Bioinformatics analysis of the factors controlling type I IFN gene expression in autoimmune disease and virus-induced immunity

Di Feng [1,2] * and Betsy J. Barnes [1,2]

[1] Department of Biochemistry and Molecular Biology, Rutgers Biomedical and Health Sciences, Newark, NJ, USA
[2] Rutgers Biomedical and Health Sciences, New Jersey Medical School-Cancer Center, University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

Patients with systemic lupus erythematosus (SLE) and Sjögren's syndrome (SS) display increased levels of type I interferon (IFN)-induced genes. Plasmacytoid dendritic cells (PDCs) are natural interferon producing cells and considered to be a primary source of IFN-α in these two diseases. Differential expression patterns of type I IFN-inducible transcripts can be found in different immune cell subsets and in patients with both active and inactive autoimmune disease. A type I IFN gene signature generally consists of three groups of IFN-induced genes – those regulated in response to virus-induced type I IFN, those regulated by the IFN-induced mitogen-activated protein kinase/extracellular-regulated kinase (MAPK/ERK) pathway, and those by the IFN-induced phosphoinositide-3 kinase (PI-3K) pathway. These three groups of type I IFN-regulated genes control important cellular processes such as apoptosis, survival, adhesion, and chemotaxis, that when dysregulated, contribute to autoimmunity. With the recent generation of large datasets in the public domain from next-generation sequencing and DNA microarray experiments, one can perform detailed analyses of cell-type specific gene signatures as well as identify distinct transcription factors (TFs) that differentially regulate these gene signatures. We have performed bioinformatics analysis of data in the public domain and experimental data from our lab to gain insight into the regulation of type I IFN gene expression. We have found that the genetic landscape of the *IFNA* and *IFNB* genes are occupied by TFs, such as insulators CTCF and cohesin, that negatively regulate transcription, as well as interferon regulatory factor (IRF)5 and IRF7, that positively and distinctly regulate *IFNA* subtypes. A detailed understanding of the factors controlling type I IFN gene transcription will significantly aid in the identification and development of new therapeutic strategies targeting the IFN pathway in autoimmune disease.

Keywords: type I interferons, bioinformatics, autoimmunity, transcriptional regulation, transcription, genetic

INTRODUCTION

Patients with autoimmune diseases, such as systemic lupus erythematosus (SLE) and Sjögren's syndrome (SS), display increased expression of type I interferon (IFN)-induced genes. Plasmacytoid dendritic cells (PDC), as natural IFN-producing cells, are considered to be a primary source of IFN-α in such diseases (1, 2). The type I IFN family consists of multiple members, including 14 IFN-α subtypes, -β, -ε, -κ, -ω, -δ, and -τ. These members may have autocrine effects on the IFN-producing cells themselves, such as PDCs, and paracrine effects on neighboring cells, as well as systemic effects on distant immune cells (3). IFNs can be added directly to cell cultures and molecular profiling performed to understand their biologic effect. For instance, the direct treatment of peripheral blood mononuclear cells (PBMCs) with 0.6 pM of IFN-α, -β, or IFN-ω led to the increased expression of about 200 genes (4). Broadly speaking, an IFN gene signature should include all of these genes. These genes can be functionally classified into antiviral pathways, apoptosis control, cell surface receptor expression, chemokine/cytokine expression, and components of IFN signaling pathways.

Although methods of bioinformatics analysis are not yet intensively used in immunology research, the field is changing fast and significant information can now be obtained from the public domain for the analysis of mechanisms controlling type I *IFN* gene expression. This report explores several elements of translational bioinformatics analysis, specifically addressing the biological questions relevant to how type I IFN expression is regulated in autoimmune disease. We collected publically available microarray gene expression datasets in Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI) and performed data mining and pathway analysis. With the growing datasets in public repository that are shared in the research community, the integrative analysis of experimental data and disease profiling data sets has become an important approach to our understanding of autoimmune disease pathology at the molecular level. In this study, we have also used human datasets from

the Encyclopedia of DNA Elements (ENCODE) to understand the epigenetic codes that control the type I IFN gene cluster. This information can be used as a reference to guide future experiments that focus on epigenetic changes in more relevant human immune cell populations such as monocytes and dendritic cells. Understanding the regulation and epigenetic control of type I IFN expression will be useful for the development of new therapeutic interventions targeting the IFN pathway in autoimmune disease.

MATERIALS AND METHODS

MATERIALS
Gene expression microarray data were retrieved from NCBI's GEO through series accession numbers GSE17762 and GSE10325. Data were loaded with GEO query and limma R packages from the Bioconductor project. Alternatively, GEO2R, an interactive web tool, was used. Next-generation sequencing datasets from multiple cell lines and cell types were retrieved from the ENCODE Project[1].

METHODS
In brief, for the analysis of microarray data, gene symbols and value of log fold changes for individual genes were extracted from NCBI's GEO and Ingenuity IPA software was used to perform pathway analysis. For next-generation sequencing datasets, ENCODE offers a few software tools for analyzing the data. One relevant tool is factor book, which organizes all the information associated with individual transcription factors (TFs) (5). Although useful, it should be noted that the current lack of information on human primary immunocytes limits one's ability to analyze individual genes/gene clusters and therefore limits the value and/or relevance of some of these datasets.

The following information provides a brief summary of methods used for the analysis of next-generation sequencing data. For example, the epigenome analysis of the *IFNA* gene cluster was performed using a variety of resources for data visualization. In brief, the genetic region was located and retrieved in UCSC genome browser using URL http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr9:21000000-21550000. Methylated/unmethylated CpGs data was retrieved using Methylation-sensitive restriction enzyme sequencing (MRE-seq) and MeDIP-seq loaded from http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=ucsfBrainMethyl. Methyl Reduced Representation Bisulfite Sequencing (RRBS) tracks were loaded from http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=wgEncodeHaibMethylRrbs, samples used include all cells in the following list: http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=342586899&c=chr9&g=wgEncodeRegTfbsClusteredV2. Histone modification data, including H3K4me3 was loaded from http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=342586899&c=chr9&g=wgEncodeReg. For the analysis of CTCF and other relevant TFs, we selected TFs and cell types by adding tracks from http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeAwgTfbsUniform. TF binding peaks were either calculated using ENCODE pre-processed data with a False Discovery Rate of 1% or mapped to human genome hg37 using CLC Genomics Workbench

---

[1]http://genome.ucsc.edu/ENCODE/

software 5.5, followed by peak calling using Model-based Analysis for ChIP-Seq (MACS).

