets defining the chromatin landscape of FLS.**

**a)** Log fold change enrichment of chromatin states as defined by ChromHMM in ATAC-seq data. **b)** Log fold change enrichment of chromatin states as defined by ChromHMM in prey fragments of Capture HiC measurements. **c)** Log fold change enrichment of chromatin states as defined by ChomHMM in consistent TAD boundaries. **d)** Basal average expression of genes (RNA-seq counts) across non-TSS, TSS and random ChromHMM annotations. TSS = transcription start site, TSS\_F = flanking TSS; TSS\_up = upstream TSS; TSS\_down = downstream TSS; Enh\_gene = enhancer genic; ZNF = zinc finger; Het\_chrom = heterochromatin



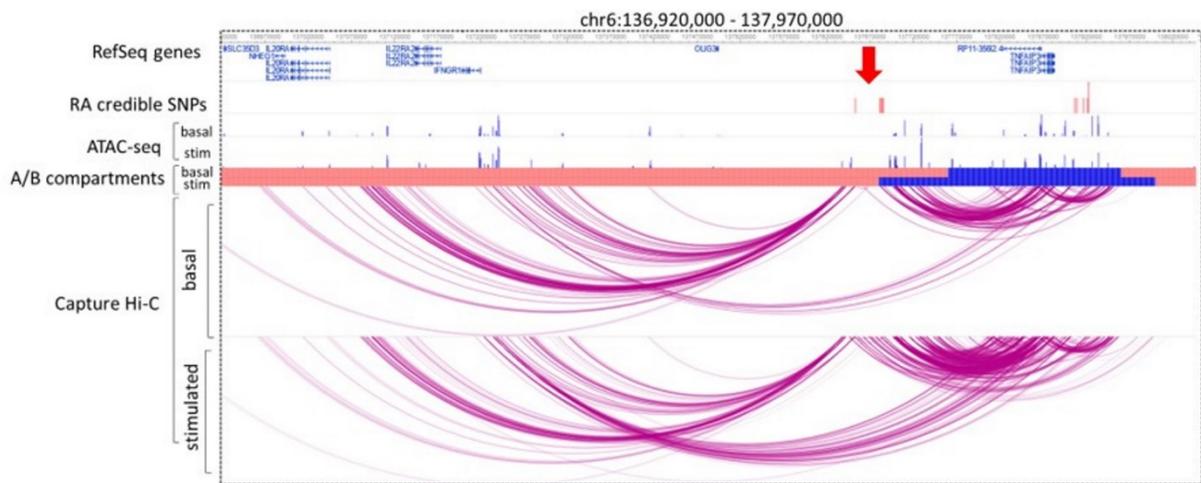
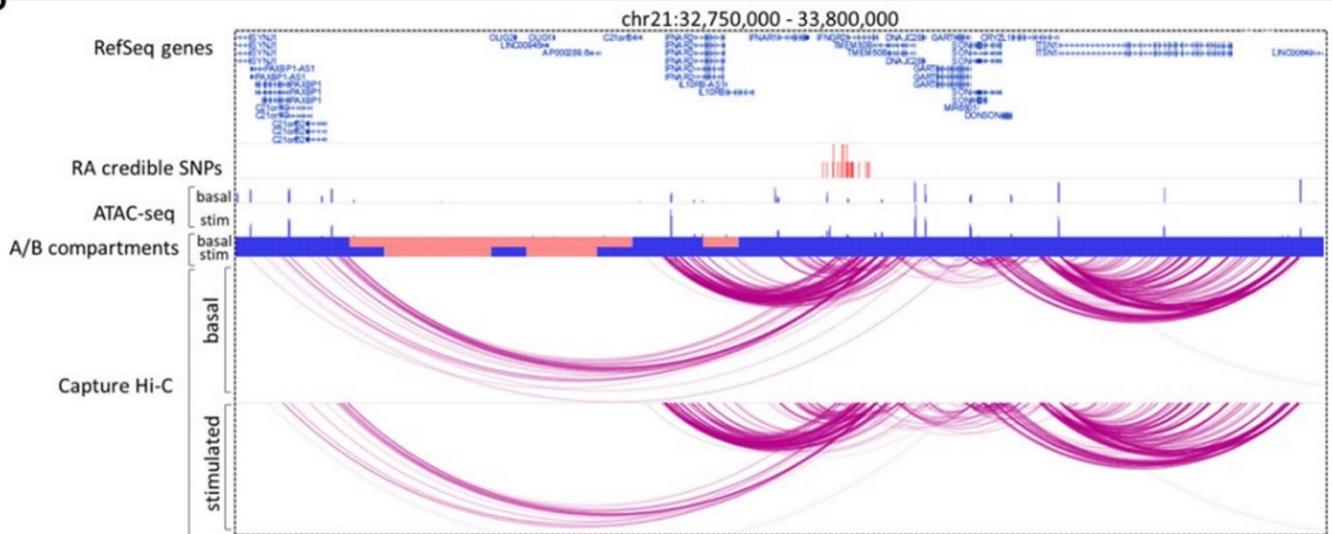
**Fig. 3. Effect of TNF stimulation on the chromatin landscape in FLS.**

**a)** Analysis of TADs boundaries after TNF stimulation in FLS by TADCompare. Number of non-differential (green) and differential TAD boundaries is shown. Differential TAD boundaries are classified as boundary position changes (complex, merge, shifted, split) or strength change (differential boundary magnitude). Complex, merged and split boundary changes represent the most disruptive changes of the 3D structure of the genome. **b)** Correlation of the loop intensity as determined by CHiC with change in the expression of nearby genes (log fold change). **c)** Graphical representation of the RNA-seq, ATAC-seq and CHiC data integration to identify transcription factor binding sites in TNF-stimulated FLS. This figure was created using BioRender. **d)** De novo DNA motif discovery identifies two motifs (motif 2 and motif 3) with high similarity to the binding sites of homeobox (TAATTA) and forkhead box transcription factors (TAAA) in the dataset with TNF-repressed genes.



