A B-cell acute lymphoblastic leukemia regulatory network defines novel therapeutic targets in *IGH-CRLF2* patients

Sana Badri[1,2,*], Beth Carella[1,3,*], Priscillia Lhoumaud[1], Dayanne M. Castro[2], Ramya Raviram[1, 4], Aaron Watters[6], Richard Bonneau[2,5,6,**], Jane A. Skok[1,**].

[1]Department of Pathology, New York University School of Medicine, New York, NY, USA.

[2]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY, USA.

[3]Division of Blood and Marrow Transplantation and Cellular Therapies, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA

[4]Department of Chemistry and Biochemistry, University of California, San Diego, CA, USA

[5]Department of Computer Science, Courant Institute of Mathematical Sciences, New York, USA.

[6]Flatiron Institute, Center for Computational Biology, Simons Foundation, New York, NY, USA.

[*] These authors contributed equally to this manuscript

[**] Correspondence and requests for materials should be addressed to:

e-mail: rb133@nyu.edu (Richard Bonneau) jane.skok@med.nyu.edu (Jane A. Skok)

ABSTRACT:

CRLF2 overexpression in B-ALL patients with an *IGH-CRLF2* translocation activates JAK-STAT, PI3K and ERK/MAPK signaling pathways. Although inhibitors of these pathways are available, investigating alternate targets could reduce treatment-associated toxicities. Comparing RNA-seq from *IGH-CRLF2* and non-translocated patients we defined a translocation gene signature. Next, we assembled a B-ALL cancer-specific regulatory network using 529 B-ALL patient samples from the NCI TARGET database coupled with priors generated from ATAC-seq peak TF-motif analysis. The network was used to infer differential changes in TF activities predicted to control *IGH-CRLF2* deregulated genes, thereby enabling identification of translocation-associated pathways and potential new therapeutic targets.

INTRODUCTION:

Acute lymphoblastic leukemia (ALL) is the most common cancer in children[1]. Historically, clinical criteria have driven risk stratification for these patients, however over time many genetic alterations have been identified as prognostic predictors. Several groups have described varied genetic signatures associated with pediatric ALL with the aim of elucidating their contributions to leukemogenesis[2,3]. Improved risk stratification based on genetic signature has altered treatment and led to significant improvements in overall survival[1]. However, about 20% of patients fail current treatment strategies or die following relapse. Moreover, adults with ALL have a worse prognosis and an average overall survival of 35-50%[1]. Therefore, it is important to gain a better understanding of the mechanisms by which genetic alterations drive leukemogenesis to refine therapies that target the disease-essential pathways involved[4].

Genetic alterations that lead to overexpression of the cytokine receptor-like factor 2 (*CRLF2*) gene have been associated with a high-risk subset of pediatric patients with B-cell acute lymphoblastic leukemia (B-ALL). The *CRLF2* gene encodes the thymic stromal lymphopoietin receptor (TSLPR) which forms a heterodimer with IL-7 receptor alpha (IL7RA) to bind TSLP[5]. Binding of TSLP to the IL7RA-TSLPR complex signals the phosphorylation of Janus kinase 1 (JAK1) and Janus kinase 2 (JAK2), leading to the activation of the JAK-STAT signaling pathway[6]. Studies have shown that stimulation of TSLP in B-ALL not only induces activation of the JAK-STAT pathway, but also activates the PI3K/mTOR[7] and ERK/MAPK signaling pathways[1].

CRLF2 overexpression occurs in 5-15% of patients with B-ALL and in 50-60% of pediatric B-ALL patients with Down syndrome[6,8]. *CRLF2* overexpression in B-ALL can occur either from a chromosomal translocation between *CRLF2* and the immunoglobulin heavy chain locus (*IGH*) on chromosome 14 (*IGH-CRLF2*) or from an interstitial deletion of the pseudoautosomal region of the X/Y chromosomes resulting in the fusion of *CRLF2* to the *P2RY8* gene (*P2RY8-CRLF2*)[9]. *CRLF2* deregulation very rarely occurs via activating mutations

of *CRLF2*. The *IGH-CRLF2* translocation occurs in precursor cells and is thought to result from aberrant rejoining during V(D)J-recombination[10,11]. This translocation is most commonly found in adolescents and adult patients and is typically associated with a poor prognoses, while the *P2RY8-CRLF2* fusion is found in younger patients[1].

CRLF2 chromosomal alterations are often accompanied by mutations in *JAK1*, *JAK2* and *Ikaros* (*IKZF1*) genes. Several groups have suggested that aberrant *CRLF2* signaling cooperates with mutant JAK and IKZF1 activity to promote the development of leukemia[9,12,13]. As a result, the focus has shifted towards the use of signal transduction inhibitors (STIs) to target JAK-STAT, PI3K and MAPK signaling pathways[1]. Although, STIs have shown promise in early clinical trials[14-16], broad application of signal transduction inhibitors is challenging due to the interconnected roles of their targets in biological processes (i.e. JAK kinases) including immunity and hematopoiesis[17]. Moreover, it has been shown that mutated JAK2 is required for the initiation of leukemia, but it is not necessary for its maintenance [18]. Therefore, there is a need to more closely investigate the genome-wide impact of the *IGH-CRLF2* alteration in B-ALL to aid in the identification of novel therapeutic targets.

To do so, we first sought to identify transcriptional regulators that control differentially expressed genes associated with the *IGH-CRLF2* translocation through a comparative analysis of translocated versus non-translocated (Non-T) *IGH-CRLF2* B-ALL samples in both patients and cell lines. While robust genome-wide changes in gene expression were separately observed in patients and cell lines, only a small subset of changes were consistent amongst the two groups. Thus, we chose to define the *IGH-CRLF2* associated gene set using only patient samples. We constructed a B-ALL transcriptional network to define the interactions between transcription factors (TFs) and the genes they regulate[19-21] using 529 B-ALL patient samples from the NCI TARGET database. We inferred the targets of TFs linked to the gene set associated with the *IGH-CRLF2* cohort using the Inferelator algorithm [22,23], along with RNA-seq, ATAC-seq, and TF-motif analysis. The network was then used to predict differential transcription

factor activity (TFA) in *IGH-CRLF2* versus non-translocated (Non-T) samples. This approach enabled the identification of ten potential regulators of differentially expressed genes (including DNMT1, EGR1, FOXP1, ZBTB7A). The differentially expressed gene targets of these TFs are enriched in anticipated and novel *CRLF2*-associated pathways. Transcription of several of these gene targets (*IL1RAP*, *LEF1*, *PPP1R13B,* etc.) change across all patients and cell line samples. It is of note that these genes have been implicated in both Acute myeloid leukemia (AML)[24,25] and ALL[26,27]. Thus, it is plausible that the genes along with their regulators could contribute to the maintenance of leukemia and be potential candidates for therapeutic targeting in *IGH-CRLF2* B-ALL. The network-based approach we applied has been similarly implemented in several other systems to infer regulatory interactions that have been experimentally validated[28-31].

RESULTS:

Genome-wide transcriptional changes in primary IGH-CRLF2 patient samples

Traditional chemotherapy is non-specific and targets rapidly dividing cells. As a result, toxicity results in injury to healthy cells, causing further morbidity and at times, can be dose-limiting [32]. To reduce overall toxicity and improve prognosis, most research has been directed towards understanding the underlying molecular pathology of the leukemia. In this study, we focused on identifying pathways associated with the *IGH-CRLF2* translocation (**Fig. 1a**) with the goal of finding new potential therapeutic targets in this subset of B-ALL patients.

Gene expression profiles associated with the *IGH-CRLF2* translocation were identified by comparing RNA-sequencing from 17 primary B-ALL patient samples, with the translocation (n=13) with non-translocated (Non-T, n=4) patients. Differential analysis performed using DESeq2[33] uncovered a total of 1,179 de-regulated genes with an adjusted p-value of less than 0.01 and $|\log_2 \text{FoldChange}| > 1$ (**Fig. 1b**). Of these 507 genes (~43%) were up-regulated and 669 (~57%) down-regulated in the *IGH-CRLF2* translocated patients. To determine whether the

gene expression changes were localized to the translocated chromosomes we analyzed the percentage of differentially expressed genes per chromosome. The results clearly demonstrate that expression changes are distributed across all chromosomes and there is no enrichment of deregulated genes on chromosome 14 or the pseudoautosomal region of X/Y (**Supplementary Fig. 1a**).

To further analyze the changes in gene expression between the *IGH-CRLF2* and the Non-T patient samples we performed a principal component analysis. This revealed a clear separation between the two groups of patients (**Supplementary Fig. 1b**). In addition, hierarchical clustering of the 1,179 differentially expressed genes not only separates the *IGH-CRLF2* and Non-T samples, but also clearly divides the *IGH-CRLF2* patients into two groups, referred to as Group 1 and Group 2 (**Fig. 1c**). These two groups could not be distinguished by the presence or absence of activating mutations in *JAK* and *IKZF1* which were found in most of the translocated cohort: 12/13 and 10/13 *IGH-CRLF2* patients, respectively (**Fig. 1c**). As expected the *IGH-CRLF2* translocation clearly results in the up-regulation of *CRLF2* ($\log_2$FoldChange= 5.53) in the translocated versus Non-T control patient samples as shown by RNA-seq tracks on IGV (**Fig. 1d**), suggesting activation of JAK-STAT signaling. Furthermore, *SOCS6*, a suppressor gene of cytokine signaling[34], is statistically significantly down-regulated ($\log_2$FoldChange= -1.68) in *IGH-CRLF2* samples (**Supplementary Fig. 1c**), further supporting the observation that JAK-STAT[1,4] signaling is activated in these samples.

Using the log2 fold changes of the 1,179 differentially expressed genes, ingenuity pathway analysis (IPA) was performed. Positive and negative z-scores indicate pathways that are activated or repressed respectively in the *IGH-CRLF2* samples. The IPA analysis (**Supplementary Fig. 1d**) identified 20 significant pathways, the majority of which (19/20) had negative z-scores, indicating that the majority of the pathways are repressed. The PI3K and ERK pathways were included in the repressed cohort which was unexpected as activation of these pathways is normally linked to *CRLF2* overexpression[7], Additionally, the JAK-STAT

signaling pathway did not emerge as being statistically enriched according to the gene set we defined, despite the fact that translocated patients have activating mutations in *JAK* genes. One possible explanation for this outcome is that heterogeneity between patient samples leads to noise in gene expression changes which blurs the analysis. In summary, analysis of RNA-seq data in primary patient samples identifies many significant transcriptional changes associated with the translocated *IGH-CRLF2* chromosomal alteration. Furthermore, hierarchical clustering of the patient samples using the differentially expressed gene list, not only clearly distinguishes *IGH-CRLF2* patients from Non-T, but also defines two distinct *IGH-CRLF2* group of patients.