RESULTS AND DISCUSSION

RELATIONSHIP BETWEEN THE TYPE I IFN GENE SIGNATURE AND CLINICAL AUTOIMMUNE BIOMARKERS
We have performed an in depth bioinformatics analysis of genes regulated by type I IFNs, as well as the mechanisms controlling type I IFN expression, in autoimmune diseases using publically available datasets. In many cases, we found that IFN-induced genes directly explain the presence of clinical biomarkers that appear in patients with autoimmune diseases. For example, we found that IFN-α increases the expression of interleukin (IL)-15 and its receptor IL-15Rα in PBMCs. IL-15, that is primarily expressed by activated monocytes and dendritic cells, binds to IL-15Rα (CD359) on accessory cells and is trans-presented to T cells that express functional IL-15Rα, composed of IL-2/15Rβ (CD122) and γc chains. Several groups have reported elevated IL-15 levels in the sera of SLE patients, however, the functional consequence of IL-15Rα activation in SLE remains to be studied (6). In addition to IL-15 and IL-15Rα, IFN-β moderately upregulates *IL-7* and *CD59* transcripts in PBMCs. IL-7 is a survival factor for naïve, early effector, and memory CD4+ and CD8+ T cells. It is primarily produced by fibroblastic reticular cells (FRCs), a mesenchymal cell population found in the stromal environment of lymphoid organs. In SLE patients, soluble (s)IL-7R concentrations were found to be elevated in the serum and raised levels of sIL-7 were detected in patients with lupus nephritis (LN) that reflected activation of kidney tissue cells (7). Receptor blockade by anti-IL-7Rα in MRL-*Fas*^lpr^ lupus mice resulted in alleviation of dermatitis, lymphadenopathy, splenomegaly, and total serum IgG2a; yet, only a marginal reduction in IgG2a autoantibodies was found (8). CD59 are glycosylphosphatidylinositol-anchored proteins with complement inhibitory properties that prevent the terminal polymerization of the membrane attack complex. Increased numbers of CD55- and CD59-lymphocytes and CD59-granulocytes were found in SLE patients as compared with controls (9).

PATHWAY ACTIVATION BY TYPE I IFNs
Type I IFNs may play a pathological role in autoimmune disease through their ability to regulate key signaling pathways important in the innate immune response. For instance, we found that IFN-α upregulates the expression of Toll-like receptors (*TLR*)-3 and *TLR-7*, as well as the critical cofactor myeloid differentiation primary response protein 88 (*MyD88*). IFN-α also enhances the expression of interferon regulatory factor (*IRF*)2, which competitively inhibits IRF1-mediated transcriptional activation of *IFNA* and *B* genes. As compared to IFN-α, the effect of IFN-β on gene expression extends to *TLR-1*, *TRAF/TANK*, *IRF4*, and *IRF1*. We also found in our analysis that the human dual specificity mitogen-activated protein kinase kinase 5 (*MAP2K5*) can be upregulated by IFN-α/IFN-β and mitogen-activated protein kinase kinase 8 (*MAP3K8*) can be induced by IFN-β. Since p38 MAPK acts up-stream of type I IFN-induced STAT (signal transducers and activators of transcription) 1 signaling (10, 11), the up-regulation of *MAP3K8* or *MAP2K5* may provide further hints toward the biologic effects of type I IFN on cells. For example, MAP3K8 has

been shown to promote the production of tumor necrosis factor (TNF)-α and IL-2 during T lymphocyte activation. It is also known that addition of IFN-α with anti-CD3 antibodies results in enhanced T helper (Th)1 responses that associate with enhanced phosphorylation of STAT1 (12).

It is well-known that IFN-α has pro-apoptotic effects in many cancer cell types including myeloma (13), renal cell carcinoma (14), and glioma (15). It is also known that monocytes stimulated with IFN-α express functional TNF-related apoptosis-inducing ligand (TRAIL), which is capable of killing myeloma cells (16). IFN-α also increases the expression of functional FasL exclusively on natural killer (NK) cells (17). The functional clustering of genes regulated by IFN-β, using DAVID tools, revealed a number of genes that control apoptosis, including *caspase 1*, *8*, and *10*, *TRAIL* (*TNFSF10*), and *FADD [Fas (TNFRSF6)-associated via death domain]*.

THE TYPE I IFN GENE SIGNATURE IN SLE B CELLS AND T CELLS

Disease biomarkers or disease gene signatures provide important clues for our understanding of disease pathogenesis and aid in the identification and development of new therapeutic strategies for treatment. High-throughput screening technologies, such as DNA microarrays, have been used to profile disease signatures in PBMCs from SLE patients (18), and subsequently, in specific subsets such as monocytes, neutrophils, T cells, and B cells. The presence of a type I IFN gene signature in PBMC of SLE patients has been recognized for nearly 35 years now (19). However, not all IFN-inducible genes that have been identified by *in vitro* assays can be detected *in vivo* in PBMCs isolated from SLE patients. About 20 IFN-inducible genes were consistently found to be highly expressed in PBMC from SLE patients (18). In our analysis of SLE B and T cells, we found that approximately 10 IFN-inducible genes were consistently and highly expressed. The gene transcriptional signatures that appear to overlap between cell types include *Mx1*, *ISGF-3*, *PRKR*, *IFIT1*, and *IFI44* in cells that have been either exposed to type I IFNs *in vivo* or *in vitro*. This gene signature has been used as a readout for the type I IFN bioassay and is considered a measure of the "IFN-α activity score" in patients with SLE and other inflammatory or autoimmune diseases (19, 20).

Intensive pathway analyses with KEGG[2], BioCarta[3], and GenMAPP[4] have shown up-regulated activation markers on SLE T cells and genes that correlate with STAT1 expression (21). Using IPA[5] analysis of independent datasets, we also found groups of genes in the network that strongly correlate with STAT1, suggesting a persistent and strong effect downstream of type I IFNs in SLE T cells. Furthermore, IFN response factor consensus sequences (ISREs) can be found up-stream of the start sites of each of the genes in the type I IFN gene signature. Our independent analysis also indicated groups of up-regulated genes in SLE T cells that can be modulated by STAT4. Genome-wide mapping of STAT4 and IRF 5 occupancy in immune cells from SLE patients by chromatin immunoprecipitation combined with next-generation sequencing (ChIP-seq) revealed the possible cooperation of high

mobility group-I/Y, specificity protein 1, and paired box 4 with IRF5 and STAT4 in transcriptional regulation (22). As noted above, IFN-regulated pathways derived from *in vitro* data do not always align with microarray datasets obtained from primary cells of SLE patients. In this regard, short-term IFN treatment has been shown to promote apoptosis signaling via TRAIL pathways. However, anti-apoptotic signatures, including elevation of caspase 8 and FADD-like apoptosis regulator (*CFLAR*), were identified in lupus T cells (21). Our bioinformatics pathway analysis identified additional genes, such as *BIRC5*, that participate in the B cell anti-apoptotic pathway in cells isolated from SLE patients. Given that apoptosis and the clearance of apoptotic material have been implicated in SLE pathogenesis, further research detailing the in depth analysis and mapping of these anti-apoptotic pathways in PBMC subsets will be of significant importance to our understanding of SLE pathogenesis.

