Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Punwani D, Zhang Y, Yu J, et al. Multisystem anomalies in severe combined immunodeficiency with mutant *BCL11B*. N Engl J Med 2016;375:2165-76. DOI: 10.1056/NEJMoa1509164

Supplementary Appendix, Multisystem Anomalies in BCL11B Mutant Human Severe Combined Immunodeficiency

Table of Contents:

Investigators:

Divya Punwani, Ph.D.[1,2], Yong Zhang, M.D., Ph.D.[2,3], Jason Yu, Ph.D.[1], Morton J. Cowan, M.D.[1], Sadhna Rana, Ph.D.[4], Antonia Kwan, M.D., Ph.D.[1], Aashish N. Adhikari, Ph.D.[5], Carlos O. Lizama, Ph.D.[6], Bryce A. Mendelsohn, M.D.[1], Shawn P. Fahl, Ph.D.[3], Ajithavalli Chellappan[4], Rajgopal Srinivasan, Ph.D.[4], Steven E. Brenner, Ph.D.[5], David L. Wiest, Ph.D.[3,7], Jennifer M. Puck, M.D.[1,7]

[1] Department of Pediatrics, University of California San Francisco School of Medicine and UCSF Benioff Children's Hospital, San Francisco, CA 94158

[2] Drs. Punwani and Zhang contributed equally.

[3] Blood Cell Development and Function Program, Fox Chase Cancer Center, Philadelphia, PA 19111-2497

[4] Innovation Labs, Tata Consulting Services, Hyderabad 50019, Telangana, India

[5] Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

[6] Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA 94143

[7] Drs. Wiest and Puck contributed equally.

Supplementary methods

Genetic and whole exome analysis

Initial Sanger sequence of B+ SCID genes *CD3D, CD3E, CDEZ, IL2RG, IL7R, JAK3, PNP, PTPRC, RMRP,* and *ZAP70* in the patient was normal. The patient and healthy father shared a heterozygous deletion within chromosome 3q25.1 (genes *CLRN1-AS1* to *P2RY12*, inclusive) of unknown significance.

Exome sequencing (HiSeq2500, Illumina) followed capture with a Nimblegen v3.0 kit with analysis to identify rare protein-altering variants based on human genome build 37 (Genome Reference Consortium GRCh37, Feb 2009). Read alignment and variant calling were performed as described.[1] Basic quality metrics are in Table S1. Using an in-house annotation tool, Varant (30 Sep 2014)[2] variants were annotated as NonSyn, StartGain, StartLoss, StopGain, StopLoss, FrameShiftInsert, FrameShiftDelete, NonFrameShiftDelete, NonFrameShiftInsert, SpliceAcceptor, SpliceDonor, CDS_not_multiple_of_3. Only variants with genotype quality (GQ) scores ≥30 and minor allele frequencies (MAF) ≤0.02 in both 1000 Genomes Phase 3 (accessed Nov 2014)[3] and Exome Sequencing Project database, v2[4] were retained for downstream analysis. Additionally, variants in the proband were required to follow compound heterozygous, homozygous recessive, X-linked recessive, or uniparental disomy inheritance models or to be *de novo* (absent in both parents). These filtering steps shortlisted 57 variants in 38 genes for further analysis (Figure S1).

No shortlisted variants were in our target gene list (Table S2) extracted from reported human primary immunodeficiency (PID) genes.[5] We therefore expanded the list to include interacting partners of the target genes, extracted from BioGRID v3.2,[6] MINT release 21 Dec 2010[7] and HPRD release 9 (13 Apr 2010),[8] as included in the Pathway Commons database v5 (Nov 2011).[9] The list was further expanded to include genes associated with T cell related phenotype terms (Table S3) in either the Mouse Genome Informatics (MGI) (12 Jan 2014),[10] with files accessed from databases:
ftp://ftp.informatics.jax.org/pub/reports/index.html,
ftp://ftp.informatics.jax.org/pub/reports/VOC_MammalianPhenotype.rpt,
ftp://ftp.informatics.jax.org/pub/reports/MGI_PhenoGenoMP.rpt ,
ftp://ftp.informatics.jax.org/pub/reports/HMD_HumanPhenotype.rpt,
ftp://ftp.informatics.jax.org/pub/reports/MPheno_OBO.ontology, or in the Human Phenotype Ontology Annotation (HPO) (28 Jan 2014):
http://compbio.charite.de/jenkins/job/hpo.annotations.monthly/lastStableBuild/artifact/annotation/ALL_SOURCES_ALL_FREQUENCIES_diseases_to_genes_to_phenotypes.txt and
http://compbio.charite.de/jenkins/job/hpo.annotations/lastStableBuild/artifact/misc/phenotype_annotation.tab[11]

The shortlisted 57 variants were matched against this expanded gene list, resulting in 5 high quality variants in 3 genes: *BCL11B*, *NRLP1*, and *USO1*. We also considered a variant in the *CAPNS1* gene with genotype quality of 27, slightly lower than our default threshold of confidence. The variant genes were scored as described below.

Gene and variant prioritization (annotations, data sources and scoring criteria)

We developed an *ad hoc* approach to ranking candidate genes and variants among the shortlisted variants, looking for genes with features including interactions and shared pathways similar to those of known target genes. Additionally, we identified variants

predicted to be deleterious and following an appropriate inheritance pattern. To prioritize genes, we created a profile of target gene characteristics based on enriched annotation terms. We obtained enriched ($p < 0.05$) annotation terms (Table 4) for the annotation categories medical conditions, pathways, genome ontology, human phenotype ontology, and mouse phenotype ontology, by uploading the target gene list to the ToppFun website (accessed 20 Oct 2013) at https://toppgene.cchmc.org/enrichment.jsp.[12] We scored the top shortlisted variants as follows:

1) The interacting partners of target genes were extracted from BioGRID v3.2,[6] MINT release 21 (Dec 2010)[7] and HPRD release 9 (13 Apr 2010)[8] databases as included in the Pathway Commons database v5 (Nov 2011).[9] A candidate gene interacting with at least one target gene based on a high-throughput study was assigned a score of 1. If the interaction was supported by at least one low-throughput experiment (biochemical studies *in vitro* or *in vivo*) it was assigned a score of 2.

2) Mammalian Phenotype Ontology (MPO) and HPO terms were sought that matched those assigned to known target genes. Gene-to-phenotype associations from MPO were from the MGI website (12 Jan 2014) ftp://ftp.informatics.jax.org/pub/reports/index.html).[10] HPO associations (28 Jan 2014) were from http://compbio.charite.de/jenkins/job/hpo.annotations.monthly/lastStableBuild/artifact/annotation/ALL_SOURCES_ALL_FREQUENCIES_diseases_to_genes_to_phenotypes.txt and http://compbio.charite.de/jenkins/job/hpo.annotations/lastStableBuild/artifact/misc/phenotype_annotation.tab).[11] A candidate gene associated with low or absent T cells was assigned a score of 2 (Table S3), reflecting the proband's phenotype. Any other phenotype of the candidate gene that matched a phenotype in the target gene profile was assigned a score of 0.5.

3) Gene ontology (GO) terms matching those assigned to known target genes were extracted from the UniProt database (Nov 2014), ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/taxonomic_divisions/uniprot_sprot_human.dat.gz.[13] A candidate gene with the following enriched GO terms was awarded a score of 1: T cell activation, T cell differentiation, T cell differentiation in thymus, regulation of T cell activation, T cell receptor signaling pathway, regulation of T cell differentiation, T cell receptor complex, T cell receptor binding.

All other GO annotations for a candidate gene that matched those in the target gene profile received a score of 0.5.

4) Gene-pathway associations were extracted from the Pathway Commons v5 (Nov 2011).[9] We used its constituent databases: NCI-NATURE from the Pathway Interaction Database,[14] HumanCyc v17.1[15] and Reactome v46. A candidate gene that shared at least one pathway with the pathways in the target gene profile was assigned a score of 1. Exceptions to this rule were enriched pathways associated with candidate genes that we deemed to be less informative. Specifically the pathways "adaptive immune system," "innate immune system," and "apoptosis" were assigned a score of 0.5.

5) Gene associations with medical conditions were extracted from Online Mendelian Inheritance in Man (OMIM) (accessed 21 Jan 2014),[17] Genetic Association Database (GAD) (18 Jan 2014), http://geneticassociationdb.nih.gov/[18] and NHGRI-GWAS (21 Jan 2014) v1 accessed from www.ebi.ac.uk/gwas.[19] A candidate gene that shared at least one condition with one of the medical conditions in the target gene profile was assigned a score of 1.

We also scored genes based on annotations of their variants, as follows:

6) Homozygous, compound heterozygous, or X-linked variants were given a score of 2. A list of putative haploinsufficient genes compiled from an automated OMIM and literature survey[20] was used to annotate the candidate genes. All *de novo* variants in putatively haploinsufficient genes were assigned a score of 2, and the remaining *de novo* variants were assigned a score of 1.

7) Variants predicted to be deleterious by any of PolyPhen2 v2.2.2 (Feb 2012) http://genetics.bwh.harvard.edu/pph2/[21] SIFT (Human_db_37_ensembl_63, Aug 2011 http://sift.jcvi.org)[22] or CADD version 1.2 (CADD score ≥20)[23] were assigned a score of 2. The PolyPhen2 and SIFT scores were extracted from dbNSFP v2.1.[24, 25] Variants predicted to be probably deleterious by PolyPhen2 or having CADD scores 14-20 were assigned a score of 1. Variants with CADD scores <14 were assigned a score of 0.5.

The variant and gene annotations and their scores are summarized in Tables S5 and S6. We also analyzed heterozygous gene deletions in the proband inherited from the father (*CLRN1-AS1, MED12L, RNA5SP145, GPR171, P2RY14, SETP11, GPR87, P2RY13, P2RY12*); however, our we found no associations between these genes and our target genes.

Manual analysis of top candidate gene BCL11B

The B-Cell CLL/Lymphoma 11B gene (*BCL11B*) harboring a heterozygous *de novo* missense mutation emerged as the best candidate (Table S6), based on the following annotations:

1) Protein-protein interactions reported in the BioGRID database revealed associations of murine Bcl11b with Notch1, Il2, and Foxp3 proteins (proteins involved in T cell development and activation). Since neither Notch1 nor Foxp3 were part of our target gene list, only the Il2-Bcl11b interaction was identified by our automated approach. The Notch1 and FoxP3 interactions were noted later based on a manual survey of the literature.[26, 27] As reported in BioGRID database, the human Notch1-BCL11B interaction and murine Foxp3-Bcl11b interaction were identified by high-throughput affinity capture experiments followed by western blotting.[28, 29] The mouse Bcl11b-Il2 interaction was supported by biochemical studies[30] and hence was given higher priority than interactions in other shortlisted genes where such support was lacking.

2) Querying the MPO for T cell related key terms found Bcl11b annotated with "decreased double-negative T cell number," or reduced number thymic T cells that express neither CD4 nor CD8. Relevance of this annotation was confirmed by literature search, as the paper[31] reported block of T cell development at CD4-CD8- double negative stage in mice with defective *Bcl11b*($^{-/-}$) gene.

3) The GO annotation for *BCL11B* "T cell differentiation in thymus" was extracted automatically from the UniProt database (Nov 2014).[13] Manual inspection revealed that it was inferred from electronic annotation. To find the original annotation resource linked to publication, we browsed the AmiGO website that gives original databases and publications for their annotations,[32] learning that the GO annotation was derived from the MGI database with support from the literature on experimental validation of the association of *BCL11B* with T cell development.[31]

4) The Pathways Commons (v5) used for the analysis did not report *BCL11B* to be associated with any of the enriched pathways used in this analysis. A manual review of the literature identified a study showing that *BCL11B* has a role in TCR/CD28 signaling to

trigger T cell activation.[37] The gene participates in the activation of *IL2* gene expression by enhancing the activation of NFKB via enhanced cot-kinase expression.