Differential analysis between the two distinct Group 1 and Group 2 patients results in 456 differentially expressed genes. Although pathway analysis results in no significant pathway enrichment we identified six differentially expressed genes involved in p53 signaling, including the proto-oncogene *MDM2* and p53 regulator *MDM4* that are down-regulated in Group 1 (**Fig. 1e**). We postulated there maybe a link between p53 signaling and relapse but were not able to confirm whether p53 signaling in one subtype provides any advantage in this context, as there were 25% (1 of 4) Group 1 patients that relapsed, compared to 33% (3 of 9) of patients in Group 2.

***IGH-CRLF2* patient samples have limited changes in chromatin accessibility**

To further investigate the impact of the *IGH-CRLF2* translocation, we hypothesized *IGH-CRLF2* transcriptional changes could be accompanied by changes in chromatin accessibility. To test this hypothesis, we performed ATAC-sequencing and compared chromatin accessibility in *IGH-CRLF2* versus Non-T patient samples (18 *IGH-CRLF2* and 6 Non-T). These patient numbers are different to the numbers used for RNA-seq analysis as some of the patient samples for RNA-seq did not pass quality control. Differential analysis of all the ATAC-seq peaks identified 162 regions that were more accessible and 126 regions with reduced accessibility in the *IGH-CRLF2* translocated condition (**Fig. 2a**). The majority of ATAC-seq

peaks were not significantly differential, indicating accessibility is for the most part stable between *IGH-CRLF2* and Non-T patient samples. Significant accessibility changes were evenly distributed across promoters, gene bodies, and intergenic regions (**Supplementary Fig. 2b**).

Differential peaks were assigned to promoters (within 3kb of TSS), UTRs, exons and introns of genes (**Fig. 2b**). About 82.6% (238) of differential ATAC-seq peaks were associated with genes (n=224), either on promoters or gene bodies. Of the genes that had at least one differential ATAC-seq peak in the promoter or gene body, 14.2% (32) were significantly differentially expressed in *IGH-CRLF2* versus Non-T patient samples. The majority of the differential ATAC-seq peaks that overlap differentially expressed genes were regulated in the same direction, with the exception of 7 ATAC-gene pairs (**Fig. 2b**). In the example shown in **Fig. 2c,** three significantly more accessible peaks were associated with overexpressed *DPP4* in *IGH-CRLF2* patients. Conversely, the down-regulated *PKIA* gene was linked to a significant reduction in accessibility (**Supplementary Fig. 2c**)**.** We also observed several instances where a significant change in accessibility was linked to a gene whose expression was not affected by the *IGH-CRLF2* translocation (**Supplementary Fig. 2d).** Importantly, the majority of transcriptional changes were not associated with significant changes in chromatin accessibility near the promoter or within the gene bodies (**Fig. 2d**), as demonstrated for the *RRAS* gene and two other examples shown in **Fig. 2e**, **Supplementary Fig. 2e-f**. Thus, we conclude that overall, the *IGH-CRLF2* translocation leads to limited genome-wide chromatin accessibility changes and accessibility remains stable even at promoter regions of differentially expressed genes. However, we postulated that differences in transcription factor (TF) binding at stable ATAC-seq peaks could influence gene expression changes and turned to patient-derived cell lines as a model to address this question.

Transcriptional changes in patient *IGH-CRLF2* derived cell lines are not recapitulated in patient samples

The patient-derived cell lines used in our analysis were MHH-CALL-4[35] (CALL) and MUTZ5[36] (MUTZ) that harbor *IGH-CRLF2* translocations, and a control non-T pre-B leukemic cell line, SMS-SB[37] (SMS). As seen in the RNA-seq tracks of **Fig. 3a**, *CRLF2* is clearly overexpressed in the *IGH-CRLF2* translocated CALL and MUTZ cell lines, compared to the Non-T SMS control. Principal component analysis (PCA) of cell line RNA-seq samples (**Supplementary Fig. 3a**) indicates that the majority of the variance (71%) lies between *IGH-CRLF2* cell lines and Non-T cell lines, while about 26% of variance separates the two translocated cell lines. Thus the PCA analysis indicates that the *IGH-CRLF2* translocation is the major cause of differences between the two conditions, consistent with what was observed in patient samples.

Differential gene expression analysis of CALL versus SMS (3,045 DE genes) and MUTZ versus SMS (2,702 DE genes) identified thousands of differentially expressed genes (**Fig. 3b**), with roughly equivalent numbers of up and down-regulated genes in each case. The number of overlapping differentially expressed genes (ie those likely to be related to the translocation event) is shown in **Fig. 3c**. As with the patient samples, the gene expression changes were distributed across all chromosomes (**Supplementary Fig. 3b**).

As shown in **Fig. 3d**, the $\log_2$ fold change of (CALL/SMS) and (MUTZ/SMS) demonstrated that about 97% of the genes were in convergent orientation (960 up-regulated, 772 down-regulated). However, of the common 1,732 convergent differentially expressed genes in the cell lines, only 221 overlapped with the differentially expressed genes found in translocated versus non-T patient samples (**Fig. 3e**). Furthermore, only 181 out of the 221 genes were in convergent orientation (**Fig. 3f**), indicating that the profile of gene expression changes in cell lines may not be the best representation of what is occurring in patients. Furthermore, no significantly enriched pathways were identified from the overlapping gene set.

To identify the gene list that best separates the translocated class from the non-translocated class, we performed PCA using the differentially expressed genes from cell lines

(Set 1) and patient samples (Set 2). The PCA results associated with the 1,792 differentially expressed genes from Set 1, demonstrate a clear separation between translocated cell lines and non-translocated cell lines. In contrast, the patient samples cluster together and do not separate according to the PC scores calculated using Set 1 genes (**Fig. 3g** – upper panel). In contrast, when the PCA associated with Set 2 was performed (1,176 differentially expressed genes in patients), we found a clear separation between cell lines and patients on the first principal component that separates the *IGH-CRLF2* from Non-T condition on the second principal component, indicating that Set 2 genes are more representative of the effect of the translocation in both primary patient samples and cell line samples (**Fig. 3g** – lower panel). In summary, we found many significant transcriptional changes in cell lines but only a subset of these were recapitulated in patients. On the other hand, the top 500 patient-associated transcriptional changes can be used to not only distinguish between patients and cell lines (PC1), but also clearly separate the *IGH-CRLF2* samples from the Non-T samples (PC2).

Chromatin accessibility changes in *IGH-CRLF2* translocated cell lines are more numerous than those in patient samples

Our analyses reveal little compatibility in expression changes between *IGH-CRLF2* and Non-T in cell lines and patient samples. To investigate this further we analyzed chromatin accessibility using ATAC-sequencing. Significant ATAC-seq peaks were called and a reference peakome of 41,111 peaks was created to include all possible ATAC-seq peaks across the three cell lines, CALL, MUTZ, and SMS. DESeq2 analysis on the cell line peakome identified 2,550 altered ATAC-seq peaks between CALL and the SMS control, and 2,718 between MUTZ and the SMS control (**Fig. 4a**). The overlap between the two comparisons identified 1,396 differentially accessible ATAC-seq peaks (**Fig. 4b**), of which 99.7 % (1392/1396) were in convergent orientation (**Fig. 4c**). This number is incompatible with the total number of differentially accessible ATAC-seq peaks (288) identified in patient samples. However, we

postulated that regions that were changing in patients should change similarly in cell lines. To determine this, we ranked all differential ATAC-seq peaks in patient samples (**Fig. 2a**) according to fold change and adjusted p-value, and took the 162 most accessible peaks and 126 least accessible peaks in *IGH-CRLF2* patient samples and calculated the ATAC signal of the cell lines at these regions. The limited alterations in accessibility in patients were weakly recapitulated in cell lines as shown in **Fig. 4d** However, compared to the stability of ATAC-seq peaks in patients, we observed more alterations in acessbility in cell lines. This finding further supports the conclusion that using cell lines as a model to study the *IGH-CRLF2* translocation is not ideal. Therefore, all the downstream analysis focused solely on the transcriptional changes identified in patient samples.

Construction of a B-ALL regulatory network using ATAC-motif derived priors

CRLF2 overexpression activates a signaling cascade that involves many TF regulators and gene targets. Although, we identified a subset of gene expression changes that could be important for the pathogenesis of the leukemia, it is not clear how all of these genes are regulated and connected. Here, our aim was to first identify which TFs could potentially be regulating the differentially expressed genes we identified, and second to infer the relationship between these TFs and their target genes. To address this, we performed TF-motif analysis, using FIMO[38], at ATAC-seq peaks that fall within promoter regions (20kb upstream of TSS) of all genes genome-wide. We used a hyper-geometric test to test for enrichment of TF-motifs at differentially expressed genes, which resulted in 102 unique significant TF motifs enriched at significantly up- and down-regulated genes with four motifs DMRTC2, DUXA, PITX1, and PITX3 were found in both sets (**Fig. 5a**). Thus, motif enrichment analysis resulted in a list of 102 unique potential TF regulators that could be important for the regulation of the *IGH-CRLF2* associated gene signature.

To better understand the relationship between the candidate TF regulators identified in **Fig. 5a** and the genes they potentially regulate, we sought to infer a transcriptional regulatory network using the Inferelator algorithm[22,23]. The main limitation of network inference is the sample size, to model transcriptional interactions in a B-ALL specific context requires a large dataset that included hundreds of B-ALL patient samples, not limited to samples with an *IGH-CRLF2* rearrangement. Making use of the TARGET initiative we analyzed 529 B-ALL RNA-seq samples and obtained a normalized gene expression matrix that could be used for the construction of a B-ALL specific regulatory network.