GENETIC LANDSCAPE OF THE TYPE I IFN CLUSTER

The human type I *IFN* gene cluster spans approximately 450 kb on chromosome 9p22. *IFNB* and *IFNε* define the boundaries of the cluster, with all other type I *IFN* genes, except *IFNk*, distributed between these borders. This gene cluster also contains *KLHL9*, which is a substrate-specific adaptor of the BCR (BTB-CUL3-RBX1) E3 ubiquitin ligase complex that functions in cell division. Studies of virus-induced type I IFN production in murine fibroblasts indicates the presence of an immediate-early response gene, *IFNA4*, which is induced rapidly and without the need for ongoing protein synthesis, and *IFNA2*, *5*, *6*, and *8*, that display delayed induction, are induced more slowly, and require cellular protein synthesis. In CpG-stimulated human PDCs, *IFNA5*, *IFNA10*, *IFNA4*, *1/13*, *21*, *14*, *16*, and *6* transcription can be detected within 2 h. *IFN21* and *IFNA16* levels are dramatically up-regulated further after 8 h suggesting an efficient positive feedback loop regulating expression of these two genes. Recent analysis of data from the ENCODE Consortium suggests that this important gene cluster may be controlled by epigenetic regulation supporting new mechanistic insight and a basis for the design of experiments focused on this aspect of type I IFN gene regulation.

Methylation

Indeed, there has already been significant data in the literature to support the mechanism(s) of epigenetic regulation in autoimmune diseases. In particular, DNA from SLE T cells was found to be less methylated than control DNA from normal T cells by measuring the cellular deoxymethylcytosine content (23). Interestingly, non-T cells from lupus patients displayed normal DNA methylation levels (24). Decreased DNA methyltransferase (DNMT) activity in lupus T lymphocyte nuclear proteins was considered to be responsible for the observed DNA hypomethylation in lupus T cells. Patients with lupus had significantly lower levels of *DNMT1* mRNA, but not *DNMT3A* or *DNMT3B*, as compared with healthy controls (25). A preliminary analysis of microarray data from immature monocyte-derived dendritic cells (MDDCs) revealed that they express abundant amounts of *DNMT1*, which is downregulated after LPS stimulation. The methylation status of DNA from SLE PDCs and the levels of *DNMT1* expression in this important IFN-α producing cell type are not currently known.

In general, hypermethylation in the promoter of a gene is associated with gene suppression, while hypomethylation is linked to gene expression; methylation within the gene body is also associated with gene expression. Two next-generation sequencing technologies have recently been developed for the analysis of gene methylation – methylated DNA immunoprecipitation sequencing (MeDIP)-seq, to detect methylated CpGs (26, 27), and MRE-seq, to detect unmethylated CpGs (28). Integrative methodologies that combine both MeDIP-seq and MRE-seq can differentiate hypermethylation, intermediate, and hypomethylation regions of DNA. An integrative analysis of *KLHL9* indicates that the CpG islands of the *KLHL9* promoter are highly hypomethylated (**Figure 1**). These islands are highly conserved since they were found to be present in virtually all cell types queried. Combining these data with ChIP-seq histone modification data in the same tissues, we found that hypomethylated CpGs of *KLHL9* are occupied by significant levels of trimethylated lysine 4 on histone H3 (H3K4me3) (**Figure 1**). Two other hypomethylated regions in the type I IFN cluster, located in the genomic region between *IFNA2* and *IFNA8*, have relative low levels of enrichment for H3K4me3 peaks (**Figure 1**). H3K4me3 is a histone modification that accumulates at the transcription-start site (TSS) of active genes and is believed to be important for transcription activation. Loss of H3K4me3 occurs at TSSs and leads to gene transcriptional inactivation as a result of promoter hypermethylation. The occupancy of H3K4me3 in the promoter of *KLHL9* may ensure the protection of CpG islands from methylation. In contrast, the other two hypomethylated sites that are located quite far from the TSSs of *IFNA2* and *IFNA8*, may not be functional for transcription. DNA methylation by RRBS from various cell types, including B cells, failed to reveal strong methylation signals in the *IFNA* gene cluster. One exception to this is that MeDIP-seq defined methylation peaks were found to be distributed between *IFNA* genes from brain tissue.

Thus far, data do not support that methylation is the likely major mechanism by which *IFNA* gene expression is suppressed in most non-IFN-producing cells. Further experimental studies will be necessary to determine whether constructive hypomethylation, as well as H3K4me3 occupancy, is important for regulating *IFNA* gene transcription in IFN-producing cells such as monocytes, and PDCs.

Chromatin structure

There are multiple *IFNA* and *IFNB* genomic regions that have open chromatin structure in an evolutionarily conserved pattern across species and most human cell types. Since DNase I hypersensitive sites (DHSs) reflect the local openness and accessibility of chromatin, chromatin structure or accessibility of *IFNA* clustering may be similar among different cells. In general, hypersensitive sites are found only in the chromatin of cells in which the associated gene is being expressed, and do not occur when the gene is inactive. Therefore, mapping DHSs within nuclear chromatin is a powerful method of identifying genetic regulatory elements (29). However, the distribution of DHSs in promoters and other gene regions of similarly expressed genes differs among different chromosomes. Furthermore, silenced genes have a more open chromatin structure than previously thought and DHSs in 3′-untranslated regions (3′-UTRs) have been shown to negatively correlate with gene expression levels (30), thus going against the standard dogma. Bioinformatics analysis of DHSs in the *IFN* gene cluster between different cell types revealed a highly conserved pattern (**Figure 2**);
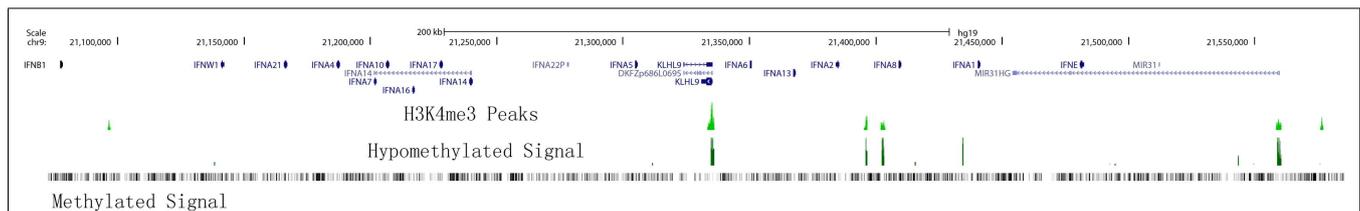