5) Disease databases used in this study did not report association of *BCL11B* with human immunodeficiency. However, mouse studies indicated involvement of *Bcl11b* in inflammatory bowel disease[38] and autoimmune encephalomyelitis.[39] Human *BCL11B* is associated with lymphoblastic leukemia.[40]

6) *BCL11B* was annotated as putatively haploinsufficient by automated text extraction,[20] based on a PubMed abstract showing *Bcl11b*$^{+/-}$ mice susceptible to γ-ray-induced thymic lymphomas.[41]

7) The *BCL11B* variant was predicted deleterious by SIFT and PolyPhen2, while CADD assigned a borderline score of 14. We confirmed the effect prediction for the *BCL11B* variant using SNAP (SNAPfun, 2 Apr 2014), http://rostlab.org/services/snap/.[42] Additional observations of conservation, allele frequency and other features confirmed the automated tools' predictions. For example, GERP++ from DBNSFP2.0[43] gave an RS score of 3.96 for the *BCL11B* variant position, with scores >2 usually regarded as evolutionarily conserved and potentially functional.

8) We also investigated the protein domains and alignments of BCL11B, a zinc finger transcription factor that binds DNA via its second and third zinc finger domains. The proband's *de novo* missense variant pN441K lies in the second zinc finger domain spanning residues 427-455, completely conserved between mice and humans, and the amino acid N441 is completely conserved in mammals (Figure S2).

Chromatin immunoprecipitation-sequencing (ChIP-seq)

A schematic representation of the workflow for the Chromatin immunoprecipitation-sequencing experiment is shown in Figure S6,A and described below.

Transduction of human HSC

WT and mutant (mut) *BCL11B* cDNAs were 3'-FLAG-tagged and subcloned into the green fluourescent protein (GFP)-expressing, lentiviral vector pLenti CMV/TO GFP-Zeo DEST (Addgene). Human $CD34^+$ HSC selected from peripheral blood mobilized stem cells using CD34 microbeads (Miltenyi Biotech) were transduced with WT or mutant *BCL11B* or pLenti CMV/TO GFP-Zeo DEST vector, with no *BCL11B* insert (empty vector, GFP control). The cells were cultured for 14 days at 37°C in HSC medium: X-Vivo-15 (Lonza) with 100ng/ml human SCF, Flt3L and thrombopoietin, and 20ng/ml IL-3 (all from Peprotech). DNA from an aliquot of cells was extracted on day 10 of the culture to determine vector copy number by quantitative PCR.

- HSC + WT *BCL11B*: VCN = 6±1
- HSC + mut *BCL11B*: VCN = 10±1
- HSC + empty vector, GFP control: VCN = 11±1

Selection of transduced HSC and ChIP-seq

GFP expressing transduced cells were sorted into three separate tubes from each sample (FACS Jazz, BD Biosciences) and DNA from cells in each of the triplicate tubes was sonicated to mean size of 250 bp and used for chromatin immunoprecipitation with the iDeal ChIP-seq kit for Transcription Factors (Diagenode), following the manufacturer's instructions. DNA co-precipitated with anti-FLAG antibody (M2, Sigma Aldrich) was sequenced by the University of California, Davis, Genome Center on a HiSeq 4000 sequencer (Illumina) to obtain 90bp single reads.

Analysis of sequences

The ChIP-seq reads were groomed using FASTQ groomer V1.04,[44] and groomed reads were mapped to the human genome (build hg38) using BWA for the Illumina v1.2.3 program.[45] The MACS callpeak v2.1.0.201511222.0 program[46] was used to determine enriched BCL11B peaks using GFP-ChIP as control. The parameters for MACS were: effective genome size = 2.45e09; tag size = 36 (use 50, sequence 36 bp); model fold 2.5; FDR (q-value) cutoff 0.05. IDR 2012[47] was used to check the reproducibility of the peaks among replicated with an FDR (q-value) cutoff 0.05. The ChIP-seq peaks were visualized using IGB v9.0.0.[48] To insure that all of the experiments were processed consistently, all of the above steps were performed as part of a Galaxy workflow, which can be found at https://usegalaxy.org/u/laiumiunix/w/chipseq. All sequences were uploaded to the Gene Expression Omnibus (GEO) database, http://www.ncbi.nlm.nih.gov/geo/; accession number GSE84941.

Results

Two BCL11B target binding sites were found in the *BCL11B* gene itself, one in exon 2 (*BCL11B*_1) and the other in exon 4 (*BCL11B*_2) of transcript variant 1 (NM_138576). BCL11B_1 showed a 25-fold increase, and BCL11B_2 showed a 30-fold increase in BCL11B binding compared to the background GFP control, while the BCL11B mutant exhibited 2- to

5-fold less binding to these sites, respectively, demonstrating that the mutant protein had a lower binding affinity for these sites compared to the WT. In contrast, a unique site in the 5' untranslated region of *TACC1* isoform 1 (NM_006283.2), encoding transforming acidic coiled-coil containing protein 1, was bound only by mutant BCL11B, but not WT, with a mean 6-fold increase compared to binding by the GFP controls.

The differential binding of WT and mutant samples for all three sites was confirmed by quantitative PCR using aliquots of the DNA obtained after immunoprecipitation with anti-FLAG antibody (shown in Figure 3H, main text). A schematic representation of ChIP-seq methods, analysis and results is shown in Figure S5.

These results suggested that our patient's mutation in the DNA binding domain of BCL11B may have altered the DNA binding specificity, causing the mutant protein to exhibit decreased binding to canonical BCL11B target sites, while simultaneously increasing binding to novel sites.

Quantitative PCR primer sequences

Human gene	Forward primer	Reverse primer
BCL11B	ATGCTATTCTTGCCTTTCATTTCAGAA	TCCAAACTCAACTTGAACTCTCATCT
CCR7	TGAGGTCACGGACGATTACAT	GTAGGCCCACGAAACAAATGAT
CCR9	ACAGCCAAATCAAGGAGGAATC	CAGCAAGCCATGACCACGA

Supplementary figures

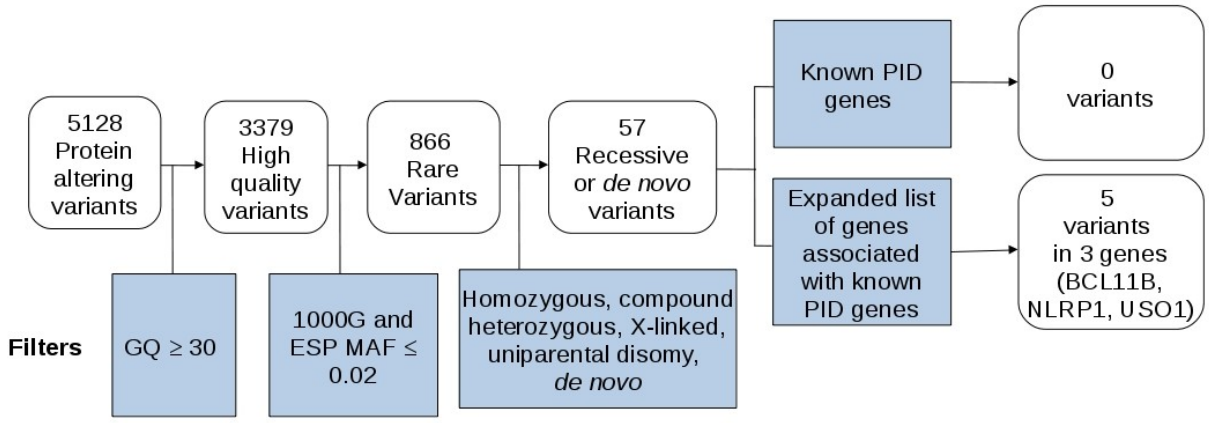


Figure S1. Gene and variant filtering approach.

Variant filtering for proband, showing sequentially applied filters for protein altering variants, quality (GQ ≥30) and minor allele frequency (MAF ≤0.02), yielding a set of 866 rare variants that were further annotated based on inheritance models. Since the proband's parents were healthy, the analysis was further restricted to variants that were inherited in a recessive manner or were de novo. This set was further divided into variants in known primary immunodeficiency genes vs. genes in the extended list that included interacting partners. Five shortlisted high quality variants were obtained in 3 genes (*BCL11B, NLRP1, USO1*).

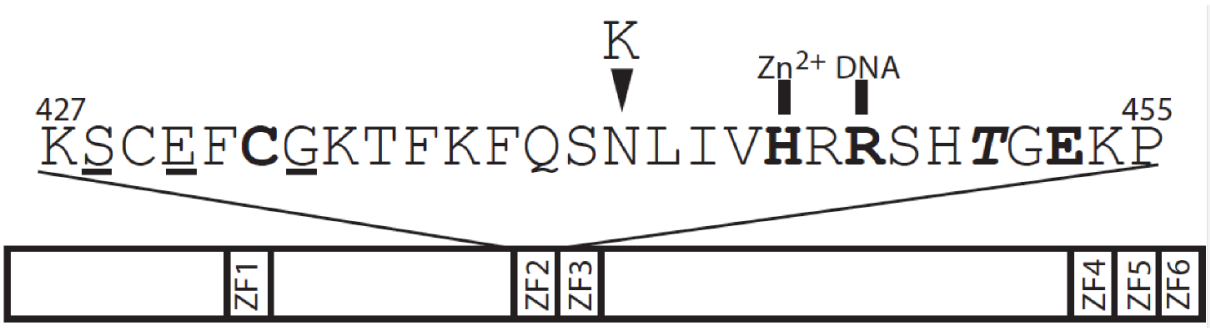


Figure S2. Human BCL11B protein sequence.

Protein NP_612808.1 diagram with zinc finger (ZF) positions and features of highly evolutionarily conserved ZF2, residues 427-455; the H445 coordinates the Zn^{+2} ion, while R447 binds to DNA backbone and nucleotide bases at target sites.[49] Patient mutation, pN441K, is shown by inverted triangle. Somatic mutation sites in T-cell acute leukemia are in **bold** and *italics*;[49,50] sites mutated in melanoma or lung cancer are underlined (COSMIC database,[51] COSM3499636 S428L, COSM3499635 E430K, and COSM349009 G433A).

