Title

Defining the sequence specificity of Ikaros DNA binding zinc fingers and their role in T cell activation and tumor suppression

Permalink

https://escholarship.org/uc/item/0nb94574

Author

Wadsworth, Sarah Elizabeth

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Defining the sequence specificity of Ikaros DNA binding zinc fingers

and their role in T cell activation and tumor suppression

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Microbiology, Immunology and Molecular Genetics

by

Sarah Elizabeth Wadsworth

2013

ABSTRACT OF THE THESIS

Defining the sequence specificity of Ikaros DNA binding zinc fingers

and their role in T cell activation and tumor suppression

by

Sarah Elizabeth Wadsworth

Master of Science in Microbiology, Immunology and Molecular Genetics

University of California, Los Angeles, 2013

Professor Stephen T. Smale, Chair

Ikaros (*Ikzf1*) is a transcription factor necessary for the development of lymphoid and other

hematopoietic lineages. Several studies have identified Ikaros as an important lymphoid tumor

suppressor both in murine models and human cases of acute lymphocytic leukemia (ALL). Few

gene targets of Ikaros regulation have been identified, and the mechanistic link to tumor

suppression is unknown. Ikaros binds DNA through a region of four tandem C2H2 DNA-binding

zinc fingers, and early characterization suggested that the first and fourth fingers have

modulatory binding that allows for targeting of different consensus sequences. However, the

binding sequence specificity of each of these fingers remained ambiguous. The Smale lab has

developed a mutant mouse strain with a deletion of the exon encoding the fourth zinc finger, $Ikzf1^{\Delta F4/\Delta F4}$, which among other phenotypes, develops spontaneous aggressive T cell lymphomas with 100% penetrance. These results suggest that the binding specificity of zinc finger 4 is necessary for regulating target genes that maintain Ikaros tumor suppressor function.   This thesis examines the DNA binding specificity of Ikaros zinc fingers 1 and 4 and identifies a putative consensus sequence for each. In addition, the kinetics and stability of binding is characterized for purified proteins without zinc fingers 1 or 4. Mice with a heterozygous deletion of the Ikaros DNA binding domain, $Ikzf1^{DN/+}$, also develop thymic lymphomas and have been observed to show a decreased threshold of stimulation for T cell activation. This phenotype is thought to have a connection with the malignant transformation of these cells. $Ikzf1^{\Delta F4/\Delta F4}$ T cells are tested by *in vitro* assays and found to exhibit the same significantly decreased threshold for activation. These assays are used to further investigate gene deregulation in activated $Ikzf1^{\Delta F4/\Delta F4}$ T cells using RNA high throughput sequencing. Many deregulated genes with potential contribution to decreased activation threshold and loss of tumor suppressor function were identified by RNA-seq analysis. In combination with RNA-seq, Ikaros ChIP-seq data could be used to validate functional targets and associate loss of Ikaros occupancy with deregulation of gene expression. Ikaros target genes have remained largely elusive in part due to difficulty with traditional chromatin immunoprecipitation (ChIP) methods, so a new technique is tested and optimized for future application to Ikaros ChIP-seq.

The thesis of Sarah Elizabeth Wadsworth is approved.

Michael F. Carey

M. Carrie Miceli

Stephen T. Smale, Committee Chair

University of California, Los Angeles

2013

Table of Contents

List of Figures

Section 3:

List of Supplementary Figures:

Introduction: The role of Ikaros in lymphopoesis and leukemogenesis

Ikaros (*IKZF1*) is a DNA binding zinc finger transcription factor that is necessary for the development of lymphoid and other hematopoietic lineages[2]. It plays a crucial role in regulating both activation and repression of gene expression to potentiate differentiation of hematopoietic precursors[3,4,5,6,7,8,9]. A recent study has shown that Ikaros is required not only for lymphoid lineage priming in the hematopoietic stem cell (HSC), but also repression of genetic mechanisms of self renewal and multipotency in differentiated progeny[10]. Ikaros interacts with chromatin remodeling and histone deacetylation complexes[11,12,13,14,15,16], and has been shown to localize with silent genes at foci of pericentromeric heterochromatin[17,18,19], suggesting a role in heritable epigenetic silencing that is incompletely understood. Alternate splicing generates many different isoforms of Ikaros, several of which contain unique combinations of the four DNA binding Zn finger domains[20,21] (Intro Figure A).

Intro Figure A: Ikaros Isoforms
Included exons are represented by divisions, zinc fingers by colored insets.
Fingers 1-4 are DNA-binding, 5 and 6 are involved in multimerization.
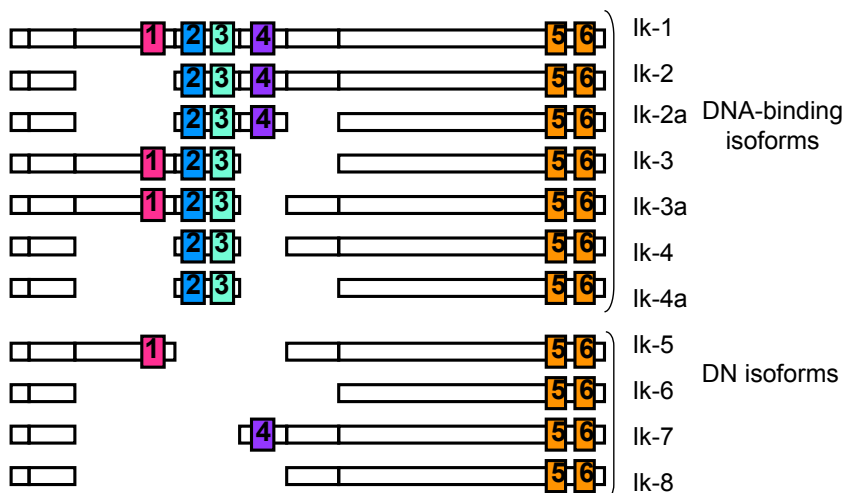
These Ikaros isoforms are each expressed at varying levels throughout lymphoid development[17], and show different DNA binding specificities and affinities, which appear dependent on modular

inclusion of the first and fourth DNA binding Zn finger exons[19,20]. From these observations, it has

been hypothesized that these isoforms may perform different gene regulatory functions.

Ikaros null (*Ikzf1*[null]) mice lack development of all B cells and fetal T cells, and postnatal T cells

are abnormal[22]. Furthermore, a homozygous knock-in mutation of a dominant negative form of

Ikaros that cannot bind DNA (*Ikzf1*[DN]) abolishes all pre- or post-natal lymphoid development[1]

(Intro Figure B).

Intro Figure B: Early block in lymphoid development of *Ikzf1*[null] and *Ikzf1*[DN]
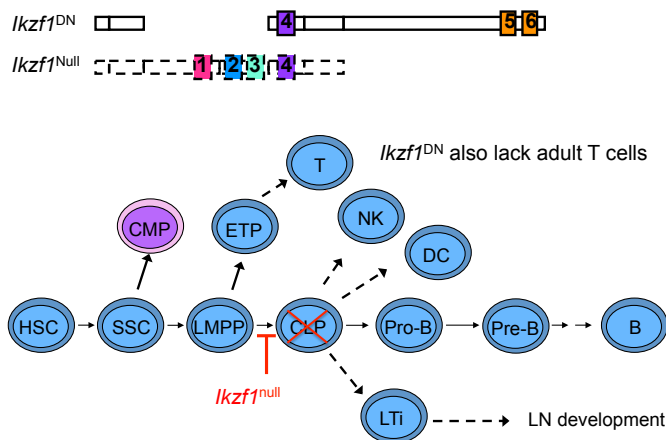
Mice that have an Ikaros heterozygous deficit (*Ikzf1*[null /+] and *Ikzf1*[DN/+]) retain the ability to

produce lymphocytes, but both develop lymphoproliferative disorders, which reliably progress

to aggressive early onset T cell leukemia and lymphoma in *Ikzf1*[DN /+] mice[21,23]. These studies

have identified Ikaros as an important lymphoid tumor suppressor, although the mechanisms by

which it operates have yet to be elucidated. In humans, recent genome analysis has revealed

heterozygous deletion of Ikaros in 84% of BCR-ABL B-progenitor acute lymphoblastic leukemia

(ALL) cases[24], and a strong association of Ikaros deletion with ALL cases that have poor

prognosis[25]. These results suggest that synergy between BCR-ABL and deletion of the Ikaros

gene are involved in malignant transformation. In human and mouse, mutations that reduce

Ikaros activity collaborate with Notch overactivation to drive T cell leukemogenesis[26,27], due to a

loss of Ikaros repression of several Notch target genes[28,29]. However, this repression does not

occur in hematopoietic precursors[27], indicating that maintenance of lineage heritable silencing is

an Ikaros tumor suppressor function. The findings of Mullighan et al.[24] support this conclusion,

as high-risk ALLs exhibited an up-regulation of HSC genes, and a downregulation of lymphocyte

genes. It therefore appears that a loss of Ikaros allows for reactivation of stem cell genetic

mechanisms that confer unlimited self-renewal, potentiating malignant transformation.

The Smale lab at UCLA has taken a novel approach to disentangle the potentially different

targets and roles of the Ikaros DNA binding isoforms in hematopoiesis and tumor suppression.

By generating knock-in mice lacking Ikaros exons that encode either the first or fourth DNA

binding Zn finger (Intro Figure C), we have limited the subset of isoforms that can be expressed.

Characterization of these strains has shown dramatically different homozygous phenotypes[30]

(Intro Figure D).

Intro Figure C: Targeted deletion of *Ikzf1* exons 4 and 6
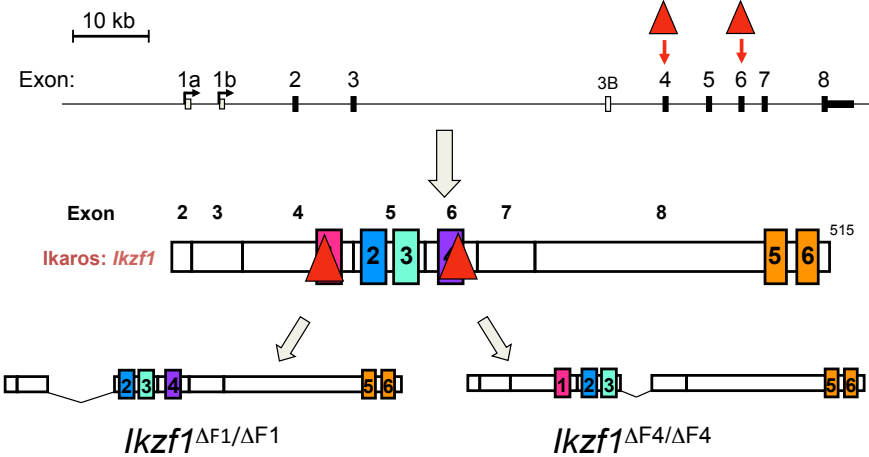
Intro Figure D: *IKZF1* ZnF1 and ZnF4 distinct essential functions

Function	*Ikzf1*[null]	*Ikzf1*[ΔF1/ΔF1]	*Ikzf1*[ΔF4/ΔF4]	Required ZnF
LN Development	−	+	−	2 3 4
Fetal B-cells	−	+$_\downarrow$	−	2 3 4
Adult B-cells	−	+$_\downarrow$	+$_\downarrow$	2 3
Thymic B-cells	na	$\downarrow$	+	1 2 3
Tumor Suppressor	−	+	−	2 3 4

Both develop mature B and T cells, in important contrast to null and DN mutants. Mice lacking

zinc finger 1 (*Ikzf1*[ΔF1/ΔF1]) have a unique deficit of peritoneal and thymic B cells, and those lacking

zinc finger 4 (*Ikzf1*[ΔF4/ΔF4]) do not develop lymph nodes, which also derive from lymphoid

progenitors. Most strikingly, *Ikzf1*[ΔF4/ΔF4] mice develop spontaneous T cell lymphomas that are

not observed in *Ikzf1*[ΔF1/ΔF1] mice. These malignancies appear as aggressive as those in mice with

an Ikaros heterozygous deficit (*Ikzf1*[null/+] and *Ikzf1*[DN/+]) and with 100% penetrance. These results

suggest that the binding specificity of ZnF4 is necessary for regulating target genes that maintain

Ikaros tumor suppressor function.

Intro Figure E: *Ikzf1*[ΔF4/ΔF4] mutants develop spontaneous thymic lymphoma[29]
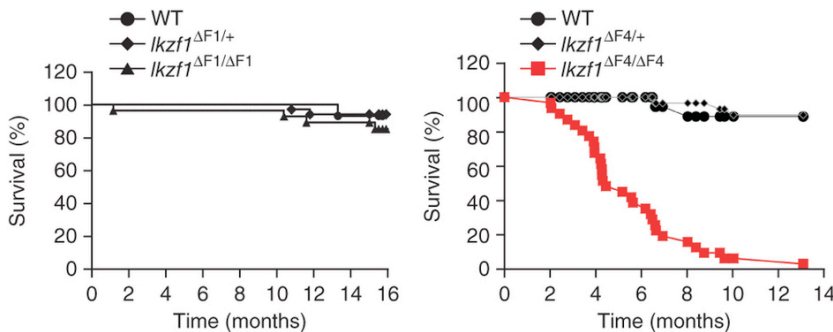
Section 1: Characterization of Ikaros zinc finger DNA binding specificity

Chapter 1: Understanding of Ikaros DNA binding

Ikaros family members have a conserved DNA binding domain at the amino N-terminus that typically contains four C2H2 zinc fingers[31]. C2H2 zinc fingers are characterized by the presence of two cysteine and two histidine residues. The "finger" structure is created by an antiparallel β-sheet, which contains a loop formed by the two cysteines, and an α-helix, which contains a loop formed by the two histidines. These two structural units are coordinated around a Zn ion by a tetrahedral arrangement of the zinc ligands, Cys2 and His2 (Figure 1a)[32]. The primary DNA contacts are made by the α-helix, which binds the major groove through hydrogen-bond interactions from amino acids at helical positions −1, 3, and 6 to three sequential bases on one strand of the DNA, and from helical position 2 to the next base in the 5' direction on the complementary strand (Figure 1a)[31].

Figure 1a: C2H2 DNA binding Zinc fingers[31]
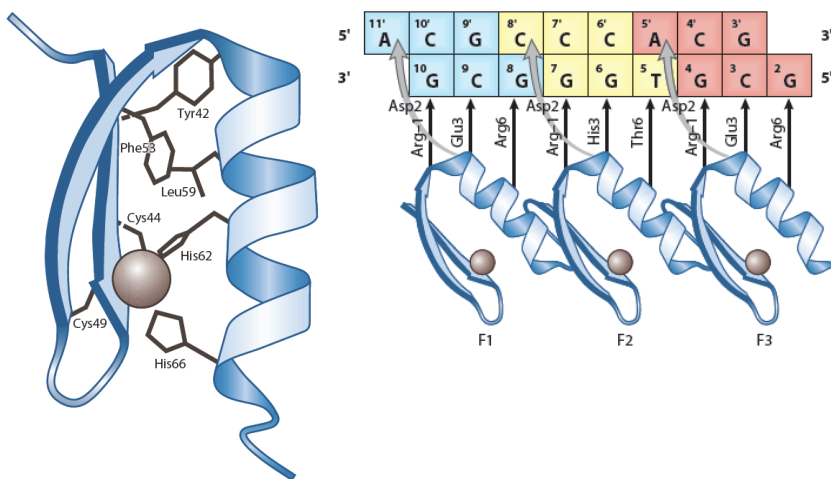
Through these interactions each Zn finger binds a sequential four base pair site, and tandem zinc fingers bind overlapping subsites that can compose a longer recognition sequence (Figure 1a). Most Zn finger DNA-binding proteins contain several tandem fingers, however, only two or three

5

are typically needed for stable DNA binding *in vitro*[31,33]. This raises the question of whether an

entire domain of tandem zinc fingers work together to recognize a single extended consensus, or

if different subsets of fingers are used to bind different target sequences. A large fraction of

both ubiquitous and developmentally specific classes of transcription factors contain multiple

DNA binding Zn fingers[31,34], and therefore this question is important for understanding

mechanisms of transcription on a broad scale.

Early studies of transcription factors such as TFIIA and CTCF showed that various combinations

of zinc fingers have differential binding specificity *in vitro*[35,36]. More recent ChIP-seq analysis of

CTCF binding sites has been able to further distinguish separate classes of binding sites for

different zinc finger combinations, and suggests that it carries out different functions based on

these DNA interactions[37]. Initial characterization of Ikaros also suggested that fingers 1 and 4

modulate binding to different DNA sequences, in contrast to the essential roles of fingers 2 and

3 in binding all sites[1,19]. However, detailed knowledge of the specific motifs recognized by

different combinations of Ikaros zinc fingers has not yet been obtained.

Previous studies have attempted to define the binding specificity of different Ikaros isoforms

and determine amino acid contacts with the DNA[18,20]. A small scale binding site selection assay

readily defined the motif bound by fingers 2 and 3, however the contacts for fingers 1 and 4

remained ambiguous[20]. This study expressed GST fusions of Ikaros isoforms (Ik-1, -2, -3, -4 and -

5) in DH5α *E. coli*, which were purified using glutathione beads. These fusion proteins were kept

bound to the beads for binding site selection with random oligonucleotides. The

oligonucleotides with affinity for the fusion proteins were eluted, PCR amplified, and repeated in

six sequential rounds of selection before being cloned and sequenced. Only fusion proteins with

isoforms Ik-1 (zinc fingers 1-4), Ik-2 (2-4), and Ik-3 (1-3) effectively selected oligonucleotides.

These sequences were aligned and base frequencies were determined for each position in a site-

specific matrix. Putative recognition sequences (Figure 1b) for Ik-1 (ZnF 1-4), Ik-2 (ZnF 2-4), and

Ik-3 (ZnF 1-3) were calculated from alignment of 24, 36, and 25 selected oligonucleotides,

respectively. This relatively small population identified a core motif TGGGAA bound by fingers 2

and 3, but confirmed very few flanking nucleotides. Base specificity on the 5' flanking end

remained especially ambiguous and did not define any consensus sequence for finger 4 contacts.

Figure 1b: Published consensus sequences from binding site selection assay[20]
Bases with a frequency 30-40% are lowercase, 40-60% are capitalized, and >60% are bold.
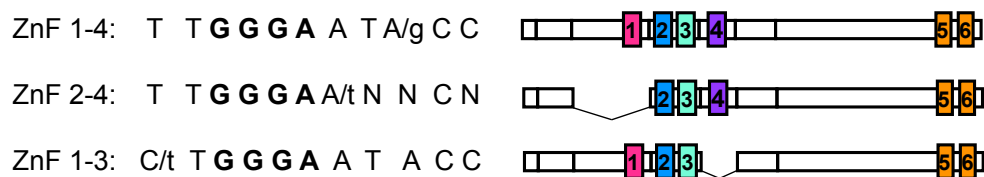
A later study used substitution mutant analysis of amino acids within fingers 2 and 3 to

interrogate critical interactions with the DNA[18]. Nuclear extracts from 293T cells transfected

with Ikaros vector constructs were used for electrophoretic mobility shift assays (EMSA) to

determine the effect of each mutation on DNA binding capability with various oligonucleotide

probes. A zinc finger homology model for fingers 2 and 3 (Figure 1c) was generated using five

solved crystal structures containing >50% common sequence with Ikaros. This strategy

confirmed contacts between the DNA sequence and key residues within the zinc finger 2 and 3

alpha helicies. In combination with the core consensus motif previously identified, this model

can help visualize the putative binding structure and emphasizes how little is understood about

the sequence specificity of fingers 1 and 4. Predictions based on probabilities from known zinc

finger-DNA binding interactions[38] also provide limited insight, as they do not match what

information has been obtained from previous Ikaros studies (Supplementary Figure 1a).