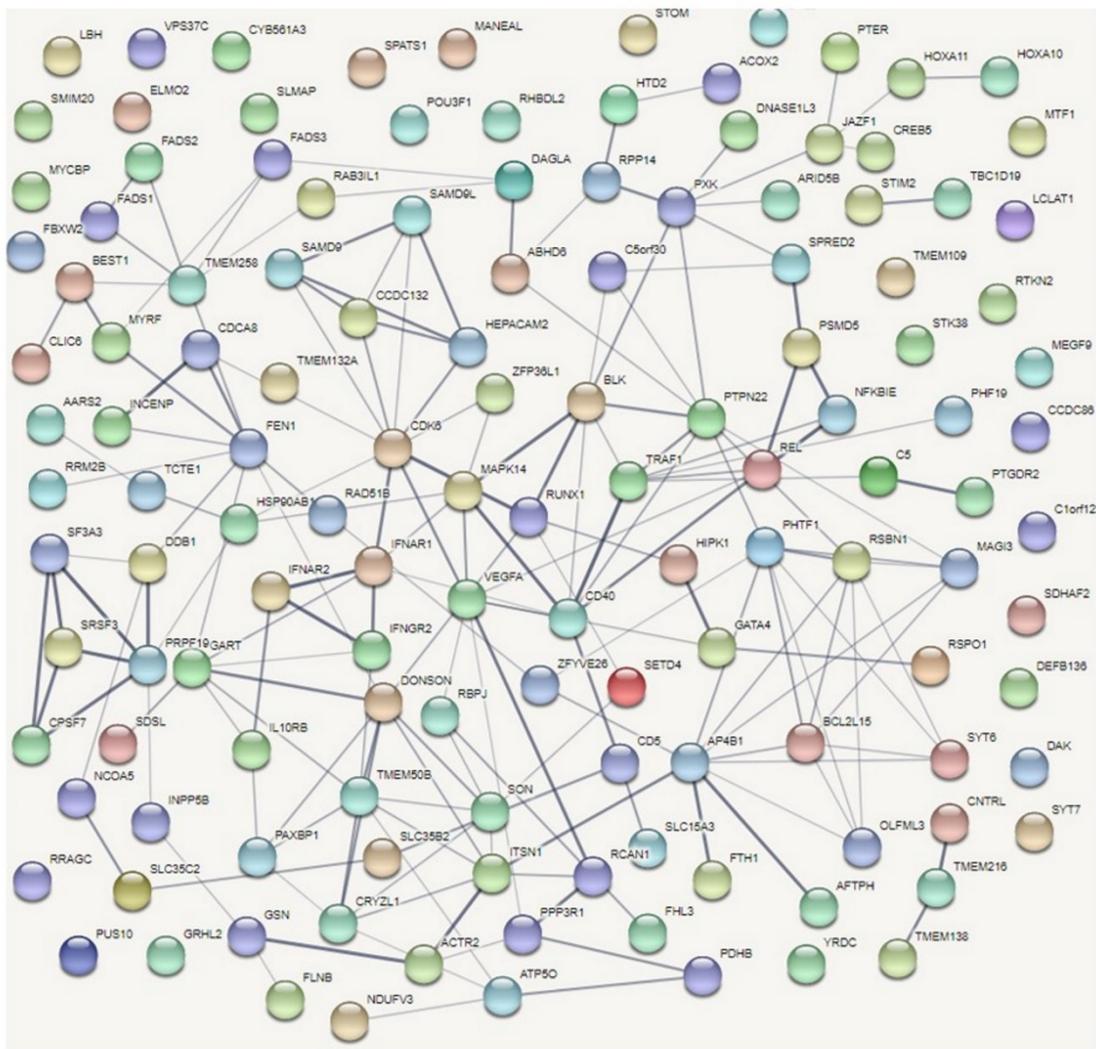
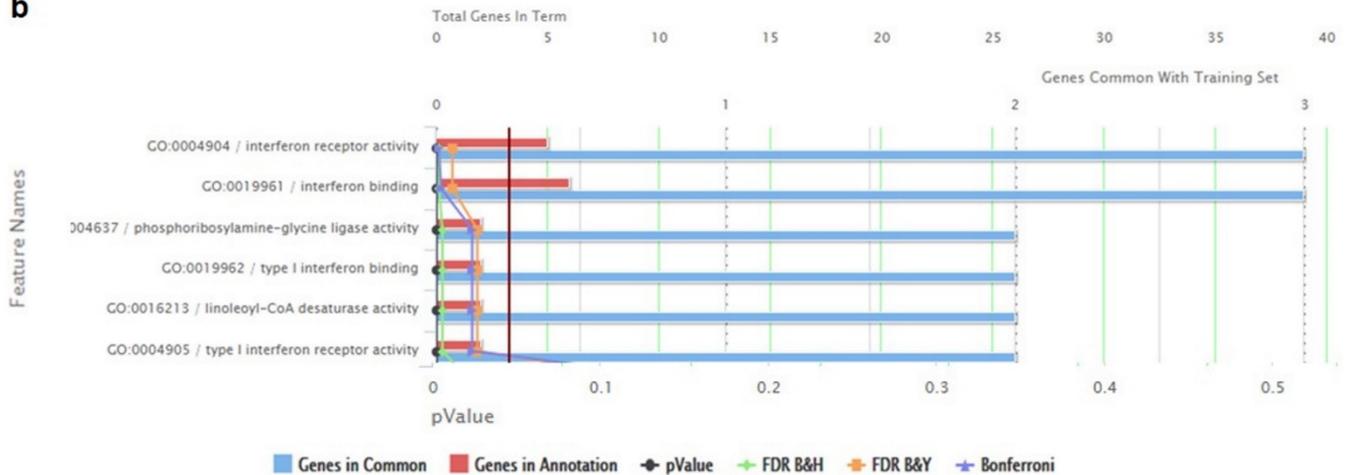
**Fig. 4. Heritability and causal SNPs in FLS.**

**a)** Partitioned heritability ( $h^2$ ) of RA attributed to active regions in each sample of FLS ( $n=7$ ) and 111 available Roadmap cell types/tissues (Epigenomics Mapping Roadmap Consortium<sup>11</sup>). **b)** The sum of posterior probability overlapping active DNA regulatory elements at each of the 73 sites, defined as the union of H3K4me3, H3K4me1 and H3K27ac marks, across FLS samples and the 111 cell types/tissues.

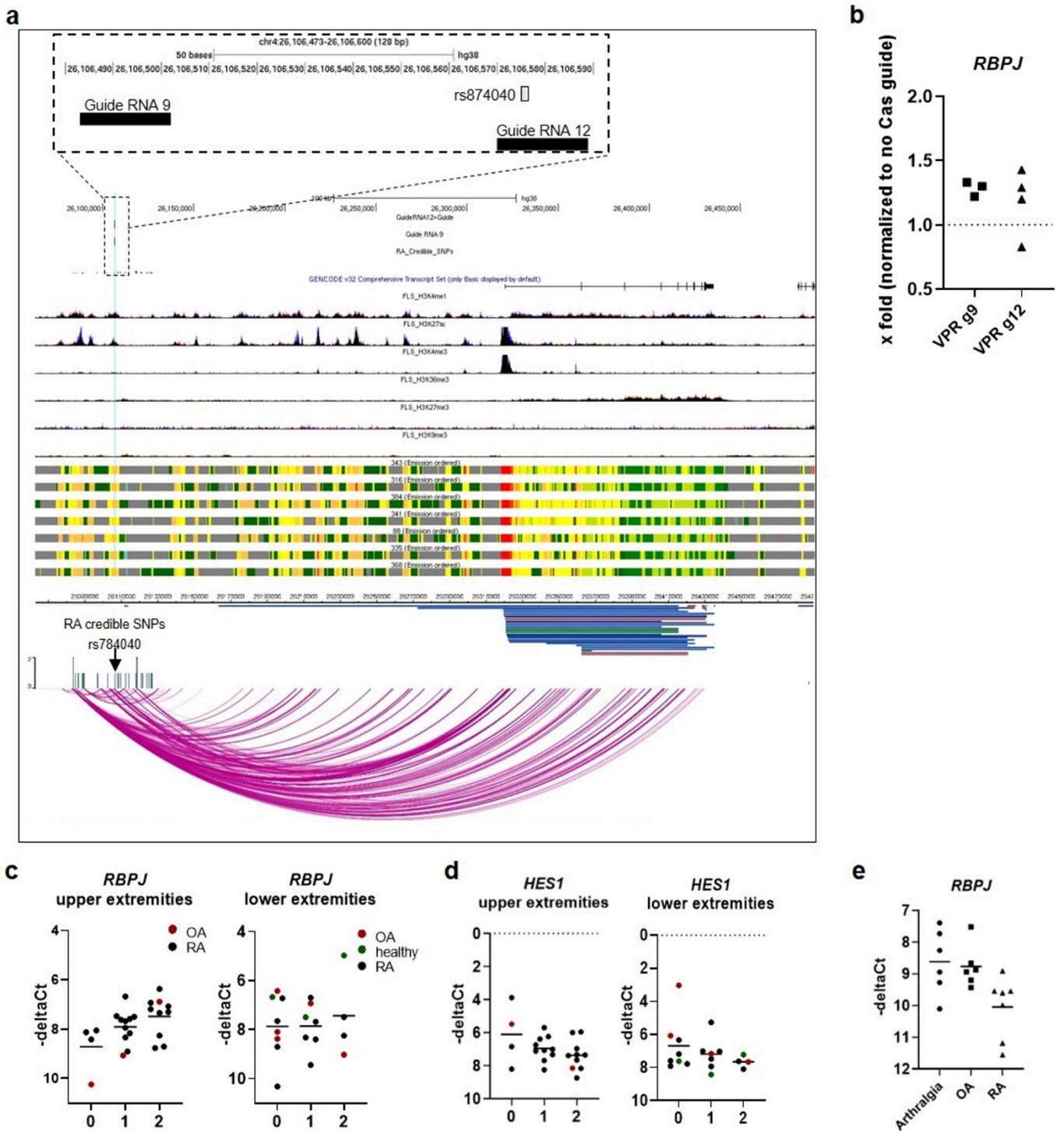
**a****b**

**Fig. 5 TNFAIP3 and IFNAR1 genetic risk loci - exemplary regions linked to TNF stimulation.**

Two exemplary risk regions where TNF stimulation had profound effects on chromatin structure and influenced the genetic regions containing RA SNPs. **a)** The TNFAIP3 region on chromosome 6q23 (red arrow) containing RA credible SNPs (red lines, rs17264332 in Supplementary Table 5) changed from closed chromatin (light red bar) to open chromatin state (blue bar) after TNF stimulation and exhibited increased interactions with the promoter of *TNFAIP3* in stimulated FLS. **b)** The genomic IFNAR2 region of the credible SNP set on chromosome 21 (rs73194058 in Supplementary Table 5) interacted with several nearby genes involved in the interferon response. These interactions were further enhanced by TNF stimulation. Chromatin at the IFNAR2 gene locus changed from a closed (light red bar) to open (blue bar) state in stimulated FLS.