Recent studies have incorporated prior information of TF target genes from different data types, like ChIP-seq, ATAC-seq, and TF-motif analysis to considerably improve network inference [28-30,39]. Here, we focused on combining chromatin accessibility data together with TF-motif analysis of the *IGH-CRLF2* cohort to generate priors and infer the B-ALL network (**Fig. 5b**). We selected only regulatory interactions of TFs that were significantly enriched at promoters of differentially expressed genes between *IGH-CRLF2* and Non-T patient samples (102 TFs – **Fig. 5a**). Transcription factor activities estimated using the ATAC-seq motif derived priors and the gene expression matrix obtained from the TARGET database were used for the Inferelator algorithm (**Fig. 5b**). Finally, a B-ALL specific regulatory network involving 102 TFs and 37,086 interactions with combined confidences > 0.5 was inferred. The number of gene targets inferred for each individual TF is shown in **Supplementary Fig. 4a**, and the number of differentially expressed gene targets in the *IGH-CRLF2* cohort are shown in **Supplementary Fig. 4b**.

Defining a sub-regulatory network with TFs affected by *CRLF2* alteration

Transcriptional regulatory networks shed light on the relationship between transcription factors (TFs) and their gene targets. Here, we constructed a B-ALL specific regulatory network involving 102 TFs that may be important regulators of the *IGH-CRLF2* gene signature. To

narrow down the list, we created a sub-network and identified TF regulators in the *IGH-CRLF2* cohort that were altered either at the mRNA level or the protein activity level. First, we analyzed the expression levels of all genes encoding the 102 TF regulators and identified five TF regulators that were significantly differentially expressed between *IGH-CRLF2* and Non-T patients (**Supplementary Fig. 4c**). Next, we determined TFs that had differences in their activity in the *IGH-CRLF2* cohort. Transcription factor activity (TFA) is commonly estimated from mRNA levels, however since posttranslational modifications can influence TF activity, use of mRNA as a proxy for TFA is not the best approach. If prior knowledge of interactions involving TFs and their target genes is available it can considerably improve TFA estimation and network inference[28,30]. Thus, to estimate the activities of the 102 TFs identified from motifs of significantly up- and down-regulated genes in the *IGH-CRLF2* versus Non-T patients, a normalized gene expression matrix was used. This was combined with priors of TF-gene interactions identified from the inferred B-ALL regulatory network (**Fig. 5b**). Sixteen TFs had significant differences in the mean estimated TFA in the *IGH-CRLF2* cohort (**Supplementary Fig. 4b**) and five had significantly altered levels of expression giving a total of twenty significant TF regulators (ZNF713 was identified in both analyses (**Fig. 5c**). TFs with no significant differences in mean TFA are shown in **Supplementary Fig. 5a**. We note that many TFs with significantly altered activity are implicated in cancer. For instance, FOXP1, a member of the forkhead family of transcription factors is known to play important roles in B-cell development and lymphoid malignancies[40] and could therefore potentially contribute to tumorigenesis in B-ALL patients with the *IGH-CRLF2* alteration.

We previously suggested that heterogeneity between patients could lead to noise in the *IGH-CRLF2* gene signature. To filter gene expression changes we focused on inferred target genes of TFs that are affected by *CRLF2* overexpression, either at the transcriptional level or TF activity level. The 6,545 regulatory interactions in **Fig. 5c** describe the relationship between significantly differential TFs and their gene targets. Of these, 387 interactions are regulatory

interactions involving differentially expressed genes (329 out of a total of 1177 unique genes - 28%) in the *IGH-CRLF2* versus Non-T patient cohort (**Fig. 6a,b**). We define the filtered list of 329 differentially expressed genes and their twenty TF regulators as the *CRLF2* specific sub-network (**Fig. 6c)**. The number of differentially expressed gene targets for each TF regulator is shown in **Supplementary Fig. 6a.** As seen in **Supplementary Fig. 6b**, several differentially expressed gene targets are regulated by more than one TF.

Differentially expressed gene targets of TFs with predicted altered activity are enriched in expected and novel *CRLF2*-associated pathways

We performed pathway analysis focusing on the differentially expressed gene targets in the *CRLF2*-altered sub-network (**Fig. 7a**) and identified 21 enriched pathways including the anticipated activation of canonical JAK-STAT, PI3K, and ERK pathways. Thus, in contrast to our initial analysis which focused on all differentially expressed genes in the *IGH-CRLF2* translocated versus Non-T patient samples, we were able to identify known *CRLF2*-associated pathways including the JAK-STAT signaling pathway. These findings validate the approach taken in order to identify a more *CRLF2*-relevant differentially expressed gene list. In **Fig. 7a,** differentially regulated genes (33) are labeled and connected with the appropriate pathway. For example, PI3K up-regulation of signaling is linked to the differential expression of *SOCS6*, *RALB*, *RRAS*, and *PTPN6*. Additionally, the FLT3 signaling[41,42] pathway, which is known to be associated with *CRLF2* overexpression, is identified as being upregulated.

The other pathways identified by our analyses could also be playing an important role in *CRLF2*-overexpressing B-ALL. As such, we wanted to define a handful of interesting genes that could have the highest potential as therapeutic targets. For this, we compared all 33 differentially expressed genes in the patient samples and in each pairwise cell line comparison (CALL versus SMS, and MUTZ versus SMS). Although, we previously found that patient-derived transcriptional changes and cell line-derived transcriptional changes do not strongly overlap,

there are some patient-derived transcriptional changes that are recapitulated in the cell lines. Thus, we compared log2 fold changes and significance of the 33 differentially expressed genes across all patient and cell line comparisons (**Fig. 7b**) and identified six common genes. Upregulation of *IL1RAP*, *LEF1*, *CD79B*, *CTGF*, and repression of *PPM1H*, *PPP1R13B* was robust across all *IGH-CRLF2* patient and cell lines. Previous studies[24-27,43,44] implicate the majority of these genes in hematologic malignancies but their affect on *IGH-CRLF2* B-ALL is not known. Therefore, analysis of the changes in TF activity and gene target mRNA level in the *CRLF2* affected sub-network identified a list of genes that are strong candidates for targeted therapies in *IGH-CRLF2* translocated leukemia.

Discussion

Patients with B-ALL who carry the IGH-CRLF2 translocation are at increased risk for refractory disease and relapse. This alteration leads to *CRLF2* overexpression and activation of JAK-STAT and other associated pathways[1]. Available treatments include non-specific cytotoxic chemotherapies which are often effective but responsible for many of the toxicities seen in these patients. The addition of Tyrosine-Kinase Inhibitors and JAK inhibitors in the treatment of leukemias has allowed for targeted treatments with fewer toxicities, however, alternate pathway-specific therapies are needed to further tailor treatment. To better study the impact of the *IGH-CRLF2* translocation, we analyzed RNA-seq from rearranged *IGH-CRLF2* and Non-T B-ALL patient samples and identified hundreds of differentially expressed genes, but were unable to link these changes to the expected pathway alterations associated with the *IGH-CRLF2* translocation. We subsequently analyzed ATAC-seq data to evaluate the effect of this genetic alternation on DNA accessibility. We found accessibility to be mostly stable near the promoters of genes with altered gene expression and conclude that chromatin accessibility is not driving changes in the transcriptional landscape. Instead, we inferred binding of specific proteins at accessible sites at differentially expressed gene promoters that influence transcription.

Leveraging hundreds of available B-ALL RNA-seq samples from the TARGET database along with prior knowledge of TF-motifs at accessible promoters, we inferred a global B-ALL network. Deeper investigation of the TF regulators and their targets in this network identified a candidate list of potential therapeutic targets.

IL1RAP, one of the significantly overexpressed candidates we identified in *IGH-CRLF2* rearranged patients, encodes a component of the interleukin complex. Upregulation of this gene is known to occur in Acute myeloid leukemia (AML), another lymphoid malignancy associated with poor prognosis[25]. Using antibodies targeting the IL1RAP receptor expressed on the surface of immature AML cells, Agerstam et al. demonstrated clear antileukemic effects in xenograft models.. Targeting IL1RAP also blocks IL-1 signaling and inhibits proliferation of human AML cells[25]. Using antibodies, RNA interference, and deletion of *IL1RAP*, a more recent study demonstrated it was possible to inhibit pathogenesis *in vivo* and *in vitro* without disrupting normal hematopoietic function[24]. *IL1RAP* plays an important role in potentiating AML cells, and there is strong evidence for its therapeutic effects in AML. However, its function in B-ALL is not known.

Other candidates, *LEF1* and *PPP1R13B* have also been implicated in cancer. Lymphoid enhancer-binding factor 1, *LEF1*, is a transcription factor that acts downstream of the Wnt/β-catenin signaling pathway. This TF can independently regulate gene expression and is necessary for stem cell maintenance and organ development[26]. Disrupted *LEF1* is associated with cancer progression and proliferation of cells in ALL, chronic lymphocytic leukemia (CLL), Burkitt lymphoma (BL), and colorectal cancer (CRC)[26]. In particular, one study clearly shows high expression of *LEF1* associated with shorter relapse-free survival (RFS) in B-ALL[27]. Overall, this gene is considered a biomarker for patient prognosis in many hematological malignancies. Though, we now provide evidence suggesting that it has a specific role in leukemogeneis in translocated *IGH-CRLF2* patients. Additionally, *PPP1R13B*, another candidate target, encodes a member of the apoptosis stimulating p53 family of proteins (ASPP), this gene is known to be

repressed in ALL and its reduced expression is due to hypermethylation of the *PPP1R13B* gene promoter[43,44].

DNA methylation is associated with transcriptional repression and is maintained by the DNMT1 protein. The DNMT1 motif was one of the top regulators identified via motif enrichment analysis at differentially expressed gene promoters. The role of DNMT1 as a transcription factor is not clear, however, it still may have regulatory potential through its function in maintaining DNA methylation and in particular for patients with the *IGH-CRLF2* translocation. For instance, Loudin et al. studied transcription and methylation profiles in overexpressing *CRLF2* ALL patients with Down syndrome. Their methylation profiles indicate high methylation levels correlate with reduced gene expression including genes involved in cytokine-receptor interactions[45]. We hypothesize that methylation could be playing a role in this context.

Overall, we have investigated the transcriptional impact of *CRLF2* overexpression in a high-risk subset of B-ALL patients. Using an integrative approach, we derived regulatory interactions that have identified strong candidates for targeted therapies in B-ALL patients with an *IGH-CRLF2* translocation or other leukemias. *IL1RAP*, *LEF1*, and *PPP1R13B* are amongst the most interesting candidates as there is strong evidence for their role in hematological malignancies. The strategy we have used here can further be adapted to elucidate important regulators responsible for additional subsets of B-ALL and other challenging pediatric malignancies.