Figure 1c: *IKZF1* ZnF2 and ZnF3 DNA binding interactions.
Integration of core consensus sequence[20] with ribbon representation of the homology model of zinc fingers
2 and 3[18]. Mutation of residues in red abolish all DNA binding. Substitution of residues in yellow or blue
disrupt binding to different subsets of sequences.
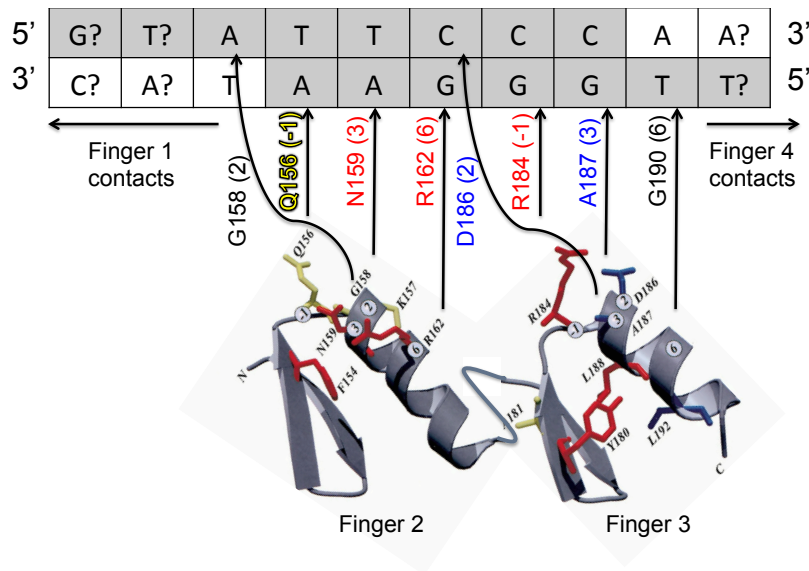
Previous attempts to identify Ikaros target genes based on gene expression from knockdown or

microarray analysis of *Ikzf1*null and *Ikzf1*DN mouse strains have yielded few verified candidates.

Loss of Ikaros causes such dramatic changes in the expression of thousands of genes that the

strategy of taking a loss of function or null phenotype, and working backward to identify direct

targets based on gene expression, has been very challenging. The Ikaros mutant strains

described in the preceding introduction were generated in the Smale lab with the goal of

identifying subsets of deregulated genes and phenotypes that can be attributed to loss of either

zinc finger 1 or 4. Early phenotypic characterization of these mutant mice, prior to the work

presented in this thesis, supported the hypothesis that fingers 1 and 4 regulate distinct sets of

target genes, and therefore different biological functions. As a result, it became a high priority to define the binding specificity of zinc fingers 1 and 4, so as to be able to identify sites bound selectively by isoforms containing different combinations of zinc fingers. Given a well characterized recognition sequence for each zinc finger, it would be possible to use the motifs for different combinations to do a bioinformatic scan of the mouse genome and identify candidate targets for isoform-dependent regulation. Bioinformatic data would validate any binding sites identified by ChIP-seq and establish a direct Ikaros target connection with altered gene expression in mutant mice.

Chapter 2: GST and His tag fusion protein purification

To determine the specificity of Ikaros zinc fingers 1 and 4, and identify motifs bound selectively

by different combinations of zinc fingers, Ikaros GST (glutathione S-transferase) fusion proteins

were created for use in *in vitro* binding assays. These constructs were designed to contain just

the Ikaros exons that encode DNA binding zinc fingers 1-4, 1-3, 2-4, or 2-3 (Figure 2a).

Figure 2a: Truncated *IKZF1* GST-fusion protein constructs
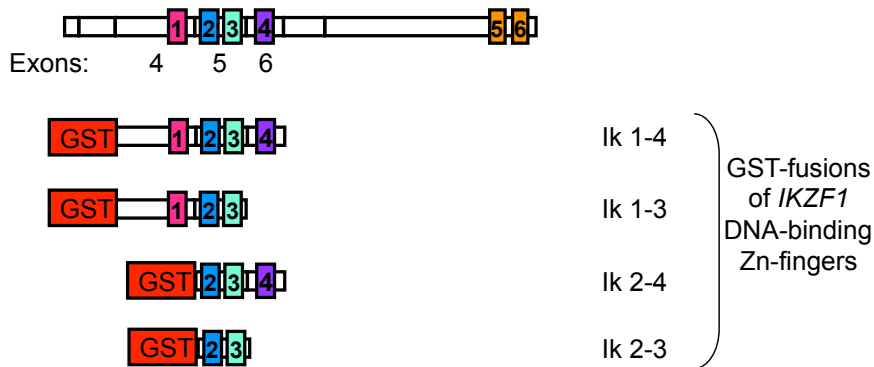
There were several reasons for omitting the C-terminal dimerization domain from these

constructs. One rationale for creating truncated fusion proteins was based on prior lab

experience with attempts to express Ikaros full-length isoforms in bacteria. Expression of the

dimerization domain forms insoluble inclusion bodies, and functional protein has never been

successfully obtained after denaturing. It is possible to express the entire protein in mammalian

cell culture (ex. 293T cells), but this is not a feasible method for large scale purification. Another

problem with including the dimerization domain is that it stabilizes binding to suboptimal sites

through cooperative binding, and the spacing between recognition sites for the two subunits can

be highly variable. There is near complete flexibility in the dimerization domain, half-sites can

be separated anywhere from 2 to 20 bases without affecting stability. In order to examine in

detail the specific base "preferences" for each zinc finger, the conditions need to be as stringent

as possible to select for the highest affinity sequence. Dimerization would bias the results and obscure these preferences.

The entire exons encoding the DNA binding zinc fingers were included in the constructs to reduce the probability of misfolding during translation, and to emulate the strategy used in creation of the $Ikzf1^{\Delta F4/\Delta F4}$ and $Ikzf1^{\Delta F1/\Delta F1}$ mice. $Ikzf1$ sequences including exons 4-6, 4-5, 5-6, or 5 alone (encoding DNA binding zinc fingers 1-4, 1-3, 2-4, or 2-3, respectively) were amplified from cDNA and subcloned into bacterial expression plasmid pGEX-4T-1 (Supplementary Figure 2a). This vector contains a lac promoter for induction of protein expression by isopropyl β-D-thiogalactopyranoside (IPTG). Proteins were expressed in Rosetta DE3 competent cells, a BL21 strain engineered to provide for translation of eukaryotic proteins by supplying tRNA codons rarely used in $E.\ coli$. The DE3 strain contains a copy of the T7 RNA polymerase gene that is also under the control of a lac promoter, such that expression of the polymerase is also induced by IPTG and serves to enhance transcription.

When the GST Ikaros fusion proteins were initially expressed under standard conditions, they were found to be insoluble. Troubleshooting this challenge began by testing the co-transformation of chaperone proteins (ex. GroEL/ES), which assist in the non-covalent folding of protein macromolecules. After determining that this had no effect, the next hypothesis was that the protein was forming an insoluble aggregate through DNA binding, but addition of DNase to the lysis buffer did not solve the insolubility. However, a titration of NaCl determined that a high salt concentration (>500 mM) would release the protein from the insoluble fraction. For those

11

applications discussed in the following chapters that allowed for use of proteins without a GST

tag, a method was developed for thrombin cleavage. The pGEX-4T-1 vector contains a thrombin

recognition site immediately after the N-terminal GST tag. Thrombin conditions were optimized

for cleavage of the protein(s) while bound to the glutathione agarose affinity column. The

resulting cleaved protein fractions were adequately pure for the necessary *in vitro* binding assay

techniques (Figure 2b) and quantified by UV-Vis spectroscopy. The final protocol for protein

expression, purification, and thrombin cleavage can be found in Appendix 1: Methods.

Figure 2b: SDS-PAGE of thrombin-cleavage GST column purification

Zn fingers:	1-4	1-3	2-4	2-3
kDa:	20	16	11	6
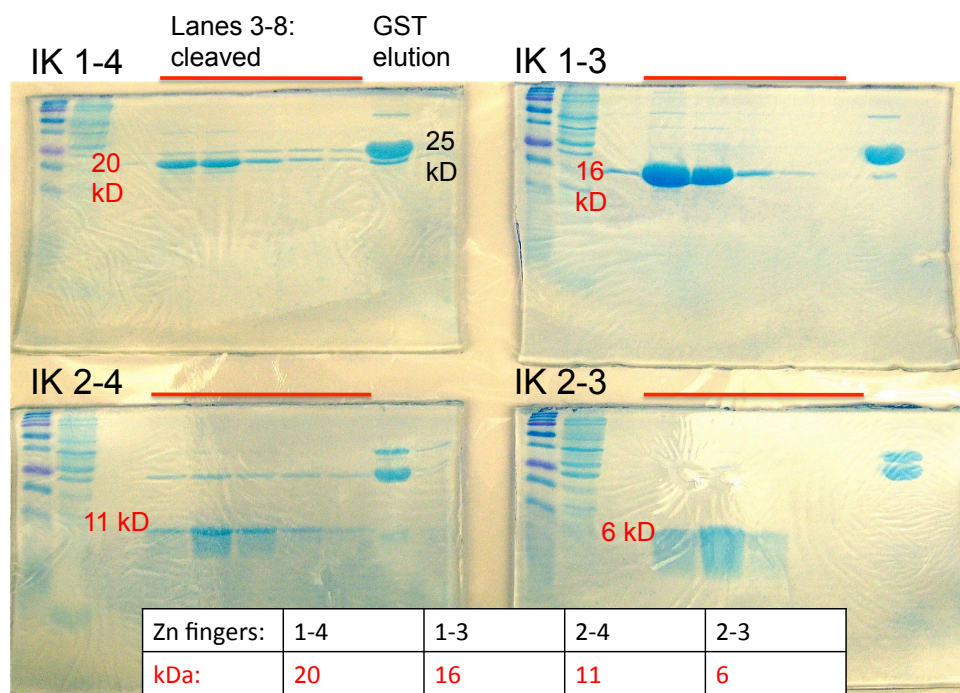
For another application, a 6X His (histidine) tag was favorable to GST. An initial approach

attempted to use a pet28 vector with an N-terminal His tag and purify the proteins using cobalt

resin. Although pure proteins were obtained, they were found to have negligible DNA binding

activity due to chelation of the zinc ions and subsequent loss of zinc finger structural integrity.

All immobilized metal affinity chromatography (IMAC) columns used for purification of His-

tagged proteins contain either Ni2+ or cobalt charged chelate resins, which makes them

unsuitable for purifying zinc finger proteins. After determining this, a C terminal 6X His tag was

instead added to the original N-terminal GST-fusion construct proteins, such that the 6X His tag

remained after cleavage and GST removal (by the same method as previously discussed). This

presence of a common epitope for antibody recognition is essential for an application discussed

at the beginning of Chapter 3.

Chapter 3: Ikaros DNA binding specificity

Almost twenty years have passed since the initial Ikaros binding site selection studies were

conducted, during which there have been great advancements in biotechnology. Among these,

microarrays have made it possible to do high-throughput screening of molecular interactions.

The Bulyk lab at Harvard has developed a protein binding microarray (PBM), a DNA microarray-

based technology that allows for rapid, high-throughput characterization of DNA consensus

binding sites[39]. A collaboration with their lab was established with the goal of obtaining detailed

knowledge of the specific motifs recognized by different combinations of Ikaros zinc fingers, and

defining the sequence specificity of fingers 1 and 4. The PBM method (Figure 3a) uses a

microarray with a library of double stranded DNA (dsDNA) oligos annealed to the surface of

individual wells. An epitope tagged DNA-binding protein of interest is washed over the

microarray and binds to recognition sequences. The array is then labeled with a fluorophore-

conjugated antibody to this epitope and is scanned for fluorescent intensity.

Figure 3a: Protein Binding Microarray (PBM)[37]
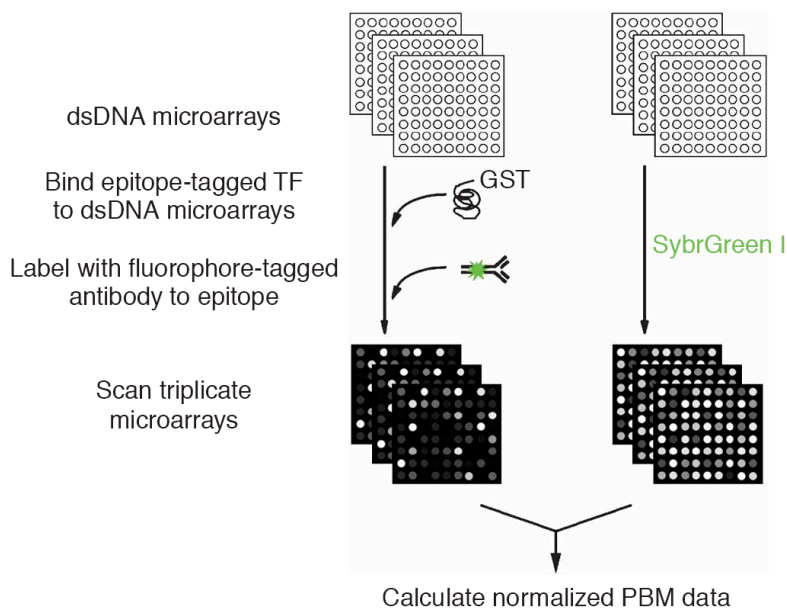
The PBM method has been utilized to define the binding specificity of many proteins, and has been validated with yeast transcription factors containing DNA-binding zinc finger domains[37]. The Ikaros GST fusion proteins described in Chapter 2 were first used in an unbiased array containing 8 base pair random oligomers. These arrays were performed and analyzed by Trevor Siggers from the Bulyk lab. Analysis of the arrays revealed three high affinity motifs, one of which demonstrated a clear sequence preference for ZnF1 (Figure 3b). Unfortunately, no motifs with ZnF4 specificity were identified, and only one motif contained more than 2 bases upstream of the core GGGAA, where ZnF4 is known to bind. Furthermore, the 5' region in this motif was recognized to contain a potential mirror binding site on the complementary strand (GGGA). Oligos containing this sequence would have double the relative local concentration of binding sites, thereby increasing the likelihood of binding and selection of this motif. The motif that showed a higher affinity for ZnF1 also contains a potential binding site on the complementary strand (GGAA). Additionally, it is possible that GST dimerization may have affected results by stabilizing binding by proteins with fewer zinc fingers, and to suboptimal sites in general.

Figure 3b: Unbiased PBM results (8-mer random oligomers)
PBM selected high affinity binding motifs; the first and third motifs have two potential Ikzf1 binding sites.

Bound with high affinity by all 4 proteins:
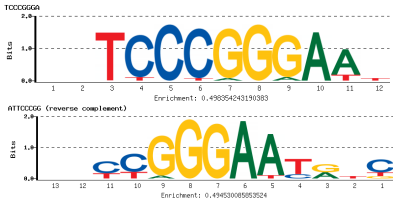
Higher affinity for proteins with ZnF1:
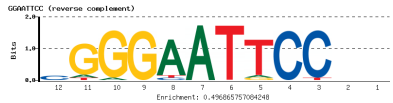
To further verify and dissect these motifs, they were subsequently tested by electrophoretic mobility shift assay (EMSA). This technique uses dsDNA oligonucleotides labeled with a radioactive isotope to visualize a shift in mobility through a polyacrylamide gel when bound by a protein. Polynucleotide kinase (PNK) is used to end-label dsDNA oligos by catalyzing the transfer of the γ-phosphate from γP^{32}-ATP to the 5′-terminus. The radioactively labeled oligos are purified from free γP^{32}-ATP using columns containing Sephadex™ G-50 resin. These columns allow DNA purification by the process of gel filtration; molecules larger than the pores in the matrix are excluded from the gel and elute first. A scintillation counter is used to determine counts per minute (cpm) and an equivalent amount of radioactivity for each probe is used in subsequent binding reactions. The oligos are mixed in a room temperature binding reaction with 1 μg (1 μL of 1mg/mL) of the purified protein at a final physiological salt concentration (50 mM) before running through an electrophoretic gel. Those oligos bound by protein will migrate more slowly due to size-restriction, resulting in the "shift" of a band representing the protein-DNA complex. For full EMSA protocol see Appendix I: Methods.

Oligos designed to contain the PBM motifs, and mutant variations of each, were screened for binding by Ikaros-GST fusion proteins using EMSA (Supplementary Figure 3a). One motif that was selected with high affinity for all proteins by PBM unexpectedly showed much higher binding affinity for proteins containing ZnF1 by EMSA. Otherwise, as predicted, mutations of key bases in the core consensus negated binding, and mutations in the flanking oligos had variable results, some reducing or increasing binding selectively for different zinc finger combinations.

Verified endogenous Ikaros binding sites frequently occur in clusters. To better understand the

protein-DNA interactions on a one-to-one ratio, 13 oligos were designed from a subset of

endogenous multi- or dimeric sites, with each of the individual sites separated. A screen of

these single site isolation oligos showed various patterns of binding to proteins containing

different combinations of zinc fingers (Supplementary Figure 3b).

EMSA was concurrently used to validate that the truncated Ikaros GST fusion proteins would

recapitulate the binding affinities seen for different *Ikzf1* isoforms, with a selection of oligos

previously designed and tested in the original binding selection study (Supplementary Figure

3c)[20]. These oligos, labeled IK-BS1 through 6, were designed based on the identified consensus

sequence, and mutations or dimeric combinations thereof (the one exception being IK-BS2,

which is a similar NF-kB site from the IL-2R promoter). Binding patterns for the IK-BS1-6 oligos

with Ikaros GST fusion proteins (Figure 3c) matched the original study results with isoforms

containing the corresponding combinations of zinc fingers. Examples of two of the binding

patterns displayed in Figure 3c are circled in red; IK-BS3 has higher affinity for proteins

containing ZnF1, while IK-BS6 shows ZnF4 specificity.

Figure 3c: Example of EMSA categorized binding patterns demonstrated with IK-BS oligos 1 through 6.
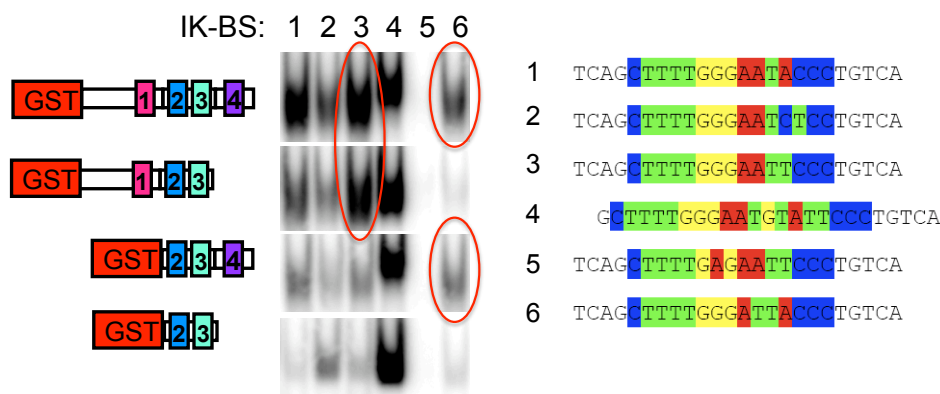
For all three of these screens (ie. the PBM motifs, endogenous isolated sites, and IK-BS1 through 6), the oligos were categorized by selectivity for different zinc finger combinations. Oligo binding patterns were categorized as one of the following: "All" (Any combination of ZnF), "Any 3" ZnF, "ZnF1" specific, "ZnF4" specific, "Only all 4" ZnF, and "None" (Supplementary Figure 3d). A position weight matrix (PWM) was created based on the combined results from all three screens to determine a putative "consensus" sequence (Figure 3b). This PWM is a matrix calculation of the frequency each base appears at a given position in the DNA sequence. Those oligos which were categorized as having ZnF1 or ZnF4 specificity were used in combination with those which bound "All" to determine favorable sequences for ZnF1 and ZnF4, respectively. However, as this is only a sampling of 40 sequences, it is not a rigorous analysis, and the results are only an informed hypothesis of what bases are recognized at each position in the consensus sequence.

Figure 3d: Position Weight Matrix (PWM) analysis
Bases shown for a frequency > 0.4

Position	1	2	3	4	5	6	7	8	9	10	11	12	13
Base:	T	TA	GT	T	G	G	G	A	A	TC	GT	TC	CT

It is important to note that results from the PBM arrays and EMSA affinity analysis suggested that the presence of dimers formed by GST might be influencing selectivity. Two of the three PBM motifs contain two potential binding sites, and EMSA images showed that protein shifts with a single-site oligo appeared the same as those with a dimeric site. Moving forward, subsequent protein purifications removed the GST domain by on-column cleavage with thrombin (as described in Chapter 2 and Appendix I). These cleaved forms were used for all further EMSA testing of this putative consensus.

Oligos were next designed to validate the putative consensus with combinations of the favored and disfavored ZnF1 and ZnF4 contacts, with each simplified to one sequence. The results are shown in Figure 3e, where "+" = favorable and "-" = unfavorable.