**a****b**

**Fig. 6 Predicted functional networks of genes that were associated with SNPs active in FLS.**  
**a)** A protein-protein interaction network was established using STRING with default settings (medium confidence). The obtained network had more interactions than expected by chance with a protein-protein-interaction enrichment p-value of  $1.28 \times 10^{-8}$ . The thickness of the lines indicates the strength of data support. Colours, distances and location on the map were assigned randomly. **b)** Functional enrichment of genes interacting with SNPs active in FLS was detected using ToppFun in default settings. Significant terms for GO molecular function are shown. FDR = false discovery rate; B&H = Benjamini-Hochberg ; B&Y = Benjamini-Yekutieli.

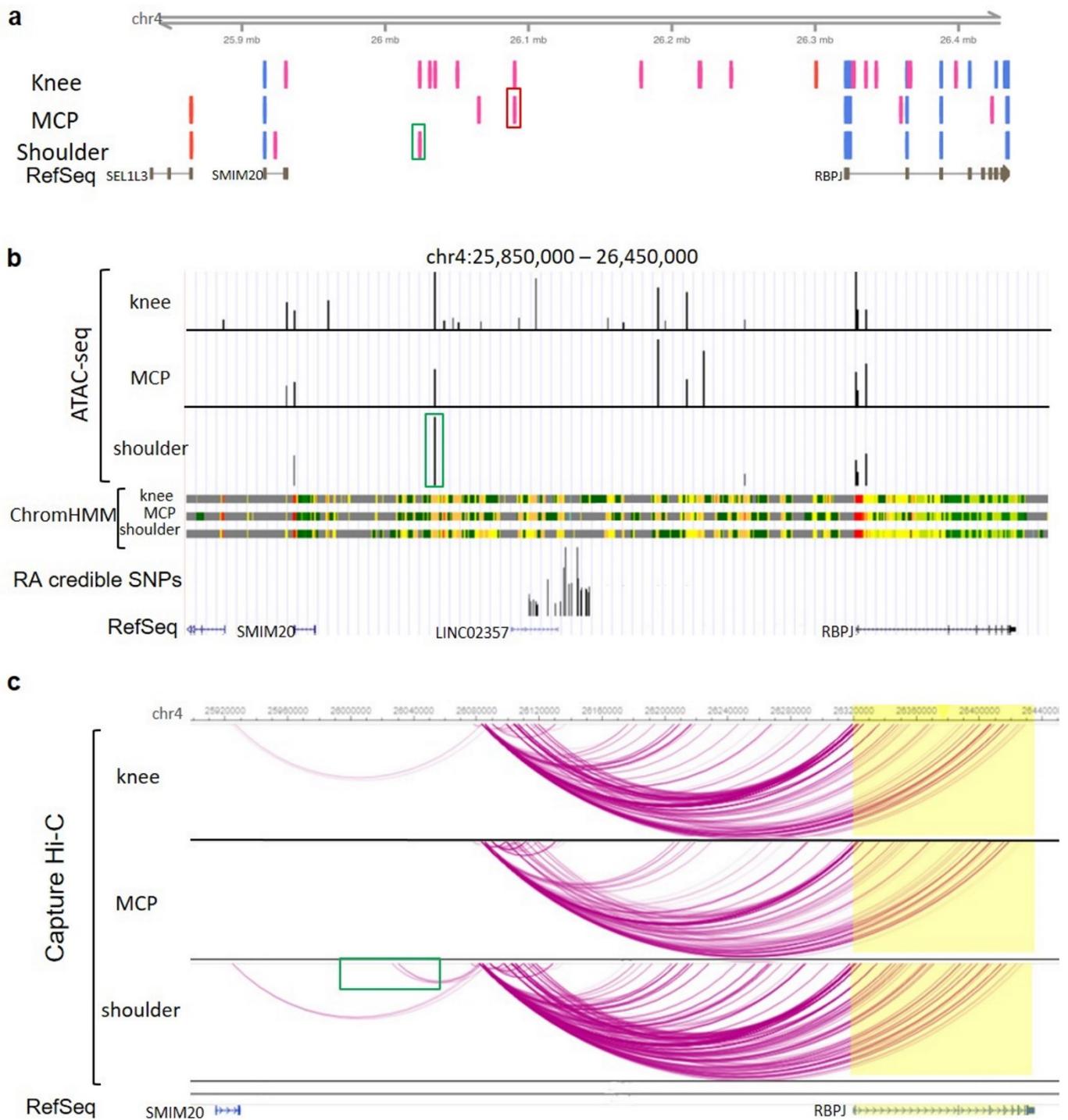


**Fig.7 *RBPJ* expression in FLS is affected by genotype and disease**

**a)** Fine mapping, epigenetic and chromatin conformation analyses at the *RBPJ* locus

indicate rs874040 as a likely causal credible SNP and *RBPJ* as a candidate causal gene in this locus. The location of credible set SNPs with rs874040 and the guide RNAs (guide RNA 9 and guide RNA 12) targeting the rs874040-containing enhancer are indicated.

**b)** *RBPJ* expression in FLS transduced with VP64-p65-Rta dCas9 (VPR) and two different guide RNAs (g9 and g12) targeting the genomic region around chr4:26106575 (rs87040). *RBPJ* expression was normalized to FLS that were transduced with respective guide RNAs but not VPR-dCas9 (set to 1).  $-\text{deltaCt}$  = cycle of threshold of *RBPJ* expression – cycle of threshold *RPLP0*. **c)** *RBPJ* expression in FLS isolated from individuals homozygous for rs874040 in the locus near the *RBPJ* gene (0), heterozygous (1) or homozygous for the wild type variant (2). Upper extremity joints included joints of the hand, elbows and shoulders; lower extremity joints included hips, knees and joints of the feet. **d)** Expression of *HES1* in the same FLS cohort. **e)** *RBPJ* expression in individuals with joint pain, but no histological signs of arthritis (arthralgia), OA and RA.



**Fig. 8 Joint-specific enhancers and chromatin interactions in the *RBPJ* locus might influence the joint-specific expression of *RBPJ*.**

**a**) CAGE measurements of active enhancers (pink bars) and active promoters (light blue bars) in FLS from knees (n=2), metacarpophalangeal (MCP) joints (n=3) and shoulders (n=2). Red box highlights enhancer used in MCP joints overlapping risk SNPs. Green box highlights main enhancer in shoulders.

**b**) Representative depiction of open chromatin as measured by ATAC-seq (black bars) in FLS from knees (n=3), MCP joints (n=2) and shoulders (n=2) and representative depiction of ChromHMM regulatory regions in FLS from knees (n=3), MCP joints (n=2) and shoulders (n=2). Green box highlights main enhancer in shoulders. **c**) Chromatin interactions in FLS from knees (n=2), MCP joints (n=2) and shoulders (n=2) as measured by CHiC. Green box highlights interaction of the shoulder enhancer with the risk locus.