FIGURE 1 | H3K4me3 peaks and methylation tracks on the type I *IFN* gene cluster. Members of the type I *IFN* gene cluster are shown and illustrated proportionally according to Human (*Homo sapiens*) Genome hg19. H3K4me3 peaks and UCSC DNA methylation tracks are shown for a human B cell line.
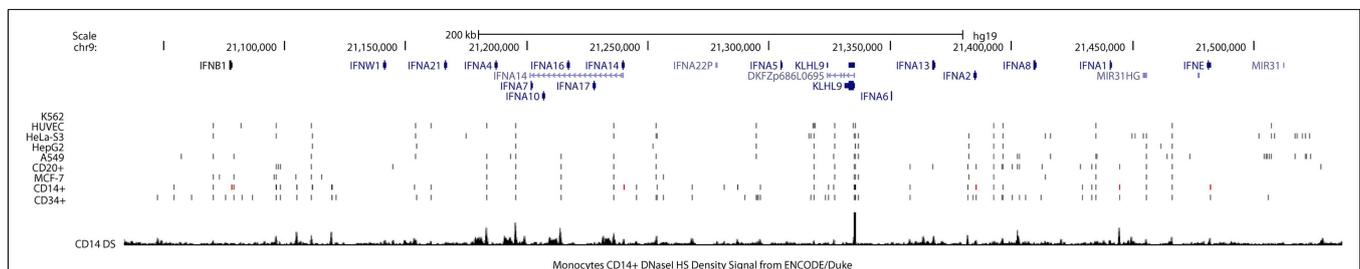

FIGURE 2 | DNase I hypersensitivity sites in the type I *IFN* gene cluster are highly conserved between cell types. Members of the type I *IFN* gene cluster are shown and illustrated proportionally according to Human (*Homo sapiens*) genome hg19. Cell lines and cell types analyzed are listed on the left side. Short vertical lines below the gene track indicate the open chromatin position marked by DNaseI hypersensitivity sites from ENCODE/OpenChromatin (Duke University) for each cell type. Red lines indicate the novel sites identified between cell types. The DNase I hypersensitivity signal peaks for CD14+ monocytes are shown at the bottom for reference to chromatin marks.

however, we found additional DHSs in CD14$^+$ monocytes that can produce type I IFNs. We also found that CD34$^+$ stem cells have more DHSs close to promoters within the *IFN* gene cluster (**Figure 2**). These data support the presence of unique cell-specific chromatin structures which may play important regulatory roles in the control of type I IFN expression.

Histone modification

Modifying the chromatin template at a particular gene locus can also serve as an important mechanism of gene transcriptional activation that exhibits cell-type specific expression patterns. The functional importance of histone acetylation in type I IFN production has been supported by studies that show increased IFN-β expression in cells treated with histone deacetylase inhibitors, such as Trichostatin A (TSA) (31), and decreased IFN-β expression in murine macrophages where the binding of bromodomain-containing BET (bromodomain and extraterminal) transcriptional regulators to acetylated histones was inhibited (32). Di- or tri-methylation of H3K9 is capable of suppressing gene expression not only passively, by inhibiting acetylation, but also actively, by recruiting transcriptional repressors of the heterochromatin protein 1 (HP1) family. We found that H3K9me2 occupancy at *IFN* and *ISG* promoters is inversely correlated with gene expression. Furthermore, human MDDCs that are capable of producing type I IFNs, as compared with human lung fibroblasts that do not, show decreased H3K9me2 occupancy at the *IFNB* promoter. In the absence of G9A, a methyltransferase for H3K9me2, non-professional IFN-producing cells were shown to be converted into potent IFN-β producers (33). Together, these data support the importance of histone modifications in the regulation of type I IFN expression.
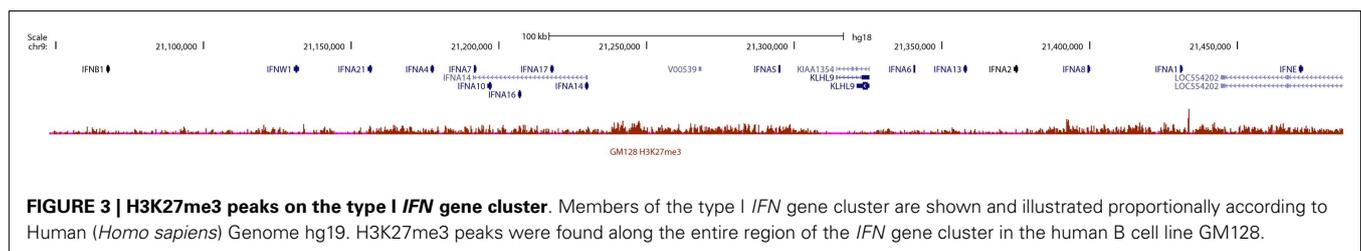
H3K27me3, on the other hand, are found to be associated with the repression of gene transcription in a cell-type specific manner. Polycomb Repressive Complex 2 (PRC2) is a histone methyltransferase that catalyzes tri-methylation of Histone 3 at Lysine 27 (H3K27me3) (34). A detailed profile of H3K27me3 peaks reveal that broad peaks at TSS are associated with transcriptional suppression while skewed peaks up-stream of the TSS may not be suppressive (35). Indeed, we found that *IFNA* regions in B cells, which are incapable of producing IFN-α, are widely occupied with H3K27me3, as shown by the substantive peaks found along the gene cluster (**Figure 3**). In contrast, ChIP-seq data from monocytes demonstrate that H3K27me3 peaks occupy some *IFNA* genes, such as *IFNA2*, *IFNA14*, and intergenic regions between *IFNA2* and *IFNA8*, while the remaining *IFNA* genes were not suppressed by H3K27me3.
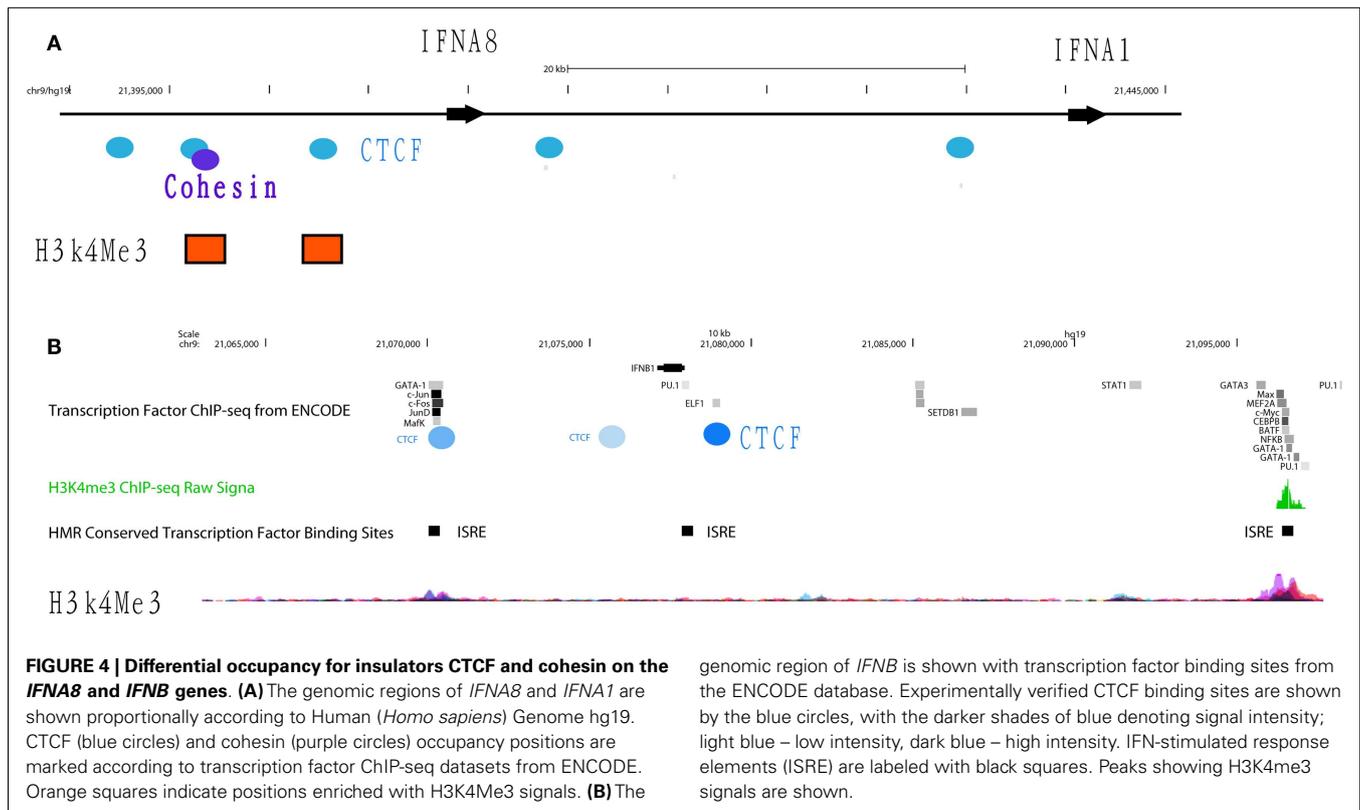
As mentioned above, current dogma holds that H3K4me3 represents a chromatin landmark that is present at the TSS for genes that are either actively transcribed or permissive for gene transcription. However, H3K4me3 are not sufficient to license cells to produce IFN-α. For example, multiple H3K4me3 occupancy peaks can be identified in the IFN regulatory regions in B cells that do not express IFN-α. A good comparison would be with PDCs, yet the histone codes are not yet available for this cell type. In PDCs, TLR-7 signaling quickly turns on transcription of *IFNB*, *IFNA2*, *IFNA8*, and *IFNA14* genes at 30 min post-stimulation with peak levels being achieved at this time point. In comparison, peak levels of *IFNA5*, *IFNA6*, *IFNA10*, *IFNA13*, and *IFNA21* were observed around 4 h post-stimulation (36). Based on our bioinformatics analysis, we reason that transcriptional suppression by H3K27me3, if it exists in PDCs in a pattern similar to that found in CD14$^+$ monocytes, may not be functional in PDCs or can quickly be replaced by H3K4me3 after TLR-7 activation. Alternatively, the *IFN* gene cluster in PDCs may not have H3K27me3 markers. It is not known whether chromatin change is necessary for *IFNA* transcriptional activation or whether chromatin status is responsible for differentially transcribed type I *IFN* genes. Further studies in human PDCs will be required to address this.