Methods

All methods are available in supplemental note section of supplemental information

Acknowledgements

The authors thank all members of the Skok and Bonneau labs and William Carroll for discussions. The Children's Oncology Group for providing us with the patient samples and

metadata. New York School of Medicine High Performance (HPC) for technical support. Adriana Heguy and the Genome Technology Center (GTC) core for sequencing efforts. The TARGET database.

Funding

This work was supported by 1R21CA188968-01A1 and 1R35GM122515 (J.S), American Society of Hematology Research Training Award for Fellows (B.C), AACR (P.L), National Cancer Center (P.L).

Author contributions

B.C, S.B, R.B and J.S designed and managed the project. B.C and P.L collected samples and performed experiments. S.B performed data analyses. D.C, R.R, and A.W aided in analysis design. B.C, R.B, and J.S secured funding. S.B and J.S wrote the manuscript. B.C, R.B and J.S revised the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Fig. 1. Genome-wide transcriptional changes in primary *IGH-CRLF2* patient samples. a, Schematic of *IGH-CRLF2* translocation between chromosome 14 and chromosome X/Y. **b**, Volcano plot of the differentially expressed genes (FDR=5%, $|\log_2$ fold change$|$ >1) between *IGH-CRLF2* patients and Non-T patients. Red points and blue points correspond to up- and down regulated genes (n=507 and 669, respectively) in the *IGH-CRLF2* patient samples. **c**, Heatmap of 1,176 differential expressed genes from the 17 patient samples, with known information corresponding to B-ALL associated variants and other metadata indicated at the top

of the heatmap. **d**, Screenshot of RNA-seq tracks for Non-T, *IGH-CRLF2* Group 1, and *IGH-CRLF2* Group 2 at *CRLF2* locus. **e**, Volcano plot of the differentially expressed genes between *IGH-CRLF2* Group 1 Group 2 patients. Red and blue points correspond to up- and down-regulated (n=77 and 379, respectively).

Fig. 2. *IGH-CRLF2* patient samples have limited changes in chromatin accessibility. a, Heatmap of 162 more accessible ATAC-seq peaks in *IGH-CRLF2* patients (purple bar) and 126 less accessible ATAC-seq peaks in *IGH-CRLF2* patients. **b**, Differentially accessible peaks (n=36) linked to differentially expressed genes (n=32) and the $\log_2$ fold change of gene expression (x-axis) plotted against the $\log_2$ fold change of ATAC-seq reads (y-axis). Linked genes are labeled and colored according to genomic annotation. **c**, IGV screenshot of ATAC-seq tracks at up-regulated *DPP4* gene across all patients. The highlighted region indicates significantly more accessible ATAC-seq peaks. **d**, Average ATAC-seq signal of Non-T (blue) and *IGH-CRLF2* (red) patients at 631 ATAC-seq peaks at up-regulated gene promoters in *IGH-CRLF2* patients (upper panel), and 706 ATAC-seq peaks at down-regulated gene promoters in *IGH-CRLF2* patients (lower panel). **e**, IGV screenshot of ATAC-seq tracks at up-regulated *RRAS* gene across all patients, orange highlighted region indicates ATAC-seq region with no significant difference.

Fig. 3. Transcriptional changes in patient I*GH-CRLF2* derived cell lines are not recapitulated in patient samples. a, IGV screenshot of RNA-seq tracks at *CRLF2* gene on chromosome Y in CALL, MUTZ, and SMS cell lines. CALL and MUTZ samples harbor the *IGH-CRLF2* translocation whereas the control cell line SMS does not. **b,** Number of differential expressed genes between *IGH-CRLF2* cell line samples (CALL and MUTZ) compared to SMS control, red and blue bars indicate up- and down-regulated genes in CALL and MUTZ. **c**, Overlap of differentially expressed gene sets between CALL and MUTZ compared to SMS

control (1792 common differentially expressed genes) **d**, Scatter plot of the 1,792 common genes plotting log_2 fold change of CALL and MUTZ (y and y-axis, respectively) and log_2 fold change of over SMS control. **e**, Overlap of differentially expressed genes (221) between cell lines (Set 1) and patient samples (Set 2). **f**, Scatterplots of log_2 fold change of RNA-seq counts of translocated cell lines (CALL and MUTZ) over SMS control (x-axis) plotted against log_2 fold change of *IGH-CRLF2* patient over Non-T patient RNA-seq counts (y-axis). Red and blue points indicate common convergent up- and down-regulated genes (n=120 and 61, respectively) in cell line and patient samples. **g**, Principal component analysis of all patient and cell line samples computed according to the differentially expressed genes in cell lines (Set 1) (upper panel) and patients (Set 2). Cell lines and patient samples are clearly separated on the first principal component, while translocated and non-translocated samples cluster according to condition on the second principal component (highlighted by red ellipses).

Fig. 4. Chromatin accessibility changes are more numerous in cell lines than patient samples. a, Number of differential ATAC-seq peaks between *IGH-CRLF2* cell line samples (CALL and MUTZ) compared to SMS control, red and blue bars indicate increased and decreased peaks in CALL and MUTZ. **b**, Overlap of ATAC-seq peaks between CALL and MUTZ indicate 1,396 -seq peaks commonly change in both CALL and MUTZ compared to SMS control. **c**, Scatter plot of the 1,396 ATAC-seq peaks showing log_2 fold change of CALL and MUTZ (y- and x-axis, respectively) over SMS control. **d**, Average ATAC-seq signal of Non-T patients (blue), *IGH-CRLF2* patients (red), SMS cell line samples (orange), CALL cell line samples (green), MUTZ cell line samples (purple) at 162 and 126 significantly increased and decreased peaks (upper and lower panels, respectively). Heatmaps show average signal 1kb on either side of the midpoint of each ATAC-seq peak, .

Fig. 5. Inferred B-ALL network identifies significant TFs in the *IGH-CRLF2* patient cohort. **a**, Significantly enriched TF motifs (102) at accessible regions 20 kb upstream of the promoters of differentially up and down expressed genes. Plot indicates $-\log_{10}$(p-value) on x-axis obtained from the hyper-geometric test performed to evaluate enrichment (p-value <0.05) of a particular TF motif. Blue and red bars indicate motif enrichment at down- and up-regulated genes, respectively. **b**, Workflow for construction of B-ALL regulatory network. Inferelator algorithm was used to infer a regulatory network with ATAC-motif derived priors and 529 RNA-seq B-ALL patient samples from the TARGET database. Transcription Factor Activity (TFA) was calculated using the prior matrix and TARGET gene expression matrix for enriched TF motifs at differentially expressed genes in the *IGH-CRLF2* cohort. The resulting network consists of 102 TFs and 37,086 significant regulatory interactions with combined confidences > 0.5. **c**, Filtered B-ALL network involving regulatory interactions between significant TFs and their gene targets. TFs are the source nodes labeled in black and the gene targets are the target nodes in blue. The size of the source nodes reflects the number of targets each TF has. TF activation and repression of a gene are represented by red and blue sedge, respectively.

Fig. 6. *CRLF2* specific sub-network depicting significant TFs and their differentially regulated target genes. **a**, Pie chart representing the percentage of differentially expressed genes controlled by TFs affected by *CRLF2* overexpression, either at the transcriptional level or TF activity level (387). **b**, Pie chart representing the percentage of differentially expressed genes (329) in the *IGH-CRLF2* sub-network controlled by TFs with significantly altered expression or activity. **c**, *CRLF2* specific sub-network depicting significant TFs and their differentially regulated target genes. TFs are labeled in black and their respective gene targets are connected to them by red and blue edges indicating the type of regulation (activation/repression).

Fig. 7. Differentially expressed gene targets of predicted significant TFs are enriched in expected and novel *CRLF2*-associated pathways. a, Significant pathways ($-\log_{10}$(B-H p-value > 0.5 & |z-score| > 1) identified from analysis of differentially expressed gene targets in the *CRLF2* specific sub-network are shown at the top of the figure, ranked by the $-\log_{10}$(Benjamini-Hochberg p-value). Z-score in the bar plot indicates sign of the pathway with red and blue representing up- and down-regulation of the pathway. Differentially expressed genes enriched in these pathways are indicated by purple bins while green bins indicate TF-gene interactions. **b**, Positive and negative $\log_2$ fold changes of all the differentially expressed genes in IGH-CRLF2 vs Non-T patients and cell lines (CALL vs SMS, MUTZ vs SMS) are indicated as red (positive) and blue (negative) bins with significantly deregulated genes labeled with white stars.

References

1 Tasian, S. K. & Loh, M. L. Understanding the biology of CRLF2-overexpressing acute lymphoblastic leukemia. *Crit Rev Oncog* **16**, 13-24 (2011).

2 Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764, doi:10.1038/nature05690 (2007).

3 Mullighan, C. G. *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110-114, doi:10.1038/nature06866 (2008).

4 Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **106**, 9414-9418, doi:10.1073/pnas.0811761106 (2009).

5 Pandey, A. *et al.* Cloning of a receptor subunit required for signaling by thymic stromal lymphopoietin. *Nat Immunol* **1**, 59-64, doi:10.1038/76923 (2000).

6 Roll, J. D. & Reuther, G. W. CRLF2 and JAK2 in B-progenitor acute lymphoblastic leukemia: a novel association in oncogenesis. *Cancer Res* **70**, 7347-7352, doi:10.1158/0008-5472.CAN-10-1528 (2010).

7 Tasian, S. K. *et al.* Aberrant STAT5 and PI3K/mTOR pathway signaling occurs in human CRLF2-rearranged B-precursor acute lymphoblastic leukemia. *Blood* **120**, 833-842 (2012).

8 Russell, L. J. *et al.* Characterisation of the genomic landscape of CRLF2-rearranged acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **56**, 363-372, doi:10.1002/gcc.22439 (2017).

9 Russell, L. J. *et al.* Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood* **114**, 2688-2698 (2009).

10 Tsai, A. G., Yoda, A., Weinstock, D. M. & Lieber, M. R. t(X;14)(p22;q32)/t(Y;14)(p11;q32) CRLF2-IGH translocations from human B-lineage ALLs involve CpG-type breaks at CRLF2, but CRLF2/P2RY8 intrachromosomal deletions do not. *Blood* **116**, 1993-1994, doi:10.1182/blood-2010-05-286492 (2010).