**Table 1.** Alterations in TAD structure after TNF stimulation and genes in stimulation specific TADs

Locus	Chr	TAD_diff	Type of Difference	Enriched_in	Expressed in FLS	Differential expressed in FLS (p adj <0.05)
9	chr1	Differential	Strength Change	Basal	RC3H1; RABGAP1L; PRDX6; KLHL20; GAS5-AS1; GAS5; ZBTB37	GAS5; PRDX6; ZBTB37; RABGAP1L; RC3H1
10	chr2	Differential	Complex	Basal	LBH; LCLAT1; YPEL5	LBH; LCLAT1
13	chr2	Differential	Merge	Stim	NPAS2	NPAS2
19	chr3	Differential	Strength Change	Basal	CMC1; SLC4A7; AZI2	CMC1
20	chr3	Differential	Split	Basal	SLMAP; FLNB; FLNB-AS1; PXX; PDHB	PXX; FLNB
21	chr3	Differential	Strength Change	Basal	PCCB; PPP2R3A; MSL2; STAG1	PCCB
22	chr4	Differential	Strength Change	Basal	WDR1; ZNF518B	
27	chr6	Differential	Strength Change	Basal	SRPK1; MAPK14; PI16; LHFPL5; BRPF3; KCTD20; STK38; SRSF3; PANDAR; CDKN1A; C6orf89; MTCH1	SRPK1; C6orf89; STK38; MTCH1; KCTD20; BRPF3; CDKN1A; SRSF3
28	chr6	Differential	Strength Change	Stim	RSPH9; VEGFA; CDC5L; TMEM63B; HSP90AB1; SLC35B2	TMEM63B; CDC5L
29	chr6	Differential	Strength Change	Stim	IFNGR1; WAKMAR2; TNFAIP3	TNFAIP3; IFNGR1
30	chr6	Differential	Strength Change	Stim	IFNGR1; WAKMAR2; TNFAIP3	TNFAIP3; IFNGR1
32	chr6	Differential	Strength Change	Basal	RPS6KA2; AFDN	
33	chr7	Differential	Strength Change	Stim	CREB5; TAX1BP1; JAZF1	JAZF1; TAX1BP1
47	chr11	Differential	Strength Change	Basal	TRIM44; FJX1; COMMD9	COMMD9
48	chr11	Differential	Strength Change	Basal	FADS2; SLC15A3; TKFC; INCENP; AHNAK; INTS5; CCDC86; PRPF19; TMEM109; TMEM132A; VPS37C; DDB1; CYB561A3; TMEM138; CPSF7; MYRF; FEN1; FADS1; FADS3; RAB31L1; BEST1; FTH1; EEF1G; TUT1; MTA2; EML3; ROM1; GANAB; LBHD1; CSKMT; UQCC3; UBXN1; LRRN4CL; HNRNPUL2-BSCL2; HNRNPUL2; TMEM179B; TMEM223; NXF1; STX5; RNU2-2P; SLC3A2	TMEM132A; FTH1; SLC15A3; FADS1; AHNAK; VPS37C; INCENP; RAB31L1; LRRN4CL; EEF1G; TMEM138; ROM1; TUT1; DDB1; FADS2; CCDC86; MTA2; TMEM109
49	chr11	Differential	Strength Change	Basal	ACAT1; ATM; CUL5; NPAT; POGLUT3	ATM; ACAT1; NPAT; CUL5
57	chr14	Differential	Strength Change	Basal	ZFYVE26; ZFP36L1	
65	chr18	Differential	Split	Basal	SPIRE1; SEH1L; CEP192; AFG3L2; CEP76	CEP76; CEP192; SPIRE1; SEH1L
72	chr22	Differential	Strength Change	Basal	MAPK1; UBE2L3; PPIL2; YPEL1; PPM1F	MAPK1; UBE2L3
73	chr22	Differential	Strength Change	Basal	JOSD1; GTPBP1; SUN2; CBX6; APOBEC3C; CBX7; RPL3; MIEF1; ATF4; RPS19BP1	GTPBP1; RPL3; SUN2; ATF4; CBX6

**Table 2.** Changes of gene expression in loci with changed chromatin interactions after TNF stimulation  
*Genes with a log2fold change > ±0.5 and padj < 0.05 are marked in red (downregulated) and in green (upregulated).*

Locus	Interacting Gene	Base Mean Expression	log2Fold Change	lfcSE	stat	p-value	padj
12	SPRED2	772.22	-0.34	0.08	-4.36	1.29E-05	7.23E-05
19	SLC4A7	3702.62	-0.23	0.13	-1.80	0.072	0.145
20	PXK	1012.38	-0.33	0.11	-3.05	0.002	0.008
<b>23</b>	<b>RBPJ</b>	<b>5319.00</b>	<b>-0.56</b>	<b>0.22</b>	<b>-2.51</b>	<b>0.012</b>	<b>0.032</b>
<b>24</b>	<b>IL6ST</b>	<b>18850.31</b>	<b>0.72</b>	<b>0.08</b>	<b>8.91</b>	<b>4.91E-19</b>	<b>2.30E-17</b>
<b>30</b>	<b>TNFAIP3</b>	<b>7841.42</b>	<b>4.57</b>	<b>0.19</b>	<b>23.43</b>	<b>2.14E-121</b>	<b>3.51E-118</b>
<b>34</b>	<b>CDK6</b>	<b>4906.40</b>	<b>0.92</b>	<b>0.12</b>	<b>7.89</b>	<b>3.05E-15</b>	<b>9.21E-14</b>
35	TNPO3	1671.71	0.05	0.07	0.76	0.449	0.597
<b>41</b>	<b>PHF19</b>	<b>1065.75</b>	<b>0.71</b>	<b>0.14</b>	<b>5.09</b>	<b>3.56E-07</b>	<b>2.74E-06</b>
<b>42</b>	<b>TRAF1</b>	<b>1041.51</b>	<b>2.80</b>	<b>0.22</b>	<b>12.65</b>	<b>1.14E-36</b>	<b>1.98E-34</b>
<b>48</b>	<b>VPS37C</b>	<b>552.68</b>	<b>0.71</b>	<b>0.07</b>	<b>10.56</b>	<b>4.76E-26</b>	<b>3.90E-24</b>
<b>52</b>	<b>CDK2</b>	<b>648.74</b>	<b>0.85</b>	<b>0.13</b>	<b>6.69</b>	<b>2.23E-11</b>	<b>3.72E-10</b>
52	RAB5B	1805.92	-0.24	0.06	-4.06	4.88E-05	2.42E-04
<b>52</b>	<b>RNF41</b>	<b>1878.98</b>	<b>-0.63</b>	<b>0.11</b>	<b>-5.62</b>	<b>1.93E-08</b>	<b>1.88E-07</b>
52	ANKRD52	1954.38	0.23	0.07	3.09	0.002	0.007
52	CNPY2	1033.14	0.37	0.07	5.30	1.16E-07	9.89E-07
52	PAN2	312.46	-0.23	0.08	-2.80	0.005	0.015
<b>52</b>	<b>STAT2</b>	<b>4951.27</b>	<b>0.74</b>	<b>0.20</b>	<b>3.77</b>	<b>1.64E-04</b>	<b>7.21E-04</b>
<b>52</b>	<b>TIMELESS</b>	<b>728.81</b>	<b>1.11</b>	<b>0.14</b>	<b>8.10</b>	<b>5.66E-16</b>	<b>1.87E-14</b>
52	GLS2	18.80	0.34	0.26	1.31	0.191	0.317
53	SLC26A10	10.97	-0.29	0.39	-0.75	0.455	0.603
53	OS9	5461.71	-0.12	0.07	-1.83	0.068	0.138
53	AGAP2	5.83	0.43	0.31	1.36	0.173	0.294
53	TSPAN31	370.51	-0.29	0.11	-2.56	0.010	0.028
53	CDK4	1630.73	0.06	0.07	0.88	0.378	0.529
53	AVIL	64.47	-0.44	0.15	-2.93	0.003	0.011
53	CTDSP2	3707.75	-0.47	0.06	-7.79	6.45E-15	1.85E-13
<b>54</b>	<b>SH2B3</b>	<b>1493.15</b>	<b>0.52</b>	<b>0.10</b>	<b>5.29</b>	<b>1.25E-07</b>	<b>1.06E-06</b>
54	ATXN2	781.23	-0.05	0.08	-0.71	0.478	0.625
64	ORMDL3	800.27	0.40	0.08	4.75	1.99E-06	1.32E-05
64	PSMD3	1842.20	0.18	0.05	3.66	2.53E-04	0.001
<b>65;68</b>	<b>IFNAR2</b>	<b>644.38</b>	<b>2.81</b>	<b>0.11</b>	<b>26.14</b>	<b>1.14E-150</b>	<b>2.92E-147</b>
<b>68</b>	<b>IFNAR1</b>	<b>2355.18</b>	<b>0.95</b>	<b>0.08</b>	<b>12.55</b>	<b>3.95E-36</b>	<b>6.53E-34</b>
<b>68</b>	<b>IFNGR2</b>	<b>1248.98</b>	<b>1.10</b>	<b>0.08</b>	<b>12.90</b>	<b>4.71E-38</b>	<b>9.03E-36</b>
68	ITSN1	1138.10	-0.24	0.11	-2.10	0.036	0.080