Figure 3e: Testing of PWM consensus by EMSA
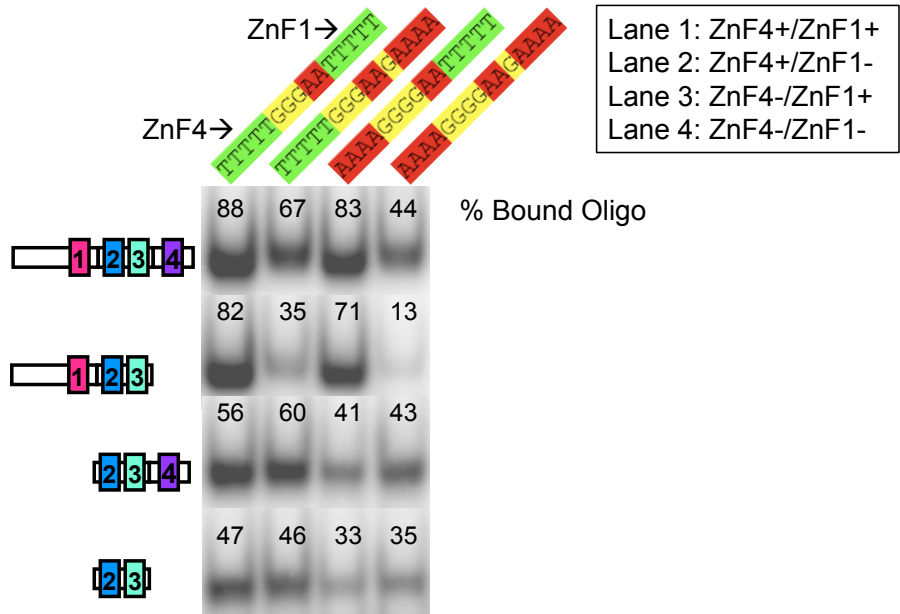
Conclusions can be drawn from each lane of the EMSA images shown in Figure 3e. Lane 1, which contains an oligo with favored bases for both ZnF1 and ZnF4, shows relatively strong binding for all of the proteins, confirming recognition of these sites. Lane 2, for which the oligo now lacks a ZnF1 recognition site, shows a substantial loss of binding for both Ik 1-4 and Ik 1-3 (88% → 67% and 82% → 35% bound, respectively), and Ik 1-3 drops slightly below Ik 2-3. However, both Ik 2-4 and Ik 2-3 remain at the same percentage, suggesting that ZnF1 is auto-inhibiting binding for both Ik 1-4 and Ik 1-3 in the absence of ZnF1 contacts. This auto-inhibition could be due to a conformational change and/or steric hindrance. There is also a striking difference in binding for Ik 1-4 versus Ik 1-3 without ZnF1 contacts (67% versus 35% bound), while both have a similar affinity when a ZnF1 recognition site is present (lane 1 or 3). This would indicate that proteins

containing ZnF1 are therefore dependent on ZnF4 for binding stability in the absence of ZnF1

contacts. Lane 3, where the oligo contains a recognition site for ZnF1, but not ZnF4, shows

"ZnF1 specificity", ie. ZnF1 is required for high affinity. The percentage bound also drops for

both Ik 2-4 and Ik 2-3, likely because of the T$\rightarrow$G change at the 5' end of the core, which is

contacted not only by ZnF4, but also ZnF3. In Lane 4, where the oligo contains neither of the

ZnF1 or ZnF4 recognition sequences, the binding of all the proteins drops to a percentage similar

or below that of Ik 2-3, indicating that neither ZnF1 or ZnF4 provide any significant benefit.

Again, Ik 1-3 appears to be auto-inhibited by ZnF1 in the absence of ZnF1 contacts, as it drops far

below that of Ik 2-3 (13% versus 35% bound). As would be predicted, Ik 2-3 maintains a similar

percentage bound for all of these oligos, since the only bases that these zinc fingers contact are

those one position 5' or 3' to either side of the core GGGAA.

In vivo, the two prominent naturally occurring isoforms are Ik-1 and Ik-2, which include ZnF 1-4

and 2-4, respectively. Results of this analysis suggest that in locations where a ZnF1 recognition

site is not present, the binding of the isoform with ZnF 1-4 may be unstable, and could act as a

dominant negative. Therefore the balance of expression of these two isoforms at different

stages of development might affect regulation of different target genes. The observation that

certain biological functions of Ikaros are maintained in *Ikzf1*$^{\Delta F4/\Delta F4}$, and therefore do not require

ZnF4, could also suggest that ZnF4 stability may not be necessary for certain mechanisms of

Ikaros transcriptional regulation (eg. recruitment of specific other factors, roles in activation

versus repression, etc.).

Although the DNA contacts for ZnF1 have been mostly confirmed, the specificity of ZnF4 remains

unclear. No recognition sequence has been identified that is truly "ZnF4 specific", ie. that will

bind with higher affinity to those proteins that contain ZnF4. One hypothesis from this

observation is that ZnF4 may primarily make nonspecific interactions with the DNA backbone,

which has been seen in studies of other zinc finger proteins[40,41]. To interrogate ZnF4 specificity,

oligos were designed based on frequencies found by PWM calculations. These sequences

started with the highest frequency bases for each position in the ZnF4 binding region and

individually substituted any bases with ambiguous preference (Supplementary Figure 3e). This

ZnF4 substitution analysis did not yield a specific "consensus", but did confirm a general

preference for thymidine at most positions: 5' N [T/A] T [T/C] T 3' (followed by GGGAA core

downstream). The only novel specific base recognition identified was the thymidine at the third

base upstream from the core. The thymidine at the position one base upstream of the core had

already been confirmed in the consensus of ZnF3, as it is contacted by the glycine at position 6

of the alpha helix. The glutamine at position 2 of the ZnF4 α-helix contacts this base pair

through the adenine on the complementary strand.

The initial unbiased PBM contained the standard 8-mer random oligo sequences for binding

selection. However, the Ikaros zinc fingers can recognize a sequence up to 13 bases in length, so

a customized chip with longer random oligomers is preferable. A subset of a PBM chip was

customized with 95 specially designed 15-mers containing the core GGGAA. These were flanked

by combinations of favored or disfavored recognition sequences for ZnF1 and ZnF4, and

individual substitutions of every base at each position. GST fusion proteins were still used for

this array, since His tagged proteins were not yet successfully purified at this time, and the PBM protocol requires a common epitope for antibody recognition. The PBM data plots of log(F), where F is mean fluorescence over 3 probe replicates, clearly show that ZnF4 does impact binding (Figure 3f). The two plots at bottom compare Ik 1-3 versus Ik 1-4, and Ik 2-3 versus Ik 2-4, showing the difference in binding affinity of a protein with or without ZnF4. These two plots have low R^2 correlation values of 0.38 and 0.30, respectively, and therefore demonstrate that ZnF4 does make a significant difference in binding. However, there was again no clear ZF4 motif revealed from this PBM analysis, further supporting the hypothesis that ZnF4 makes relatively nonspecific interactions with the DNA backbone.

Figure 3f: Log fluorescence plots for custom PBM with GST tagged proteins
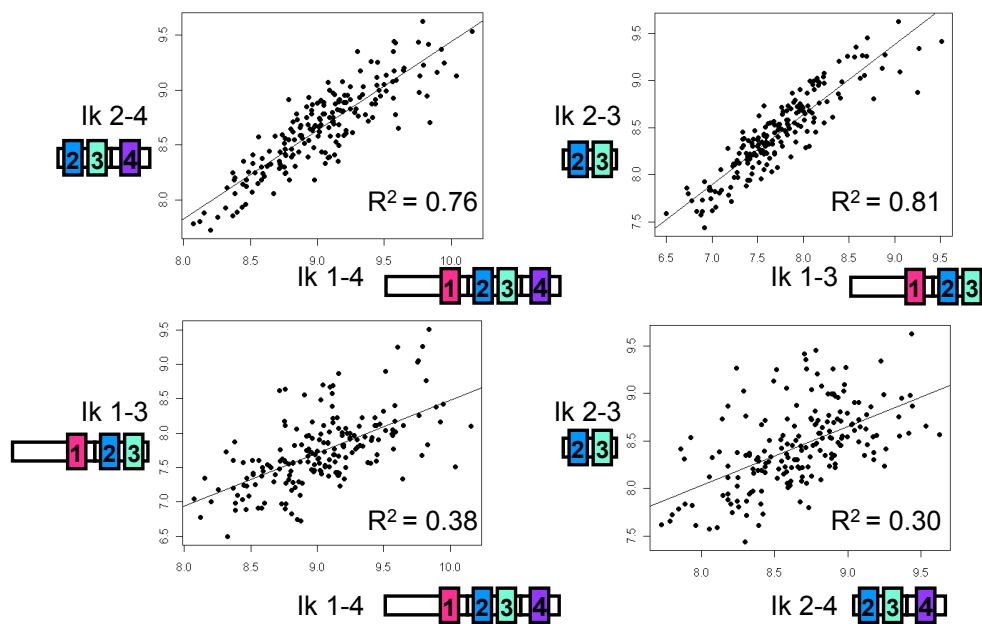
One unexplained result from this array was the absence of significant ZnF1 specificity, as shown by the R^2 values of 0.76 and 0.81 for the log(F) plots comparing Ik 2-4 versus Ik 1-4, and Ik 2-3 versus Ik 1-3, respectively. This would not be expected from the set of oligos used in the array, which included sequences containing a ZnF1 recognition site without preferred contacts for

ZnF4. Such oligos should have shown preferential binding by proteins containing ZnF1. It is

possible that there was not a large enough fraction of these sequences to be significant.

The PBM raw fluorescence values will vary between proteins, so to directly compare one protein

to another, it is necessary to normalize the data. This is done by calculating a z-score, which

takes into account the average and the total range of values within the data set for a given

protein. The z-score is a dimensionless value that represents the number of standard deviations

a data point is from the mean of a population. A positive z-score is above the mean, a negative

z-score below the mean. It is calculated by subtracting the mean of a population from an

individual raw data point, and then dividing by the standard deviation ($z = (x-\mu)/\sigma$, where μ is

the mean and σ is the standard deviation of the population).

Once the z-scores have been calculated, these values can be compared between all of the

proteins for every oligo sequence used in the array. Software that analyzes patterns changes

between data sets, such as Cluster, can be used to group the oligos based on binding affinity for

the different proteins. The heat map in Figure 3g visualizes the z-score values on a red-yellow-

green (high-to-low) color scale, and clusters the data into 6 groups based on binding patterns.

Group 2 shows high affinity for all proteins, while group 6 shows low affinity for all proteins.

Groups 3 and 4 show higher affinity for proteins with finger 4, as was seen with the GST fusions

by EMSA. Groups 1 and 5 unexpectedly showed higher affinity for proteins without finger 4, a

selectivity which had not been observed by EMSA. Upon closer examination, all of the

sequences in this group either contained a poor ZnF4 recognition site or a base pair substitution of the 5' glutamine in the core GGGAA, therefore disrupting ZnF4 contact.

Figure 3g: Cluster heatmap of custom PBM z-scores
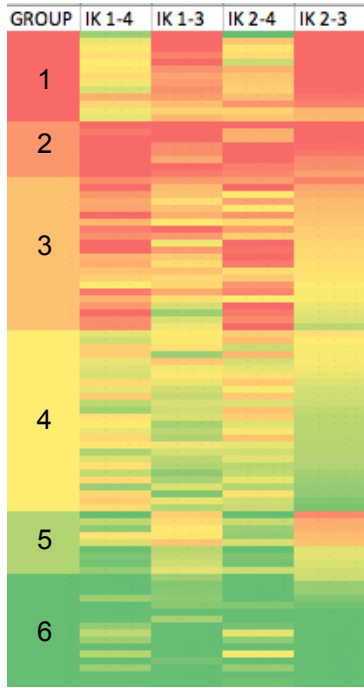
As described in Chapter 2, proteins have been purified with a 6X His tag for antibody recognition in future PBM arrays. The next PBM approach was to design a full custom array with the same format as the previous smaller set of sequences, that is containing the GGGAA core, but this time with random 5 base-pair flanking regions. This method will be able to thoroughly test ZnF1 and ZnF4 specificity with a larger population of sequences more likely to result in a statistically significant consensus motif. This array has been run in duplicate using the aforementioned His tagged proteins, and is awaiting analysis by the Siggers lab.

The Hes1 promoter illustrates an example of an endogenous site that requires ZnF4 for binding stability. Hes1 is a Notch target gene that is repressed by Ikaros, and transcripts are elevated in

thymocytes from *Ikzf1*$^{\Delta F4/\Delta F4}$ mice (Figure 3h). This derepression is likely to play a role in

malignant transformation of thymic lymphomas, as the *Ikzf1*$^{\Delta F4/\Delta F4}$ mutation and BCR-ABL, a

constitutively expressed mutation of Notch, are seen to have a synergistic effect[29]. In the EMSA

shown at the right of Figure 3h, the oligo in lane 1 contains the full Hes1 promoter with a

dimeric *Ikzf1* binding site, while lanes 2 and 3 contain each of the half sites individually. Isolation

of the two half-sites into separate oligos shows that one has a poor recognition sequence for

ZnF1, and therefore Ik 1-3 does not bind well without ZnF4. Binding of Ik 1-3 dimers to this

endogenous tandem site also appears unstable, and this is supported by the QPCR data that

shows derepression of Hes1 transcription *in vivo*.

Figure 3h: *In vitro* binding to Hes1 promoter and correlating *in vivo* thymocyte Hes1 transcript levels

(H. Schjerven, unpublished data)
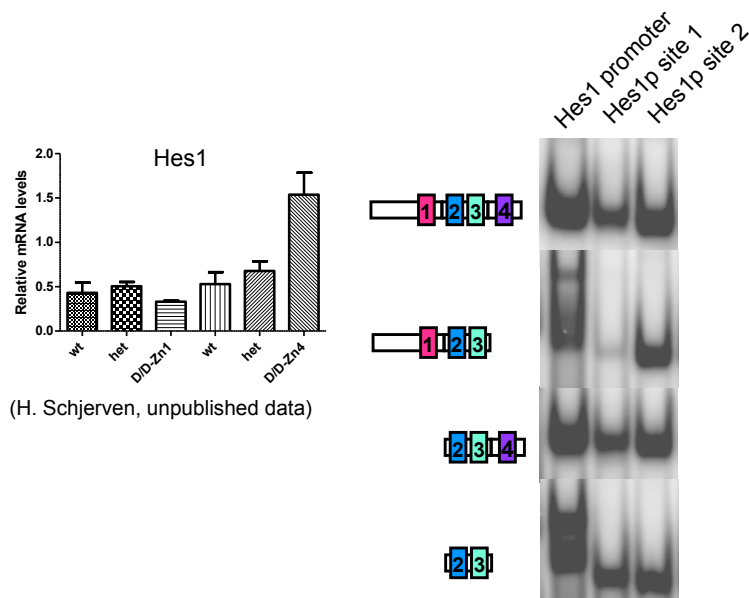
Recent progress with Ikaros chromatin immunoprecipitation (ChIP) methods has resulted in

successful identification of novel binding sites throughout the genome. Some of the sequences

most enriched for Ikaros binding contained domains with a string of repeated recognition sites.

Among these were regions that lost binding signals in thymocytes specifically from either

Ikzf1$^{\Delta F4/\Delta F4}$ or *Ikzf1*$^{\Delta F1/\Delta F1}$ mice. These sites represent *in vivo* examples of specificity for ZnF4 and

ZnF1. Oligos were designed to isolate a single site from three of these repetitive regions, and

used for EMSA to validate *in vitro* binding affinity with different ZnF combinations (Figure 3i).

Ikaros ChIP signal was present in both mutants for region Snx29, and this positive control shows

strong binding for each of the ZnF combinations. Region Nr3c2 lost Ikaros ChIP signal specifically

in *Ikzf1*$^{\Delta F1/\Delta F1}$, and shows loss of binding for Ik 2-4, which lacks ZnF1. Region Snx25 lost Ikaros

ChIP signal specifically in *Ikzf1*$^{\Delta F4/\Delta F4}$, and shows loss of binding for Ik 1-3, which lacks ZnF4, and

lower affinity for Ik 1-4 than Ik 2-4, supporting the hypothesis that ZnF1 will auto-inhibit binding

when a ZnF1 recognition site is not present. Each of these EMSA results therefore recapitulates

the Ikaros ChIP data.

Figure 3i: *In vitro* binding validation of ChIP target sequences[29]
Below are nucleotides flanking the GGGAA core recognition sequence.
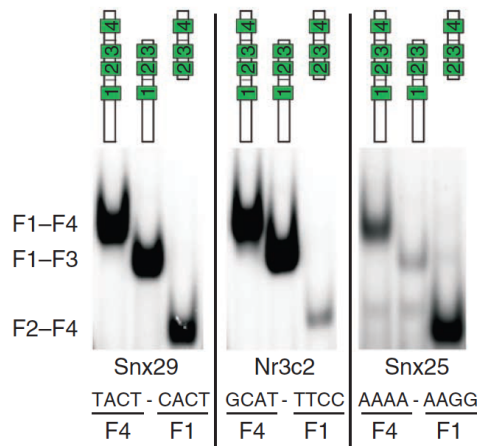
Chapter 4: Kinetics and stability of DNA binding

Following the EMSA results indicating that ZnF4 plays an important role in the stability of Ikaros

DNA binding, one primary goal became to look more closely at the kinetics of this interaction.

An optical phenomenon known as surface plasmon resonance (SPR) has been utilized to develop

biosensors that detect molecular interactions in real time. This technology, produced by

Biacore®, can characterize both the binding affinity and kinetics of unlabeled reactants. The

method involves immobilizing a ligand (eg. DNA) to a sensor chip, and injecting a binding partner

(eg. protein) across the surface (Figure 4a[42]). When a binding interaction occurs, the associated

increase in mass causes a proportional increase in refractive index, which is observed as a shift in

the index of refraction from a light source reflected off the chip surface, which can be detected

and measured in resonance units (RU). As this is measured in real time, the interaction produces

an RU curve that represents the binding association, equilibrium, and dissociation (Figure 4a[41]).

From these curves several thermodynamic properties of the binding interaction can be

determined.

Figure 4a: Surface Plasmon Resonance (SPR)[41]
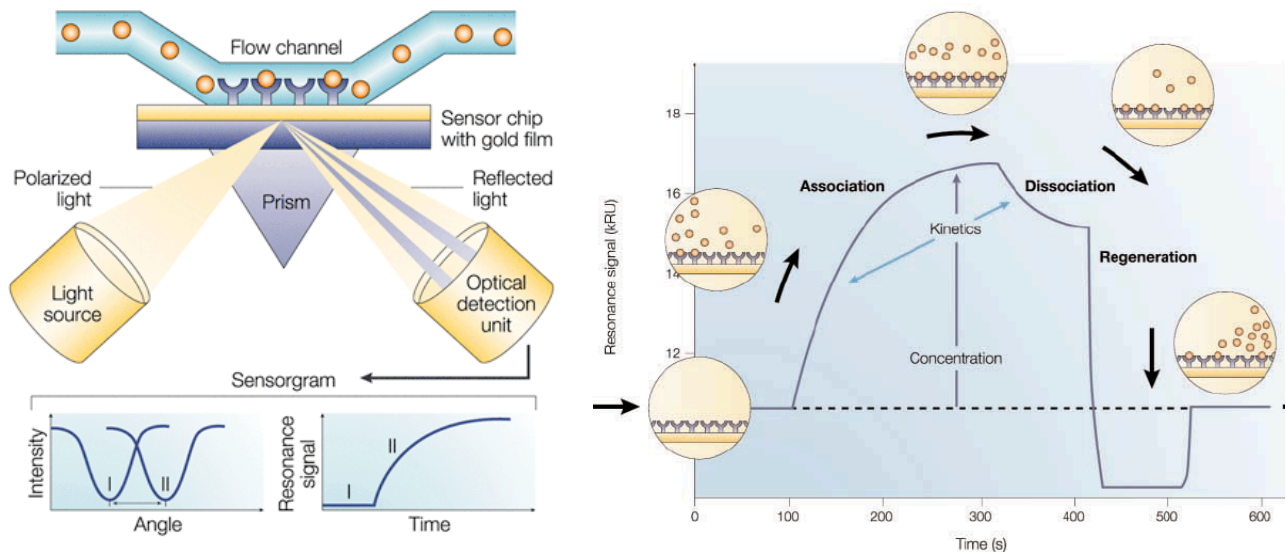
Many studies have used SPR to examine the DNA binding interactions of zinc finger proteins and other structurally related families[40,43,44,45]. For the purpose of this research, the SPR biosensor chip was labeled with DNA oligos containing Ikaros binding sites. Single stranded DNA oligos were commercially synthesized with a 5' biotin conjugation and annealed to complementary strands to form biotinylated dsDNA. These oligos were used to label a chip with a matrix of carboxymethylated dextran pre-immobilized with streptavidin on the surface (Figure 4b). For full labeling protocol, see Appendix I: Methods.