11 Vesely, C. *et al.* Genomic and transcriptional landscape of P2RY8-CRLF2-positive childhood acute lymphoblastic leukemia. *Leukemia* **31**, 1491-1501, doi:10.1038/leu.2016.365 (2017).

12 Mullighan, C. G. *et al.* Rearrangement of CRLF2 in B-progenitor–and Down syndrome–associated acute lymphoblastic leukemia. *Nature genetics* **41**, 1243-1246 (2009).

13 Harvey, R. C. *et al.* Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood* **115**, 5312-5321 (2010).

14 Tasian, S. K. *et al.* Potent efficacy of combined PI3K/mTOR and JAK or ABL inhibition in murine xenograft models of Ph-like acute lymphoblastic leukemia. *Blood* **129**, 177-187, doi:10.1182/blood-2016-05-707653 (2017).

15 Loh, M. L. *et al.* A phase 1 dosing study of ruxolitinib in children with relapsed or refractory solid tumors, leukemias, or myeloproliferative neoplasms: A Children's Oncology Group phase 1 consortium study (ADVL1011). *Pediatr Blood Cancer* **62**, 1717-1724, doi:10.1002/pbc.25575 (2015).

16 Maude, S. L. *et al.* Targeting JAK1/2 and mTOR in murine xenograft models of Ph-like acute lymphoblastic leukemia. *Blood* **120**, 3510-3518, doi:10.1182/blood-2012-03-415448 (2012).

17 Springuel, L., Renauld, J.-C. & Knoops, L. JAK kinase targeting in hematologic malignancies: a sinuous pathway from identification of genetic alterations towards clinical indications. *Haematologica* **100**, 1240-1253, doi:papers3://publication/doi/10.3324/haematol.2015.132142 (2015).

18 Kim, S. K. *et al.* JAK2 is dispensable for maintenance of JAK2 mutant B-cell acute lymphoblastic leukemias. *Genes Dev* **32**, 849-864, doi:10.1101/gad.307504.117 (2018).

19 Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. & Guthke, R. Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems* **96**, 86-103, doi:10.1016/j.biosystems.2008.12.004 (2009).

20 Alsadeq, A. *et al.* Zeta-chain-associated protein kinase-70 (ZAP-70) promotes acute lymphoblastic leukemia (ALL) infiltration into the central nervous system by enhancing chemokine receptor 7 (CCR7) expression. *Blood* **126**, 901 (2015).

21 Chai, L. E. *et al.* A review on the computational approaches for gene regulatory network construction. *Comput Biol Med* **48**, 55-65, doi:10.1016/j.compbiomed.2014.02.011 (2014).

22 Bonneau, R. *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* **7**, R36, doi:10.1186/gb-2006-7-5-r36 (2006).

23 Madar, A., Greenfield, A., Ostrer, H., Vanden-Eijnden, E. & Bonneau, R. The Inferelator 2.0: a scalable framework for reconstruction of dynamic regulatory network models. *Conf Proc IEEE Eng Med Biol Soc* **2009**, 5448-5451, doi:10.1109/IEMBS.2009.5334018 (2009).

24 Mitchell, K. *et al.* IL1RAP potentiates multiple oncogenic signaling pathways in AML. *J Exp Med* **215**, 1709-1727, doi:10.1084/jem.20180147 (2018).

25 Agerstam, H. *et al.* Antibodies targeting human IL1RAP (IL1R3) show therapeutic effects in xenograft models of acute myeloid leukemia. *Proc Natl Acad Sci U S A* **112**, 10786-10791, doi:10.1073/pnas.1422749112 (2015).

26 Santiago, L., Daniels, G., Wang, D., Deng, F. M. & Lee, P. Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment. *Am J Cancer Res* **7**, 1389-1406 (2017).

27 Kuhnl, A. *et al.* Overexpression of LEF1 predicts unfavorable outcome in adult patients with B-precursor acute lymphoblastic leukemia. *Blood* **118**, 6362-6367, doi:10.1182/blood-2011-04-350850 (2011).

28 Miraldi, E. R. *et al.* Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Res* **29**, 449-463, doi:10.1101/gr.238253.118 (2019).

29 Castro, D. M., de Veaux, N. R., Miraldi, E. R. & Bonneau, R. Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput Biol* **15**, e1006591, doi:10.1371/journal.pcbi.1006591 (2019).

30 Arrieta-Ortiz, M. L. *et al.* An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. *Mol Syst Biol* **11**, 839, doi:10.15252/msb.20156236 (2015).

31 Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**, 289-303, doi:10.1016/j.cell.2012.09.016 (2012).

32 Vagace, J. M. & Gervasini, G. Chemotherapy toxicity in patients with acute leukemia. *INTECH Open Access Publisher* (2011).

33 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

34 Croker, B. A., Kiu, H. & Nicholson, S. E. SOCS regulation of the JAK/STAT signalling pathway. *Semin Cell Dev Biol* **19**, 414-422, doi:10.1016/j.semcdb.2008.07.010 (2008).

35 Tomeczkowski, J. *et al.* Absence of G-CSF receptors and absent response to G-CSF in childhood Burkitt's lymphoma and B-ALL cells. *British journal of haematology* **89**, 771-779 (1995).

36 Meyer, C. *et al.* Establishment of the B cell precursor acute lymphoblastic leukemia cell line MUTZ-5 carrying a (12:13) translocation. *Leukemia* **15**, 1471-1474 (2001).

37 Smith, R. G., Dev, V. G. & Shannon, W. A., Jr. Characterization of a novel human pre-B leukemia cell line. *J Immunol* **126**, 596-602 (1981).

38 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

39 Greenfield, A., Hafemeister, C. & Bonneau, R. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* **29**, 1060-1067, doi:10.1093/bioinformatics/btt099 (2013).

40 Koon, H. B., Ippolito, G. C., Banham, A. H. & Tucker, P. W. FOXP1: a potential therapeutic target in cancer. *Expert Opin Ther Targets* **11**, 955-965, doi:10.1517/14728222.11.7.955 (2007).

41 Schwartzman, O. *et al.* Suppressors and activators of JAK-STAT signaling at diagnosis and relapse of acute lymphoblastic leukemia in Down syndrome. *Proc Natl Acad Sci U S A* **114**, E4030-E4039, doi:10.1073/pnas.1702489114 (2017).

42 Zhang, S. *et al.* Essential role of signal transducer and activator of transcription (Stat)5a but not Stat5b for Flt3-dependent signaling. *J Exp Med* **192**, 719-728, doi:10.1084/jem.192.5.719 (2000).

43 Agirre, X. *et al.* ASPP1, a common activator of TP53, is inactivated by aberrant methylation of its promoter in acute lymphoblastic leukemia. *Oncogene* **25**, 1862-1870, doi:10.1038/sj.onc.1209236 (2006).

44 Bergamaschi, D. *et al.* ASPP1 and ASPP2: common activators of p53 family members. *Mol Cell Biol* **24**, 1341-1350, doi:10.1128/mcb.24.3.1341-1350.2004 (2004).

45 Loudin, M. G. *et al.* Genomic profiling in Down syndrome acute lymphoblastic leukemia identifies histone gene deletions associated with altered methylation profiles. *Leukemia* **25**, 1555-1563, doi:10.1038/leu.2011.128 (2011).