lfcSE = standard error for log2 fold change, stat = *statistic* value for the null hypothesis, padj = p-value adjusted for multiple testing using Benjamini-Hochberg

All rights reserved. No reuse allowed without permission.

**Table 3.** Posterior probability and cellular context of category 1 and 2 loci

Locus	Chr	Number of credible SNPs (>posterior prior 0.001)	Posterior probability top SNP	total posterior probability top 3 SNPs	SNP category	Number of SNPs in 30% of FLS enhancers (posterior>0.001)	Number of SNPs in 30% of T cell enhancers (posterior>0.001)	Number of SNPs in 30% of B cell enhancers (posterior>0.001)	FLS, T, B, ALL, NONE (based on posteriors of SNPs in enhancers)	Top posterior rs numbers
4	chr1	10	0.53	0.94	1	1	0	0	FLS	rs4839319, rs4839318, rs77227025
5	chr1	2	0.91	0.99	1	0	0	0	B CELL	rs6679677, rs2476601
7	chr1	10	0.25	0.62	1	0	0	0	NONE	rs624988, rs771587, rs12137270, rs12405671, rs11586238
10	chr2	9	0.21	0.59	1	2	3	2	ALL	rs10175798, rs10173253, rs906868, rs7579944, rs1355208
17	chr2	4	0.43	0.92	1	0	2	0	T CELL	rs231724, rs231723, rs231775
20	chr3	13	0.34	0.82	1	4	2	3	ALL	rs73081554, rs185407974, rs180977001, rs35677470, rs114584537
24	chr5	1	1	1	1	0	1	0	T CELL	rs7731626
29	chr6	8	0.23	0.67	1	0	0	0	NONE	rs17264332, rs11757201, rs6920220, rs6927172
30	chr6	15	0.42	0.86	1	2	5	3	NONE	rs58721818, rs61117627
32	chr6	9	0.32	0.71	1	0	5	0	T CELL, B CELL	rs1571878, rs3093017, rs10946216
40	chr9	3	0.37	0.98	1	0	0	0	NONE	rs11574914, rs2812378, rs10972201
42	chr10	6	0.73	0.87	1	0	0	0	T CELL	rs706778, rs10795791
44	chr10	10	0.21	0.49	1	0	10	0	T CELL	rs537544, rs568727, rs570613, rs570730, rs7897792
46	chr10	10	0.37	0.73	1	9	1	0	ALL	rs12764378, rs71508903, rs77509998
60	chr15	5	0.26	0.74	1	0	0	0	NONE	rs919053, rs8026898, rs7170107, rs16953656, rs8043362
62	chr16	9	0.29	0.82	1	0	0	0	B CELL	rs13330176, rs2139492, rs2139493
66	chr19	2	0.6	1	1	0	0	0	NONE	rs74956615, rs34536443
70	chr21	19	0.56	0.93	1	3	1	0	FLS	rs8133843, rs8129030, rs9979383
73	chr22	9	0.29	0.74	1	0	0	2	B CELL	rs2069235, rs909685, rs9611155
1	chr1	81	0.29	0.38	2	1	1	1	NONE	rs187786174, rs60733400, rs876938
2	chr1	13	0.29	0.52	2	0	0	0	NONE	rs2240336, rs12737739, rs13375202
3	chr1	36	0.43	0.51	2	1	0	0	FLS	rs28411352, rs28489009, rs2306627
6	chr1	29	0.22	0.42	2	0	2	4	B CELL	rs1217404, rs2476604, rs1217420
11	chr2	13	0.18	0.4	2	1	1	1	ALL	rs34695944, rs56095903, rs67574266, rs13031237, rs13031721
13	chr2	17	0.27	0.63	2	0	0	1	NONE	rs9653442, rs6712515, rs11676922, rs1160542, rs10865035
18	chr3	63	0.27	0.55	2	0	0	0	NONE	rs4452313, rs4416363, rs7617779
19	chr3	11	0.21	0.48	2	1	1	0	T CELL, B CELL	rs9310852, rs4680838, rs9880772, rs1353286
22	chr4	49	0.2	0.29	2	0	0	0	NONE	rs7660626, rs13142500, rs6831973
31	chr6	13	0.17	0.49	2	0	2	2	B CELL	rs2451258, rs2485363, rs654690, rs1994564, rs212389
33	chr7	47	0.24	0.47	2	3	1	2	ALL	rs186735625, rs57585717, rs2158624
36	chr8	20	0.08	0.23	2	3	1	0	FLS, B CELL	rs2736337
38	chr8	11	0.15	0.39	2	1	0	0	FLS	rs678347, rs507201, rs657425
43	chr10	17	0.27	0.45	2	0	3	2	T CELL, B CELL	rs947474, rs10796038, rs10796040
45	chr10	44	0.36	0.54	2	0	2	0	NONE	rs12413578, rs144536148, rs186856025
47	chr11	59	0.47	0.53	2	4	2	5	NONE	rs12574838, rs331463
51	chr11	17	0.12	0.36	2	2	8	0	NONE	rs4936059, rs11221402, rs7106876
54	chr12	22	0.22	0.51	2	0	1	1	T CELL, B CELL	rs10774624, rs3184504, rs7310615
56	chr14	12	0.13	0.37	2	3	2	0	ALL	rs1950897, rs911263, rs1885013, rs2104047, rs3784099
57	chr14	21	0.22	0.45	2	7	10	6	ALL	rs7146217, rs36045050, rs11158764
58	chr15	17	0.25	0.48	2	1	12	2	T CELL, B CELL	rs8032939, rs8043085, rs4924273
61	chr16	18	0.08	0.23	2	0	0	0	NONE	rs11075010, rs1579258, rs4584833
63	chr17	39	0.49	0.67	2	0	0	6	NONE	rs7224929, rs58483057, rs2071456
67	chr20	11	0.32	0.59	2	1	1	1	ALL	rs4239702, rs4810485, rs1883832
71	chr21	37	0.32	0.6	2	0	26	0	T CELL	rs1893592, rs225433, rs11203203
72	chr22	103	0.24	0.26	2	9	20	22	T CELL, B CELL	rs11089637, rs11089620, rs5754387

All rights reserved. No reuse allowed without permission.