Figure 4b: Surface of an SPR biosensor chip labeled with biotinylated DNA
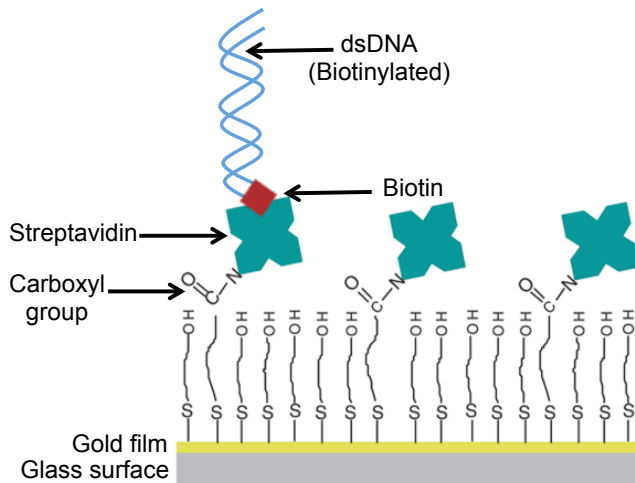
One advantage of the Biacore technology is that the binding interactants need not be labeled with an epitope or fluorescent marker. For this study, the cleaved Ikaros proteins could be used instead of the fusions to GST, which is bulky (26 kD) and would likely interfere with an accurate assessment of the binding interactions. Each Biacore chip has 4 channels, and the ideal setup has a separate control (unlabeled) channel to pair with each labeled test channel. This is so the protein can flow through only one control and test channel per run, which yields more precise results than several in series. Four sequences were used for labeling the chips, the same as were shown in Figure 3e, which contain combinations of the favored and disfavored ZnF1 and ZnF4

contacts (ZnF1+/ZnF4+, etc.). These were chosen for consistency and the ability to make direct

comparisons to the previously analyzed differences in binding to each of the zinc finger

combinations. Since 4 sequences were chosen, two chips were used, both of which were

labeled in channels 2 and 4, with channels 1 and 3 used as controls for each, respectively.

Protein concentrations were determined using the Bradford assay, and samples were run on the

Biacore® model T-100, with each protein injected in a two-fold titration series of increasing

concentration. The chip was regenerated to remove all protein between each injection to avoid

variability due to residual signal. RU curves were processed and analyzed using a program called

Scrubber, developed by the Biomolecular Interaction Facility at the University of Utah. This

program was chosen as an alternative to the Biacore company software (BIA Evaluation)

because it has a more intuitive interface, and was previously used by another lab member.

Using the Scrubber software, RU curves are evaluated by fitting algorithms that compare the raw

data to binding models of second order reactions (Example shown in Figure 4c).

Figure 4c: Second order reaction modeling using Scrubber
Example shown is a titration of IK 1-3 (green); plots of the modeling equations are shown in orange.
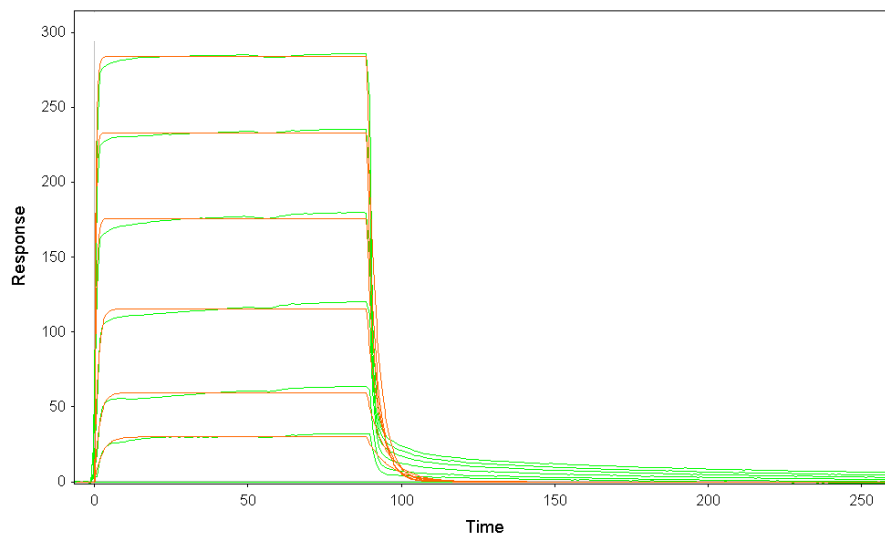
(The plots of RU curve models for titrations of each of the individual proteins with the ZnF1+ZnF4+ oligo can be found in Supplementary Figure 4a-d). A second order reaction occurs when a reaction A+B $\rightleftharpoons$ C (reactants A and B in equilibrium with complex C) has a rate proportional to the product of the two reactant concentrations r=k[A][B]. From these models several thermodynamic constants can be determined, including the apparent binding affinity, half life ($t_{1/2}$) and dissociation constant (K_D).

The dissociation constant K_D can be used to quantify the affinity between a ligand (L) such as DNA, and a protein (P). For the reaction P+L $\rightleftharpoons$ C, the dissociation constant K_D = ([P][L])/[C], where [P], [L] and [C] represent molar concentrations of the protein, ligand and complex, respectively. The dissociation constant for a ligand corresponds to the molar concentration at which half the protein is bound, that is when the concentration of protein bound to ligand [C], equals the concentration of protein with no ligand bound [P]. The lower the dissociation constant, the higher the affinity between the protein and ligand. For example, a nanomolar (nM) dissociation constant represents a tighter binding reaction than a micromolar (μM) dissociation constant.

Another useful calculation for understanding the kinetics of a binding reaction is the half life, $t_{1/2}$. The general definition of half life is the amount of time required for any quantity to drop to half of its original value. In the context of a protein-DNA ligand binding reaction (P + L $\rightleftharpoons$ C), this is the time required for half of the protein-DNA complexes that were present at equilibrium to "fall apart". When the concentration of a ligand is fixed, such as in the Biacore chip DNA

30

immobilization, the half life of a second order reaction, for which P + L ⇌ C, depends on the initial protein concentration P_0. Therefore, to appropriately compare the binding by each protein to different oligos, thermodynamic constants were calculated from one concentration where RU curves for all of the proteins could be fitted to modeling equations (500 nM). Figure 4d shows the RU curves for all of the proteins at 500 nM, along with tables containing RU max values and calculations of $t_{1/2}$ (sec) and K_D. From layering the RU curves for each oligo on the same plot, the qualitative aspects of the binding by each of the proteins can be directly compared (To avoid confusion, plots of the modeling equations are not shown). The max RU changes show the differences in apparent binding affinity, and the relative slopes of the dissociation curves suggest differences in stability of each binding reaction.

Figure 4d: First round SPR data for all proteins at 500 nM
RU curves/max values and calculations of $t_{1/2}$ (sec) and K_D

ZnF4+ ZnF1+ — TTTTTGGGAATTTTT

	RU_{max}	$t_{1/2}$ (sec)	K_D
IK 1-4	396	80.53	40 nM
IK 1-3	230	9.04	55 nM
IK 2-4	220	36.44	67 nM
IK 2-3	34	6.75	50 nM

Running Buffer
IK 1-4 – 500 mM
IK 1-3 – 500 mM
IK 2-4 – 500 mM
IK 2-3 – 500 mM

ZnF4+ ZnF1- — TTTTTGGGAAGAAAA

	RU_{max}	$t_{1/2}$ (sec)	K_D
IK 1-4	370	46.96	39 nM
IK 1-3	128	15.52	80 nM
IK 2-4	224	83.43	32 nM
IK 2-3	57	7.65	120 nM

Running Buffer
IK 1-4 – 500 mM
IK 1-3 – 500 mM
IK 2-4 – 500 mM
IK 2-3 – 500 mM

ZnF4- ZnF1+ — AAAAGGGGAATTTTT

	RU_{max}	$t_{1/2}$ (sec)	K_D
IK 1-4	399	62.71	28 nM
IK 1-3	178	3.17	210 nM
IK 2-4	180	17.29	333 nM
IK 2-3	22	3.68	272 nM

Running Buffer
IK 1-4 – 500 mM
IK 1-3 – 500 mM
IK 2-4 – 500 mM
IK 2-3 – 500 mM

ZnF4- ZnF1- — AAAAGGGGAAGAAAA

	RU_{max}	$t_{1/2}$ (sec)	K_D
IK 1-4	270	16.40	155 nM
IK 1-3	52	8.92	51 nM
IK 2-4	187	29.74	122 nM
IK 2-3	32	5.82	77 nM

Running Buffer
IK 1-4 – 500 mM
IK 1-3 – 500 mM
IK 2-4 – 500 mM
IK 2-3 – 500 mM

Several of the conclusions from EMSA analysis are recapitulated by comparing the Biacore

results for each protein at 500 mM. New observations can also be made, most importantly, that

the t $_{\frac{1}{2}}$ of DNA binding for proteins that include ZnF4 is always 2-20 fold higher (varies depending

on the oligo sequence) than those without ZnF4. This shows that ZnF4 increases binding stability

regardless of whether the oligo contains a putative ZnF4 recognition site, reinforcing the

hypothesis that binding contacts may be largely nonspecific interactions. Another important

observation is that Ik 1-4 has a much longer t $_{\frac{1}{2}}$ than Ik 2-4 when the oligo contains a ZnF1

recognition sequence. However, the t $_{\frac{1}{2}}$ of Ik 2-4 becomes much longer than the t $_{\frac{1}{2}}$ of Ik 1-4

when the ZnF1 contacts are lost, further supporting the idea that there may be auto-inhibition

by ZnF1.

For an oligo containing both ZnF1 and ZnF4 recognition sites (ZnF1+ZnF4+), maximum RU values

represent an approximate highest capacity reference for each protein. These values range from

RU_{max}= 396 for Ik 1-4, down to RU_{max}= 34 for Ik 2-3. However, mass causes a proportional

increase in the refractive index shift observed by SPR, and when this is taken into account, Ik 1-4

and Ik 2-4 have a nearly identical level of binding, while Ik 1-3 shows ~75% of this level, and Ik 2-

3 ~30%. This more closely correlates with the binding percentages seen by EMSA. Each of the

proteins has a low K_D value within a narrow range (40-67 nM), indicating a relatively similar high

affinity for this oligo. By EMSA analysis, an oligo that lacks a ZnF1 recognition site but maintains

a ZnF4 recognition site (ZnF1- ZnF4+) showed significant loss of binding for Ik 1-3, and some loss

for Ik 1-4, while Ik 2-4 and Ik 2-3 remained the same. This suggested that ZnF1 was auto-

inhibiting binding in the absence of ZnF1 contacts. With Biacore, Ik 1-3 loses almost 50% of

RU$_{max}$ value, while the rest of the proteins stay relatively the same. Although Ik 1-4 drops only

26 RU, the t $_{½}$ drops by half, and is much shorter than for Ik 2-4 (47 sec versus 83 sec,

respectively). As previously mentioned, this further supports the idea that ZnF1 is inhibiting the

stability provided by ZnF4.

By Biacore, an oligo that maintains a ZnF1 recognition site but lacks a ZnF4 recognition site

(ZnF1+ ZnF4-) did not demonstrate the same ZnF1 specificity for high affinity binding as was

shown by EMSA. However, RU$_{max}$ did drop ~20-30% for all of the proteins except Ik 1-4, and the

t $_{½}$ of Ik 2-4 drops by half to 17 sec. When compared to t $_{½}$= 63 sec for Ik 1-4, this indicates that

ZnF1 does provide a considerable advantage for binding to this oligo. For an oligo that contains

neither of the ZnF1 or ZnF4 recognition sequences (ZnF1-ZnF4-), EMSA showed a much lower

level of binding for Ik 1-3 than any of the other proteins. The same is true for the Biacore RU$_{max}$

data once mass is taken into account, again supporting the hypothesis of ZnF1 inhibition. It is

also important to note that the RU$_{max}$ value of Ik 1-3 drops 75% from ZnF1- ZnF4+ to ZnF1-ZnF4-,

confirming the effect of the T$\rightarrow$G change at the 5' end of the core on recognition by ZnF3. The

only considerable change in RU$_{max}$ for the Ik 1-4 protein is binding to the ZnF1-ZnF4- oligo,

showing a ~30% decrease in RU$_{max}$ and a ~60% decrease in t $_{½}$ from the ZnF1-ZnF4+ oligo,

suggesting that there is some specificity involved in ZnF4 stability.

Shortly after running the untagged proteins, variations with a 6XHis tag were successfully

purified and tested in the Biacore assay. These proteins nonspecifically bound to the chip, and

were unable to be removed by standard regeneration buffers. The proteins interacted with the

matrix on the surface of the chip, as this binding occurred in all of the channels, even the controls. Most ionic binding is counteracted by the NaCl in the running buffer, but hydrophobic surfaces can be exposed in partially denatured proteins and interact with the surface of the chip. This is best prevented with detergent, which had not been added to the buffer, but was never previously a problem, so the effect was exacerbated in some way by the presence of a 6XHis tag. Histidine has a positively charged imidazole functional group and is a chelator of zinc, so additional $ZnCl_2$ was present in the buffer to prevent chelation of the zinc finger proteins. It is possible that interactions between the 6XHis and $ZnCl_2$ caused protein to bind to the carboxymethylated dextran matrix on the chip surface. The binding was irreversible and these chips were unable to be salvaged by any attempts to clean with high detergent solutions, so new chips were labeled for future use.

Two key protocol modifications were made for Biacore experiments run with the newly labeled chips. To prevent any recurrence of the proteins "sticking" to the chip, a low concentration of HBPS (HEPES + Tween 20) detergent was added to the running buffer. This is reflected in the protocol included in Appendix I. Additionally, to better emulate the conditions used in the binding reaction for EMSA, the analysis compartment temperature was changed from 4°C to 25°C. It is relevant to note that the chip is located in the analysis compartment, separately from where the proteins are stored prior to injection, so no additional degradation should have occurred. The "room temperature" 25°C binding condition is also a more standard setting used by the majority of Biacore studies. The addition of detergent to the running buffer and the higher binding temperature dramatically decreased both RU_{max} and $t_{1/2}$ for most of the proteins

(Figure 4e). RU_{max} binding levels for the same concentration range as the previous experiment were low enough that only data from the chip with ZnF4 recognition sites was analyzed. Shown in Figure 4e are RU curves for duplicate runs of each of th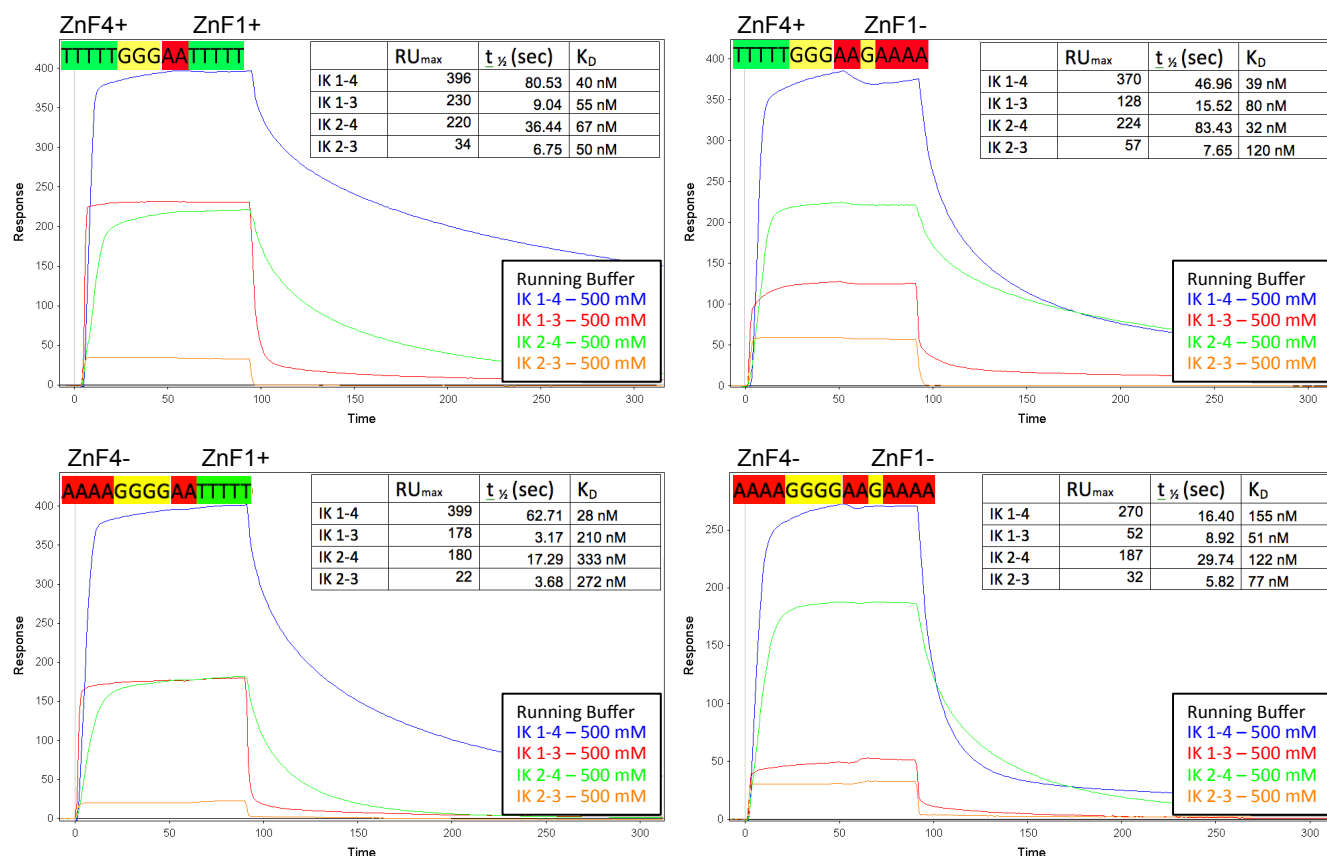e proteins at 1 uM, and model equations based on their averages. (The plots of RU curve models for titrations of each of the individual proteins with the ZnF1+ZnF4+ oligo can be found in Supplementary Figure 4e-h).

Figure 4e: Second round SPR data (1 µM protein concentrations)
Model equations for duplicates are shown in orange

	RU_{max} (avg)	$t_{1/2}$ (sec)	K_D
IK 1-4	388	18.0	9.5 nM
IK 1-3	108	1.3	173.4 nM
IK 2-4	194	8.8	31.8 nM
IK 2-3	32	0.6	63.8 nM

Running Buffer
IK 1-4 – 1 µM
IK 1-3 – 1 µM
IK 2-4 – 1 µM
IK 2-3 – 1 µM

	RU_{max} (avg)	$t_{1/2}$ (sec)	K_D
IK 1-4	327	9.3	20.0 nM
IK 1-3	31	1.8	8.7 uM
IK 2-4	221	16.7	12.3 nM
IK 2-3	50	0.2	306.0 nM

Running Buffer
IK 1-4 – 1 µM
IK 1-3 – 1 µM
IK 2-4 – 1 µM
IK 2-3 – 1 µM
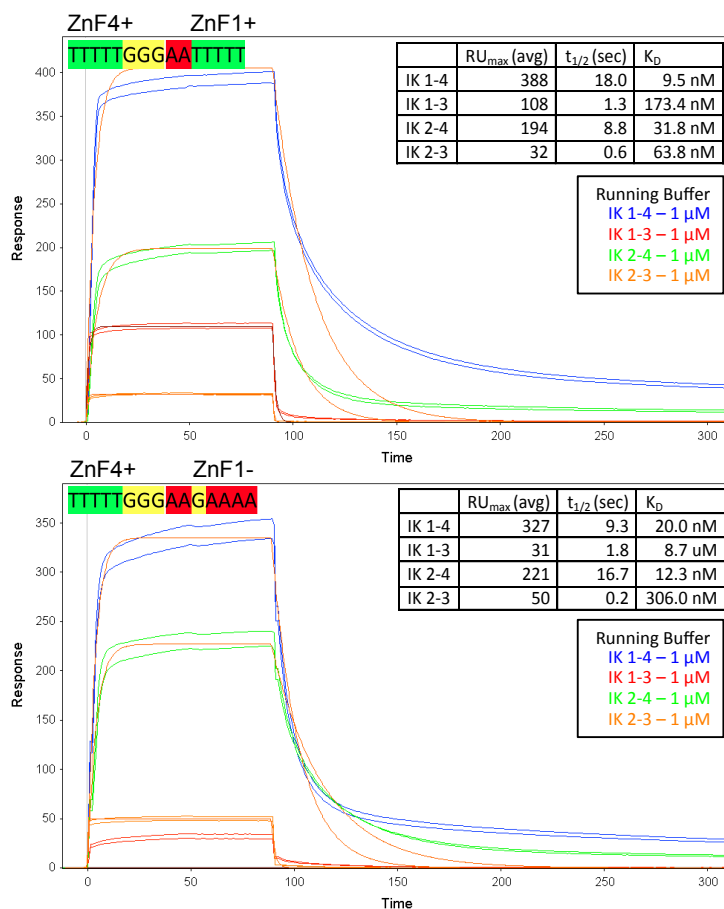
Although this data is from a twofold higher concentration (1 µM versus 500 mM), RU_{max} values for all of the proteins except Ik 1-3 are comparable to the first set of experiments. By comparison, RU_{max} values are about twofold lower for Ik 1-3, and this time the raw value drops below Ik 2-3 for the oligo without a ZnF1 site, even before adjusting for mass. The t ½ is greatly

decreased for all the proteins due to higher binding temperature and detergent, but the same patterns are observed. Ik 1-4 has a much longer t $_{\frac{1}{2}}$ than Ik 2-4 when a ZnF1 recognition site is present, and again the reverse is true without a ZnF1 site. Dramatic changes in K_D reflect the lower binding affinity for Ik 1-3, it is about threefold higher than the first experiment for the oligo with a ZnF1 site, and 100X higher for the oligo without a ZnF1 site.