Transcription factors regulating basal repression of IFNA gene expression

The transcriptional repressor CTCF (11-zinc finger protein) or CCCTC-binding factor is thought to regulate the 3-dimensional (3D) structure of chromatin by binding strands of DNA together and forming DNA loops (37). CTCF represses gene expression by blocking the interaction between enhancers and promoters (38). This phenomenon may serve as a chromatin barrier to block the spread of heterochromatin structures and set boundaries between active chromatin regions marked by histone H2A acetyl Lys5 (H2AK5ac) and repressive regions marked by H3K27me3 (39). The cohesin complex, consisting of cohesion proteins SMC1, SMC3, SCC3, and the α-kleisin SCC1, may contribute to CTCF-mediated repression. Many CTCF/cohesin binding sites are located at promoter regions suggesting a joint regulatory role for these factors (40). Although most cohesin sites overlap with CTCF, a significant proportion of each factor's sites are independent of the other, implying CTCF-independent functions of cohesin as well as cohesin-independent CTCF functions. Bioinformatics analysis of CTCF ChIP-seq data from ENCODE cell lines identified several CTCF insulators that are basally located in the promoters and intergenic regions of *IFNA5*, *A1*, *A2*, and *A8* (**Figure 4A**). In gum fibroblast cells (AG09319), we found five CTCF binding sites in the region covering *IFNA14*, *A17*, *A16*, *A10*, *A7*, and *A4*. There is only one CTCF/SMC3 binding site in the *IFNB* gene. The regulatory region of the *IFNA2* gene contains two CTCF binding sites. The second CTCF site yields co-binding with SMC3, suggesting

FIGURE 3 | H3K27me3 peaks on the type I *IFN* gene cluster. Members of the type I *IFN* gene cluster are shown and illustrated proportionally according to Human (*Homo sapiens*) Genome hg19. H3K27me3 peaks were found along the entire region of the *IFN* gene cluster in the human B cell line GM128.

FIGURE 4 | Differential occupancy for insulators CTCF and cohesin on the *IFNA8* and *IFNB* genes. (A) The genomic regions of *IFNA8* and *IFNA1* are shown proportionally according to Human (*Homo sapiens*) Genome hg19. CTCF (blue circles) and cohesin (purple circles) occupancy positions are marked according to transcription factor ChIP-seq datasets from ENCODE. Orange squares indicate positions enriched with H3K4Me3 signals. **(B)** The genomic region of *IFNB* is shown with transcription factor binding sites from the ENCODE database. Experimentally verified CTCF binding sites are shown by the blue circles, with the darker shades of blue denoting signal intensity; light blue – low intensity, dark blue – high intensity. IFN-stimulated response elements (ISRE) are labeled with black squares. Peaks showing H3K4me3 signals are shown.

the cohesin complex may function in this *IFN* genomic region. Based on these data, we speculate that CTCF may indeed function as an *IFNA* suppressor and block promoter activation.