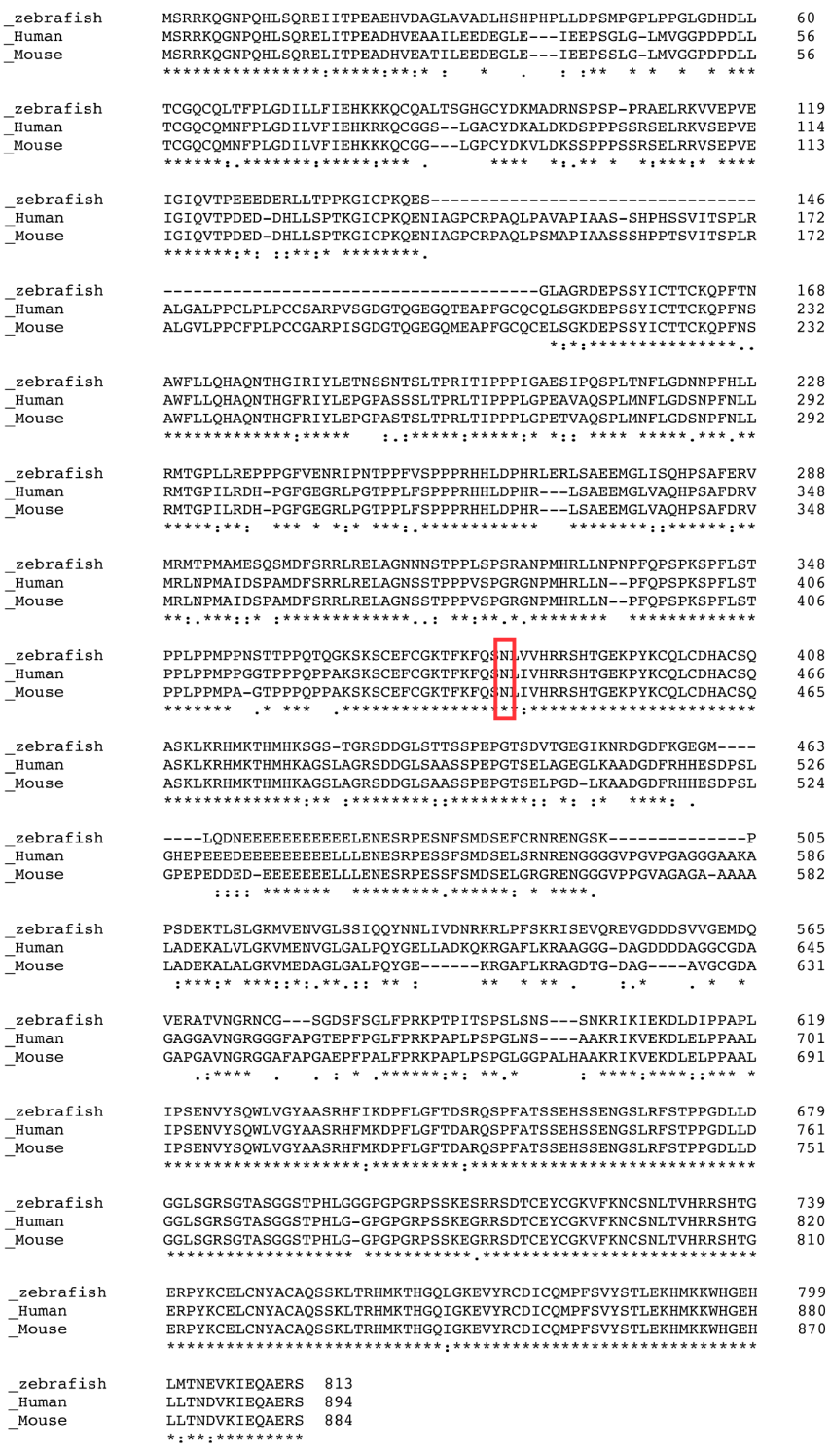


Figure S3. Alignment of zebrafish bcl11b, human BCL11B and mouse Bcl11b proteins.

Human and zebrafish proteins are 73% identical, indicated by "*" beneath shared residues. The DNA binding zinc finger domains 2 and 3, including the pN441 (outlined in red) mutated to K in the SCID patient, are conserved between human, mouse and zebrafish.

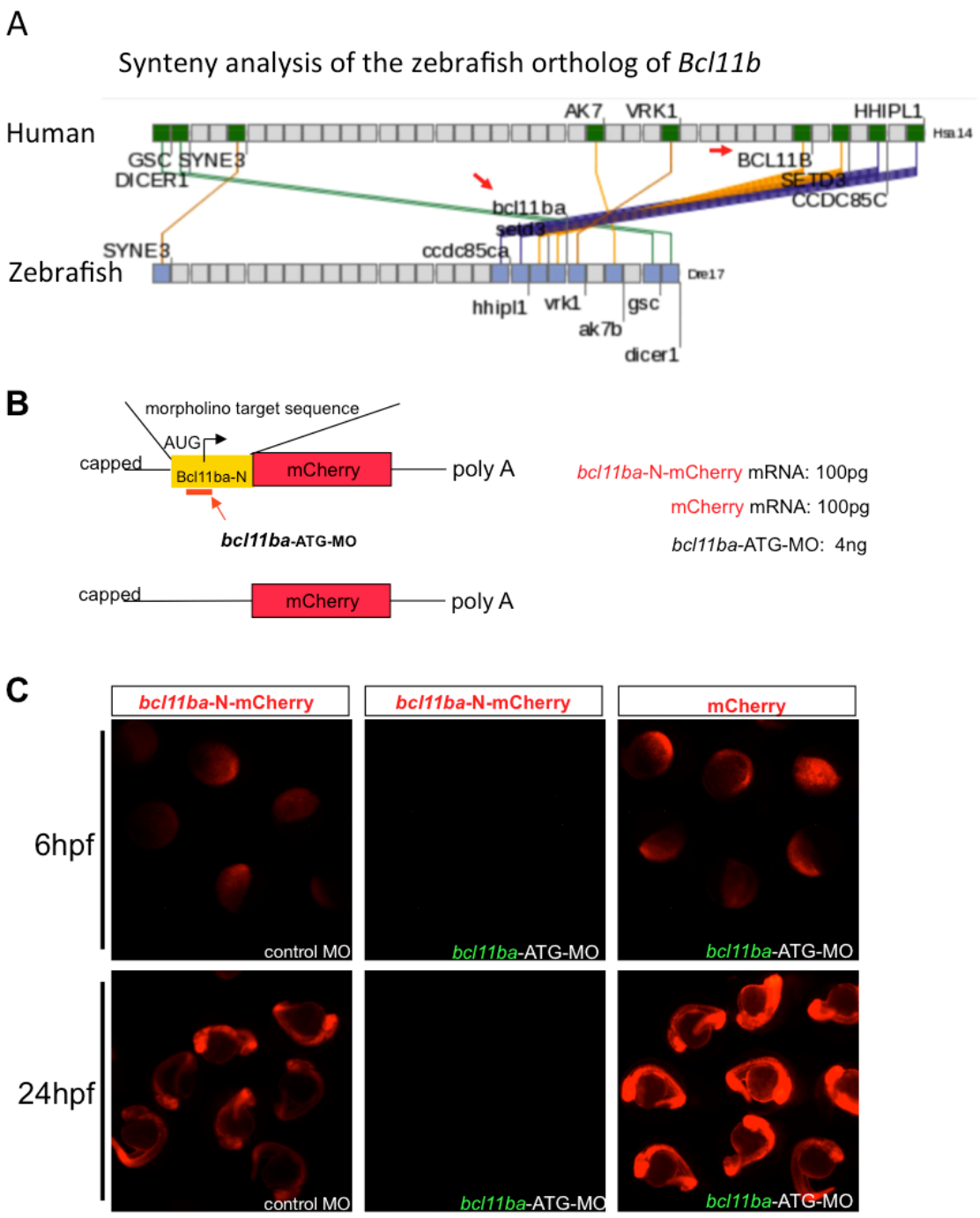
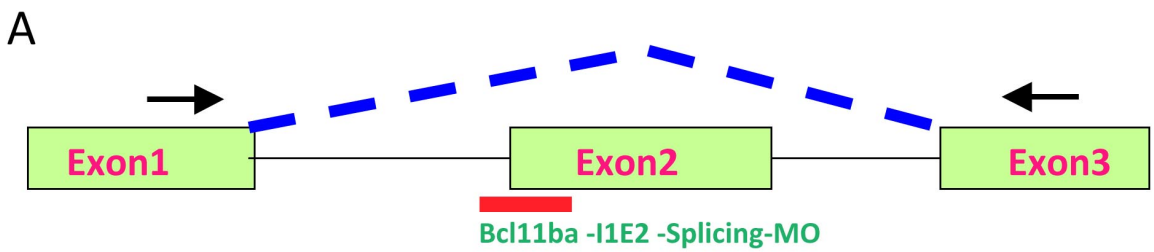
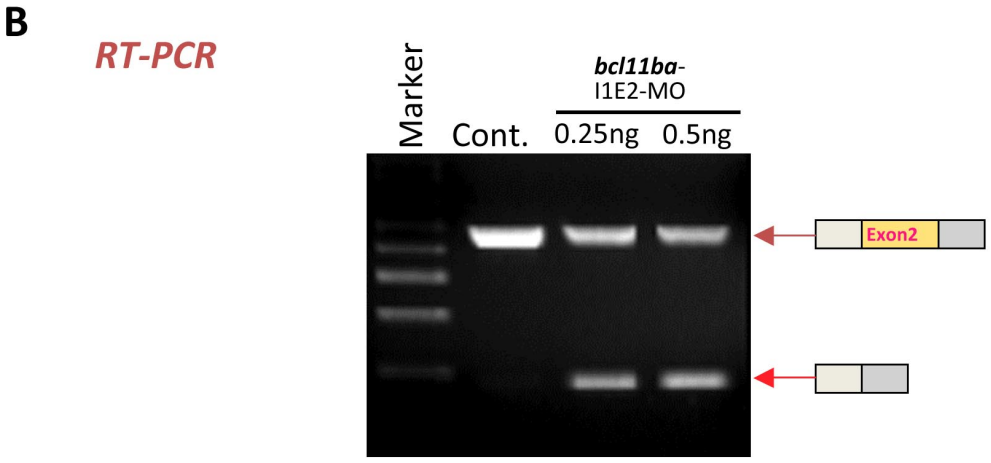


Figure S4: Site-specific effect of bcl11ba-MO injected into zebrafish.

A, Chromosomal synteny between zebrafish bcl11b and human BCL11B orthologs; B, Schematic diagrams of bcl11ba-N-mCherry fluorescent reporter mRNAs, the upper one containing the bcl11ba-ATG-MO target sequence (yellow box) fused in-frame with mCherry; C, 100pg bcl11ba-N-mCherry or control mCherry mRNA injected with a standard control morpholino or bcl11ba-ATG-MO. Embryos were photographed at 6 or 24 hpf.


A
Exon1
Exon2
Exon3
Bcl11ba -I1E2 -Splicing-MO


B
RT-PCR
Marker
bcl11ba-
I1E2-MO
Cont.
0.25ng
0.5ng
Exon2

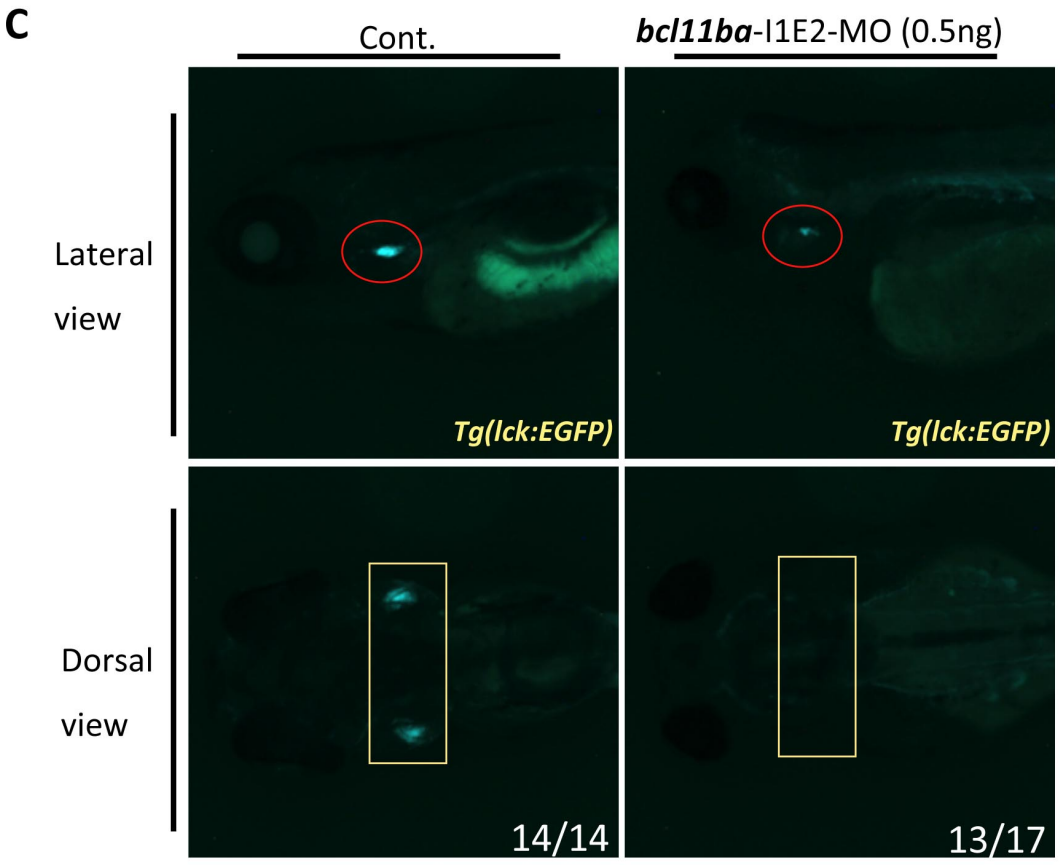

C
Cont.
bcl11ba-I1E2-MO (0.5ng)
Lateral
view
Tg(lck:EGFP)
Tg(lck:EGFP)
Dorsal
view
14/14
13/17

Figure S5: Effect of splice site bcl11b morpholino on T cell development in zebrafish.