**Table 4.** Loci which were assigned to FLS based on posterior probability

Locus	Chr	SNP category	FLS, T, B, ALL, NONE (based on posteriors of SNPs in enhancers)	Top posterior rs numbers	Candidate genes (proximal)	Candidate genes (interacting)	FLS basal proximal genes	FLS basal distal genes	FLS stim. proximal genes	FLS stim. distal genes
3	chr1	2	FLS	rs28411352, rs28489009, rs2306627	POU3F1; MANEAL	RSPO1; RHBDL2	INPP5B; YRDC; C1orf122; MTF1; SF3A3; FHL3	MTF1; GNL2; RRAGC; MYCBP; YRDC; C1orf122; SF3A3	YRDC; C1orf122; MTF1; SF3A3	MTF1; CDCA8; RRAGC; MYCBP; YRDC; C1orf122
4	chr1	1	FLS	rs4839319, rs4839318, rs77227025	PTPN22	RLIMP2; SYT6; BCL2L15	MAGI3; RSNB1; AP4B1; PHTF1; AP4B1-AS1; HIPK1; OLFML3	HIPK1; AP4B1-AS1; OLFML3	MAGI3; AP4B1-AS1; HIPK1; OLFML3	HIPK1; AP4B1-AS1; OLFML3
10	chr2	1	ALL	rs10175798, rs10173253, rs906868, rs7579944, rs1355208			LBH	LCLAT1	LBH	LCLAT1
11	chr2	2	ALL	rs34695944, rs56095903,			REL		REL	REL; PUS10
12	chr2	3	FLS	rs1858037, rs1858036, rs11673987, rs11126035, rs906577			SPRED2	ACTR2; SPRED2	SPRED2	ACTR2; SPRED2; AFTPH; PPP3R1
20	chr3	1	ALL	rs73081554, rs185407974, rs180977001, rs35677470, rs114584537	DNASE1L3; HTD2	ALCOX2	FLNB; FLNB-AS1; ABHD6; RPP14; PXX	PXX; SLMAP; FLNB; FLNB-AS1; PDHB	FLNB; FLNB-AS1; PXX	PXX; FLNB; FLNB-AS1; PDHB; SLMAP
23	chr4	3	FLS	rs34046593, rs36020664, rs11933540, rs6448434, rs6448432		RBPJ; SMIM20; TBC1D19; STIM2		RBPJ		RBPJ
25	chr5	3	FLS	rs2561477		PDZPHIP	C5orf30			
27	chr6	3	FLS	rs2234067, rs1885205, rs916287, rs4713969, rs879036				STK38; SRSF3		STK38; SRSF3; MAPK14
28	chr6	3	ALL		TCTE1	SPATS1; VEGFA	AARS2; NFKBIE	HSP90AB1; SLC35B2	NFKBIE	HSP90AB1; SLC35B2; NFKBIE
33	chr7	2	ALL	rs186735625, rs57585717, rs2158624		HOXA11; HOTAIRM1	JAZF1	JAZF1; HOXA10; HOXA11-AS; CREB5	JAZF1	JAZF1; HOXA11-AS; CREB5
34	chr7	3	FLS, B CELL	rs4272, rs8179, rs42034		SAND9; HEPACAM2; VPS50	CDK6	CDK6; SAMD9; SAMD9L	CDK6	CDK6; SAMD9; SAMD9L
36	chr8	2	FLS, B CELL	rs2736337	BLK	TDH; GATA4; DEFB				
38	chr8	2	FLS	rs678347, rs507201, rs657425	GRHL2			RRM2B		
41	chr9	3	ALL	rs10985070		C5	C5; PHF19; TRAF1	PHF19; MEGF9; CNTRL; FBXW2; PSMD5; GSN; GSN-AS1	PHF19; TRAF1	PHF19; TRAF1; CNTRL; FBXW2; GSN; GSN-AS1; PSMD5; STOM
46	chr10	1	ALL	rs12764378, rs71508903,		RTKN2	ARID5B	ARID5B	ARID5B	ARID5B
48	chr11	3	ALL	rs968567, rs7943728, rs61896141, rs61897793, rs61897795	CD5	PTGDR2	FADS2; SLC15A3; TKFC; INCENP; CCDC86; PRPF19; TMEM109; TMEM132A; VPS37C; DDB1; CYB561A3; TMEM138; CPSF7; MYRF; FEN1; FADS1; FADS3; BEST1; FTH1	FADS2; FADS1; MYRF; FEN1; INCENP; FADS3; TMEM109; RAB31L1; CCDC86; PRPF19; TMEM109	FADS2; SLC15A3; TKFC; INCENP; CCDC86; PRPF19; TMEM109; TMEM132A; VPS37C; DDB1; CYB561A3; TMEM138; CPSF7; MYRF; FEN1; FADS1; FADS3; BEST1; FTH1	FADS2; FADS1; MYRF; FEN1; VPS37C; INCENP; FADS3; BEST1; FTH1; CCDC86; PRPF19; TMEM109; TMEM132A; SLC15A3
56	chr14	2	ALL	rs1950897, rs911263, rs1885013, rs2104047, rs3784099			RAD51B	ZFP36L1		ZFP36L1
57	chr14	2	ALL	rs7146217, rs36045050, rs11158764			RAD51B; ZFYVE26; ZFP36L1	ZFP36L1	ZFYVE26; ZFP36L1	ZFP36L1
67	chr20	2	ALL	rs4239702, rs4810485, rs1883832			CD40	SLC35C2; ELMO2		ELMO2; SLC35C2; NCOA5
68	chr21	3	ALL	rs73194058, rs11702844, rs11700997		IL10RB	SON; IFNGR2; ITSN1; TMEM50B; GART; DONSON; CRYZL1	IFNGR2; GART; SON; ITSN1; IFNAR1; PAXBP1-AS1; IL10RB-DT; DONSON; NDUUF3; ATP5PO	SON; IFNGR2; ITSN1; GART; DONSON	IFNGR2; GART; SON; ITSN1; IFNAR2; IFNAR1; PAXBP1; IL10RB-DT; IL10RB; DONSON; PTER
69	chr21	3	FLS	rs7278771, rs7283600, rs147868091			RCAN1	RCAN1	RCAN1	RCAN1
70	chr21	1	FLS	rs8133843, rs8129030, rs9979383		SET4; PPP1R2P2		RUNX1		RUNX1; CLIC6