Within the *IFNA* gene cluster, we have yet to identify any other TF in the ENCODE datasets that basally occupies the promoter regions between TSSs and the proximal CTCF sites. In contrast, multiple TFs, such as NF-κB and PU.1 (in B cells), do constitutively occupy regions up-stream of CTCF sites that control individual *IFNA* genes. CTCF binding sites are not conserved but cell-type specific. While the majority of cells show CTCF occupancy up-stream of the *IFNA2* gene, binding is absent in fibroblast cells. Similarly, at the *IFNB* promoter, CTCF binding was identified in some B cells lines, HeLa cells, MCF-7, and osteoblast cells, but not in any fibroblast cell lines or A549 lung carcinoma cells. Lack of binding of this insulator may render fibroblast cells to produce type I IFNs upon the appropriate stimulation, such as viral infection, thus supporting that CTCF binding to the *IFNA* gene may be regulated. In this regard, dexamethasone treatment in A549 cell lines induces CTCF to bind to the *IFNA8* promoter. Finally, the discrepancy of CTCF binding patterns in Epstein–Barr virus (EBV)-transformed B cell lines suggests that viral infection may interfere with CTCF function. It is known that CTCF/cohesin occupancy is essential for IFN-gamma (*IFNg*) gene transcription (41). Thus, this complex may have a similar function and be important for regulating *IFNB* gene transcription via maintaining the 3D chromatin structure of the *IFNB* locus in fibroblast cells (**Figure 4B**). Based on these data, we propose that the DNA regions in the *IFN* gene cluster that contain CTCF occupancy may be subject to control by this factor to ensue *IFNA* transcription during viral or viral-like challenges in IFN-producing cells. This

region may also be used as a landmark to demarcate the promoter region that spans from a TSS to the CTCF binding sites and enhancer regions located up-stream of CTCF binding site.

Transcription factors that regulate induction of IFNA gene expression

Interferon regulatory factors, as their name suggests, have been long known to regulate type I *IFN* gene expression (42). Of the nine mammalian IRF family members currently identified to date, IRF7 has garnered the most attention for its role in regulating *IFNA* gene expression (3). IRF7 is highly expressed in human PDCs and allows bypass of the classic autocrine feedback loop that is regulated by IFN-β (43). IRF7 was also shown to be required for murine PDCs to produce an antiviral IFN immune response (44). Similarly, IRF5 has been implicated in the regulation of type I *IFN* gene expression (45). Early data in human cell lines revealed the regulation of type I *IFN* genes and IFN-stimulated genes (ISGs) by IRF5 in response to virus (46). Later data in mice supported these findings. For example, splenic PDCs from mice lacking *Irf5* were shown to produce less type I IFNs in response to virus infection (47). IRF5 has also been recently reported to regulate IFN-β production in myeloid dendritic cells downstream of the mitochondrial antiviral-signaling protein (MAVS) (48). Furthermore, recent studies demonstrate that IRF5 and NF-κB p50 are key co-regulators of IFN-β and IL-6 expression in TLR9-mediated activation of human PDCs (49). Although both of these IRF family members have been implicated as key regulators of IFN-α production, no ChIP-seq data is available to support these findings. Interestingly, data from the aforementioned STAT4/IRF5 ChIP-seq datasets in PBMCs did not support the direct regulation of type I

IFN expression by IRF5 since no peaks were detected in the IFN gene cluster after immunoprecipitation with anti-IRF5 antibodies (22). In this case, PBMCs were stimulated with either IFNa2 or SLE immune complexes before immunoprecipitation with anti-IRF5 or anti-STAT4 antibodies. In the case of IRF7, a cursory review of the literature and publicly available datasets indicate that no ChIP-seq data is currently available for this TF. We have recently performed IRF5 and IRF7 ChIP-seq in human PDCs stimulated with virus. Our unpublished data indicate that these two TFs bind to different regions in the *IFNA* gene cluster (**Figure 5**). These data support the distinct and differential roles for IRF5 and IRF7 in type I *IFN* gene regulation (45, 49). With regard to autoimmune diseases such as SLE and SS that display a pathogenic type I *IFN* gene signature, determination of the mechanisms by which these two IRF family members cooperatively and distinctly regulate *IFNA* subtype expression in the critical IFN-α producing cell types will be important for the design of new therapeutic strategies targeting these two factors.

CONCLUDING REMARKS

With the recent generation of large datasets in the public domain from next-generation sequencing and DNA microarray experiments, others and we have begun to perform detailed analyses of cell-type specific gene signatures as well as identify distinct TFs that differentially regulate these gene signatures in a cell type- and disease-specific manner. This report describes a sample workflow and method of integrative analysis to inspect, clean, and model data from GEO and ENCODE with the goal of highlighting information and knowledge discovery at the gene cluster level. We demonstrate that this method can extract valuable information including downstream pathway analysis, DNA methylation, chromatin structure, histone modification, and TF binding to a gene of interest (in our case, type I *IFNA*s). This report summarizes data from our bioinformatics analysis of the type I *IFN* gene cluster
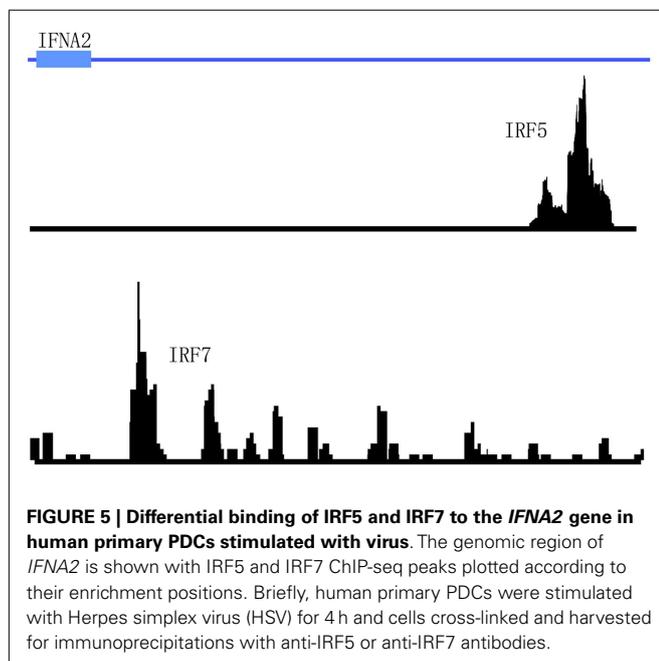
using data in the public domain and experimental unpublished data from our lab (**Tables 1** and **2**). We have found that the genetic landscape of the *IFNA* and *IFNB* genes are occupied by TF, such as insulator CTCF and cohesin, that negatively regulate transcription, as well as IRF5 and IRF7, that positively and distinctly regulate the *IFNA* subtypes. This information can be used as a reference to guide future experiments that focus on proving and/or disapproving these novel regulatory mechanisms that control type I *IFN* expression. A detailed understanding of the factors controlling type I *IFN* gene transcription will significantly aid in the identification and development of new therapeutic strategies targeting the IFN pathway in autoimmune disease.

Table 1 | Results from computational pathway analysis of microarray data sets.

Genes and pathways	*Ex vivo* type I IFN treatment	In SLE patients
IL-15 and its receptor IL-15Rα	Up-regulated	Up-regulated
IL-7	Up-regulated	Up-regulated
CD59	Up-regulated	Up-regulated
MAP kinase	MAP kinase (ERK2) activity at up-stream of STAT1, MAP2K5, MAP2K5 are up-regulated	Unknown
TLR pathway (*TLR-3, 7, 1, TRAF/TANK, IRF4,* and *IRF1*)	Up-regulated	TLR-7 up-regulated[a]
STAT	STAT1	STAT1, STAT4
Apoptotic pathways	Up-regulated *caspase 1, 8,* and *10, TRAIL, FADD*	Up-regulated anti-apoptotic genes including *BIRC5*

[a]*Indicates data from Ref. (50).*
The following list of genes and pathways were predicted to be active in PBMCs treated with type I IFN ex vivo and in SLE patients.