A, Targeting site of bcl11ba-I2E3-splicing morpholino, red, to mediate knockdown by excision of exon 2 from the *bcl11ba* transcript; B, 0.25 or 0.5ng bcl11ba-I2E3-splicing morpholino injected into 1-cell stage WT embryo, confirming mis-splicing of bcl11ba mRNA by RT-PCR at 5dpf; C, 0.5ng bcl11ba-I2E3-splicing morpholino injected into Tg(lck:EGFP) embryos to monitor T lineage development. T cells failed to appear in the thymi of splicing morphants (red circles: lateral view, yellow boxes: dorsal view). Numbers in lower right of pictures denote the proportion of embryos with the depicted phenotype.

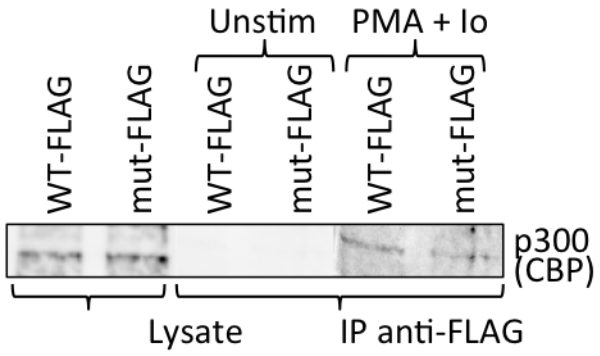
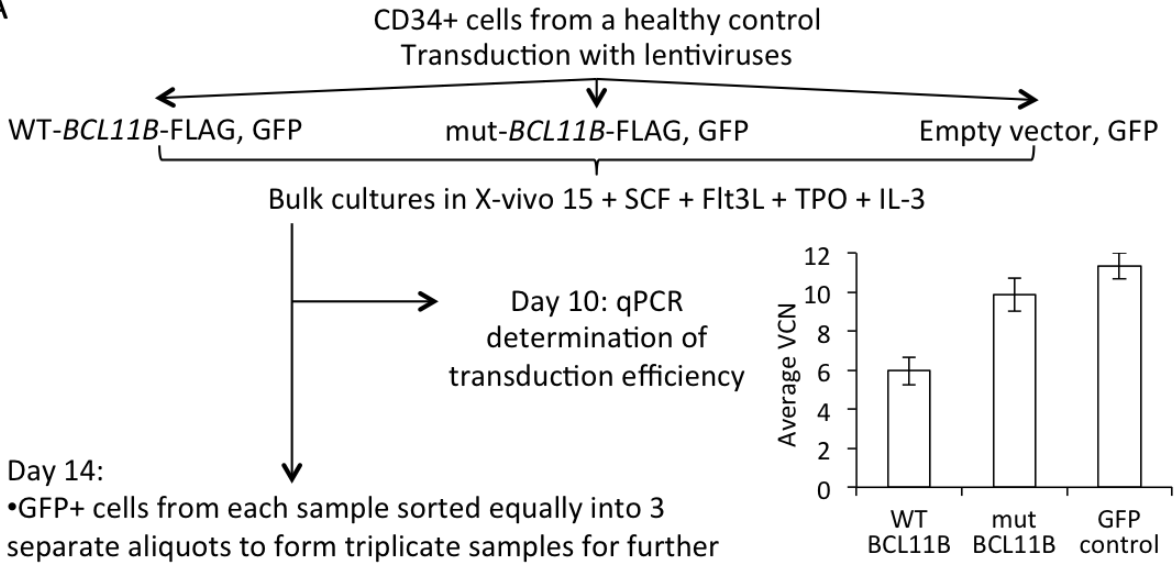


Figure S6: pN441K mutation does not affect binding to CBP.

Western blot detecting association of p300 (CBP) with BCL11B in nuclear protein extracts from Jurkat cells transfected with equal amounts of FLAG-tagged WT or mutant *BCL11B* cDNA plasmids. Left 2 lanes, nuclear lysates from unstimulated cells; middle and right lanes, nuclear lysates from unstimulated or PMA and ionomycin stimulated (PMA+Io) cells, respectively, immunoprecipitated with anti-FLAG antibody and blotted with anti-p300.

A



Day 14:

•GFP+ cells from each sample sorted equally into 3 separate aliquots to form triplicate samples for further treatment. Therefore:

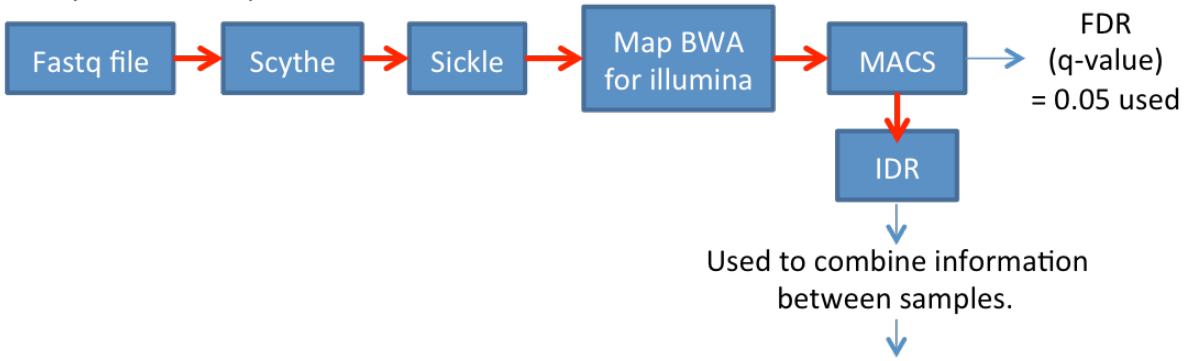
- WT-*BCL11B*, selected GFP+ cells: 3 samples
- mut-*BCL11B*, selected GFP+ cells: 3 samples
- GFP control, selected GFP+ cells: 3 samples

•Sorted cells sonicated to obtain 200 – 400 bp DNA fragments.

•DNA used for immunoprecipitation (IP) with anti-FLAG antibody.

•DNA obtained after IP subjected to library prep and sequencing.

•Analysis of the sequences:



- 2 sites in *BCL11B (BCL11B_1, BCL11B_2)* detected in WT, with significantly decreased binding by mut BCL11B.
- 1 site in *TACC1* detected in mut, with no binding at that site by WT BCL11B.

Differences in amount of binding at above sites verified by qPCR of DNA obtained after immunoprecipitation with anti-FLAG antibody on Day 14 (see Figure 3H in main text).

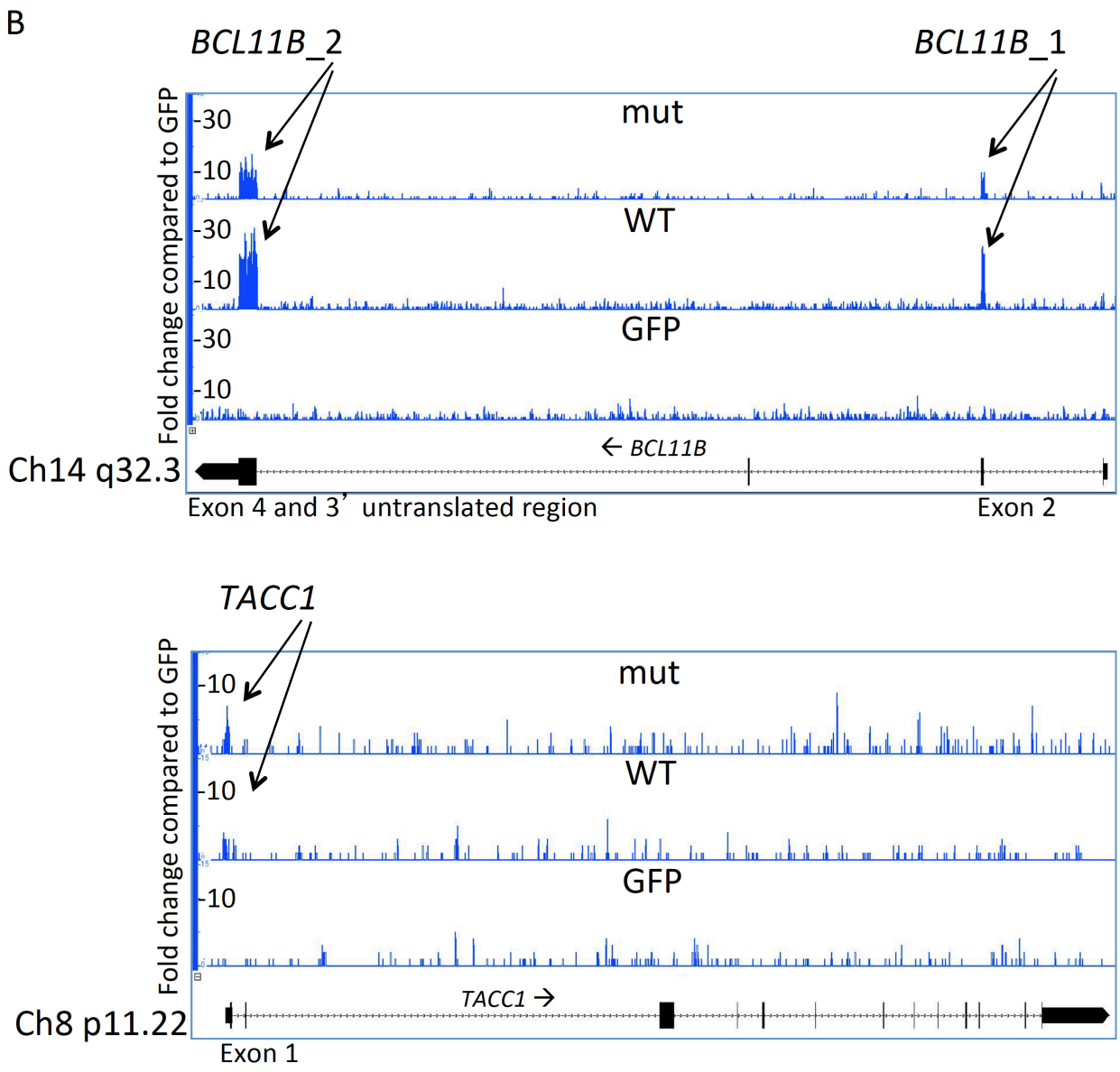


Figure S7: ChIP-seq method and results.