**Table 5.** Loci which were assigned to T and/or B cells based on posterior probability

Locus	Chr	SNP category	FLS, T, B, ALL, NONE (based on posteriors of SNPs in enhancers)	Top posterior rs numbers	Jurkat T cells proximal genes	Jurkat T cells distal genes	GM12878 B cells proximal genes	GM12878 B cells distal genes
5	chr1	1	B CELL	rs6679677, rs2476601	RSBN1; PHTF1; PTPN22	RSBN1; PHTF1; PTPN22; AP4B1; HIPK1-AS1; HIPK1; OLFML3	RSBN1; PHTF1; PTPN22	RSBN1; PTPN22; PHTF1; AP4B1-AS1; BCL2L15; AP4B1; HIPK1-AS1; HIPK1; OLFML3; LRIG2
6	chr1	2	B CELL	rs1217404, rs2476604, rs1217420	RSBN1; PHTF1; PTPN22; AP4B1-AS1	RSBN1; PHTF1; PTPN22; AP4B1; HIPK1-AS1; HIPK1; OLFML3	RSBN1; PHTF1; PTPN22; AP4B1-AS1	PTPN22; RSBN1; PHTF1; AP4B1-AS1; BCL2L15; AP4B1; HIPK1-AS1; HIPK1; OLFML3; LRIG2; FRMD8; ARL13B; STX19
8	chr1	3	B CELL	rs4657041, rs1801274, rs6671847	FCGR2A	FCER1G; NDUFS2; SDHC; MPZ; CFAP126; FCGR2A; FCGR2B; FCRLA; FCRLB; RN7SL466P; DUSP12; ATF6; PCP4L1; ADAMTS4	FCGR2A	SDHC; MPZ; FCGR2A; FCGR2B; FCRLA; FCRLB; RN7SL466P; DUSP12; ATF6; CFAP126; RNU6-481P; ADAMTS4; NDUFS2
10	chr2	1	ALL	rs10175798, rs10173253, rs906868, rs7579944,	LBH	LBH	LBH	LBH; LCLAT1
11	chr2	2	ALL	rs34695944, rs56095903, rs67574266, rs13031237, rs13031721	LINC01185; REL; RNU4-51P	LINC01185; REL; RNU4-51P; PUS10; PAPOLG; RN7SL632P; RNA5SP95; B3GNT2; RNU6-612P	LINC01185; REL; RNU4-51P	LINC01185; REL; RNU4-51P; PUS10; PAPOLG; RN7SL632P; RNA5SP95; B3GNT2; RNU6-612P
14	chr2	3	T CELL	rs13426947, rs3024859, rs7568275, rs11889341,	STAT4	STAT4; RNU6-959P; MYO1B	STAT4	STAT4; RNU6-959P
16	chr2	3	T CELL	rs1980421, rs1980422, rs7588874, rs7422494,	CD28; RNU6-474P; CTLA4	CD28; RNU6-474P; CTLA4; RAPH1; ABI2	CD28; RNU6-474P; CTLA4	CD28; CTLA4; RNU6-474P; RAPH1; PRKG1
17	chr2	1	T CELL	rs231724, rs231723, rs231775	CTLA4	CTLA4; CD28; RNU6-474P; RAPH1	CTLA4	CD28; CTLA4; RNU6-474P; RAPH1
19	chr3	2	T CELL, B CELL	rs9310852, rs4680838, rs9880772, rs1353286	EOMES	LINC02084; EOMES; LINC01967; CMC1; AZI2; ZCWPW2; NEK10; LINC01980	EOMES	LINC02084; EOMES; CMC1; AZI2; ZCWPW2
20	chr3	1	ALL	rs73081554, rs185407974, rs180977001, rs35677470, rs114584537	FLNB; DNASE1L3; FLNB-AS1; ABHD6; RPP14; HTD2; PXK	DNASE1L3; ABHD6; RPP14; HTD2; PXK; PDHB; KCTD6; ACOX2; FAM107A; FAM3D-AS1; FAM3D; FLNB	FLNB; DNASE1L3; FLNB-AS1; ABHD6; RPP14; HTD2; PXK	DNASE1L3; ABHD6; RPP14; HTD2; PXK; FLNB; FLNB-AS1; PDHB; KCTD6; ACOX2; FAM107A; FAM3D-AS1; FAM3D
24	chr5	1	T CELL	rs7731626	ANKRD55	ANKRD55; RNA5SP184; IL6ST	ANKRD55	ANKRD55; RNU6-299P; RNA5SP184; IL31RA; IL6ST
28	chr6	3	ALL		TMEM151B; AARS2; NFKBIE; TCTE1	TMEM151B; AARS2; TCTE1; HSP90AB1; SLC35B2; MIR4647; NFKBIE; SPATS1; CAPN11; TMEM63B; RN7SL811P	TMEM151B; AARS2; NFKBIE; TCTE1	TMEM151B; AARS2; TCTE1; NFKBIE; HSP90AB1; SLC35B2; MIR4647; MRPL14; TMEM63B; SPATS1; TRIM38; CAPN11; MTX2
31	chr6	2	B CELL	rs2451258, rs2485363, rs654690, rs1994564, rs212389		RSPH3; TAGAP; SYTL3; C6orf99		RSPH3; TAGAP; SYTL3; C11orf44
32	chr6	1	T CELL, B CELL	rs1571878, rs3093017, rs10946216	CCR6	CCR6; RPS6KA2	CCR6	CCR6; SFT2D1; RPS6KA2; RNASET2; MIR3939; FGFR10P; GPR31; PPIL4
33	chr7	2	ALL	rs186735625, rs57585717, rs2158624	JAZF1; JAZF1-AS1; RNU6-979P	JAZF1; RNU6-979P; JAZF1-AS1; HOTTIP; HOXA1; HOTAIRM1; HOXA3; HOXA-AS2; HOXA4; HOXA5; HOXA6; HOXA-AS3; HOXA7; HOXA9; HOXA10-AS; MIR196B; HOXA10; HOXA11; HOXA11-AS; EVX1-AS; EVX1	JAZF1; JAZF1-AS1; RNU6-979P	JAZF1; RNU6-979P; JAZF1-AS1; HOTTIP; HOXA11; HOXA11-AS; HOXA1; HOTAIRM1; HOXA3; HOXA-AS2; HOXA4; HOXA5; HOXA-AS3; HOXA7; HOXA9; HOXA10-AS; MIR196B; HOXA10; EVX1-AS; EVX1; HOXA6; PARP9
34	chr7	3	FLS, B CELL	rs4272, rs8179, rs42034	CDK6	CDK6; PEX1; RBM48; FAM133B; CDK6-AS1; SAMD9; VPS50; HEPACAM2	CDK6	CDK6; FAM133B; SAMD9; CDK6-AS1