Fig. 1

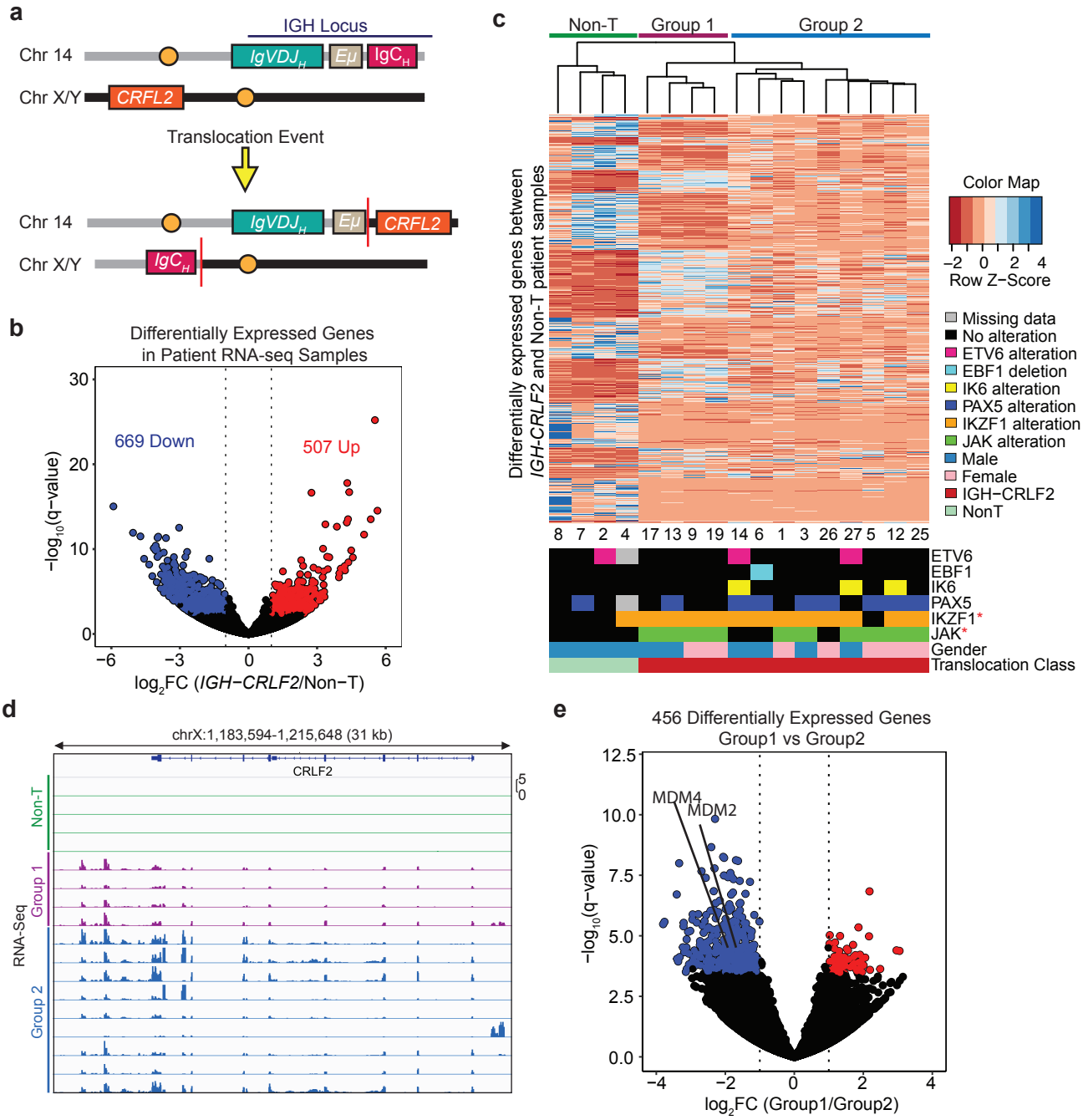

a
IGH Locus
Chr 14
IgVDJ_H
Eμ
IgC_H
Chr X/Y
CRFL2
Translocation Event
Chr 14
IgVDJ_H
Eμ
CRFL2
Chr X/Y
IgC_H
b
Differentially Expressed Genes
in Patient RNA-seq Samples
669 Down
507 Up
−log10(q−value)
log2FC (IGH−CRLF2/Non−T)
c
Non-T
Group 1
Group 2
Differentially expressed genes between
IGH-CRLF2 and Non-T patient samples
Color Map
Row Z−Score
Missing data
No alteration
ETV6 alteration
EBF1 deletion
IK6 alteration
PAX5 alteration
IKZF1 alteration
JAK alteration
Male
Female
IGH−CRLF2
NonT
8 7 2 4 17 13 9 19 14 6 1 3 26 27 5 12 25
ETV6
EBF1
IK6
PAX5
IKZF1*
JAK*
Gender
Translocation Class
d
chrX:1,183,594-1,215,648 (31 kb)
CRLF2
RNA-Seq
Non-T
Group 1
Group 2
e
456 Differentially Expressed Genes
Group1 vs Group2
MDM4
MDM2
−log10(q−value)
log2FC (Group1/Group2)

Fig.2

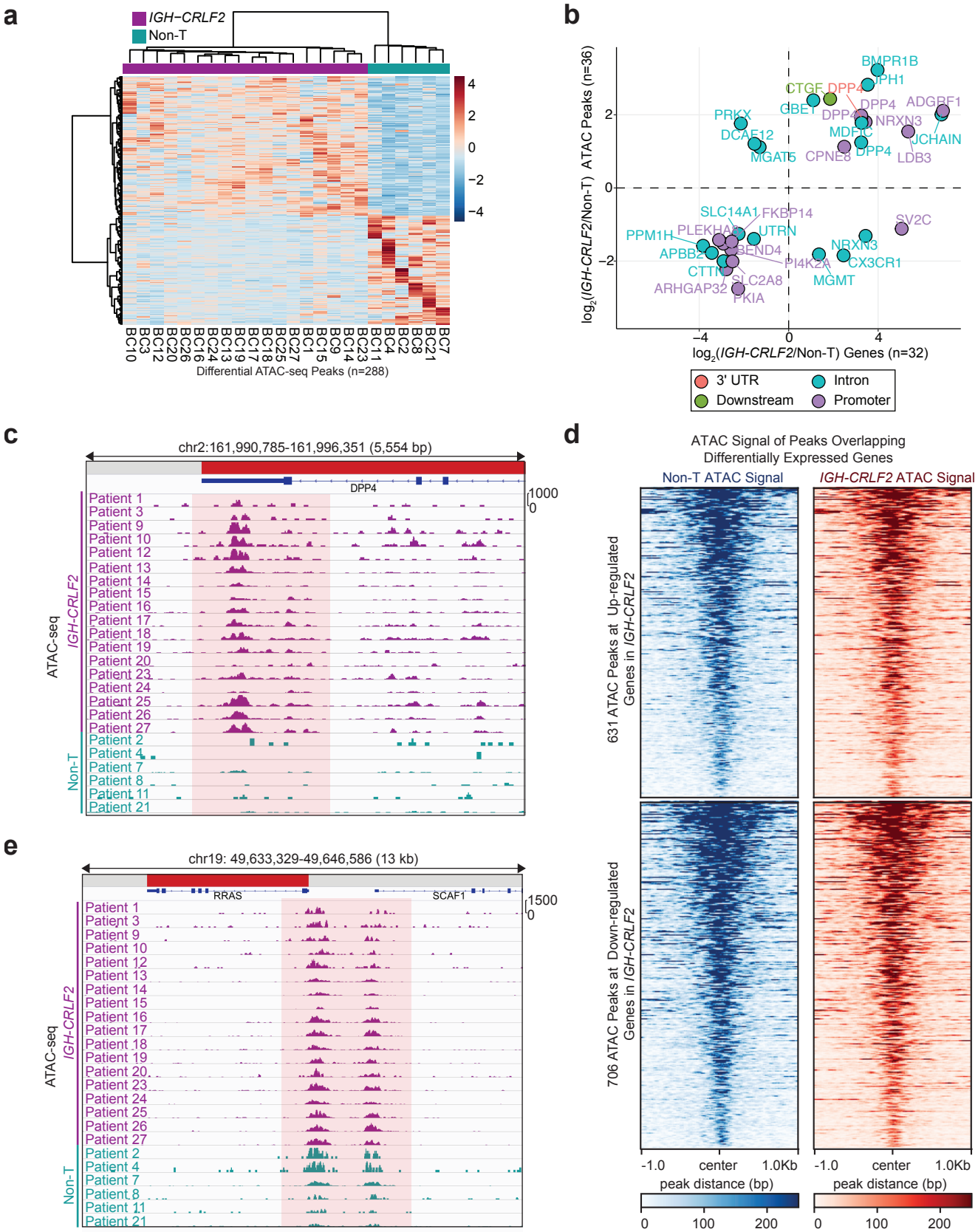

a
IGH−CRLF2
Non-T
Differential ATAC-seq Peaks (n=288)
b
log2(IGH-CRLF2/Non-T) ATAC Peaks (n=36)
log2(IGH-CRLF2/Non-T) Genes (n=32)
3' UTR
Downstream
Intron
Promoter
c
chr2:161,990,785-161,996,351 (5,554 bp)
DPP4
ATAC-seq
IGH-CRLF2
Non-T
d
ATAC Signal of Peaks Overlapping Differentially Expressed Genes
Non-T ATAC Signal
IGH-CRLF2 ATAC Signal
631 ATAC Peaks at Up-regulated Genes in IGH-CRLF2
706 ATAC Peaks at Down-regulated Genes in IGH-CRLF2
peak distance (bp)
e
chr19: 49,633,329-49,646,586 (13 kb)
RRAS
SCAF1
ATAC-seq
IGH-CRLF2
Non-T