A, Schematic representation of the design, execution and analysis of ChIP-seq, carried out using DNA from control human HSC transduced with FLAG-tagged WT or mutant *BCL11B* or empty vector (GFP control), immunoprecipitated (IP) with anti-FLAG antibody and sequenced; B, Aligned reads from sequences obtained after IP in representative samples transduced with WT *BCL11B*, mutant *BCL11B* or GFP control. Top panel: two species-conserved, canonical BCL11B binding sites within the *BCL11B* locus itself demonstrate reduced affinity in samples overexpressing pN441K mutant BCL11B and no signal in GFP controls. Bottom panel: a unique DNA binding site within *TACC1* for pN441K mutant BCL11B, with no significant binding at this site in samples overexpressing WT BCL11B or GFP.

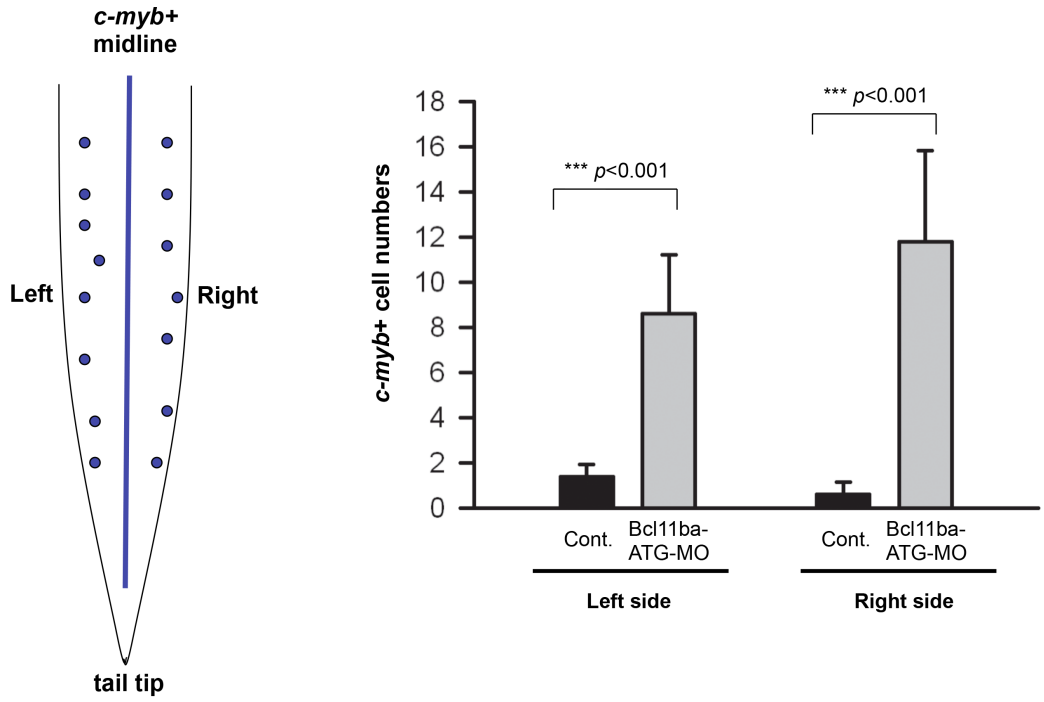


Figure S8: Bcl11b knockdown causes HSC displacement.

Embryos injected with bcl11b-ATG-MO to block translation of *bcl11b* mRNA, followed by c-myb WISH staining at 36hpf (1.5dpf) to examine the localization of hematopoietic stem and progenitor cells (HSPC). Displaced HSPC to left and right of the midline were quantified in double-blinded experiments. ** p <0.001.

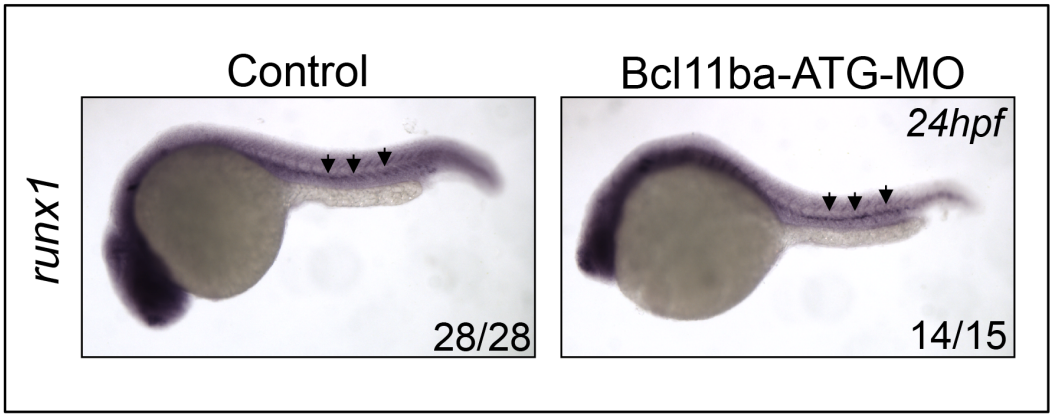


Figure S9: Bcl11b knockdown does not impair HSC emergence.

Embryos injected with Bcl11b-ATG-MO to block translation of *bcl11b* mRNA, following which the emergence of HSC was monitored at 24hpf by performing WISH with a probe reactive with *runx1*, a transcription factor required for HSC emergence. The numbers on the pictures denote the fraction of embryos with the depicted phenotype.

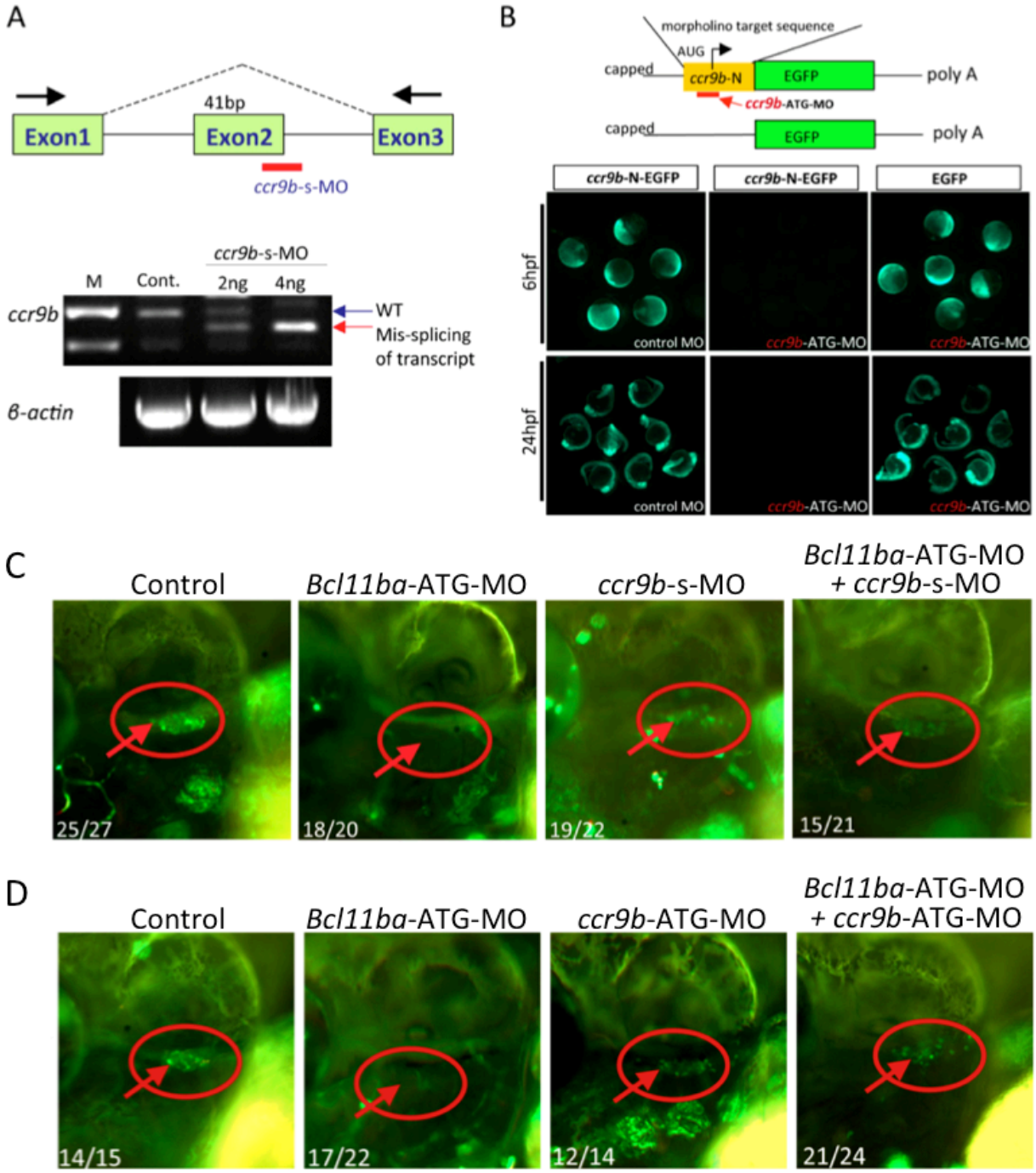


Figure S10. Effect of ccr9b dysregulation on thymic seeding in bcl11b morphants.

A. Diagram of the *ccr9b*-splice site morpholino targeting the exon2-intron2 junction (red bar). 2 or 4ng *ccr9b*-splice site morpholino was injected into 1-cell stage WT embryos, following which mis-splicing of *ccr9b* mRNA was confirmed by RT-PCR. *β-actin* mRNA was used as an internal loading control. Red arrow indicates the mis-splicing of *ccr9b* transcript.

B. Top, diagram of the *ccr9b*-N-EGFP fluorescent reporter containing or lacking the *ccr9b*-ATG-MO target sequence (orange box) fused in-frame with EGFP. Bottom, 100pg *ccr9b*-N-

EGFP or control EGFP mRNA was injected with standard control MO or *ccr9b*-ATG-MO. The embryos were photographed at 6 or 24hpf.
C. Rescue of thymic seeding in bcl11ba morphants by ccr9 knockdown. *Tg(cd41:EGFP)* were injected at the 1-cell stage with either 4ng of Bcl11b-ATG-MO, 2ng *ccr9b*-s-MO, or both together, following which the effect on thymic seeding of *cd41*+ cells was assessed at 5dpf (thymus location indicated by red circle and arrow).
D. Effect of morpholino (ccr9b-ATG-MO) knockdown of ccr9b either alone or in combination with Bcl11ba-ATG-MO, on thymic seeding at 5dpf in transgenic *Tg(cd41:EGFP)* zebrafish embryos (lateral view, red circles). Numbers refer to the fraction of embryos exhibiting the depicted phenotype.

Supplementary tables

Table S1: Exome quality measures

A) Mapping

Exome	#Reads	#Duplicates (%)	#Unmapped Reads (%)	#RPDC[a]	MQ[b] <20	#ROS[c]	#RSS[d]	Inferred insert size statistics[e]			
								shortest	longest	median	mean
Proband	83,095,910	2,552,454 (3.1%)	837,867 (1.0%)	124,238	12,953,119 (15.7%)	33,779,597 (85.4%)	4,569	101	228,180,394	233	2,399
Mother	11,3706,386	3,892,916 (3.4%)	1,402,017 (1.2%)	212,436	17,544,739 (15.6%)	46,211,524 (86.1%)	6,079	101	244,804,760	227	2,963
Father	116,877,772	3,673,700 (3.1%)	1,019,119 (0.9%)	209,487	18,375,087 (15.9%)	47470967 (85.1%)	6,127	101	243,566,370	227	2,661

[a]RPDC: Reads with pairs mapped to different chromosomes
[b]MQ: Reads with low mapping quality <20
[c]ROS: Read pairs with mapping quality (MQ) ≥20 and duplicates excluded mapped to opposite strands
[d]RSS: Read pairs mapped to the same strand.
[e]Inferred insert sizes calculated on read pairs with MQ ≥20, both mapped (duplicates excluded).