Table 2 | Results from the computational analysis of ENCODE next-generation sequencing data on the type I *IFN* gene cluster.

Epigenetic markers	Factors that affect type I IFN gene cluster
Chromatin structure	Monocytes display more DNase I hypersensitivity sites within gene cluster
Methylation	Methylation not found in non-IFN-producing cells; hypomethylated CpG island identified in cluster
Histone modification	H3K4Me3, H3K27me3, H3K9me2
Conserved transcription factor binding site	HMR conserved transcription factor binding sites computed with the Transfac Matrix Database (v7.0) identified ISRE sites
Transcription factors	IRF5, IRF7
Insulator	CTCF, SMC3, cohesin complex

FIGURE 5 | Differential binding of IRF5 and IRF7 to the *IFNA2* gene in human primary PDCs stimulated with virus. The genomic region of *IFNA2* is shown with IRF5 and IRF7 ChIP-seq peaks plotted according to their enrichment positions. Briefly, human primary PDCs were stimulated with Herpes simplex virus (HSV) for 4 h and cells cross-linked and harvested for immunoprecipitations with anti-IRF5 or anti-IRF7 antibodies.

REFERENCES

1. Farkas L, Beiske K, Lund-Johansen F, Brandtzaeg P, Jahnsen FL. Plasmacytoid dendritic cells (natural interferon-alpha/beta-producing cells) accumulate in cutaneous lupus erythematosus lesions. *Am J Pathol* (2001) 159:237–43. doi:10.1016/S0002-9440(10)61689-6

2. Hagberg N, Berggren O, Leonard D, Weber G, Bryceson YT, Alm GV, et al. IFN-alpha production by plasmacytoid dendritic cells stimulated with RNA-containing immune complexes is promoted by NK cells via MIP-1beta and LFA-1. *J Immunol* (2011) 186:5085–94. doi:10.4049/jimmunol.1003349

3. Fitzgerald-Bocarsly P, Feng D. The role of type I interferon production by dendritic cells in host defense. *Biochimie* (2007) 89:843–55. doi:10.1016/j.biochi.2007.04.018

4. Waddell SJ, Popper SJ, Rubins KH, Griffiths MJ, Brown PO, Levin M, et al. Dissecting interferon-induced transcriptional programs in human peripheral blood cells. *PLoS One* (2010) 5:e9753. doi:10.1371/journal.pone.0009753

5. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* (2012a) 22:1798–812. doi:10.1101/gr.139105.112

6. Baranda L, De La Fuente H, Layseca-Espinosa E, Portales-Perez D, Nino-Moreno P, Valencia-Pacheco G, et al. IL-15 and IL-15R in leucocytes from patients with systemic lupus erythematosus. *Rheumatology (Oxford)* (2005) 44:1507–13. doi:10.1093/rheumatology/kei083

7. Badot V, Luijten RK, Van Roon JA, Depresseux G, Aydin S, Van Den Eynde BJ, et al. Serum soluble interleukin 7 receptor is strongly associated with lupus nephritis in patients with systemic lupus erythematosus. *Ann Rheum Dis* (2013) 72:453–6. doi:10.1136/annrheumdis-2012-202364

8. Gonzalez-Quintial R, Lawson BR, Scatizzi JC, Craft J, Kono DH, Baccala R, et al. Systemic autoimmunity and lymphoproliferation are associated with excess IL-7 and inhibited by IL-7Ralpha blockade. *PLoS One* (2011) 6:e27528. doi:10.1371/journal.pone.0027528

9. Alegretti AP, Mucenic T, Merzoni J, Faulhaber GA, Silla LM, Xavier RM. Expression of CD55 and CD59 on peripheral blood cells from systemic lupus erythematosus (SLE) patients. *Cell Immunol* (2010) 265:127–32. doi:10.1016/j.cellimm.2010.07.013

10. David M, Petricoin E III, Benjamin C, Pine R, Weber MJ, Larner AC. Requirement for MAP kinase (ERK2) activity in interferon alpha- and interferon beta-stimulated gene expression through STAT proteins. *Science* (1995) 269:1721–3. doi:10.1126/science.7569900

11. Goh KC, Haque SJ, Williams BR. p38 MAP kinase is required for STAT1 serine phosphorylation and transcriptional activation induced by interferons. *EMBO J* (1999) 18:5601–8. doi:10.1093/emboj/18.20.5601

12. Monteleone G, Pender SL, Wathen NC, MacDonald TT. Interferon-alpha drives T cell-mediated immunopathology in the intestine. *Eur J Immunol* (2001) 31:2247–55. doi:10.1002/1521-4141(200108)31:8<2247::AID-IMMU2247>3.0.CO;2-4

13. Lindkvist A, Ivarsson K, Jernberg-Wiklund H, Paulsson-Karlsson Y. Interferon-induced sensitization to apoptosis is associated with repressed transcriptional activity of the hTERT promoter in multiple myeloma. *Biochem Biophys Res Commun* (2006) 341:1141–8. doi:10.1016/j.bbrc.2006.01.068

14. Clark PE, Polosukhina DA, Gyabaah K, Moses HL, Thorburn A, Zent R. TRAIL and interferon-alpha act synergistically to induce renal cell carcinoma apoptosis. *J Urol* (2010) 184:1166–74. doi:10.1016/j.juro.2010.04.064

15. Roth W, Wagenknecht B, Dichgans J, Weller M. Interferon-alpha enhances CD95L-induced apoptosis of human malignant glioma cells. *J Neuroimmunol* (1998) 87:121–9. doi:10.1016/S0165-5728(98)00079-4

16. Gomez-Benito M, Balsas P, Bosque A, Anel A, Marzo I, Naval J. Apo2L/TRAIL is an indirect mediator of apoptosis induced by interferon-alpha in human myeloma cells. *FEBS Lett* (2005) 579:6217–22. doi:10.1016/j.febslet.2005.10.007

17. Kirou KA, Vakkalanka RK, Butler MJ, Crow MK. Induction of Fas ligand-mediated apoptosis by interferon-alpha. *Clin Immunol* (2000) 95:218–26. doi:10.1006/clim.2000.4866

18. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* (2003) 100:2610–5. doi:10.1073/pnas.0337679100

19. Crow MK, Kirou KA, Wohlgemuth J. Microarray analysis of interferon-regulated genes in SLE. *Autoimmunity* (2003) 36:481–90. doi:10.1080/08916930310001625952