If future Biacore experiments were to be conducted at the same conditions, it would be necessary to use a titration of higher protein concentration. It would be informative to use a channel with nonspecific DNA that does not contain an Ikaros binding site, and compare this to a control with no DNA. This might reveal nonspecific binding signal, and potentially a difference in proteins dependent on inclusion of ZnF4. Even though ZnF4 by itself is insufficient to bind DNA as observed by EMSA, Biacore is more sensitive, and it might also be interesting to purify ZnF4 alone and test whether there is any significant interaction with a specific putative ZnF4 recognition sequence and compare to nonspecific DNA.

Chapter 5: Summary and potential future methods

This section outlined the experimental approach used to examine Ikaros DNA binding specificity. While several conclusions were made, other theories need further investigation. It can be confidently determined that ZnF1 binds a specific sequence and confers high binding affinity upon recognition. A protein without ZnF1 will have a higher affinity for a site that does not contain a ZnF1 specific sequence. This suggests auto-inhibition of binding by ZnF1 when a recognition site is not present. ZnF4 is important for stability of binding to most target sequences, especially when ZnF1 cannot bind. A site not recognized by ZnF1 will have very low affinity for a protein that does not include ZnF4. However, it is unclear whether ZnF4 makes specific contact with more than one base, and it is hypothesized that the majority of its binding affinity may be due to nonspecific interactions with the DNA backbone.

Future directions for this project would include testing of any binding motifs identified by PBM using the 6XHis tagged proteins. These can be interrogated by the same point mutation strategy to determine any unknown specific base contacts for ZnF4. Another priority goal for understanding the mechanism of Ikaros DNA binding would be to obtain 3D structures. All of the truncated proteins containing the different zinc finger combinations are 20 kd or less, and therefore are small enough for their structure to be readily determined in solution by NMR. In contrast to the intensive method of structural analysis by x-ray crystallography, NMR makes it possible to examine many structure variations with less investment. For studies of proteins such as Ikaros, that contain tandem DNA binding zinc finger domains, this method is ideal, as it provides the ability to look at the structural configuration of how each combination of zinc

fingers binds to different DNA target sequences. To pursue structural analysis by NMR, a

collaboration was established with the Feigon lab at UCLA, whom specialize in this field.

NMR structural analysis typically uses isotopically labeled proteins. These labeled proteins are

normally produced by growth in minimal media, with a 15N isotope as the only provided source

of nitrogen. This method labels nitrogen in the peptide bond for every amino acid of a protein.

Each residue except proline has an amide proton attached to this nitrogen, and the correlation

between the nitrogen and the amide produces a peak in the NMR spectra. This is observed by

15N HSQC (Heteronuclear Single Quantum Coherence), one of the most frequently used

experiments in protein NMR. Labeling with 15N is fairly inexpensive, and the HSQC spectrum

can be acquired in a relatively short time and used to screen for optimal conditions and

determine whether there are unstructured elements in the protein. These unstable regions

cause peaks to overlap, and therefore make it difficult to assign resonances to a particular

residue in the protein. HSQC can also detect protein-ligand binding interactions from the

resulting shifts in the resonance peaks. This is especially useful for studying the conformation of

a protein bound to DNA versus in a free state.

Following consultation with the Feigon lab, several modifications were made to the design of the

protein truncations and method of expression. New primers were designed to delete the 150

bp region of *Ikzf1* exon 4, upstream of ZnF1, and the 18 bp region of exon 6, downstream of

ZnF4. The aim of this strategy was to eliminate regions without stable secondary structure,

which cannot be resolved by NMR, and would interfere with the accuracy of the rest of the

structure. Protein fragments containing just ZnF2 or ZnF3 were also designed, with the goal of using structures determined for each of these zinc fingers alone as a starting point for modeling the final structures. Growth of these strains in minimal media (M9) was found to be significantly slower than in LB (Lysogeny Broth) medium, which has been optimized for bacterial growth (to reach an OD = 0.6 for induction took ~11 hrs versus 3.5 hrs, respectively) (Recipe for M9 media in Appendix I). Proteins were found to be largely insoluble, and it was also more difficult to obtain the high expression levels needed to purify an amount sufficient for NMR. Induction was tested at a range of temperatures (37°C, 32°C, 25°C, and 18°C), but none yielded significantly more soluble protein. After independently expressing a batch of the original constructs in LB and experiencing the same problem with insolubility, the cause was determined to be supplementation with $ZnSO_4$ instead of $ZnCl_2$ (as suggested by the Feigon lab).

If the NMR 3D structure project is to be continued, it will still be necessary to optimize expression and purification of the proteins on a larger scale that uses both GST affinity chromatography (with thrombin cleavage), followed by HPLC (High-performance liquid chromatography) size exclusion columns to remove contamination by thrombin (37 kd). The main purpose of this study would be to determine zinc finger conformations when bound to different DNA target sequences with and without ZnF1 and ZnF4 recognition sites. For this experimental setup, oligos must also be tested in a range of lengths to optimize NMR spectra. Oligos that are too long will have unstructured tails, while those that are too short can have unstable binding. An appropriate starting point would be two additional bases on either end of the recognition sequence.

Section 2: Investigation of T cell activation threshold in Ikaros mutant mice

Chapter 6: The unidentified link between Ikaros and T cell activation

The loss of Ikaros tumor suppressor function in heterozygous *Ikzf1*[null/+] and *Ikzf1*[DN/+] mice leads to clonal expansion and neoplastic transformation of T cells, causing development of leukemia and lymphoma in *Ikzf1*[DN/+] mice[21,22]. This coincides with a significantly lower threshold of T cell activation and proliferation[46]. *Ikzf1*[ΔF4/ΔF4] mice also develop aggressive T cell lymphoma, and it was therefore of great interest to determine whether these mice have a lower T cell activation threshold. If so, this would suggest that key target genes downstream of T cell receptor (TCR) signaling are deregulated by loss of ZnF4 and may play an important role in preventing malignant clonal expansion.

Splenic T cells can be activated *in vitro* with different concentrations of plate bound α-CD3, which targets the constant domain of the TCR, and α-CD28, which interacts with the coreceptor. Lower α-CD3/α-CD28 concentration provides fewer TCR engagements, and activation occurs only in cells with a lower threshold for response. Activation is also more difficult to initiate at lower cell numbers due to limited cell-cell contacts and access to paracrine factors, giving an advantage to cells with a lower threshold. A range of cell densities and α-CD3/α-CD28 concentrations can be analyzed to determine the threshold conditions that allow for activation. T cell activation leads to exponential proliferation, which can be quantified in several ways. One traditional method is the tritiated thymidine (thymidine linked to the tritium radioactive isotope, ^{3}H) incorporation assay. The ^{3}H-thymidine is incorporated into the newly formed DNA of replicating cells, and the subsequent increase in radioactivity can be measured in CPM. This

method was utilized in the study that showed decreased T cell activation threshold in *Ikzf1*[null/+]

and *Ikzf1*[DN/+] mice[45] (Figure 6a), and proliferation was measured over a range of both α-CD3 and

splenocyte concentrations.

Figure 6a: Decreased T cell activation threshold in *Ikzf1*[null/+] and *Ikzf1*[DN/+] mice[45]
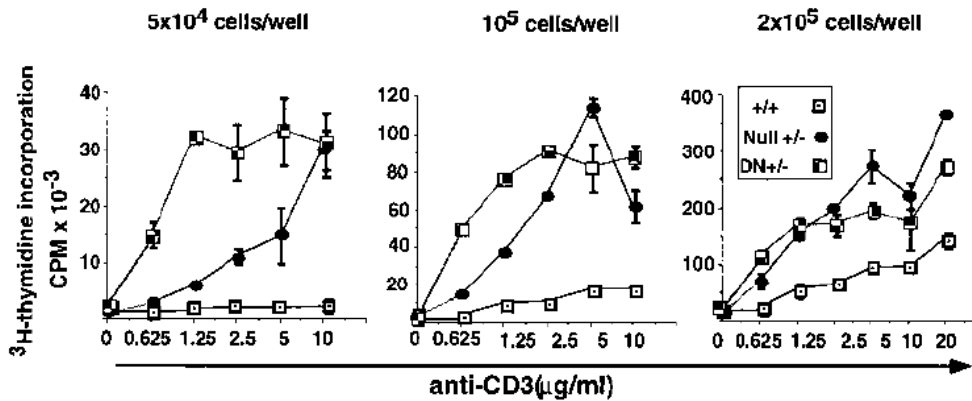
Negligible proliferation was seen for the wildtype splenocytes for both the lowest and

intermediate cell densities, at all α-CD3 concentrations. At the lowest cell density, only the

Ikzf1[DN/+] splenocytes showed any detectable proliferation for the lower concentrations of α-CD3.

However, at an intermediate cell density, the proliferation by *Ikzf1*[null/+] splenocytes was

comparable to those from *Ikzf1*[DN/+]. This data indicates a lowered threshold for T cell activation

in both the *Ikzf1*[DN/+] and *Ikzf1*[null/+] strains, but one that is more significant in *Ikzf1*[DN/+], coinciding

with the more severe lymphoma phenotype. It is important to note that since this study used

only plate bound α-CD3 for activation, these results can also be interpreted as a loss of

requirement for CD28 costimulation in these *Ikzf1* mutant strains. Data presented in this section

shows that *Ikzf1*[ΔF4/ΔF4] T cells have a lowered threshold of activation similar to that observed in

Ikzf1[null/+] and *Ikzf1*[DN/+].

Chapter 7: T cell proliferation and activation assays

In the years following initial characterization of the lowered T cell activation threshold in

$Ikzf1^{null/+}$ and $Ikzf1^{DN/+}$ mutants, using fluorescence activated cell sorting (FACS) technology has

become the prevalent method of analyzing proliferation. Working with a fluorescent marker is

preferable to radioactivity (^{3}H-thymidine), and FACS provides a more accurate quantification of

proliferative expansion. The most commonly used method for analyzing cell proliferation is

CFSE (carboxyfluorescein succinimidyl ester) staining. This technique uses a cell soluble

molecule, CFDA-SE (carboxyfluorescein diacetate succinimidyl ester), which is non-fluorescent

until it enters the cytoplasm, where esterases remove the acetate groups. This converts it to a

fluorescent ester, CFSE, whose succinimidyl ester group covalently binds to intracellular amines.

CFSE is therefore retained within cells due to the stability of the covalent interaction, and can be

used to monitor lymphocyte proliferation, since the fluorescent intensity is decreased by half

with each cell division. When done precisely, every division appears as a distinct peak, and this

method can be used to measure the fraction of cells that have undergone each number of cell

divisions.

Results using CFSE to observe divisions *in vitro* have shown proliferation of $Ikzf1^{\Delta F4/\Delta F4}$ T cells at

dramatically lower concentrations of α-CD3/28 than wildtype. For initial trial studies, whole

splenocytes were plated at a range of cell densities and α-CD3/28 concentrations to find optimal

conditions for activation in all cell types. Wildtype cells were found to proliferate well at a

density of 2 x 10^5 cells/well and an α-CD3/28 concentration of 1 µg/ml (for each), so this was

used as a starting point for titration of α-CD3/28. Shown in Figure 7a are representative CFSE

results from a decreasing threefold titration series (1, 0.3 and 0.1 µg/ml). Whole splenocytes

were stained with CFSE, plated at a concentration of 2×10^5 cells/well, and analyzed after

stimulation for either 48, 60, 72, or 90 hours (Complete splenocyte isolation and CFSE protocol

in Appendix I). CFSE fluorescence lies in the FITC channel and can bleed into the PE range, so

cells were stained with CD4 PerCP Cy5.5 and CD8 APC for FACS analysis.

Figure 7a: CFSE histograms of *in vitro* stimulated T cells from whole splenocytes.
Stimulated with either 0, 0.1, 0.3, or 1 µg/ml α-CD3/28.
Histograms shown for CD8+ at 60 hrs, and CD4+ at 72 hrs.
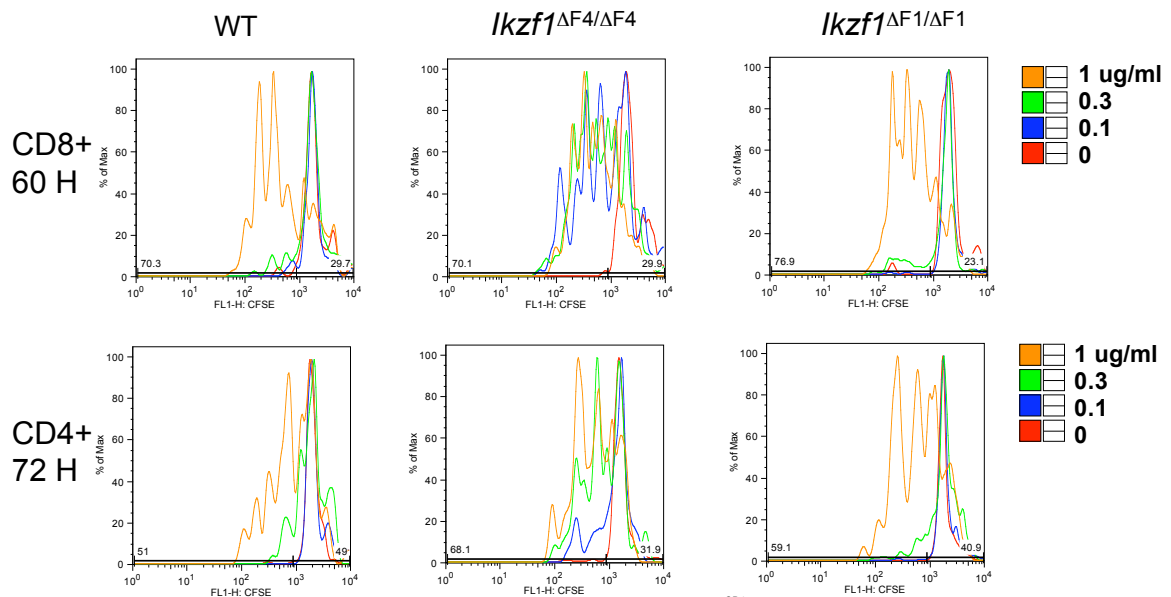
Minimal division was seen for any of the genotypes at 48 hours, and after 72 hours the cells

have divided enough times that the CFSE signal is too dilute to be detected by FACS

Approximately 5-6 divisions is the maximum that can typically be seen, and is true for the

histograms above. At the highest α-CD3/28 concentration (1 µg/ml, shown in orange), 5-6 lower

intensity CFSE peaks (decreasing to the left of the plot) can be seen for all of the genotypes. It is

clear from this data that the *Ikzf1*$^{\Delta F4/\Delta F4}$ T cells (both CD8+ and CD4+) are activated by lower

concentrations of α-CD3/28 than the other genotypes. Histograms show multiple divisions for

the *Ikzf1*$^{\Delta F4/\Delta F4}$ cells at 0.1 or 0.3 µg/ml (shown in blue and green, respectively), while the other

genotypes have negligible proliferation at either concentration. Additionally, the $Ikzf1^{\Delta F4/\Delta F4}$

CD8+ cells show a faster and more robust response to stimulation than CD4+. Repetitions of the

above experimental conditions confirmed these results, as shown by summary plots in Figure 7b

(5 replicates, standard deviation error bars).

Figure 7b: % CFSE low *in vitro* stimulated T cells from whole splenocytes.
Stimulated for either 60 or 72 hrs with a range from 0 - 1 µg/ml α-CD3/28.
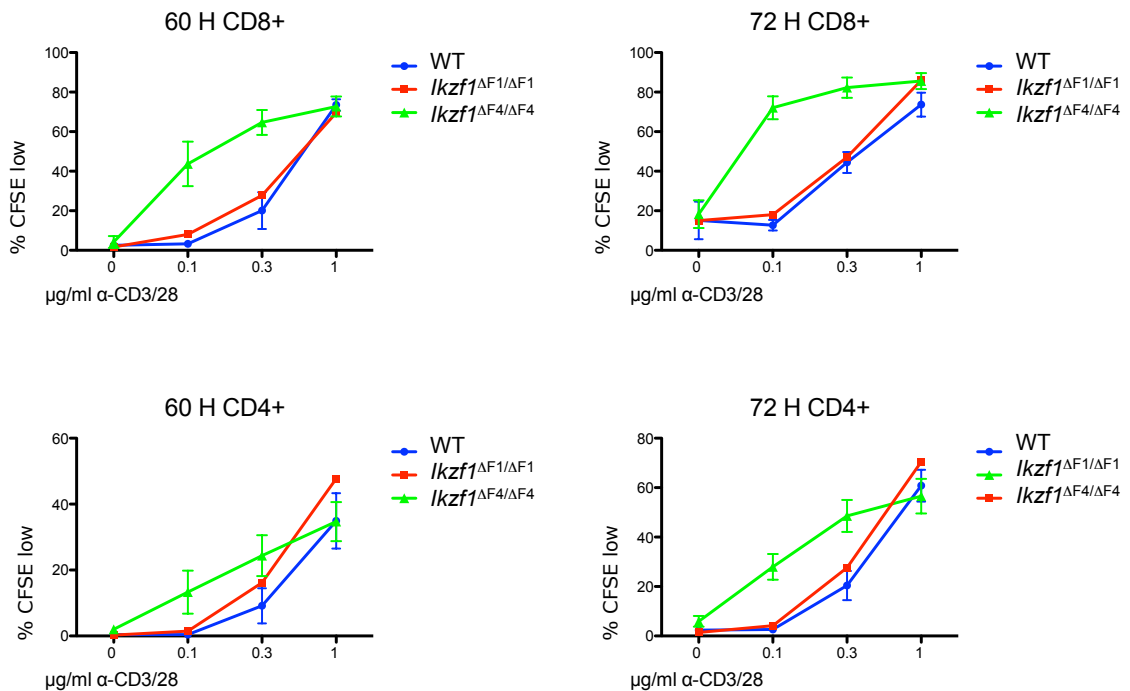
For the above summary graphs, histograms were bisected into two gated populations, CFSE high

(undivided), and CFSE low, which included any divided cell with an intensity lower than that for

unstimulated (0 hr).   The % CFSE low (divided) cells are in the same range for all genotypes at

the highest α-CD3/28 concentration (1 µg/ml), demonstrating that the conditions are consistent.

The level of $Ikzf1^{\Delta F4/\Delta F4}$ response to 0.3 µg/ml α-CD3/28 is similar to 1 µg/ml, while the other two

genotypes drop by half. Furthermore, consistent with the histograms in Figure 7a, $Ikzf1^{\Delta F4/\Delta F4}$

cells respond to the lowest concentration of α-CD3/28 (0.1 µg/ml), while the other genotypes do

not. This represents a tenfold decrease in activation threshold for $Ikzf1^{\Delta F4/\Delta F4}$ T cells,

approximately the same as was seen for $Ikzf1^{null/+}$ and $Ikzf1^{DN/+}$ mice[45] (Figure 6a).