B) Coverage: High quality (≥20) reads in the Nimblegen v3.0 capture region, excluding duplicates and reads not properly paired. The right 4 columns list the fraction of capture positions covered by at least 1, 15, 30 and 45 reads, respectively.

Exome	Mean	Sdev	Median	Mode	≥1X	≥15X	≥30X	≥45X
Proband	64.6	42.1	59.0	0.0	0.97	0.93	0.86	0.69
Mother	88.9	59.7	81.0	0.0	0.97	0.93	0.90	0.82
Father	91.3	61.7	83.0	0.0	0.97	0.94	0.90	0.83

C) Ti/Tv ratios: Only single nucleotide variants in the Nimblegen v3.0capture region that were marked PASS had GQ≥30.

Exomes	All[a]	Known[b]	Novel[c]
Proband	2.63	2.64	2.27
Mother	2.63	2.63	2.27
Father	2.62	2.63	2.19

[a]For reference, Ti/Tv ratio for 1000 genomes (phase 1) across the capture region is 2.67.
[b]Known variants, those that occur in 1000 genomes (phase 1).
[c]Novel variants, those that are not in 1000 genomes.

D) Inheritance validation: Variants with PASS annotation and GQ >30 in the proband assessed for inheritance from parents.

Source of proband variant	SNVs		INDELs	
	Known	Novel[a]	Known	Novel
Father	10332	361	536	514
Mother	10334	313	522	504
Both parents	15521	33	835	533
One of the parents[b]	3950	5	227	435
De novo	116	5	23	247
Total	40253	717	2143	2233

[a]Variants were deemed novel if absent in 1000 Genomes database.
[b]Instances in which unambiguous assignment of parent not possible.

Table S2: List of genes involved in primary immunodeficiencies[5] or T cell development.

ADAMTS8	CIITA	JAK3	RMRP
AIRE	CORO1A	LCK	SH2D1A
AK2	CYBB	LIG4	STAT5B
ATM	DCLRE1C	NBN	STIM1
BTK	DKC1	NHEJ1	TAP1
CD247	DOCK8	ORAI1	TAP2
CD3D	FOXN1	PNPLA2	TAPBP
CD3G	IKBKG	PRKDC	TBX1
CD40LG	IL2RA	PTPRC	WAS
CD8A	IL2RG	RAG1	XIAP
CD8B	IL7R	RAG2	ZAP70
CHD7	ITK	RFXANK	ZBTB1

Table S3: Terms associated with T cell related phenotypes Mouse Genome Informatics and Human Phenotype Ontology Annotation, used to extend the target gene list.

MGI & HPO T cell related terms	Scores of the genes associated with each term
absent CD4 T cell	2
absent CD8 T cell	2
absent T cell	2
decreased CD4 T cell	2
decreased CD8 T cell	2
decreased double-positive T cell number	2
decreased double-negative T cell number	2
decreased single-positive T cell number	2
decreased T cell number	2
T-lymphopenia	2
abnormal differentiation of T cells	0.5
absent lymphocyte	0.5
decreased lymphocyte cell number	0.5
T cell lymphoma	0.5
T cell leukemia	0.5

Table S4: Terms obtained by using the target gene list as input to the ToppFun program.

a) Enriched disease terms

Severe combined immunodeficiency
Acute lymphoid leukemia
Agammaglobulinemia, X-linked; XLA
Alpha/beta T-cell lymphopenia with gamma/delta T-cell expansion, severe cytomegalovirus infection, and autoimmunity
Ataxia Telangiectasia
Atypical Mycobacteriosis, Familial, X-linked 1; AMCBX1
Autoimmune polyendocrinopathy syndrome, type 1
Bare lymphocyte syndrome 2
Bruton type agammaglobulinemia
Cd8 deficiency, familial
CHARGE association
CHARGE Syndrome
Combined immunodeficiency, X-linked
DiGeorge Syndrome
Ectodermal Dysplasia, Anhidrotic, with Immunodeficiency, Osteopetrosis, and Lymphedema
Granulomatous Disease, Chronic, X-Linked
Growth hormone insensitivity with immunodeficiency
Hyper-IgM Immunodeficiency Syndrome, Type 1
Hypogammaglobulinemia and Isolated growth hormone deficiency, X-linked
Hypohidrotic ectodermal dysplasia with immune deficiency
Immune dysfunction with T-cell inactivation due to calcium entry defect 1, 2
Immunodeficiency 17
Immunodeficiency 19
Immunodeficiency 22
Immunodeficiency 26 with or without neurologic abnormalities
Immunodeficiency 8
Immunodeficiency due to defect in CD3-gamma
Immunodeficiency due to Defect in CD3-Zeta
Immunodeficiency with hyper IgM type 1
Immunodeficiency without anhidrotic ectodermal dysplasia
Immunologic Deficiency Syndromes
Interleukin 2 Receptor, Alpha, Deficiency of
LIG4 Syndrome
Lymphoproliferative syndrome 1
Lymphoproliferative Syndrome, X-Linked, 2
Microcephaly, normal intelligence and immunodeficiency
Nijmegen Breakage Syndrome
Omenn syndrome
Polyglandular autoimmune syndrome, type 1
Stromal Interaction Molecule 1; STIM1
T-cell immunodeficiency, congenital alopecia, nail dystrophy
T-Lymphocytopenia
Wiskott-Aldrich Syndrome
X-linked lymphoproliferative syndrome 1
X-linked severe combined immunodeficiency

b) Enriched pathway terms

TCR signaling in naive CD8+ T cells
Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor
Adaptive Immune System
Antigen processing and presentation
Apoptosis
ATM mediated phosphorylation of repair proteins
ATM mediated response to DNA double-strand break
ATM Signaling Pathway
B Cell Receptor Signaling Pathway
BARD1 signaling events BCR signaling pathway
Calcium signaling in the CD4+ TCR pathway
Canonical NF-kappaB pathway
CD40L Signaling Pathway
Cell Cycle: G2/M Checkpoint Costimulation by the CD28 family
CTL mediated immune response against target cells
CXCR4-mediated signaling events
Cytokine Signaling in Immune system
DNA Repair Double-Strand Break Repair
Downstream signaling in naive CD8+ T cells
Downstream TCR signaling Hematopoietic cell lineage
Homologous recombination Homologous Recombination Repair
Homologous recombination repair of replication-independent double-strand breaks
IL 17 Signaling Pathway
IL 2 signaling pathway
IL 4 signaling pathway
IL-2 Receptor Beta Chain in T cell Activation
IL-2 Signaling pathway IL-4 signaling Pathway
IL-7 Signal Transduction
IL-7 Signaling Pathway
IL-9 Signaling Pathway
IL12 and Stat4 Dependent Signaling Pathway in Th1 Development
IL12 signaling mediated by STAT4
IL12-mediated signaling events
IL2 signaling events mediated by PI3K
IL2 signaling events mediated by STAT5
IL2-mediated signaling events
IL4-mediated signaling events
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell
Interleukin 4 (IL-4) Pathway
Interleukin receptor SHC signaling Interleukin-2 signaling Interleukin-3, 5 and GM-CSF signaling
Interleukin-7 signaling
JAK-STAT signaling Jak-STAT signaling pathway
JAK/STAT signaling pathway Lck and Fyn tyrosine kinases in initiation of TCR Activation
Nef and signal transduction Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters
NO2-dependent IL 12 Pathway in NK cells
Non-homologous end-joining Nonhomologous End-joining (NHEJ)
PD-1 signaling Phosphorylation of CD3 and TCR zeta chains
Recruitment of repair and signaling proteins to double-strand breaks
Role of Tob in T-cell activation Signaling by Interleukins
Signaling by the B Cell Receptor (BCR)
Stathmin and breast cancer resistance to antimicrotubule agents
T cell activation
T cell receptor signaling pathway
T Cell Signal Transduction
T Cytotoxic Cell Surface Molecules
TCR signaling TCR signaling in naive CD4+ T cells
TCR Signaling Pathway Th1/Th2 Differentiation
The Co-Stimulatory Signal During T-cell Activation
Translocation of ZAP-70 to Immunological synapse
Innate Immune system

c) Enriched gene ontogeny (GO) terms

activated T cell proliferation
activation of immune response
adaptive immune response adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains anatomical structure homeostasis antigen processing and presentation
antigen processing and presentation of endogenous antigen
antigen processing and presentation of endogenous peptide antigen
antigen processing and presentation of exogenous peptide antigen via MHC class I
antigen processing and presentation of exogenous peptide antigen via MHC class
TAP-dependent antigen processing and presentation of peptide antigen via MHC class I antigen receptor-mediated signaling pathway apoptotic process apoptotic signaling pathway
B cell cytokine production
B cell differentiation
B cell homeostatic proliferation
B cell lineage commitment B cell mediated immunity
B cell proliferation B cell receptor signaling pathway
calcium-mediated signaling
CD40 receptor binding
CD8 receptor binding
cell activation involved in immune response
cell-type specific apoptotic process
cellular response to cytokine stimulus
cellular response to DNA damage stimulus cellular response to growth hormone stimulus cellular response to interleukin-15 cellular response to interleukin-4
cytokine production involved in immune response
cytokine-mediated signaling pathway
DNA damage checkpoint
DNA integrity checkpoint
DNA ligase (ATP) activity
DNA ligase activity
DNA ligation involved in DNA recombination
DNA metabolic process DNA recombination
DNA repair
DNA-dependent protein kinase activity
double-strand break repair
double-strand break repair via homologous recombination
double-strand break repair via nonhomologous end joining
FasL biosynthetic process Fc receptor mediated stimulatory signaling pathway
Fc receptor signaling pathway Fc-gamma receptor signaling pathway
Fc-gamma receptor signaling pathway involved in phagocytosis
G2 DNA damage checkpoint hematopoietic or lymphoid organ development
hematopoietic progenitor cell differentiation
hemopoiesis
homeostasis of number of cells
homeostatic process
immature B cell differentiation
immune effector process
immune response-activating cell surface receptor signaling pathway
immune response-activating signal transduction
immune response-regulating cell surface receptor signaling pathway
immune response-regulating cell surface receptor signaling pathway involved in phagocytosis
immune response-regulating signaling pathway
immune system development
immunoglobulin mediated immune response
immunoglobulin production
immunoglobulin production involved in immunoglobulin mediated immune response immunoglobulin V(D)J recombination
inflammatory response
innate immune response
interleukin-10 production
interleukin-12 production
interleukin-2 binding
interleukin-2 receptor activity
interleukin-2-mediated signaling pathway
interleukin-4-mediated signaling pathway
interleukin-7 binding
interleukin-7 receptor activity
interleukin-7-mediated signaling pathway
isotype switching
JAK-STAT cascade
JAK-STAT cascade involved in growth hormone signaling pathway
leukocyte activation involved in immune response
leukocyte apoptotic process leukocyte homeostasis
leukocyte mediated cytotoxicity
leukocyte mediated immunity
leukocyte proliferation
lymph node development
lymphocyte activation involved in immune response
lymphocyte apoptotic process
lymphocyte costimulation
lymphocyte homeostasis
lymphocyte mediated immunity
lymphocyte proliferation
lymphoid progenitor cell differentiation
MHC class I protein binding
MHC class II biosynthetic process
MHC protein binding
mitotic DNA damage checkpoint
mitotic DNA integrity checkpoint
mitotic G2 DNA damage checkpoint
mitotic G2/M transition checkpoint
natural killer cell mediated cytotoxicity
natural killer cell mediated immunity
negative regulation of adaptive immune response
negative regulation of apoptotic process
negative regulation of cell killing
negative regulation of defense response
negative regulation of FasL biosynthetic process
negative regulation of immune effector process
negative regulation of immune response
negative regulation of immune system process
negative regulation of innate immune response
negative regulation of leukocyte activation
negative regulation of leukocyte apoptotic process
negative regulation of leukocyte mediated cytotoxicity
negative regulation of lymphocyte activation
negative regulation of lymphocyte apoptotic process
negative regulation of lymphocyte mediated immunity
negative regulation of natural killer cell mediated cytotoxicity
negative regulation of natural killer cell mediated immunity
negative regulation of thymocyte apoptotic process
non-recombinational repair
nucleotide-binding oligomerization domain containing signaling pathway
pattern recognition receptor signaling pathway
peptidyl-tyrosine modification
peptidyl-tyrosine phosphorylation