Fig. 3

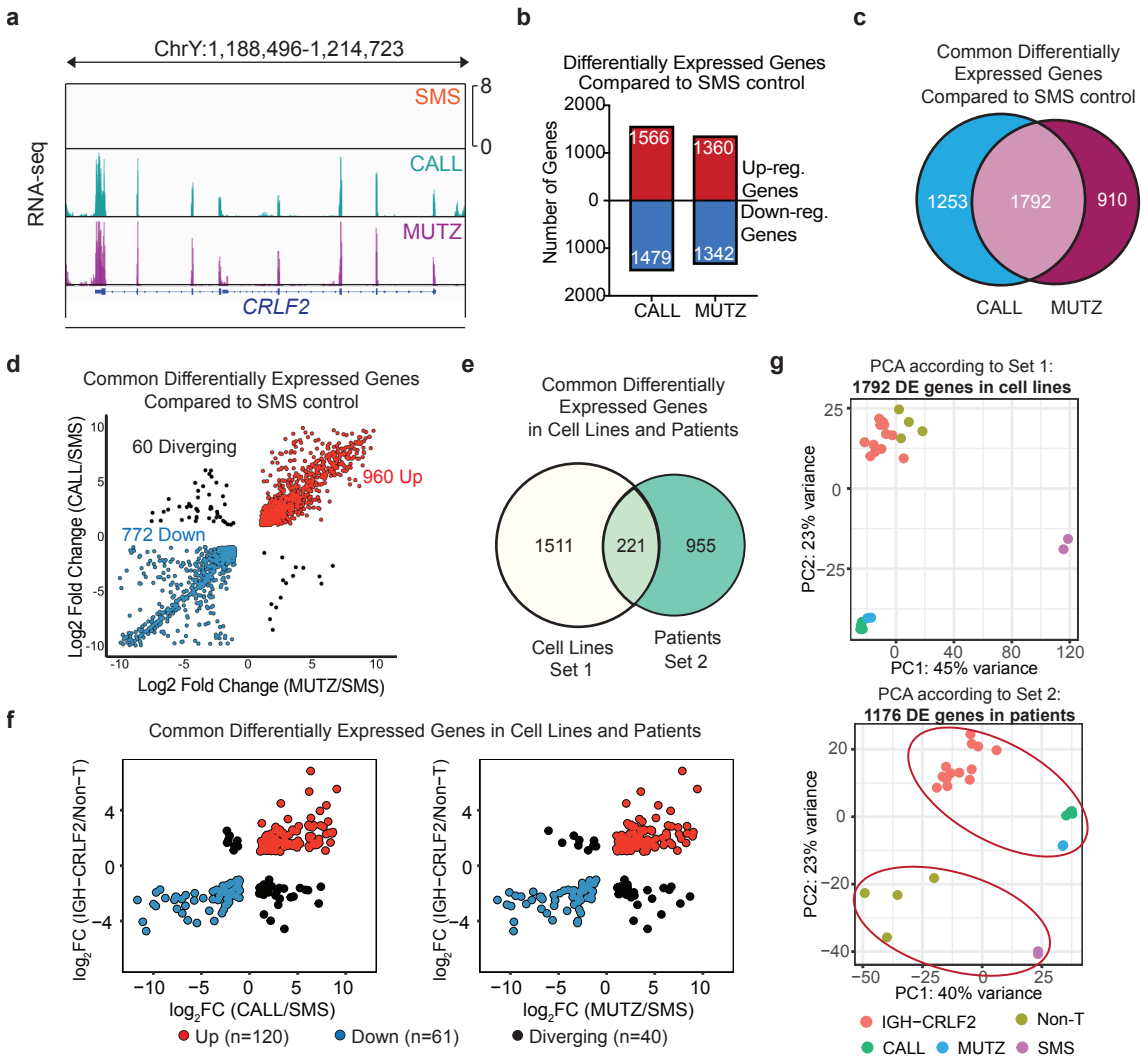

a
ChrY:1,188,496-1,214,723
RNA-seq
SMS
CALL
MUTZ
8
0
CRLF2
b
Differentially Expressed Genes
Compared to SMS control
Number of Genes
2000
1000
0
1000
2000
1566
1360
1479
1342
Up-reg.
Genes
Down-reg.
Genes
CALL
MUTZ
c
Common Differentially
Expressed Genes
Compared to SMS control
1253
1792
910
CALL
MUTZ
d
Common Differentially Expressed Genes
Compared to SMS control
60 Diverging
960 Up
772 Down
Log2 Fold Change (CALL/SMS)
Log2 Fold Change (MUTZ/SMS)
e
Common Differentially
Expressed Genes
in Cell Lines and Patients
1511
221
955
Cell Lines
Set 1
Patients
Set 2
g
PCA according to Set 1:
1792 DE genes in cell lines
PC2: 23% variance
PC1: 45% variance
PCA according to Set 2:
1176 DE genes in patients
PC2: 23% variance
PC1: 40% variance
IGH−CRLF2
Non-T
CALL
MUTZ
SMS
f
Common Differentially Expressed Genes in Cell Lines and Patients
log2FC (IGH−CRLF2/Non−T)
log2FC (CALL/SMS)
log2FC (MUTZ/SMS)
Up (n=120)
Down (n=61)
Diverging (n=40)

Fig. 4

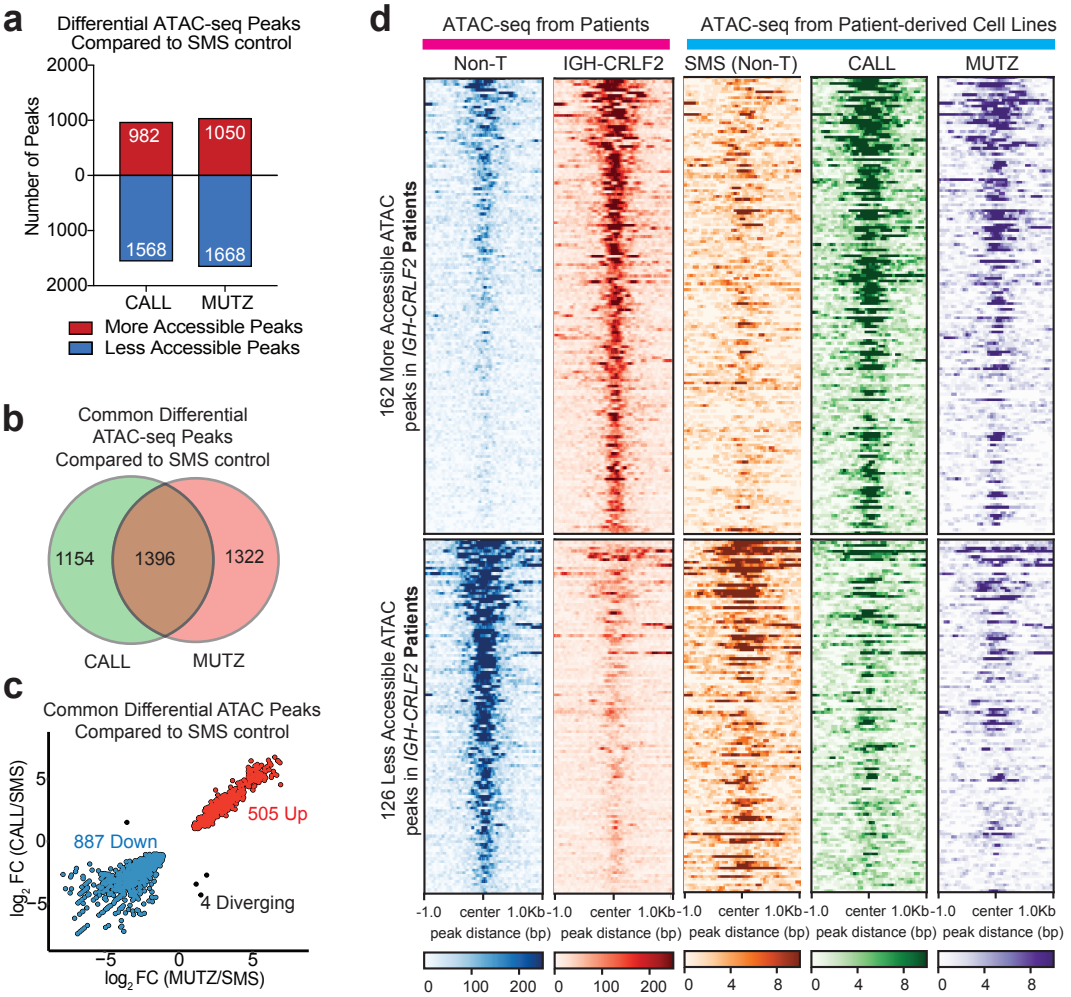

a
Differential ATAC-seq Peaks
Compared to SMS control
Number of Peaks
2000
1000
0
1000
2000
982
1050
1568
1668
CALL
MUTZ
More Accessible Peaks
Less Accessible Peaks
b
Common Differential
ATAC-seq Peaks
Compared to SMS control
1154
1396
1322
CALL
MUTZ
c
Common Differential ATAC Peaks
Compared to SMS control
log2 FC (CALL/SMS)
log2 FC (MUTZ/SMS)
505 Up
887 Down
4 Diverging
d
ATAC-seq from Patients
ATAC-seq from Patient-derived Cell Lines
Non-T
IGH-CRLF2
SMS (Non-T)
CALL
MUTZ
162 More Accessible ATAC peaks in IGH-CRLF2 Patients
126 Less Accessible ATAC peaks in IGH-CRLF2 Patients
-1.0 center 1.0Kb
peak distance (bp)

Fig. 5

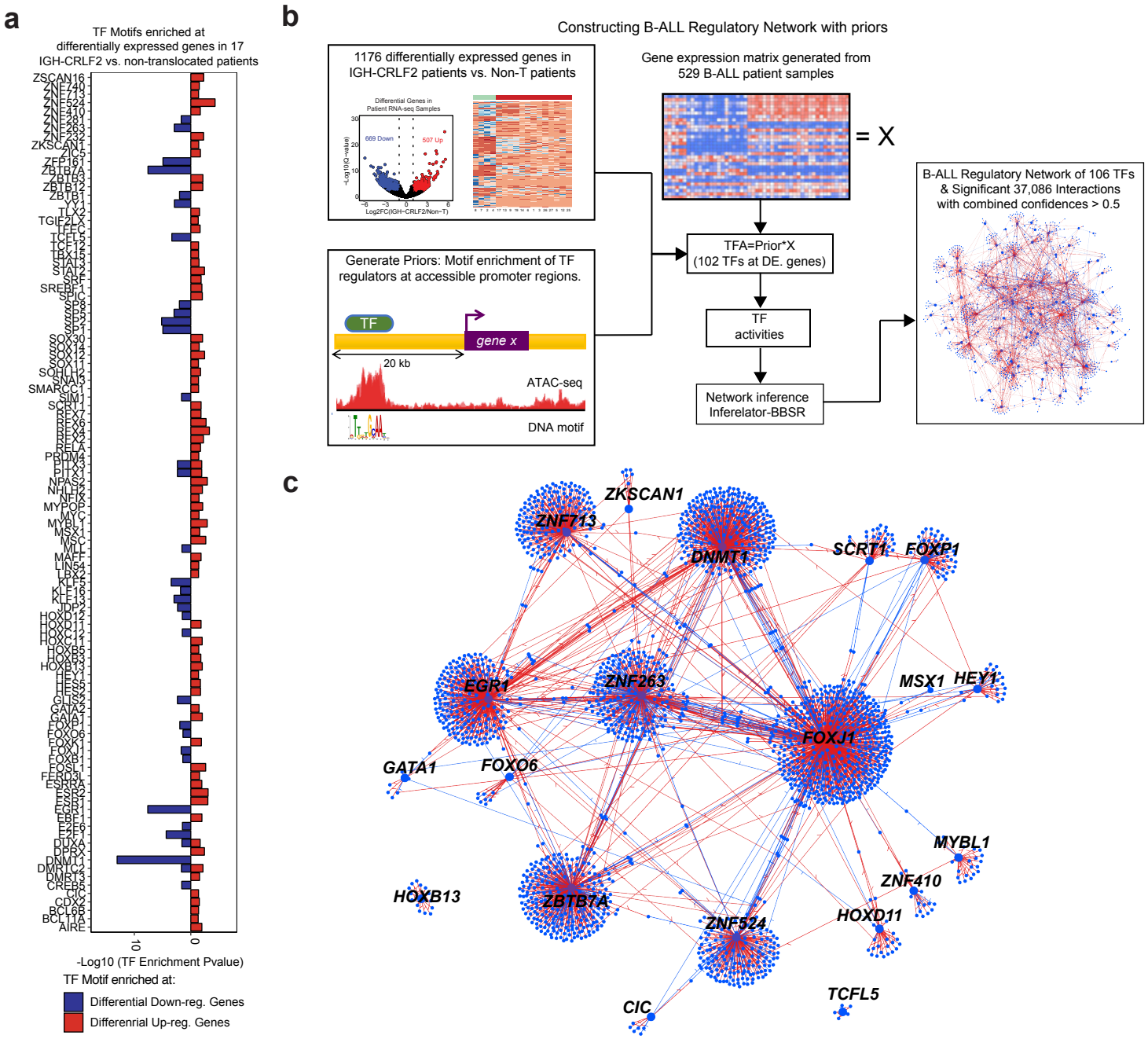

a
TF Motifs enriched at differentially expressed genes in 17 IGH-CRLF2 vs. non-translocated patients
-Log10 (TF Enrichment Pvalue)
TF Motif enriched at:
Differential Down-reg. Genes
Differenrial Up-reg. Genes
b
Constructing B-ALL Regulatory Network with priors
1176 differentially expressed genes in IGH-CRLF2 patients vs. Non-T patients
Differential Genes in Patient RNA-seq Samples
669 Down
507 Up
Log2FC(IGH-CRLF2/Non-T)
Gene expression matrix generated from 529 B-ALL patient samples
= X
Generate Priors: Motif enrichment of TF regulators at accessible promoter regions.
TF
gene x
20 kb
ATAC-seq
DNA motif
TFA=Prior*X
(102 TFs at DE. genes)
TF activities
Network inference
Inferelator-BBSR
B-ALL Regulatory Network of 106 TFs & Significant 37,086 Interactions with combined confidences > 0.5
c
ZKSCAN1
ZNF713
DNMT1
SCRT1
FOXP1
EGR1
ZNF263
MSX1
HEY1
FOXJ1
GATA1
FOXO6
MYBL1
ZNF410
HOXB13
ZBTB7A
HOXD11
ZNF524
CIC
TCFL5