For future analysis, the focus was directed to CD8+ T cells, which show a more dramatic

difference in proliferation for $Ikzf1^{\Delta F4/\Delta F4}$. This difference is also evident at an earlier time point,

which increases the likelihood of observing changes in earlier stages of activation. Activation of

T cells in response to TCR stimulation is characterized by upregulation of expression for cell

surface markers such as CD25 (IL-2 receptor) and CD69 (signal-transmitting receptor) in the first

few hours. This upregulation of these proteins can first be detected on the cell surface after

about 12 hours, and has reached a peak level by 18-20 hours. The conditions for Figure 7c are

the same as previously used for CFSE, whole splenocytes were plated at a density of 2 x 10^5

cells/well and stimulated with the a range of α-CD3/28 concentrations (0.1, 0.3 and 1 µg/ml).

Figure 7c: Upregulation of CD8+ T cell surface activation markers (CD25, CD69) from whole splenocytes.
Stimulated for 20 hrs *in vitro* with a range from 0.1 - 1 µg/ml α-CD3/28.
Density scatterplots shown only for 0.1 µg/ml α-CD3/28.

45
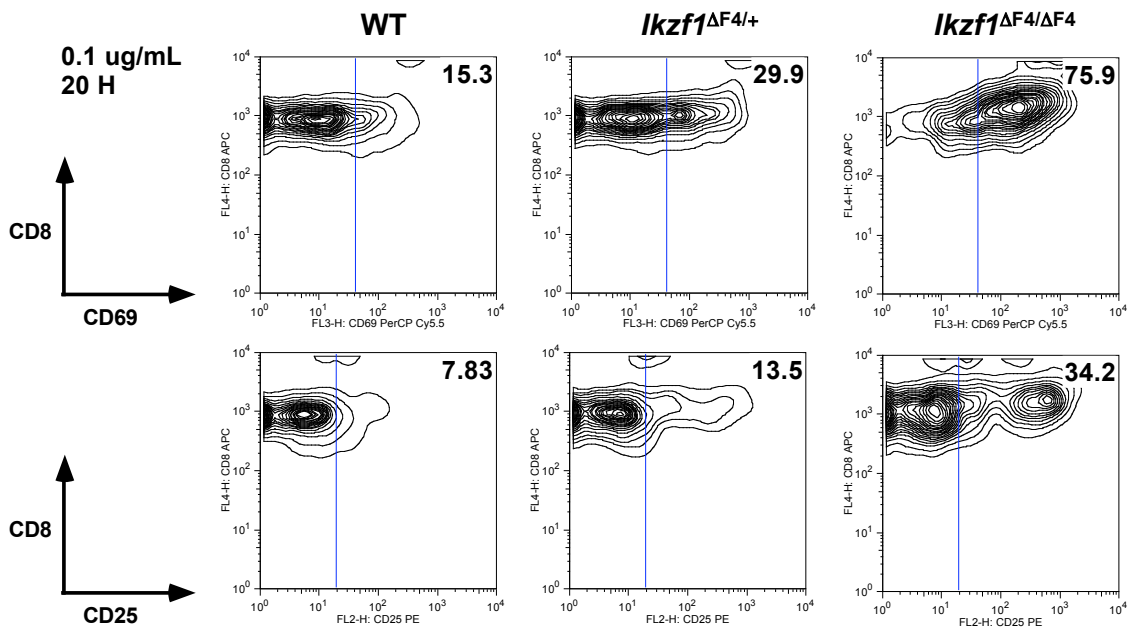
Figure 7c continued:
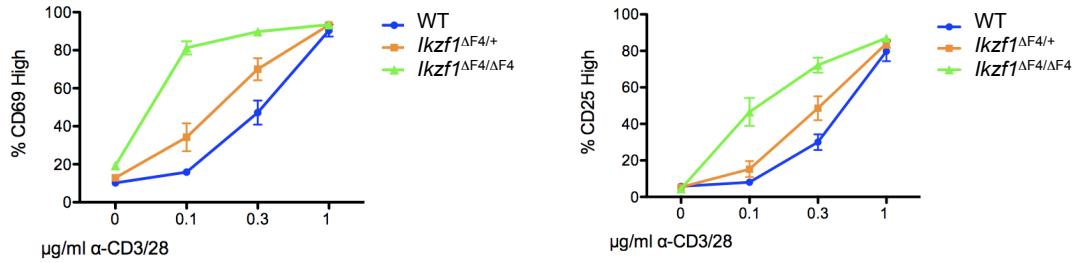
Figure 7c shows the upregulation of CD25 and CD69 for wildtype, $Ikzf1^{\Delta F4/+}$ and $Ikzf1^{\Delta F4/\Delta F4}$ CD8+

T cells. Representative density scatterplots are shown only for 0.1 µg/ml α-CD3/28 in order to

emphasize the greatest difference between the genotypes. Percentage of cells that are CD25 or

CD69 high is displayed at the upper right hand corner of the plots, and show 4-5 fold higher

upregulation for both markers in $Ikzf1^{\Delta F4/\Delta F4}$. Interestingly, upregulation appears to be

proportional to number of wildtype copies of Ikaros, as $Ikzf1^{\Delta F4/+}$ heterozygotes show an

intermediate level of activation for both markers. Summary graphs (5 replicates, standard

deviation error bars) show similar curves for upregulation as was seen for proliferation, namely

that only $Ikzf1^{\Delta F4/\Delta F4}$ responds to 0.1 µg/ml α-CD3/28, and the level of $Ikzf1^{\Delta F4/\Delta F4}$ response to 0.3

µg/ml is similar to that for 1 µg/ml, while the other two genotypes drop by half. All genotypes

have the same level of response to 1 µg/ml at 20 hours, and earlier time points (not shown)

confirmed that the kinetics of upregulation occur normally in the $Ikzf1^{\Delta F4/\Delta F4}$ and $Ikzf1^{\Delta F4/+}$ cells.

To determine if the lower activation threshold is T cell independent, or caused by changes in

other splenic paracrine factors, MACS was used to deplete other cell populations. The MACS

(magnetic activated cell sorting) technique uses streptavidin-coated magnetic beads to isolate

cells with biotin-conjugated antibodies bound to cell surface markers. Cells are first treated

with the desired cocktail of biotin-conjugated antibodies, then incubated with the streptavidin-

46

coated magnetic beads and run over a magnetic column to collect the cells recognized by any of

the antibodies. This method can be used to deplete one or more types of cells from a

population, and for the following experiments a cocktail was used to remove B cells,

erythrocytes, macrophages, and CD4+ T cells from whole splenocytes, leaving approximately 70-

80% pure CD8+ T cells (Supplementary Figure 7a) (Complete MACS protocol in Appendix I).

These MACS CD8+ T cell enriched wildtype and $Ikzf1^{\Delta F4/\Delta F4}$ populations were plated at a density

of 2 x 10^5 cells/well, an approximately 8 fold relative increase in numbers of CD8+ T cells, as

these represent about 10% of whole splenocytes. Cells were stimulated with the same range of

α-CD3/28 concentrations (0.1, 0.3 and 1 µg/ml) for 18 hours.

Figure 7d: Upregulation of cell surface activation markers (CD25, CD69) on CD8+ T cells, MACS enriched from whole splenocytes.
Stimulated for 18 hrs *in vitro* with a range from 0.1 - 1 µg/ml α-CD3/28.
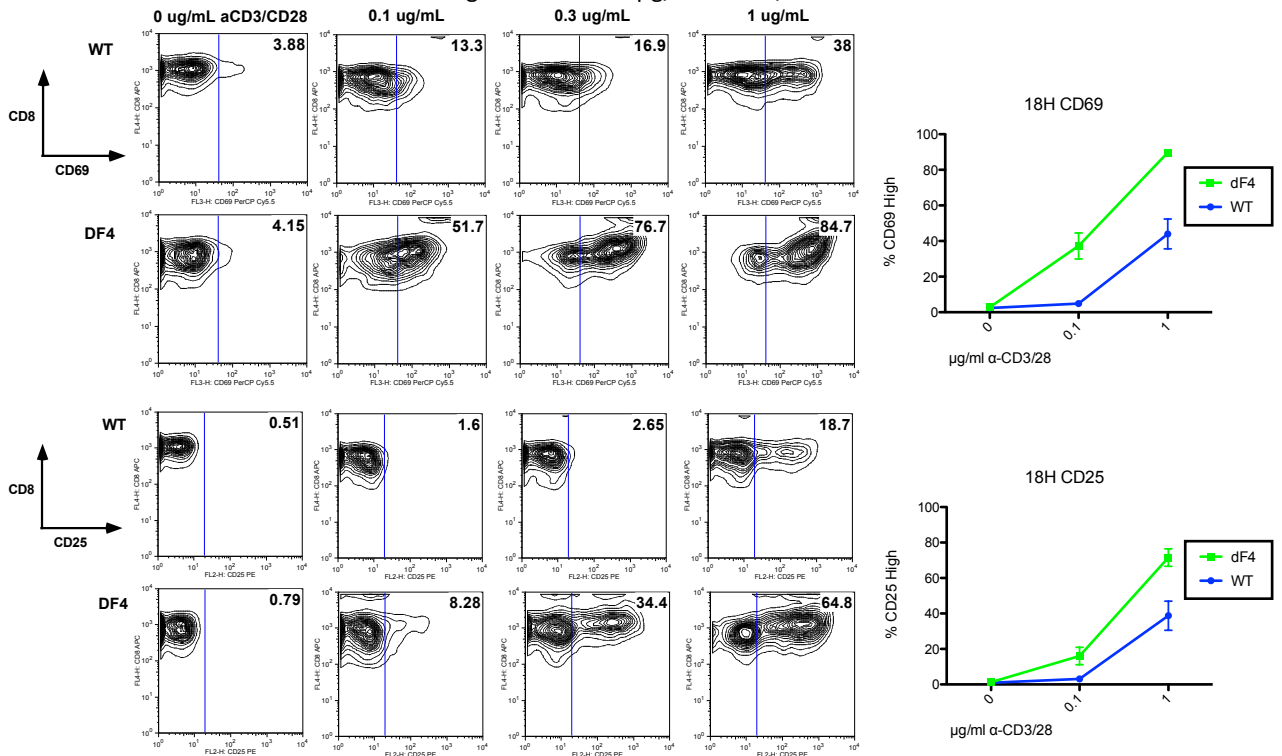
Summary graphs in Figure 7d (5 replicates, standard deviation error bars) show that both

wildtype and $Ikzf1^{\Delta F4/\Delta F4}$ cells were affected by the loss of supporting cell contacts and secreted

growth factors. When compared to whole splenocytes, activation dropped approximately 50% for MACS enriched wildtype CD8+ T cells at 1 µg/ml α-CD3/28. The same was true for *Ikzf1*$^{\Delta F4/\Delta F4}$ cells at 0.1 µg/ml α-CD3/28, but a similar level of activation was maintained for 1 µg/ml. This showed that the difference in activation threshold is maintained for MACS depleted populations, and is therefore T cell autonomous. However, cells from both genotypes need more highly concentrated stimulation by α-CD3/28 in the absence of other supporting cells and factors.

To eliminate the possibility that altered populations of mature and memory cells might be the source of cells with a lower activation threshold, naïve CD8$^+$ T cells were isolated by FACS sorting. Naïve T cells were sorted for the lineage profile CD8a$^+$Gr-1$^-$CD62L$^+$CD44Low, using a FACSAria II. The first sorting gate selected CD8a$^+$Gr-1$^-$ cells, the second isolated the fraction of this population that was CD62L$^+$CD44Low (Figure 7e).

Figure 7e: Representative staining showing FACS sorting gates for the naïve CD8 T cell lineage profile CD8a$^+$Gr-1$^-$CD62L$^+$CD44Low, and populations of naïve, mature and memory CD8 T cells
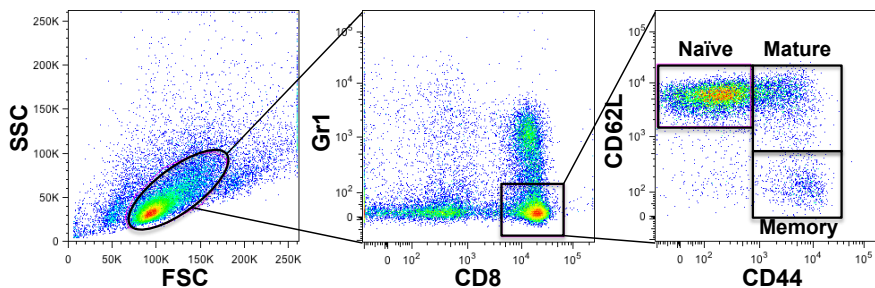
The Gr-1 antigen (or Ly-6g/Ly-6c) is a dendritic cell marker, used for sorting to distinguish T cells from CD8α^+ lymphoid dendritic cells. Gr-1 can also be expressed on memory T cells, another reason for selection of the Gr-1$^-$ population. L-selectin, or CD62-L, is a lymphnode "homing receptor" that is expressed on naïve and mature T cells, but not memory cells. CD44 is a marker for mature and effector memory T-cells (and is also expressed in the thymus). Figure 7e includes

48

a representative plot of CD44 versus CD62L, indicating the populations of naïve, mature, and memory cells. Memory cells are in the lower right quadrant (CD62⁻CD44high), mature cells in the upper right (CD62⁺CD44high), and naïve cells in the upper left (CD62⁺CD44low).

In vitro activation testing with FACS sorted naïve CD8 T cells (using the same conditions as previous experiments) found that antibody concentrations would need to be increased to obtain the same level of activation. As shown in Figure 7f, FACS sorted naïve CD8+ T cells were stimulated with increased concentrations of either 0.5 or 2 µg/ml α-CD3/28. FACS sorting removes all of the supporting cells, and a further decrease in activation threshold would be expected from MACS depletion. Cells may also have reduced responsiveness to activation due to the extended length of time for cell sorting, an additional 3 hours.

Figure 7f: Upregulation of cell surface activation markers (CD25, CD69) in FACS sorted naïve CD8 T cells (CD8+, Gr1-, CD62LHigh, CD44Low). Stimulated for 18 hrs *in vitro* with either 0.5 or 2 µg/ml α-CD3/28.
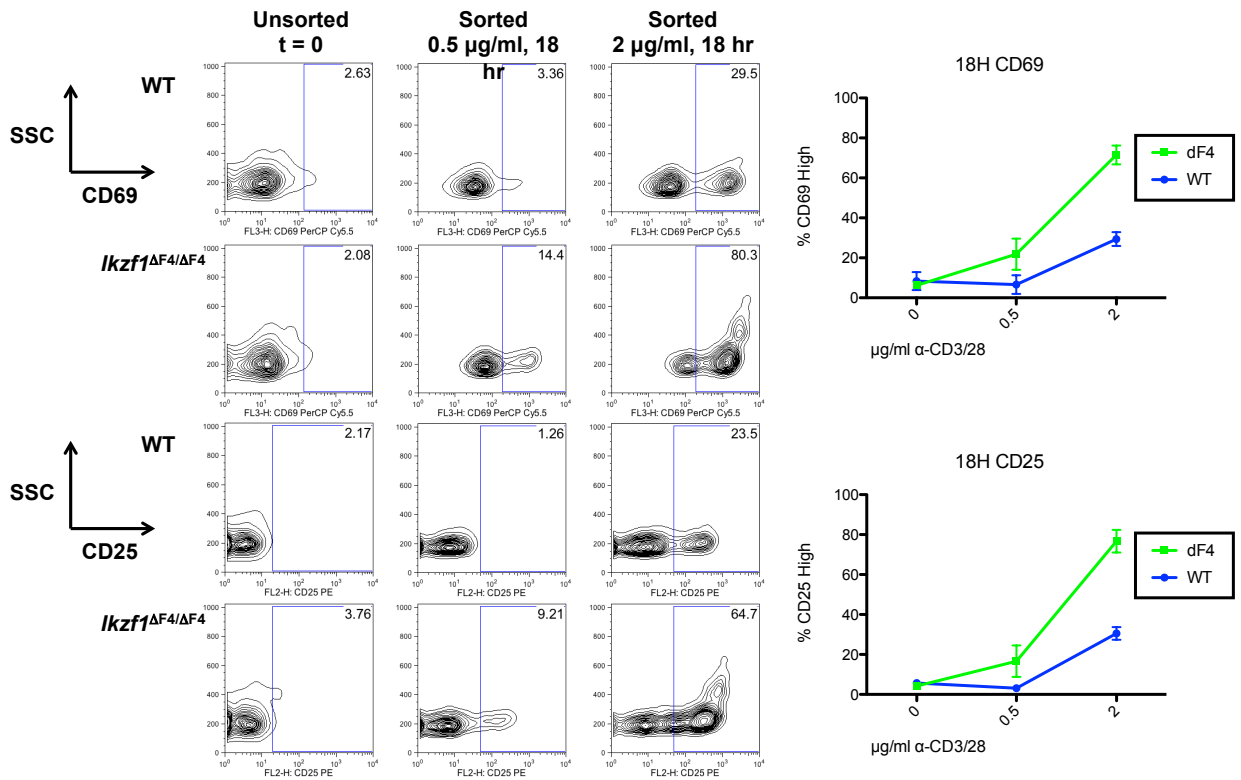
An intermediate concentration, 1 µg/ml α-CD3/28, was tested and used for stimulation in

samples for RNA-seq (as discussed in Chapter 8), as this can show the best difference in

activation between WT and *Ikzf1*$^{\Delta F4/\Delta F4}$ FACS sorted naïve CD8 T cells (Figure 7g). Scatterplots of

FSC versus SSC (forward scatter, side scatter) are also included to illustrate the difference in cell

morphology associated with activation seen in *Ikzf1*$^{\Delta F4/\Delta F4}$ but not WT cells. Volume and density

increases for "blasting cells" that have been activated and begin to initiate division in the first 36

hours, which can be seen as an increase in FSC and SSC (highlighted in Figure 7g).

Figure 7g: Example upregulation of cell surface activation markers (CD25, CD69) in FACS sorted naïve CD8 T cells (CD8+, Gr1-, CD62LHigh, CD44Low). Stimulated for 20 hrs in vitro with 1 µg/ml a-CD3/28. Samples represented were used for RNA-seq; protocol described in Chapter 8.
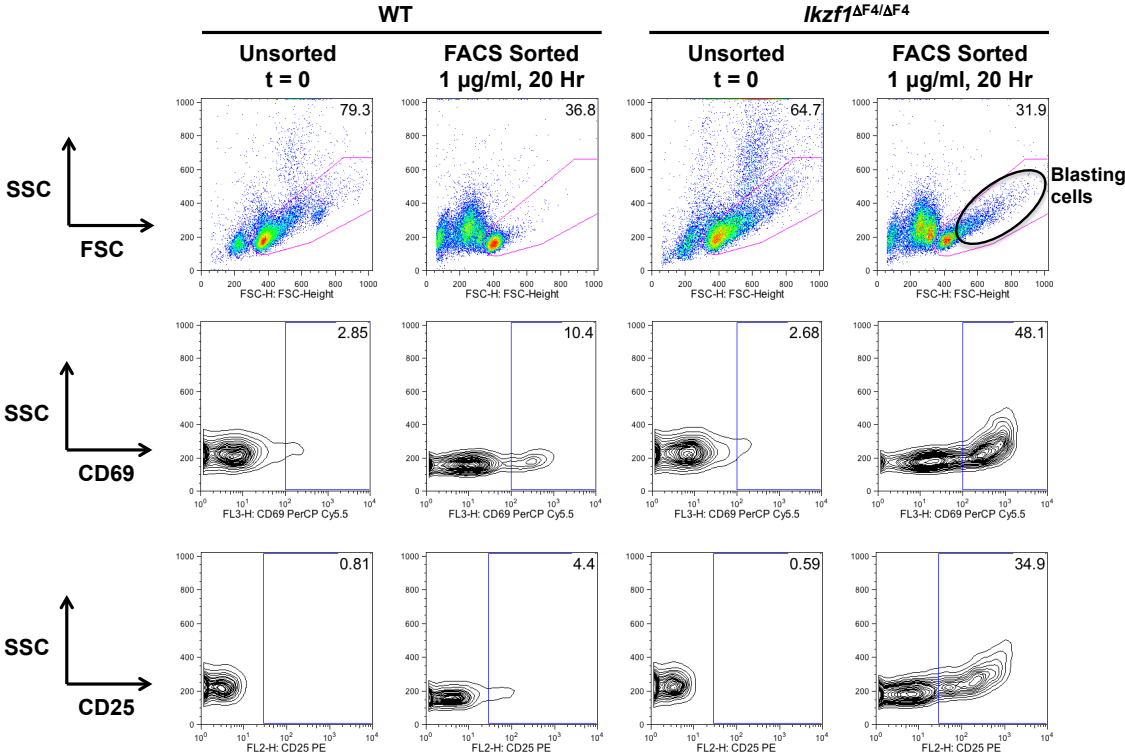
Chapter 8: High throughput RNA sequencing (RNA-seq)

Given the dramatic decrease in activation threshold seen in *Ikzf1*$^{\Delta F4/\Delta F4}$ T cells, the next logical

step was to investigate the changes in gene regulation associated with this phenotype. To this

end, high throughput RNA sequencing (RNA-seq) was pursued through the UCLA Broad Stem Cell

Research Center High Throughput Sequencing Core. This facility houses one Illumina HiSeq 2500

and three Illumina HiSeq 2000 sequencing instruments, all capable of producing up to 600 Gb of

data output in a single run. Each run generates 3 billion single reads (6 billion paired end) with

80% reaching a maximum length of 100 bp each. To utilize these instruments, cDNA sequencing

libraries were prepared with the Illumina True-Seq system. An overview of this method is shown

in Figure 8a (For full protocol, see Appendix I).