positive regulation of adaptive immune response
positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
positive regulation of antigen processing and presentation of peptide antigen via MHC class I positive regulation of antigen receptor-mediated signaling pathway
positive regulation of apoptotic signaling pathway
positive regulation of B cell activation
positive regulation of B cell differentiation
positive regulation of B cell proliferation
positive regulation of calcium-mediated signaling
positive regulation of cell activation
positive regulation of cell differentiation
positive regulation of cell killing
positive regulation of extrinsic apoptotic signaling pathway
positive regulation of hematopoietic stem cell migration
positive regulation of immune effector process
positive regulation of immune response
positive regulation of immune system process
positive regulation of leukocyte activation
positive regulation of leukocyte differentiation
positive regulation of leukocyte mediated cytotoxicity
positive regulation of leukocyte mediated immunity
positive regulation of leukocyte proliferation
positive regulation of lymphocyte activation
positive regulation of lymphocyte differentiation
positive regulation of lymphocyte mediated immunity
positive regulation of lymphocyte proliferation
positive regulation of MHC class I biosynthetic process
positive regulation of MHC class II biosynthetic process
pre-B cell allelic exclusion
pre-B cell differentiation
pro-B cell differentiation
production of molecular mediator of immune response programmed cell death
protection from natural killer cell mediated cytotoxicity
regulation of adaptive immune response
regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains regulation of antigen receptor-mediated signaling pathway
regulation of apoptotic process
regulation of B cell activation
regulation of B cell cytokine production
regulation of B cell differentiation
regulation of B cell mediated immunity
regulation of B cell proliferation
regulation of calcium ion transport
regulation of calcium-mediated signaling
regulation of cytokine production
regulation of DNA recombination
regulation of FasL biosynthetic process
regulation of immune effector process
regulation of immune response
regulation of immune system process
regulation of innate immune response
regulation of interleukin-10 production
regulation of interleukin-12 production
regulation of leukocyte activation
regulation of leukocyte apoptotic process
regulation of leukocyte differentiation
regulation of leukocyte mediated cytotoxicity
regulation of leukocyte mediated immunity
regulation of leukocyte proliferation
regulation of lymphocyte activation
regulation of lymphocyte apoptotic process
regulation of lymphocyte differentiation
regulation of lymphocyte mediated immunity
regulation of lymphocyte proliferation
regulation of MHC class II biosynthetic process
regulation of mononuclear cell proliferation
regulation of multicellular organismal development
regulation of natural killer cell mediated cytotoxicity
regulation of natural killer cell mediated immunity
regulation of production of molecular mediator of immune response
regulation of programmed cell death regulation of protein autophosphorylation
regulation of protein linear polyubiquitination
regulation of response to biotic stimulus
regulation of response to stress regulation of T cell activation
regulation of T cell differentiation.
regulation of thymocyte apoptotic process
response to biotic stimulus response to external biotic stimulus
response to gamma radiation
response to interleukin-15
response to interleukin-2
response to interleukin-4 response to interleukin-9
response to ionizing radiation
response to other organism
response to wounding
semicircular canal development
signal transduction in response to DNA damage
signal transduction involved in cell cycle checkpoint
signal transduction involved in DNA damage checkpoint
signal transduction involved in DNA integrity checkpoint
signal transduction involved in G2 DNA damage checkpoint
signal transduction involved in mitotic cell cycle checkpoint
signal transduction involved in mitotic DNA damage checkpoint
signal transduction involved in mitotic DNA integrity checkpoint
signal transduction involved in mitotic G2 DNA damage checkpoint
somatic cell DNA recombination
somatic diversification of immune receptors
somatic diversification of immune receptors via germline recombination within a single locus somatic diversification of immunoglobulins
somatic diversification of immunoglobulins involved in immune response
somatic recombination of immunoglobulin gene segments
somatic recombination of immunoglobulin genes involved in immune response
T cell differentiation
T cell differentiation in thymus
T cell homeostasis
T cell receptor binding
T cell receptor complex
T cell receptor signalling pathway
T/G mismatch-specific endonuclease activity
TAP binding
TAP1 binding
TAP2 binding
telomere maintenance
telomere organization
thymocyte apoptotic process thymus development toll-like receptor TLR1:
TLR2 signaling pathway toll-like receptor TLR6:
TLR2 signaling pathway tyrosine phosphorylation of STAT protein
tyrosine phosphorylation of Stat5 protein V(D)J recombination

d) Enriched human phenotype (HPO) terms

Severe T lymphocytopenia
Abnormality of B cell number
Abnormality of B cell physiology
Abnormality of B cells
Abnormality of bone marrow cell morphology
Abnormality of CD4+ T cells
Abnormality of CD8+ T cells
Abnormality of cells of the lymphoid lineage
Abnormality of cellular immune system
Abnormality of granulocytes
Abnormality of humoral immunity
Abnormality of immune system physiology
Abnormality of leukocytes
Abnormality of lymphocytes
Abnormality of myeloid leukocytes
Abnormality of T cells
Abnormality of the immune system
Abnormality of the intestine
Aplasia of the thymus
Aplasia/Hypoplasia of the thymus
B lymphocytopenia
Cellular immunodeficiency
Combined immunodeficiency
Decreased number of CD4+ T cells
Decreased number of CD8+ T cells
Hypergammaglobulinemia
Hyperglobulinemia
Hypogammaglobulinemia
IgG deficiency
Immunodeficiency
Immunoglobulin abnormality
Impaired memory B-cell generation
Lymphadenopathy
Lymphoma
Lymphopenia
Severe B lymphocytopenia
Severe combined immunodeficiency
T lymphocytopenia Thyroiditis

e) Enriched mouse phenotype terms

decreased T cell number
abnormal abdominal lymph node morphology
abnormal adaptive immunity
abnormal alpha-beta intraepithelial T cell morphology
abnormal alpha-beta T cell morphology
abnormal alpha-beta T cell number
abnormal antigen presentation via MHC class I
abnormal B cell activation
abnormal B cell differentiation
abnormal B cell morphology
abnormal B cell number
abnormal B cell physiology
abnormal B lymphocyte
antigen presentation
abnormal B-1 B cell morphology
abnormal B-1 B cell number
abnormal bone marrow cell morphology/development
abnormal bone marrow cell number
abnormal bone marrow hematopoietic cell morphology
abnormal CD4-positive helper T cell morphology
abnormal CD4-positive T cell differentiation
abnormal CD4-positive, alpha beta T cell morphology
abnormal CD4-positive, alpha beta T cell number
abnormal CD4-positive, alpha-beta memory T cell number
abnormal CD4-positive, a-b T cell physiology abnormal CD8-positive, a-b T cell morphology
abnormal CD8-positive, alpha-beta memory T cell morphology
abnormal CD8-positive, alpha-beta memory T cell number
abnormal CD8-positive, alpha-beta T cell differentiation
abnormal CD8-positive, alpha-beta T cell number
abnormal CD8-positive, alpha-beta T cell physiology
abnormal cell-mediated immunity
abnormal chemokine secretion
abnormal chromosome stability
abnormal class switch recombination
abnormal cytokine secretion
abnormal cytotoxic T cell cytolysis
abnormal cytotoxic T cell physiology
abnormal definitive hematopoiesis
Abnormal differentiation of T cells
abnormal DNA repair
abnormal double-negative T cell morphology
abnormal double-positive T cell morphology
abnormal effector T cell morphology
abnormal effector T cell number
abnormal follicular B cell morphology
abnormal gamma-delta intraepithelial T cell morphology
abnormal gamma-delta T cell differentiation
abnormal gamma-delta T cell morphology
abnormal gamma-delta T cell number
abnormal hematopoietic cell morphology
abnormal hematopoietic cell number
abnormal hematopoietic stem cell morphology
abnormal hematopoietic system physiology
abnormal hemopoiesis abnormal humoral immune response
abnormal Immununoglobulin levele
abnormal immature B cell morphology
abnormal immature B cell number
abnormal immune cell physiology
abnormal immune organ physiology
abnormal immune serum protein physiology
abnormal immune system cell morphology
abnormal immune system morphology abnormal immune system organ morphology abnormal immune system physiology
abnormal immunoglobulin heavy chain V(D)J recombination
abnormal immunoglobulin level
abnormal immunoglobulin V(D)J recombination
abnormal innate immunity
abnormal interferon secretion
abnormal interferon-gamma secretion
abnormal interleukin secretion
abnormal interleukin-10 secretion
abnormal interleukin-13 secretion
abnormal interleukin-2 secretion
abnormal interleukin-4 secretion
abnormal interleukin-6 secretion
abnormal intraepithelial T cell morphology
abnormal intraepithelial T cell number
abnormal leukocyte cell number
abnormal leukocyte morphology
abnormal leukocyte physiology
abnormal leukopoiesis
abnormal level of surface class I molecules
abnormal lymph node B cell domain morphology
abnormal lymph node cell ratio
abnormal lymph node morphology
abnormal lymph node secondary follicle morphology
abnormal lymph node size
abnormal lymph organ size
abnormal lymphocyte cell number
abnormal lymphocyte morphology