Fig. 6

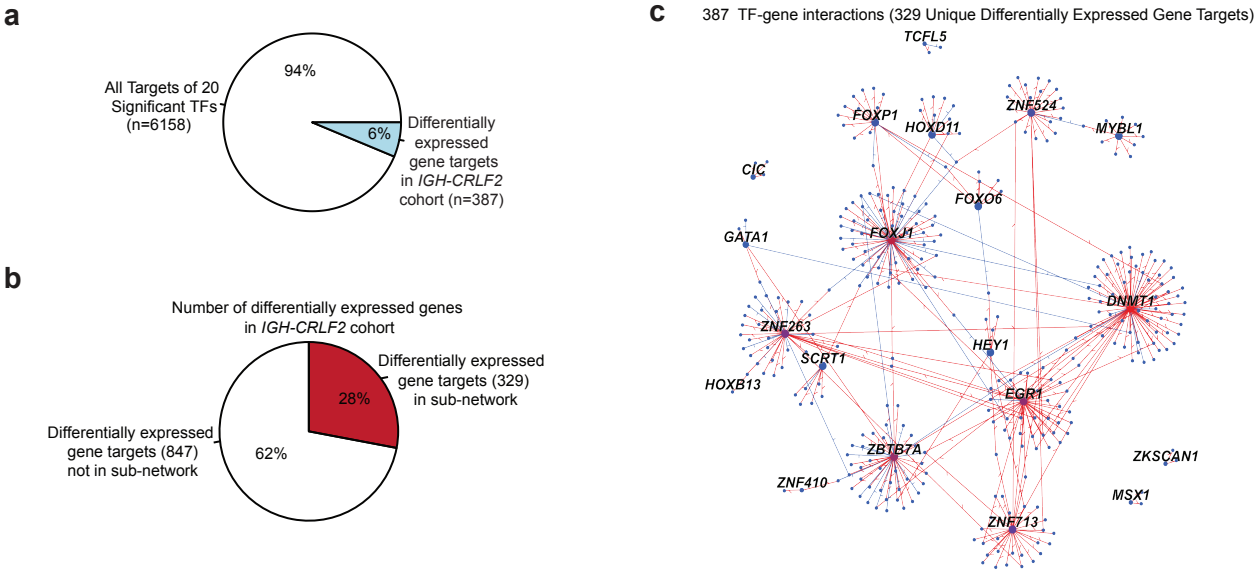
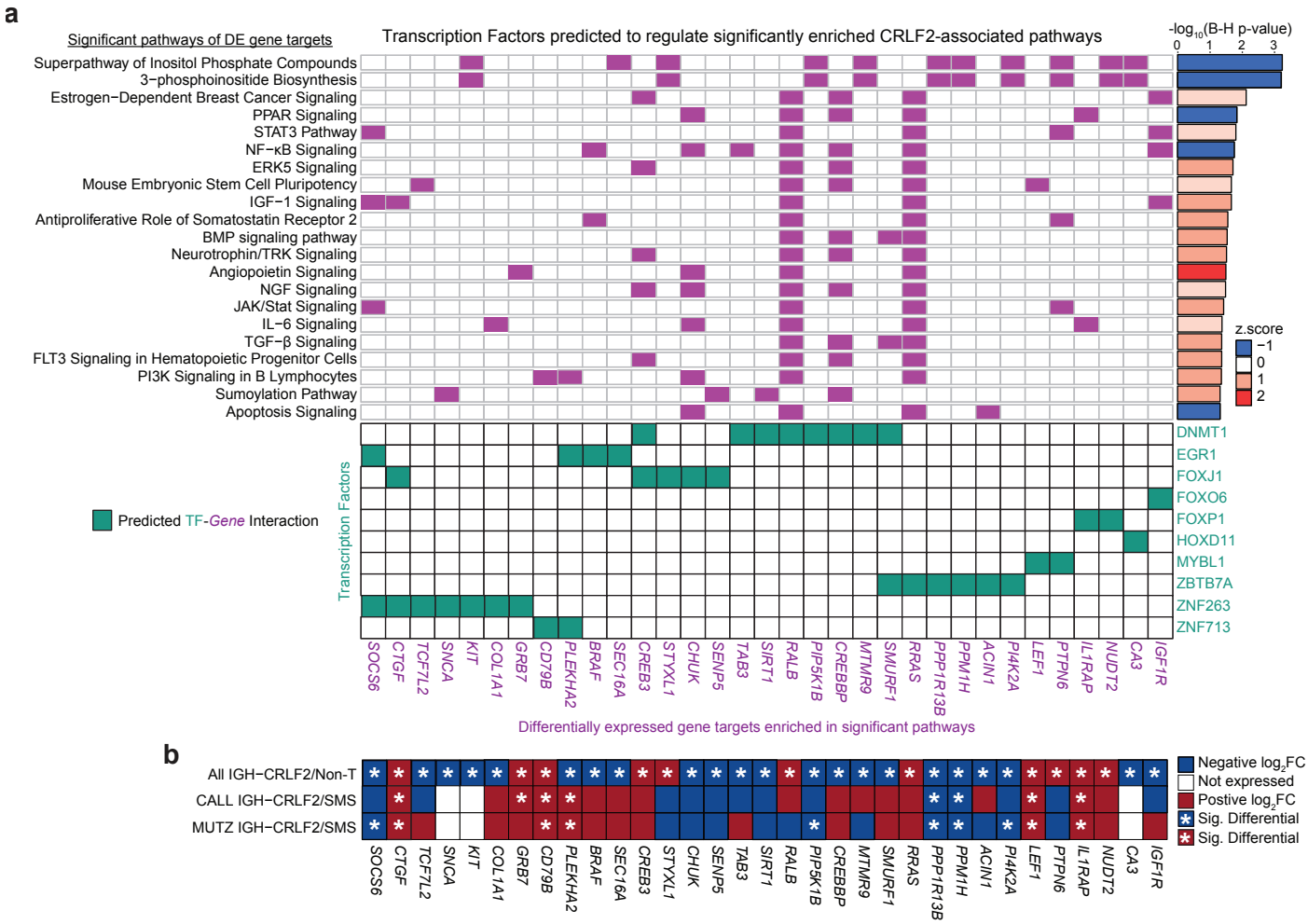

a
94%
6%
All Targets of 20 Significant TFs (n=6158)
Differentially expressed gene targets in IGH-CRLF2 cohort (n=387)
b
Number of differentially expressed genes in IGH-CRLF2 cohort
28%
62%
Differentially expressed gene targets (329) in sub-network
Differentially expressed gene targets (847) not in sub-network
c
387 TF-gene interactions (329 Unique Differentially Expressed Gene Targets)
TCFL5
FOXP1
HOXD11
ZNF524
MYBL1
CIC
FOXO6
GATA1
FOXJ1
DNMT1
ZNF263
HEY1
SCRT1
HOXB13
EGR1
ZBTB7A
ZKSCAN1
ZNF410
MSX1
ZNF713

Fig. 7


a
Significant pathways of DE gene targets
Transcription Factors predicted to regulate significantly enriched CRLF2-associated pathways
-log10(B-H p-value)
0 1 2 3
Superpathway of Inositol Phosphate Compounds
3−phosphoinositide Biosynthesis
Estrogen−Dependent Breast Cancer Signaling
PPAR Signaling
STAT3 Pathway
NF−κB Signaling
ERK5 Signaling
Mouse Embryonic Stem Cell Pluripotency
IGF−1 Signaling
Antiproliferative Role of Somatostatin Receptor 2
BMP signaling pathway
Neurotrophin/TRK Signaling
Angiopoietin Signaling
NGF Signaling
JAK/Stat Signaling
IL−6 Signaling
TGF−β Signaling
FLT3 Signaling in Hematopoietic Progenitor Cells
PI3K Signaling in B Lymphocytes
Sumoylation Pathway
Apoptosis Signaling
z.score
−1
0
1
2
Predicted TF-Gene Interaction
Transcription Factors
DNMT1
EGR1
FOXJ1
FOXO6
FOXP1
HOXD11
MYBL1
ZBTB7A
ZNF263
ZNF713
SOCS6
CTGF
TCF7L2
SNCA
KIT
COL1A1
GRB7
CD79B
PLEKHA2
BRAF
SEC16A
CREB3
STYXL1
CHUK
SENP5
TAB3
SIRT1
RALB
PIP5K1B
CREBBP
MTMR9
SMURF1
RRAS
PPP1R13B
PPM1H
ACIN1
PI4K2A
LEF1
PTPN6
IL1RAP
NUDT2
CA3
IGF1R
Differentially expressed gene targets enriched in significant pathways

b
All IGH−CRLF2/Non-T
CALL IGH−CRLF2/SMS
MUTZ IGH−CRLF2/SMS
Negative log2FC
Not expressed
Postive log2FC
Sig. Differential
Sig. Differential
SOCS6
CTGF
TCF7L2
SNCA
KIT
COL1A1
GRB7
CD79B
PLEKHA2
BRAF
SEC16A
CREB3
STYXL1
CHUK
SENP5
TAB3
SIRT1
RALB
PIP5K1B
CREBBP
MTMR9
SMURF1
RRAS
PPP1R13B
PPM1H
ACIN1
PI4K2A
LEF1
PTPN6
IL1RAP
NUDT2
CA3
IGF1R