20. Burgi Mde L, Prieto C, Etcheverrigaray M, Kratje R, Oggero M, Bollati-Fogolin M. WISH cell line: from the antiviral system to a novel reporter gene assay to test the potency of human IFN-alpha and IFN-beta. *J Immunol Methods* (2012) 381:70–4. doi:10.1016/j.jim.2012.04.010

21. Lyons PA, McKinney EF, Rayner TF, Hatton A, Woffendin HB, Koukoulaki M, et al. Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. *Ann Rheum Dis* (2010) 69:1208–13. doi:10.1136/ard.2009.108043

22. Wang C, Sandling JK, Hagberg N, Berggren O, Sigurdsson S, Karlberg O, et al. Genome-wide profiling of target genes for the systemic lupus erythematosus-associated transcription factors IRF5 and STAT4. *Ann Rheum Dis* (2013) 72:96–103. doi:10.1136/annrheumdis-2012-201364

23. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum* (1990) 33:1665–73. doi:10.1002/art.1780331109

24. Strickland FM, Richardson BC. Epigenetics in human autoimmunity. Epigenetics in autoimmunity – DNA methylation in systemic lupus erythematosus and beyond. *Autoimmunity* (2008) 41:278–86. doi:10.1080/08916930802024616

25. Zhu X, Liang J, Li F, Yang Y, Xiang L, Xu J. Analysis of associations between the patterns of global DNA hypomethylation and expression of DNA methyltransferase in patients with systemic lupus erythematosus. *Int J Dermatol* (2011) 50:697–704. doi:10.1111/j.1365-4632.2010.04804.x

26. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, et al. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling. *PLoS One* (2012) 7:e50233. doi:10.1371/journal.pone.0050233

27. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, et al. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* (2012) 7:617–36. doi:10.1038/nprot.2012.012

28. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* (2010) 28:1097–105. doi:10.1038/nbt.1682

29. Cockerill PN. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J* (2011) 278:2182–210. doi:10.1111/j.1742-4658.2011.08128.x

30. Wang YM, Zhou P, Wang LY, Li ZH, Zhang YN, Zhang YX. Correlation between DNase I hypersensitive site distribution and gene expression in HeLa S3 cells. *PLoS One* (2012b) 7:e42414. doi:10.1371/journal.pone.0042414

31. Shestakova E, Bandu MT, Doly J, Bonnefoy E. Inhibition of histone deacetylation induces constitutive derepression of the beta interferon promoter and confers antiviral activity. *J Virol* (2001) 75:3444–52. doi:10.1128/JVI.75.7.3444-3452.2001

32. Nicodeme E, Jeffrey KL, Schaefer U, Beinke S, Dewell S, Chung CW, et al. Suppression of inflammation by a synthetic histone mimic. *Nature* (2010) 468:1119–23. doi:10.1038/nature09589

33. Fang TC, Schaefer U, Mecklenbrauker I, Stienen A, Dewell S, Chen MS, et al. Histone H3 lysine 9 di-methylation as an epigenetic signature of the interferon response. *J Exp Med* (2012) 209:661–9. doi:10.1084/jem.20112343

34. Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, et al. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* (2008) 10:1291–300. doi:10.1038/ncb1787

35. Young MD, Willson TA, Wakefield MJ, Trounson E, Hilton DJ, Blewitt ME, et al. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res* (2011) 39:7415–27. doi:10.1093/nar/gkr416

36. Birmachu W, Gleason RM, Bulbulian BJ, Riter CL, Vasilakos JP, Lipson KE, et al. Transcriptional networks in plasmacytoid dendritic cells stimulated with synthetic TLR 7 agonists.

BMC Immunol (2007) **8**:26. doi:10.1186/1471-2172-8-26

37. Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* (2007) **17**:400–7. doi:10.1016/j.gde.2007.08.005

38. Hou C, Zhao H, Tanimoto K, Dean A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* (2008) **105**:20398–403. doi:10.1073/pnas.0808506106

39. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* (2009) **19**:24–32. doi:10.1101/gr.082800.108

40. Lee BK, Iyer VR. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J Biol Chem* (2012) **287**:30906–13. doi:10.1074/jbc.R111.324962

41. Sekimata M, Perez-Melgosa M, Miller SA, Weinmann AS, Sabo PJ, Sandstrom R, et al. CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma

locus. *Immunity* (2009) **31**:551–64. doi:10.1016/j.immuni.2009.08.021

42. Cham CM, Ko K, Niewold TB. Interferon regulatory factor 5 in the pathogenesis of systemic lupus erythematosus. *Clin Dev Immunol* (2012) **2012**:780436. doi:10.1155/2012/780436

43. Izaguirre A, Barnes BJ, Amrute S, Yeow WS, Megjugorac N, Dai J, et al. Comparative analysis of IRF and IFN-alpha expression in human plasmacytoid and monocyte-derived dendritic cells. *J Leukoc Biol* (2003) **74**:1125–38. doi:10.1189/jlb.0603255

44. Honda K, Yanai H, Negishi H, Asagiri M, Sato M, Mizutani T, et al. IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* (2005) **434**:772–7. doi:10.1038/nature03464

45. Barnes BJ, Moore PA, Pitha PM. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *J Biol Chem* (2001) **276**:23382–90. doi:10.1074/jbc.M101216200

46. Barnes BJ, Field AE, Pitha-Rowe PM. Virus-induced heterodimer formation between IRF-5 and IRF-7 modulates assembly of the IFNA enhanceosome

in vivo and transcriptional activity of IFNA genes. *J Biol Chem* (2003) **278**:16630–41. doi:10.1074/jbc.M212609200

47. Yasuda K, Nundel K, Watkins AA, Dhawan T, Bonegio RG, Ubellacker JM, et al. Phenotype and function of B cells and dendritic cells from interferon regulatory factor 5-deficient mice with and without a mutation in DOCK2. *Int Immunol* (2013) **25**:295–306. doi:10.1093/intimm/dxs114

48. Lazear HM, Lancaster A, Wilkins C, Suthar MS, Huang A, Vick SC, et al. IRF-3, IRF-5, and IRF-7 coordinately regulate the type I IFN response in myeloid dendritic cells downstream of MAVS signaling. *PLoS Pathog* (2013) **9**:e1003118. doi:10.1371/journal.ppat.1003118

49. Steinhagen F, McFarland AP, Rodriguez LG, Tewary P, Jarret A, Savan R, et al. IRF-5 and NF-kappaB p50 co-regulate IFN-beta and IL-6 expression in TLR9-stimulated human plasmacytoid dendritic cells. *Eur J Immunol* (2013) **43**:1896–906. doi:10.1002/eji.201242792

50. Komatsuda A, Wakui H, Iwamoto K, Ozawa M, Togashi M, Masai R, et al. Up-regulated expression of Toll-like receptors mRNAs in peripheral blood mononuclear cells from patients with systemic lupus

erythematosus. *Clin Exp Immunol* (2008) **152**:482–7. doi:10.1111/j.1365-2249.2008.03646.x