Figure 8a: Illumina Tru-Seq RNA-seq library preparation

The Illumina True-Seq RNA-seq library preparation method starts with an input of 100ng-1ug of

Total RNA, which is poly-A selected in the first step with magnetic beads to isolate the

messenger RNA (mRNA) fraction. RNA is primed with random oligonucleotides and the first

complementary DNA (cDNA) strand is synthesized. Second strand synthesis (step 4 of Figure 8a)

was modified to allow for strand-specific sequencing, which produces a library with only the

antisense strand. Because transcription can occur in either direction, knowing which DNA strand

the RNA molecule originates from helps to resolve ambiguities for genes whose exons overlap

on opposite strands, and to more accurately measure the expression levels of a given transcript.

To achieve strand-specific sequencing, dUTP/dVTP is incorporated in second strand synthesis.

The double stranded cDNA is then further processed following the standard steps 5 and 6 from

Figure 8a: end repair, A-tailing, and adapter ligation. Prior to amplification (step 7), the dUTP

strand is selectively degraded by treatment with USER (Uracil-Specific Excision Reagent) Enzyme.

To minimize cost, samples were multiplexed, ie. combined in a single lane of sequencing. This

method is possible by labeling each sample with an adapter containing a specific sequence, or

barcode, that can then be used to separate sequencing data corresponding to each sample.

As described in Chapter 7, naïve CD8+ T cells were FACS sorted for the lineage profile CD8a$^+$Gr-1$^-$

CD62L$^+$CD44Low. On average, only about 1 x 10^6 naïve CD8$^+$ T cells can be isolated from a single

spleen by FACS sorting. This presents complications for an *in vitro* activation time course when

needing to isolate a minimum of 100 ng Total RNA for Illumina sequencing library preparation.

Mice with a ΔF4 mutation in *Ikzf1* cannot be successfully bred as homozygotes due to early

onset of lymphoma, and from heterozygote breeding, less than one in four will be *Ikzf1*$^{ΔF4/ΔF4}$

(lower than 25% due to a fraction of embryonic death). It is therefore rare to obtain multiple

gender matched *Ikzf1*$^{ΔF4/ΔF4}$ mice from the same litter, and a large number of breeding cages are

needed to have age matched litters. The most that can be predicted from a colony with ~15

breeding pairs is two gender and age matched ΔF4 mice every 2-3 weeks. Because of this, it was

necessary to optimize RNA isolation from the lowest number of cells possible. Cells were lysed

using Trizol reagent, the aqueous phase isolated, and initially purified using either Qiagen

RNeasy Minelute or Micro columns. This method was insufficient to yield enough RNA to make

a sequencing library for more than one time point, and so phenol/chloroform extraction was

used for all future samples. Optimization using phenol/chloroform extraction determined that

at least 5×10^5 cells are needed to obtain approximately 100 ng of Total RNA, the minimum

needed for Illumina sequencing library preparation. Since 2×10^5 cells/well are needed for

consistent activation, 6×10^5 cells were used for each sample. This meant that a maximum of

three time points could be obtained from two mice, which yield a total of approximately 2×10^6

naïve CD8$^+$ T cells. Due to this restriction, cells for RNA isolation were stimulated with a

concentration of 1 μg/ml α-CD3/28, as can show the best difference in activation for FACS

sorted naïve cells. Cells were collected for RNA isolation without stimulation (t = 0), at 30 min

and 2 hours. These time points were chosen to capture upregulation of early activation

response genes, whose expression does not require chromatin remodeling or new protein

translation, and late response genes, the majority of which require one or both of these events.

Sequence output from the majority of current NGS (next-generation sequencing) machines is

formatted as a QSEQ file. QSEQ is a plain-text file format for sequence reads, composed of one

tabular line in the following format: Machine name, Run number, Lane number, Tile number, X

coordinate, Y coordinate, Index, Read Number, Sequence, Quality, Filter. To begin analyzing the

sequencing data, this file must first be processed by a series of UNIX platform command-line

functions that remove the unnecessary columns and group sequences by barcodes

corresponding to different samples. (A summary workflow for formatting and analysis is shown

in Figure 8b, for full protocol, see Appendix I.)

Figure 8b: RNA-seq formatting and analysis workflow
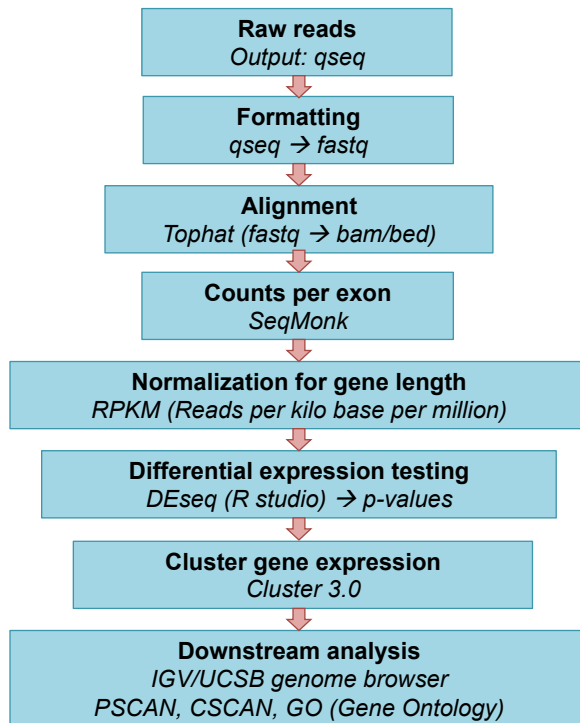
Sequencing data is next converted from a QSEQ to FASTQ format. FASTQ is a text-based format

for combining a FASTA sequence and its corresponding quality data that typically uses four lines

per sequence (Line 1: @Sequence identifier, 2: raw sequence, 3: +Sequence identifier, 4: quality

scores). This was performed using a pre-existing workflow program on the Galaxy server,

http://galaxy.hoffman2.idre.ucla.edu/, a collaborative resource established at UCLA to meet the

demand for streamlined NGS analysis needed by an increasing number of laboratories.   The

FASTQ format can then be used to align the sequencing data with a reference genome, which

provides reference seqences for gene annotation. This experimental data was aligned to the

mouse (*mus musculus)* mm9 genome (build NCBI37) using the program Tophat, also provided on

the Galaxy server. Tophat is a version of the standard Bowtie alignment software that has been

adapted to identify splice junctions, an important function for analyzing RNA-seq data. Tophat

generates two output files, BAM and BED. The primary file for further analysis is the list of

accepted read alignments in BAM format. The second file, which contains a BED graph track of

reported junctions can be visualized on a genome browser such as UCSC or IGV (Integrative

Genome Viewer). The BAM file is then used to count the number of reads for every gene exon.

This was performed in the program SeqMonk with the same reference build NCBI37. SeqMonk

generates a probe report containing the number of reads for every gene exon, which can be

used to sum up both the number of reads for each gene and the gene lengths (using UNIX

command line). The resulting text file is opened in Excel, where each of the gene annotation

numbers can be automatically looked up and reported using a corresponding file containing

associated gene names. To continue with comparison of gene expression, it is necessary to

calculate RPKM (Reads per kilo base per million). RPKM= $(10^9 \times C)/(N \times L)$, where C= number of

mapped reads for a given gene, N=number of total mapped reads in that sample, and L= length

of the gene in bp (exons only, ie. cDNA length). This RPKM calculation normalizes sequencing

data to remove bias for longer genes, which are more likely to produce sequencing reads,

therefore resulting in higher counts. Once RPKM has been determined, this value can be used

for direct comparison of expression levels between genes. One way to visualize these

differences in expression levels is to use a clustering software such as Cluster 3.0, which groups

genes based on similar expression patterns within a set of samples. Groups of differentially

expressed genes can then be analyzed for enrichment of specific transcription factor binding

sites in promoters for genes using PSCAN, a database of known transcription factor binding

motifs. A similar program, CSCAN, utilizes a server including an extensive collection of genome-

wide ChIP-Seq experiments performed on transcription factors, histone modifications, RNA

polymerases, etc. PSCAN and CSCAN help identify which transcription factors are likely to be

common regulators of a group of genes with a similar expression profile. Another useful

characterization for understanding commonalities between a group of coexpressed genes is

Gene Ontology (GO). GO analysis programs cross-reference annotated gene functions in public

databases and find similar biological mechanisms within a group of genes. The DAVID

Bioinformatics Database generated by the NIH was used for analysis of these sequencing data.

To reiterate, FACS sorted naïve CD8+ T cells for this RNA-seq study were stimulated with a

concentration of 0.5 µg/ml α-CD3/28, as this shows a difference in activation between WT and

ΔF4. Cells from one experiment were collected for RNA isolation without stimulation (t = 0), at

30 min and 2 hours. For a second experiment with lower cell yield, cells were collected at just t =

0 and 30 min. The first step in downstream analysis of this data set was to use Cluster 3 for

broad characterization of groups of differentially expressed genes among the set of all 14,555

genes with RPKM>1. A clustering heatmap of z-scores for expression is shown in Figure 8c, with

a green-yellow-red coloring scheme where green = bottom 10 percentile, yellow = mean, and

red = top 10 percentile. Groups 2 and 4 have been removed from the figure as they showed no

significant temporal change or differences between genotypes. Groups 1,6, and 10 exhibit

similar patterns of temporal expression change for both ΔF4 and WT samples. Group 3 clearly

shows higher expression in ΔF4 samples at t=0 to 30m, and Group 9 for all time points. Group 5

shows higher expression in WT samples at t=0 to 30m. These three groups (3, 5 and 9) were

further analyzed for potential common transcriptional regulators (PSCAN, CSCAN) and known

biological functions (gene ontology).

Figure 8c: Cluster heatmap for 14,555 genes with RPKM>1
Group 3 shows higher expression in ΔF4 at t=0 to 30m, Group 9 for all
Group 5 shows higher expression in WT at t=0 to 30m
(Groups 2 and 4 removed from figure: no significant temporal change or genotype differences)
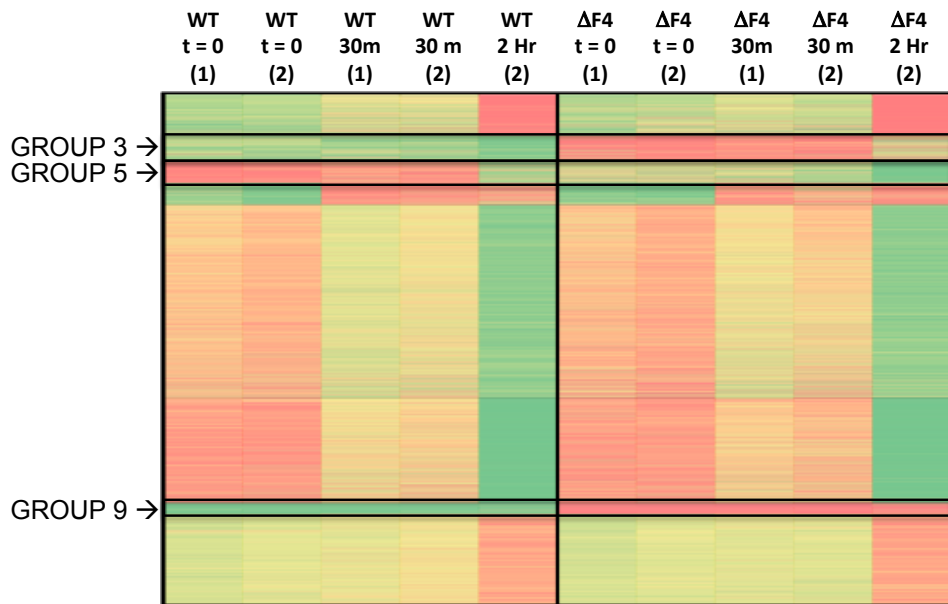
Results from all ΔF4 and WT high expression clusters shown in this chapter were compared to

narrow the field of analysis. Many transcription factors were enriched in both ΔF4 and WT high

clusters. These were regarded as generally associated with activation, because WT does show

some activation at the stimulation conditions used for these experiments. Some of the results

with weaker p-values are conflicting between groups of the same genotype or have functions

counterintuitive to the observed phenotype. All such occurrences were eliminated from further

analysis. Transcription factors of interest are those that are uniquely enriched for ΔF4 high

expression clusters, and are therefore not merely associated with typical levels of activation.

Another important group of factors are those that have an enrichment of target genes that are

more highly expressed in WT, indicating aberrant repression or loss of activation in ΔF4. For

relative brevity, those transcription factors identified from the groups in Figure 8c that also

reappear later on in more stringent analyses will be highlighted for discussion. These factors

represent the strongest putative candidates for a role in decreased activation threshold and loss

of tumor suppression function in ΔF4 mice.

Figure 8d: PSCAN and CSCAN results for over/under-represented TF binding sites (p-value <0.05) in promoters and *cis*-regulatory regions for genes in Groups 3, 5, and 9 from Figure 8a. (CSCAN replicates removed). Group 3 (335 genes, ΔF4 High, early), Group 9 (147 genes, ΔF4 High, All), Group 5 (271 genes, WT High, early)

PSCAN

Group 3 (ΔF4 High, early)

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
PLAG1	4.76	8.34E-07
Zfp423	4.64	1.51E-06
MZF1_1-4	4.34	6.66E-06
TLX1::NFIC	4.30	7.25E-06
NFKB1	4.28	8.15E-06
SP1	4.25	9.83E-06
TFAP2A	3.80	6.84E-05
Spz1	3.18	6.88E-04
INSM1	3.18	6.88E-04
PPARG::RXRA	3.16	7.14E-04
RELA	3.15	7.53E-04
Klf4	3.06	1.04E-03
CTCF	3.03	1.15E-03
Hand1::Tcfe2a	2.85	2.07E-03
Myf	2.75	2.87E-03
HNF4A	2.71	3.22E-03
ZEB1	2.70	3.32E-03
NF-kappaB	2.57	4.89E-03
Tcfcp2l1	2.57	4.89E-03
MZF1_5-13	2.57	4.92E-03
EWSR1-FLI1	2.51	5.82E-03
Mycn	2.49	6.15E-03
Myc	2.39	8.04E-03
Zfx	2.27	1.14E-02
REST	2.26	1.14E-02
ESR1	2.22	1.27E-02
EBF1	2.18	1.41E-02
Pax5	2.10	1.74E-02

Group 9 (ΔF4 High, All)

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
RELA	4.53	2.77E-06
Tcfcp2l1	4.09	2.03E-05
EWSR1-FLI1	3.03	1.18E-03
MZF1_1-4	2.99	1.37E-03
NF-kappaB	2.87	2.00E-03
NFKB1	2.78	2.69E-03
REL	1.94	2.55E-02
MZF1_5-13	1.78	3.68E-02
NFATC2	1.71	4.29E-02

Group 5 (WT High, early)

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
PLAG1	5.11	1.42E-07
INSM1	4.28	8.49E-06
SP1	4.25	1.01E-05
Egr1	3.98	3.20E-05
RREB1	3.69	1.05E-04
Klf4	3.64	1.28E-04
TFAP2A	3.04	1.16E-03
Pax5	2.75	2.88E-03
Zfx	2.42	7.63E-03
MZF1_5-13	2.15	1.53E-02
Myf	2.14	1.57E-02
ESR1	2.11	1.69E-02
NFKB1	2.10	1.73E-02
MZF1_1-4	1.99	2.28E-02
CTCF	1.94	2.56E-02
ZEB1	1.76	3.87E-02

CSCAN

Group 5 (WT High, early)

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 271 ENTRIES	OVER/ UNDER-REPRESENTED
BHLHE40_(NB100-1800)	CH12	115/271	9.77E-05	79	OVER
COREST_(sc-30189)	MEL	21/271	2.05E-03	8	OVER
MAZ_(ab85725)	CH12	126/271	6.83E-03	96	OVER
p300_(SC-584)	MEL	14/271	1.04E-02	5	OVER
n-Myc	mESC	16/265	2.43E-02	32	UNDER

Group 3 (ΔF4 High, early)

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 335 ENTRIES	OVER/ UNDER-REPRESENTED
HCFC1_(NB100-68209)	MEL	51/335	1.12E-08	101	UNDER
Nrf2	CH12	12/335	5.27E-08	45	UNDER
p300_(SC-584)	CH12	54/335	3.71E-07	24	OVER
Smc1a	mESC	87/330	3.47E-06	51	OVER
GCN5	CH12	64/335	2.17E-05	105	UNDER
FLI1_(sc-356)	Megakaryo	10/335	6.54E-04	31	UNDER
NRSF	C2C12	32/335	3.50E-03	58	UNDER
SMC3_(ab9263)	CH12	69/335	3.69E-03	44	OVER
NELFe	MEL	83/335	1.15E-02	115	UNDER
Rad21	CH12	66/335	2.21E-02	44	OVER

Group 9 (ΔF4 High, All)

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 147 ENTRIES	OVER/ UNDER-REPRESENTED
HCFC1_(NB100-68209)	MEL	10/147	2.45E-10	45	UNDER
GCN5	CH12	12/147	1.01E-09	46	UNDER
ZNF-MIZD-CP1	CH12	13/147	3.93E-07	42	UNDER
Mxi1_(AF4185)	MEL	16/147	9.04E-07	45	UNDER
TBP	CH12	19/147	1.23E-06	49	UNDER
NELFe	MEL	21/147	4.04E-06	50	UNDER
IRF4	NFS-201	16/147	4.38E-06	44	UNDER
Nrf2	CH12	2/147	1.42E-05	20	UNDER
TBP	MEL	25/147	1.94E-05	54	UNDER
Max	CH12	27/147	2.16E-05	56	UNDER
c-Myc	CH12	13/147	8.06E-05	36	UNDER
E2F4	CH12	1/147	5.47E-04	14	UNDER
SIN3A_(NB600-1263)	MEL	25/147	7.33E-04	49	UNDER
Suz12	mESC	22/147	9.10E-04	9	OVER
FLI1_(sc-356)	Megakaryo	1/147	1.19E-03	13	UNDER
Rad21	MEL	18/147	8.73E-03	7	OVER

From PSCAN analysis, just one transcription factor was identified in more than one WT high

group but none of the ΔF4, and can therefore be said to show "high WT specific" promoter

enrichment. This factor is EGR1, or Early Growth Response 1, which is seen to be enriched in the

promoters of genes which are more highly expressed in WT at t=0 and 30m. This means that

these EGR1 target genes are expressed at lower than normal levels in the ΔF4 mutant at t=0 and

30m, suggesting a priming for activation, as these genes are downregulated in WT at 2 hrs.

Evidence suggests that EGR1 is a tumor suppressor, which logically fits with loss of function in

the ΔF4 mutant.