abnormal lymphocyte physiology
abnormal lymphopoiesis
abnormal marginal zone B cell morphology
abnormal mature B cell morphology
abnormal mature B cell number
abnormal mature gamma-delta T cell morphology
abnormal memory B cell morphology
abnormal memory B cell number
abnormal memory T cell morphology
abnormal memory T cell number
abnormal monocyte morphology
abnormal mononuclear cell morphology
abnormal myeloid leukocyte morphology
abnormal natural killer cell mediated cytotoxicity
abnormal negative T cell selection
abnormal positive T cell selection
abnormal pre-B cell morphology
abnormal regulatory T cell morphology
abnormal regulatory T cell number
abnormal regulatory T cell physiology
abnormal single-positive T cell number
abnormal T cell activation
abnormal T cell apoptosis
abnormal T cell differentiation
abnormal T cell morphology
abnormal T cell number
abnormal T cell physiology
abnormal T cell proliferation
abnormal T cell receptor alpha chain V-J recombination
abnormal T cell receptor beta chain V(D)J recombination
abnormal T cell receptor delta chain V(D)J recombination
abnormal T cell receptor V(D)J recombination
abnormal T cell selection
abnormal T cell subpopulation ratio
abnormal T-helper 1 cell morphology
abnormal T-helper 1 physiology
abnormal T-helper cell morphology
abnormal thymocyte activation
abnormal thymocyte apoptosis
abnormal thymus cell ratio
abnormal thymus cortex morphology
abnormal thymus corticomedullary boundary morphology
abnormal thymus epithelium morphology
abnormal thymus lobule morphology
abnormal thymus medulla morphology
abnormal thymus morphology
abnormal thymus physiology
abnormal thymus size
abnormal thymus weight
abnormal transitional stage B cell morphology
absent B cells
absent CD4 T cell
absent CD4-positive, alpha beta T cells absent CD4-positive, alpha beta T cells
absent CD8 T cell absent CD8-positive, a-b T cells absent g-d T cells absent lymph node germinal center
absent lymph nodes
absent lymphocyte
absent mature B cells
absent mature gamma-delta T cells
absent pre-B cells
absent spleen germinal center
absent T cell absent thymus cortex
absent thymus medulla arrested B cell differentiation
arrested T cell differentiation
athymia
autoimmune response
chromosomal instability
chromosome breakage
decreased activated T cell number
decreased apoptosis
decreased B cell number
decreased B cell proliferation
decreased B-1 B cell number
decreased bone marrow cell number
decreased CD4 T cell
decreased CD4-positive, alpha beta T cell number
decreased CD8 T cell
decreased CD8-positive, alpha-beta T cell number
decreased cell proliferation
decreased cellular sensitivity to ionizing radiation
decreased cytotoxic T cell cytolysis
decreased DN4 thymocyte number
decreased double-negative T cell number
decreased double-positive T cell number
decreased follicular B cell number
decreased gamma-delta T cell number
decreased hematopoietic cell number
decreased hematopoietic stem cell number
decreased immature B cell number
decreased immunoglobulin level
decreased inflammatory response
decreased interferon-gamma secretion
decreased interleukin-13 secretion
decreased interleukin-2 secretion
decreased interleukin-6 secretion
decreased level of surface class I molecules
decreased lymphocyte decreased lymphocyte cell number
decreased marginal zone B cell number
decreased mature B cell number
decreased memory T cell number
decreased NK T cell number
decreased pre-B cell number
decreased regulatory T cell number
decreased single-positive T cell number
decreased T cell
decreased T cell apoptosis
decreased T cell number
decreased T cell proliferation
decreased thymocyte apoptosis
decreased thymocyte number
decreased thymus weight
decreased transitional stage B cell number
defective assembly of class I molecules enlarged lymph nodes
impaired humoral immune response
impaired natural killer cell mediated cytotoxicity
increased apoptosis
increased B cell derived lymphoma incidence
increased cellular sensitivity to gamma-irradiation
increased cellular sensitivity to ionizing radiation
increased germinal center B cell number
increased hemolymphoid system tumor incidence
increased sensitivity to induced cell death increased susceptibility to autoimmune disorder increased susceptibility to infection induced chromosome breakage
low lymphocyte
lymph node hypoplasia
lymphoid hypoplasia
lymphopenia
small lymph nodes
small spleen
small thymus
spontaneous chromosome breakage
T cell associated leukemia
T cell Lymphoma thymus atrophy
thymus hyperplasia
thymus hypoplasia, thymus cortex hypoplasia

Table S5: High priority variants from exome sequence analysis.

Inheritance model	Chromosome & coordinate Gene	Ref/Alt allele	Amino acid change HGVS name cDNA position & Exon	MAF in: 1000 Genomes ESP	Effect Prediction: SIFT Polyphen CADD	GT:GQ:DP proband mother father	Prioritization Score
De novo	14:99641850 *BCL11B*	A/C	N441K NM_138576.3:c.1323T>G (BCL11B_i001) NM_138576.3:n.1818T>G, Exon4	Absent Absent	D PP2D 14	0/1:99:48 0/0:99:52 0/0:99:56	9
Compound heterozygous pair	17:5425076 *NLRP1*	A/G	M1184T NM_014922.4:c.3551T>C 4106, Exon4	0.02 0.0033	D PP2B 17.63	0/1:99:39 0/1:99:63 0/0:99:58	6.5
	17:5442790 *NLRP1*	C/T	V939M NM_014922.4:c.2815G>A 3370, Exon10	0.01 0.018	D PP2D 22.8	0/1:99:57 0/0:99:52 0/1:99:75	
Compound heterozygous pair	4:76649893 *USO1*	C/A	5' UTR variant NM_003715.2:c.-111C>A (CDS_not_multiple_of_3), Exon1	0.0046 Absent	NA[a] NA 9.5	0/1:99:33 0/0:99:50 0/1:99:62	4
	4:76730201 *USO1*	G/A	Coding exonic NM_003715.2:c.2349G>A (CDS_not_multiple_of_3), Exon19	0.0027 0.0061	NA NA 3.3	0/1:99:66 0/1:99:110 0/0:99:120	
De novo	19:36631940 *CAPNS1*	GGGC/G	Deletion of glycine codon 20 (NonFrameShift) NM_001003962.1:c.45_47delCGG 182, Exon2	Absent Absent	NA NA 20.6	0/1:27:25 0/0:45:20 0/0:39:60	3.5

[a]NA, not applicable

Table S6 Best candidate genes and annotations.

Gene Prioritization	**Gene/variant annotations**	**Databases used/ prediction methods**	**Candidate Gene and Variant scores**			
			BCL11B	***NLRP1***	***USO1***	***CAPNS1***
	Protein interactions with known PID genes	BIOGRID, MINT, HPRD	+2 [Interacts with PID gene: Il2][30]	0	+1 [Interacts with Xiap][35]	+1 [Interacts with Il2rg][36]
	Mammalian Phenotypes shared with known PID genes	MGI, HPO	+2 [Decreased T cell number MP:0005018][31]	0.5 [Decreased lymphocyte number MP:0005016][33]	0	0
	Gene Ontology terms shared with known PID genes	UniProt	+1 [T cell differentiation in thymus][32]	+0.5 [Innate immune response][13]	0	0
	Gene pathways shared with known PID genes	Pathway Commons	0	+0.5 [Innate immune system][9]	0	0
	Associated diseases shared with known PID genes	OMIM, GAD, NHGRI-GWAS	0	+1 [Leukemia][34]	0	0
Variant Prioritization	Inheritance model	Custom scripts	+2 [de novo and putative haplo-insufficient][20]	+2 [Compound Heterozygous]	+2 [Compound heterozygous]	+1 [De novo]
	Predicted deleterious effect	[SIFT, PolyPhen2, CADD]	+2 [D, PP2D,14]	+2 [D,PP2B,17] [D,PP2D,23]	+0.5 [NA, NA,9.5] [NA, NA,9.5]	+2 [NA, NA,20.6]
	Total Score		9	6.5	3.5	4

References

1. Patel JP, Puck JM, Srinivasan R, et al. Nijmegen Breakage Syndrome Detected by Newborn Screening for T Cell Receptor Excision Circles (TRECs). J Clin Immunol 2015;35:227-33.
2. http://compbio.berkeley.edu/proj/varant/
3. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
4. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [December 15, 2011]
5. Al-Herz W, Bousfiha A, Casanova JL, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. Front Immunol 2011;2:54.
6. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 2006;34:D535-9.
7. Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 2012;40:D857-61.
8. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human Protein Reference Database - 2009 Update. Nucleic Acids Res 2009;37:D767-72.
9. Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 2011;39:D685-90.
10. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res 2015;43:D726-36.
11. Kohler S, Doelken SC, Mungall CJ, et al. The Human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014;42:D966-D974.
12. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 2009;37:W305-311.
13. Magrane M and the UniProt Consortium. UniProt knowledgebase: a hub of integrated protein data. Database 2011;bar009.
14. Schaefer CF, AnthonyK, Krupa S, et al. PID: The Pathway Interaction Database. Nucleic Acids Res 2009;37:D674-9.
15. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, and Karp PD. Computational prediction of human metabolic pathways from the complete human genome. Genome Biology 2004;6:R2 1-17.
16. Milacic M, Haw R, Rothfels K, et al. Annotating cancer variants and anti-cancer therapeutics in reactome. Cancers (Basel) 2012;4:1180-211.
17. Online Mendelian Inheritance in Man, OMIM McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web URL: http://omim.org/
18. Becker KG, Barnes KC, Bright TJ and Wang SA. The Genetic Association Database. Nat Genet 2004;36:431-432.
19. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014;42:D1001-1006.
20. Dang VT, Kassahn KS, Marcos AE, Ragan MA. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. Eur J Human Genet 2008;16:1350-7.
21. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

22. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protocols 2009;4:1073-81.
23. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310-5.
24. Liu X, Jian X, and Boerwinkle E. dbNSFP: A lightweight database of human non-synonymous SNPs and their functional predictions. Hum Mutat 2011;32:894-899.
25. Liu X, Jian X, and Boerwinkle E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat 2013;34:E2393-E2402.
26. Radtke F, MacDonalad HR, andTacchini-Cottier F. Regulation of innate and adaptive immunity by NOTCH. Nat Rev Immunol 2013;13:427-437.
27. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor FoxP3. Science, 2003;299:1057-61.
28. Yatim A, Benne C, Sobhian B, et al. NOTCH1 nuclear interactome reveals key regulators of its transcriptional activity and oncogenic function. Mol Cell 2012;48:445-58.
29. Rudra D, deRoos P, Chaudhry A, et al. Transcription factor FoxP3 and its protein partners form a complex regulatory network. Nat Immunol 2012;13:1010-9.
30. Cismasiu VB, Ghanta S, Duque J, et al. BCL11B participates in the activation of IL2 gene expression in CD4+ T lymphocytes. Blood 2006;8:2695-2702.
31. Wakabayashi Y, Watanabe H, Inoue J, et al. Bcl11B is required for differentiation and survival of αβ T lymphocytes. Nat Immunol 2003;6:533-39.
32. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25:288-9.
33. Masters SL, Gerlic M, Metcalf D, et al. NLRP1 inflammasome activation induces pyroptosis of hematopoietic progenitor cells. Immunity 2012;37:1009-23.
34. Enjuanes A, Benavente Y, Bosch F, et al. Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia. Cancer Res 2008;24:10178-86.
35. Havugimana PC, Hart GT, Nepusz T, et al. A census of human soluble protein complexes. Cell 2012;150:1068-81.
36. Cataldo F, Peche LY, Klaric E, et al. CAPNS1 regulates USP1 stability and maintenance of genome integrity. Mol Cell Biol 2013;33:2485-96.
37. Cismasiu VB, Duque J, Paskaleva E, et al. BCL11B enhances TCR/CD28-triggered NF-κB activation through up-regulation of Cot kinase gene expression in T-lymphocytes. Biochemical Journal, 2009;417:457-466.
38. Vanvalkenburgh J, Albu DI, Bapanpally C, et al. Critical role of Bcl11b in suppressor function of T regulatory cells and prevention of inflammatory bowel disease. J Exp Med. 2011;208:2069-81.
39. Califano D, Sweeney KJ, Le H, et al. Diverting T helper cell trafficking through increased plasticity attenuates autoimmune encephalomyelitis. J Clin Invest. 2014;124:174-87.
40. Gutierrez A, Kentsis A, Sanda T, et al. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood 2011;118:4169-73.
41. Kamimura K, Ohi H, Kubota T, et al. Haploinsufficiency of Bcl11b for suppression of lymphomagenesis and thymocyte development. Biochem Bioph Res Co 2007;355:538-42.
42. Bromberg Y, Guy Y and Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397-2398.
43. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLOS Comput Biol 2010;6:e1001025.

44. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A; Galaxy Team. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26:1783-5.
45. Li, H, Durban R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60
46. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
47. Qunhua Li, James B. Brown, Haiyan Huang, and Peter J. Bickel. Measuring reproducibility of high-throughput experiments Ann. Appl. Stat. 2011;5:1752-1779.
48. Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics. 2009;25:2730-1.
49. Gutierrez A, Kentsis A, Sanda T, et al. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood 2011;118:4169-73.
50. Neumann M, Vosberg S, Schlee C, et al. Mutational spectrum of adult T-ALL. Oncotarget 2015;6:2754-66.
51. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805-11.