Locus	Chr	SNP category	FLS, T, B, ALL, NONE (based on posteriors of SNPs in enhancers)	Top posterior rs numbers	Jurkat T cells proximal genes	Jurkat T cells distal genes	GM12878 B cells proximal genes	GM12878 B cells distal genes
35	chr7	3	B CELL	rs3778753, rs12539741, rs12706861	TNPO3; IRF5; RN7SL306P		TNPO3; IRF5; RN7SL306P	
36	chr8	2	FLS, B CELL	rs2736337	BLK	BLK; FAM167A; MTMR9; FAM167A-AS1; LINC00208; GATA4; C8orf49; NEIL2; FDF1; CTSB; DEFB136; DEFB135; DEFB134	BLK	BLK; FAM167A; AGPAT5; MIR4659A; MIR4659B; XKR6; LINC00529; MTMR9; FAM167A-AS1; LINC00208; GATA4; C8orf49; NEIL2; FDF1; CTSB; PRAG1
39	chr8	3	T CELL, B CELL	rs1516971	LINC00824; MIR1208; RN7SKP226	LINC00824; CCDC26	LINC00824; MIR1208; RN7SKP226	LINC00824; PCAT1
41	chr9	3	ALL	rs10985070	C5; PHF19; TRAF1	PHF19; TRAF1; C5; MEGF9; FBXW2; B3GNT10; PSMD5; CNTRL; RAB14; RN7SL181P; GSN; STOM	C5; PHF19; TRAF1	PHF19; TRAF1; C5; RGS3; AKNA; CDK5RAP2; MEGF9; FBXW2; B3GNT10; PSMD5; CNTRL; RAB14; RN7SL181P; GSN; STOM
42	chr10	1	T CELL	rs706778, rs10795791	IL2RA	IL2RA; LINC02656; RBM17; PFKFB3; LINC02649	IL2RA	IL2RA; LINC02656; FBH1; IL15RA; RBM17; GDI2; ANKRD16; PFKFB3; RN7SKP78; MIR3155A; MIR3155B; LINC02649; TASOR2
43	chr10	2	T CELL, B CELL	rs947474, rs10796038, rs10796040	PRKCO; LINC02656	IL2RA; LINC02656; RBM17; PFKFB3; RN7SKP78; MIR3155A; MIR3155B; LINC02649	PRKCO; LINC02656	LINC02656; IL2RA; RBM17; PFKFB3; LINC02649
44	chr10	1	T CELL	rs537544, rs568727, rs570613, rs570730, rs7897792	GATA3	RBM17	GATA3	RBM17
46	chr10	1	ALL	rs12764378, rs71508903, rs77509998	ARID5B	ARID5B; RTKN2; LINC02621	ARID5B	ARID5B; RTKN2; LINC02621
48	chr11	3	ALL	rs968567, rs7943728, rs61896141, rs61897793, rs61897795	FADS2; SLC15A3; PGA3; TKFC; DAGLA; INCENP; MS4A8; MS4A18; MS4A15; MS4A10; CCDC86; PTGDR2; ZP1; PRPF19; TMEM109; TMEM132A; CD6; RNU6-933P; CD5; VPS37C; PGA4; PGA5; VWCE; DDB1; CYB561A3; TMEM138; TMEM216; CPSF7; SDHAF2; RN7SL23P; PPP1R32; MIR4488; LRRC10B; SYT7; MYRF-AS1; MYRF; TMEM258; MIR611; FEN1; FADS1; MIR1908; FADS3; MIR6746; RAB31L1; RNU6-1243P; BEST1; FTH1; LINC02733; SCGB1D1; SCGB2A1; SCGB1D2; SCGB2A2; SCGB1D4	FADS2; FADS1; MIR1908; CD5; VPS37C; FADS3; MIR6746; LINC02733; MS4A15; MS4A10; CCDC86; PTGDR2; ZP1; PRPF19; TMEM109; TMEM132A; SLC15A3; CD6; RNU6-933P; TMEM216; VWCE; PGA5; CPSF7; SDHAF2	FADS2; SLC15A3; PGA3; TKFC; DAGLA; INCENP; MS4A8; MS4A18; MS4A15; MS4A10; CCDC86; PTGDR2; ZP1; PRPF19; TMEM109; TMEM132A; CD6; RNU6-933P; CD5; VPS37C; PGA4; PGA5; VWCE; DDB1; CYB561A3; TMEM138; TMEM216; CPSF7; SDHAF2; RN7SL23P; PPP1R32; MIR4488; LRRC10B; SYT7; MYRF-AS1; MYRF; TMEM258; MIR611; FEN1; FADS1; MIR1908; FADS3; MIR6746; RAB31L1; RNU6-1243P; BEST1; FTH1; LINC02733; SCGB1D1; SCGB2A1; SCGB1D2; SCGB2A2; SCGB1D4	TMEM258; MIR611; FEN1; FADS2; FADS1; MIR1908; CD5; VPS37C; PTPN2; FADS3; MIR6746; RNU6-1243P; BEST1; FTH1; LINC02733; INCENP; MS4A15; MS4A10; CCDC86; PTGDR2; ZP1; PRPF19; TMEM109; TMEM132A; SLC15A3; CD6; RNU6-933P; PGA5; VWCE; TMEM216; MS4A18
50	chr11	3	T CELL, B CELL	rs10790268	DDX6	DDX6; CXCR5; RNU6-376P; BCL9L; MIR4492; UPK2; FOXR1; ARCN1; PHLDB1; MIR6716; TREH; RN7SL688P	DDX6	DDX6; CXCR5; TREH; RNU6-376P; BCL9L; MIR4492; UPK2; RN7SL688P; FOXR1; COG8; VPS4A; PDF; GRK3
54	chr12	2	T CELL, B CELL	rs10774624, rs3184504, rs7310615	CUX2; SH2B3; ATXN2; BRAP; ALDH2; MIR6760; RNA5SP373; PHETA1; LINC02356; ATXN2-AS; ACAD10; MIR6761	LINC02356; SH2B3; ATXN2; ATXN2-AS; BRAP; ACAD10; ALDH2; MIR6761; MAPKAPK5; TMEM116; ERP29; NAA25; MIR3657; TRAFD1; HECTD4; RN7SKP71; RPL6; MAPKAPK5-AS1; MIR6861; CUX2; PHETA1	CUX2; SH2B3; ATXN2; BRAP; ALDH2; MIR6760; RNA5SP373; PHETA1; LINC02356; ATXN2-AS; ACAD10; MIR6761	LINC02356; SH2B3; ATXN2; ATXN2-AS; BRAP; ACAD10; ALDH2; TMEM116; ERP29; NAA25; MIR3657; HECTD4; RPL6; MIR6761; MAPKAPK5; MIR6861; RN7SKP71; MAPKAPK5-AS1; TRAFD1; RPH3A; CUX2; PHETA1

All rights reserved. No reuse allowed without permission.

Locus	Chr	SNP category	FLS, T, B, ALL, NONE (based on posteriors of SNPs in enhancers)	Top posterior rs numbers	Jurkat T cells proximal genes	Jurkat T cells distal genes	GM12878 B cells proximal genes	GM12878 B cells distal genes
56	chr14	2	ALL	rs1950897, rs911263, rs1885013, rs2104047, rs3784099	RAD51B	RAD51B; ZFP36L1; MAGOHB; RNU6-921P	RAD51B	RAD51B; ZFP36L1; RNU6-921P; RN7SL108P
57	chr14	2	ALL	rs7146217, rs36045050, rs11158764	RAD51B; ZFYVE26; RN7SL706P; RN7SL108P; RNU6-921P; ZFP36L1	RAD51B; ZFP36L1; RNU6-921P; MAGOHB	RAD51B; ZFYVE26; RN7SL706P; RN7SL108P; RNU6-921P; ZFP36L1	RAD51B; ZFP36L1; RN7SL108P; RNU6-921P
58	chr15	2	T CELL, B CELL	rs8032939, rs8043085, rs4924273	RASGRP1	RASGRP1; FAM98B	RASGRP1	RASGRP1; FAM98B
59	chr15	3	B CELL	rs35187679, rs7181071, rs71404230	DRAIC	DRAIC	DRAIC	DRAIC; SUCLG2-AS1; PAK5
62	chr16	1	B CELL	rs13330176, rs2139492, rs2139493		EMC8; COX4I1		EMC8; COX4I1; IRF8; MIR6774; LINC02132; FOXF1; MTHFSD; RNU1-103P; PTP4A3
67	chr20	2	ALL	rs4239702, rs4810485, rs1883832	CD40	CD40; MMP9; SLC12A5-AS1; SLC12A5; NCOA5; CDH22; SLC35C2; ELMO2	CD40	CD40; MMP9; SLC12A5-AS1; SLC12A5; NCOA5; ELMO2; SLC35C2
68	chr21	3	ALL	rs73194058, rs11702844, rs11700997	SON; IFNGR2; ITS1N1; TMEM50B; DNAJC28; GART; MIR6501; DONSON; CRYZL1	IFNGR2; TMEM50B; DNAJC28; GART; SON; MIR6501; DONSON; CRYZL1; LINC01548; IFNAR2; IFNAR1	SON; IFNGR2; ITS1N1; TMEM50B; DNAJC28; GART; MIR6501; DONSON; CRYZL1	IFNGR2; TMEM50B; DNAJC28; GART; SON; MIR6501; DONSON; CRYZL1; IFNAR2; IL10RB; IFNAR1; SYNJ1; IL10RB-DT; RABEP1
71	chr21	2	T CELL	rs1893592, rs225433, rs11203203	TMPRSS3; UBASH3A; RNU6-1149P	UBASH3A; ZBTB21; ZNF295-AS1	TMPRSS3; UBASH3A; RNU6-1149P	UBASH3A; ABCG1; TFF3; TFF2; TMPRSS3; RNU6-1149P
72	chr22	2	T CELL, B CELL	rs11089637, rs11089620, rs5754387	MAPK1; IGLV10-54; HIC2; TMEM191C; RN7SKP221; RIMBP3C; UBE2L3; YDJC; CCDC116; SDF2L1; MIR301B; MIR130B; PPIL2; YPEL1; RN7SL280P; RNA5SP493; PPM1F; TOP3B; IGLV4-69; IGLV8-61	UBE2L3; RIMBP3C; YDJC; CCDC116; HIC2; MIR301B; MIR130B; PPIL2; MAPK1; AIFM3; LZTR1; SDF2L1; PPM1F	MAPK1; IGLV10-54; HIC2; TMEM191C; RN7SKP221; RIMBP3C; UBE2L3; YDJC; CCDC116; SDF2L1; MIR301B; MIR130B; PPIL2; YPEL1; RN7SL280P; RNA5SP493; PPM1F; TOP3B; IGLV4-69; IGLV8-61	UBE2L3; RIMBP3C; YDJC; CCDC116; HIC2; SDF2L1; MIR301B; MIR130B; STAR9; HAUS2
73	chr22	1	B CELL	rs2069235, rs909685, rs9611155	SYNGR1	SYNGR1; MIEF1; CBX7; TAB1	SYNGR1	SYNGR1; TAB1; MGAT3; MIEF1; ATF4; RPS19BP1