Several transcription factors were identified from PSCAN in more than one group with ΔF4 high

expression, but in neither of the WT groups, and can therefore be described as having "high ΔF4

specific" promoter enrichment. These include EWSR1-FLI1, TLX1::NFIC, Hand1::Tcfe2a, RELA/

NF-kappaB, Tcfcp2l1, Zfp423, and EBF1. Strikingly, three of these factors have known

association with acute lymphoblastic leukemia (ALL), and are therefore strong candidates for a

primary role in driving malignant transformation in ΔF4 T cells: EWSR1-FLI1, TLX1::NFIC, and

Hand1::Tcfe2a. One of the most significantly enriched factors is EWSR1-FLI1, Ewing sarcoma

breakpoint region 1-Friend leukaemia virus integration 1. EWSR1-FLI1 is a t(11;22)(q24;q12)

translocation of the FLI1 proto-oncogene, an ETS transcription factor, with EWSR1, a potent

transcriptional activator. This translocation is observed in the majority of Ewing sarcoma cases,

as well as B-cell lymphoma[47]. Another FLI1 proto-oncogene t(4;11)(q21;q23) translocation has

been associated with both ALL and AML (acute myeloid leukemia), and is present in

approximately 10% of ALL patients[48]. It is likely that these target genes can also be targeted by

FLI1 alone, as abnormal expression is a predictor of poor prognosis in AML patients[49]. However,

it is also probable that one of these translocations could also occur in ΔF4 due to aberrant

activity of the VJD recombinase, since the sequences at sites of breakage in the t(4;11)(q21;q23)

translocation suggest that the VDJ recombinase is directly involved[47]. Another highly enriched

factor is TLX1::NFIC, a heterodimer between TLX1 and NFIC. TLX1, or T-cell leukemia homeobox

1, is known to have ectopic expression associated with T-cell ALL. The third factor associated

with ALL is Hand1::Tcfe2a, a heterodimer between Hand1 and Tcfe2a (transcription factor E2a),

also known as TCF3. Translocations of TCF3 have been associated with pre-B-cell ALL (t(1;19),

childhood leukemia (t(19;19)), and acute leukemia (t(12;19)). It has also been shown to directly

enhance Hes1 expression, which is increased in ΔF4, as shown later in this chapter and

previously in Chapter 3. Hes1 is a Notch1 target gene, and BCR-ABL, a constitutively expressed

mutation of Notch, is seen to have a synergistic effect with the $Ikzf1^{\Delta F4/\Delta F4}$ mutation in malignant

transformation of thymic lymphomas.

Other factors were also identified with oncogenic properties and/or relating to the decrease in

activation threshold: RELA, Tcfcp2l1, Zfp423, and EBF1. There is a general enrichment observed

for NF-kappaB family members in ΔF4 high groups, most specifically RELA (v-rel avian

reticuloendotheliosis viral oncogene homolog A), a component of the most prevalent NFkB

dimer and crucial to immune cell activation. One potentially oncogenic factor is Tfcp2l1,

transcription factor CP2-like 1. TFCP2 is induced by the oncogene AEG-1 and associated with

hepatocellular carcinoma. Two factors which are known to interact, and likely to work in

tandem are Zfp423 (Zinc finger protein 423) and EBF1 (early B-cell factor 1). These show the

same patterns of enrichment in these samples, and have multiple roles in signal transduction

during development, which may affect maturation of activated T cells.

Potential contributors to the loss of tumor suppression are also identified by CSCAN as

"underrepresented" for groups highly expressed in ΔF4. These include transcription factors

IRF4, Mxi1 and SIN3A. IRF4, or Interferon Regulatory Factor 4, is a lymphocyte specific IRF which

negatively regulates Toll-like-receptor (TLR) signaling that is central to activation. This role fits

with an underrepresentation of IRF4 targets among genes that are highly expressed in ΔF4.

Mxi1 (MAX Interactor 1, Dimerization Protein) is a transcriptional repressor reported to

negatively regulate MYC oncogenic activity by competing for MAX, a protein that binds to MYC

and is required for its function. SIN3A (SIN3 Transcription Regulator Family Member A) interacts

with MXI1 to repress MYC responsive genes and shows the same underrepresented CSCAN

pattern as Mxi1.

Gene ontology analysis was performed on the groups 3, 5 and 9 from Figure 8c, and the output

is summarized in Supplementary Figure 8a. Genes highly expressed in group 3 (ΔF4 high, early)

showed a unique enrichment for functions in cytokine production and hematopoiesis. Cytokine

growth factors are essential for activation, and genes involved in hematopoetic cell

development are likely to play a role in activation of mature markers.   Genes highly expressed in

group 9 (ΔF4 high, all time points) had an unusual enrichment of cadherin and the membrane

fraction in general. This increase in cell adhesion transmembrane proteins involved in signaling

can be interpreted as an enhancement for signal transduction, which logically corresponds to a

lower number of TCR engagements necessary for activation. Both groups 3 and 9 had a

significant enrichment for expected biological functions in regulation of apoptosis and T cell

activation (genes are listed in Supplementary Figure 8a). Two examples that stand out from the

regulation of apoptosis category in group 3 are BCL2-like2 and c-Kit. BCL2-like2 (B-Cell CLL/Lymphoma 2) is a mitochondrial membrane protein that suppresses apoptosis in lymphocytes by controlling the mitochondrial membrane permeability and inhibiting caspase activity. Constitutive expression of BCL2, as is seen in the translocation of BCL2 to the Ig heavy chain locus, is associated with B-cell lymphoma. The c-kit oncogene is a tyrosine-protein kinase that acts as cell-surface receptor for the cytokine SCF (stem cell factor), and plays an essential role in the regulation of cell survival and proliferation. Signaling triggered by SCF through c-Kit activates a multitude of pathways: AKT1, RAS, RAF1, MAP kinases (MAPK1/ERK2, MAPK3/ERK1), STAT family members STAT1, STAT3, STAT5A and STAT5B.

Since the most striking expression differences between WT and ΔF4 were observed to be present already in resting cells without stimulation, a more thorough focused analysis was conducted on the differential expression of genes at t=0. This was performed using the DESeq software package, which provides statistical analysis of RNA-seq differential expression based on the free open-source R language and Bioconductor programs[50]. The DEseq method tests for differential expression using negative binonial distribution and an estimator for the distribution's variance. This works from the assumption that a gene will have an average number of reads relative to sequencing depth for a given sample, and calculates statistical probability (p-value) for the observed deviance from this average. R/Bioconductor command lines used for experimental replicates are included in Appendix I.

Using the p-values generated by DEseq for the comparison of the two replicate WT and ΔF4 samples at t=0, a more stringent downstream analysis was conducted. Parameters were set for RPKM>1, DEseq p-value<0.05, and greater than 50 fold higher expression in ΔF4 versus WT at t=0 for both data sets. Only 25 gene promoters were identified with >50 fold higher expression in ΔF4 versus WT at t=0. Due to this small sample size, low p-values for PSCAN output were unexpected, but still observed for EWSR1-FLI1, Zfp423, TLX1::NFIC, Tcfcp2l1, all of which were previously discussed in the category of "high ΔF4 specific" enrichment.

Figure 8e: PSCAN Z-score heatmap for overrepresented TF binding sites (p-value <0.05) within promoters of individual genes with >50 fold higher expression in ΔF4 versus WT at t=0 for both data sets (DESeq p-value <0.05 for duplicates)

TF NAME	EWSR1-FLI1	Zfp423	MZF1_5-13	Znf143	TLX1::NFIC	Tcfcp2l1	SPI1	E2F1
Z-SCORE	4.33	3.43	3.38	3.29	2.53	2.44	2.16	1.89
P-VALUE	7.20E-06	3.03E-04	3.57E-04	4.91E-04	5.69E-03	7.39E-03	1.54E-02	2.95E-02
INDIVIDUAL GENE PROMOTER Z-SCORES:								
Ctnnd1	3.31	1.67	-1.17	3.57	2.62	0.02	1.85	0.19
Entpd1	3.14	0.38	1.29	1.30	0.91	-0.67	-0.23	-1.16
Kcnip3	3.00	1.89	1.72	1.20	0.58	-0.72	0.05	-0.01
Nrp1	2.31	-0.78	1.88	0.56	-0.48	0.19	0.12	1.21
Trim16	2.29	-0.10	1.88	-0.37	-0.46	0.42	-0.23	0.06
Megf10	2.12	-0.12	0.96	-0.50	-0.22	2.13	0.12	1.09
Tmem204	1.67	-0.53	0.74	0.67	1.72	-0.97	0.42	-0.08
Ctnnd1	1.65	1.67	1.29	3.57	2.62	0.02	1.85	0.19
Cacnb3	1.60	0.02	1.88	1.15	0.68	1.28	0.53	2.24
Pdzrn3	1.20	2.54	-1.09	-0.16	0.07	0.17	1.85	0.85
Obsl1	1.08	0.25	0.38	-0.85	-0.31	2.07	0.42	0.95
Cd163l1	0.75	1.30	0.21	-0.45	-0.51	-0.05	0.12	0.65
Cacnb3	0.68	0.76	1.20	-0.07	-1.04	0.69	0.02	1.21
Tmem176b	0.55	3.17	2.36	1.52	0.84	0.92	1.85	-0.01
Tmem176b	0.55	3.17	0.38	1.52	-0.21	1.58	-1.31	-0.01
Evc	-0.17	2.23	0.29	-0.14	-0.29	0.08	0.14	2.24
Dlg5	-0.20	-0.16	0.34	-0.16	0.79	0.97	-0.85	-0.84
Kcnip3	-0.25	-0.34	0.29	1.87	1.56	2.08	0.05	-0.01
Tmem176b	-0.28	1.67	2.36	1.52	0.84	3.02	1.85	-0.01
Gzmm	-0.31	-0.88	0.06	0.63	1.22	0.82	0.14	0.19
Kirrel	-0.32	1.09	0.76	2.16	0.02	-1.37	0.05	-1.40
Lima1	-0.35	-0.88	-1.30	-0.56	2.05	-1.12	1.85	-0.01
Ctnnd2	-0.40	0.20	0.22	-0.81	0.21	-0.06	-0.85	1.87
Evc2	-0.46	0.23	-0.36	-0.81	0.49	0.07	0.53	-0.18
Lima1	-1.46	-1.30	0.38	0.11	-1.06	0.64	0.53	0.19

With a p-value of 7.2 x 10^{-6}, EWSR1-FLI1 was the most significantly enriched factor seen in this PSCAN analysis. 15 of the 25 genes with greater than 50 fold higher expression in ΔF4 versus WT have high z-scores indicating promoter binding sites for EWSR1-FLI1. It is important to note that since correlation does not equal causation, many of these highly upregulated genes may not be

involved in the ΔF4 activation phenotype. However, some have biological functions that would logically contribute to a decreased activation threshold.   One of the most promising examples is Ctnnd1, catenin (cadherin-associated protein), delta 1. This gene has the highest z-score for EWSR1-FLI1 promoter binding and is one of the most dramatically upregulated, with 300-400X higher expression in ΔF4 versus WT at t=0 (and increased to ~1800X higher at t=30).   Ctnnd1 functions in cell adhesion and ligand-induced receptor signaling through the EGF, PDGF, CSF-1 and ERBB2 receptors. This increase in cell adhesion and enhanced signal transduction logically fits with a lower number of TCR engagements necessary for activation. Importantly, Ctnnd1 is implicated in more than one cause of malignant cell transformation, primarily by the proto-oncogene SRC (v-src avian sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog). It also binds and inhibits the transcriptional repressor ZBTB33, which can allow activation of Wnt signaling target genes involved in cell proliferation and oncogenic activity.

Other differentially expressed EWSR1-FLI1 target genes that may play a supporting role in the reduced activation threshold include Nrp1, Trim16, Tmem204, Cacnb3 and Pdzrn3. Nrp1 (Neuropilin 1) binds many ligands and various types of co-receptors, and is involved in several different types of signaling pathways that affect cell survival, migration, and attraction. Trim16 (Tripartite Motif Containing 16) is a B box zinc finger protein family which binds interleukin-1, is involved in cell growth, and associated with squamous cell carcinoma. Tmem204 (Transmembrane Protein 204), like Ctnnd1, plays a role in cell adhesion. Cacnb3 (Calcium Channel, Voltage-Dependent, Beta 3 Subunit) plays a role in the regulation of transcription factors and calcium transport during signal transduction. Pdzrn3 (PDZ Domain Containing Ring

Finger 3) is an E3 ubiquitin-protein ligase associated with ovarian cancer.   Dlg5 (Discs, Large

Homolog 5) is another candidate gene with greater than 50 fold higher expression in ΔF4, which

is not an EWSR1-FLI1 target, but has high z-scores for TLX1::NFIC and Tcfcp2l1. Dlg5 is a

membrane-associated guanylate kinase (MAGUK), a family of proteins which interact with

components of adherens junctions at sites of cell-cell contacts and function as scaffolding

molecules to regulate signal transduction.

To generate a larger pool of genes for enrichment analysis, parameters were extended to

include those with greater than 10 fold higher expression in ΔF4 versus WT (RPKM>1, DEseq p-

value<0.05). A total of 132 genes met these criteria, and PSCAN/CSCAN analysis for this group is

shown in Figure 8f. As would be predicted, all the same transcription factors appear in PSCAN as

were seen in the greater than 50 fold group, with the addition of two more with "high ΔF4

specific" enrichment, RELA and EBF1.

Figure 8f: PSCAN and CSCAN output for over/under-represented TF binding sites within promoters and *cis*-regulatory elements of genes with >10 fold higher expression in ΔF4 versus WT at t=0 for both data sets (p-value <0.05)

PSCAN

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
Tcfcp2l1	5.08	1.71E-07
EWSR1-FLI1	4.38	5.65E-06
MZF1_5-13	3.64	1.31E-04
NFKB1	3.55	1.88E-04
znf143	3.36	3.81E-04
MZF1_1-4	3.18	7.17E-04
RELA	3.15	7.96E-04
NF-kappaB	3.06	1.07E-03
ESR1	2.97	1.44E-03
TLX1::NFIC	2.94	1.61E-03
PLAG1	2.78	2.66E-03
SP1	2.73	3.11E-03
TFAP2A	2.71	3.29E-03
INSM1	2.46	6.80E-03
Pax5	2.30	1.06E-02
REST	2.27	1.13E-02
RREB1	1.76	3.87E-02
EBF1	1.75	3.99E-02
Zfp423	1.65	4.91E-02

CSCAN

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 132 ENTRIES	OVER/ UNDER-REPRESENTED
HCFC1_(NB100-68209)	CH12	13/132	2.78E-11	50	UNDER
GCN5	CH12	11/132	1.69E-08	42	UNDER
IRF4	NFS-201	12/132	9.11E-07	39	UNDER
TBP	CH12	17/132	6.40E-06	44	UNDER
Max	CH12	22/132	8.09E-06	50	UNDER
ZNF-MIZD-CP1_(ab65767)	CH12	13/132	1.51E-05	38	UNDER
c-Myc	CH12	11/132	1.29E-04	33	UNDER
E2F4	CH12	0/132	1.51E-04	13	UNDER
Nrf2	CH12	3/132	7.53E-04	18	UNDER
Mxi1_(AF4185)	MEL	19/132	1.03E-03	41	UNDER
NELFe	MEL	23/132	1.44E-03	45	UNDER
TBP	MEL	27/132	5.61E-03	48	UNDER
USF2	CH12	2/132	7.21E-03	14	UNDER
SIN3A_(NB600-1263)	MEL	24/132	8.29E-03	44	UNDER
E2F1	mESC	16/131	2.19E-02	31	UNDER
MyoD_(sc-32758)	C2C12	9/132	2.38E-02	3	OVER
CHD2_(AB68301)	CH12	18/132	2.51E-02	35	UNDER
COREST_(sc-30189)	CH12	8/132	3.72E-02	21	UNDER
tcf3	mESC	4/131	4.75E-02	1	OVER

GO analysis was also performed with input genes from the group with greater than 10 fold higher expression in ΔF4 versus WT (Supplementary Figure 8b). Surprisingly, none of the resulting categories had any clear connection to activation, cell cycle, or apoptosis. However, there may be a less obvious underlying role, or there may be too few genes involved to generate a significant cluster.

The DEseq package also has the capability to generate p-values for gene variance between two conditions without replicates. This method assumes that the mean is a good predictor of the dispersion. Given two samples with different conditions, the majority of genes can be expected to have comparable expression, while a minority are altered by the condition variable. The estimated variance should not be greatly affected by this relatively small pool of differentially expressed genes. Furthermore, these genes will only increase the dispersion estimate, and the p-values generated will be less significant. Since there is only one set of data for t = 2 Hrs, this DEseq method without replicates was used for comparison of differentially expressed genes throughout the time course (R/Bioconductor command lines included in Appendix I). Given that p-values from DEseq with experimental replicates will be lower, p-values for both of these experiments were calculated separately (ie. without replicates) to obtain p-values in an appropriately comparable range for setting cutoff parameters.

P-values were generated by DEseq from the individual comparisons of t = 0 expression to either 30 min or 2 hrs for both WT and ΔF4 samples analysis, and used to analyze upregulation during the time course. Parameters were set for RPKM>1, with greater than 5 fold upregulation (at 30

min or 2 hrs) in either WT or ΔF4, and a corresponding p-value<0.05. A group of 359 genes met

these criteria, and a clustering heatmap of their z-scores for expression is shown in Figure 8g.

Group 1 (34 genes) shows higher upregulation in ΔF4 samples at 30 min, and group 5 (19 genes)

at both 30 min and 2 hrs. Group 3 (38 genes) shows higher upregulation in WT at 30 min.   A few

genes stand out among those in group 3, which would have lower than normal upregulation in

ΔF4: Cdkn1a, Cdkn1c, Atf3, and Tgif1. Cdkn1a and Cdkn1c (Cyclin-Dependent Kinase Inhibitor 1A

and 1C) are negative regulators of cell proliferation. They bind to and inhibit cyclin-dependent

kinase (Cdk) activity, preventing phosphorylation of Cdk substrates, and blocking cell cycle

progression. Lower expression of Cdkn1a/c at early time points (eg. 30 min), as seen in ΔF4,

would give these cells an advantage for initiating cell division in response to activation. Atf3

(CAMP-Dependent Transcription Factor ATF-3) and Tgif1 (TGFB-Induced Factor Homeobox) are

both transcriptional repressors. Atf3 binds and represses transcription at the cAMP response

element (CRE), while Tgif1 is a transcriptional corepressor of SMAD2.

Figure 8g: Cluster heatmap for 359 genes with RPKM>1, FOLD>5 (for either WT or ΔF4 at 30 m or 2H for either experiment), corresponding p-value<0.05 (calculated w/o replicates)
Group 1 shows higher upregulation in ΔF4 at 30m; Group 5 in ΔF4 at t=30m,2H and Group 3 in WT at 30m
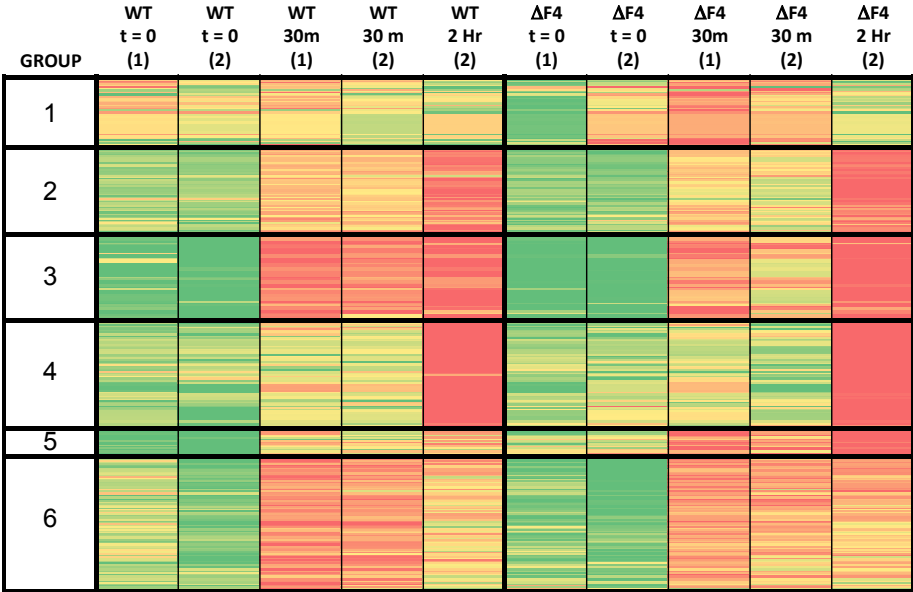
Many genes in group 1 (ΔF4 high at 30m) have biological functions that are likely to be involved in the ΔF4 decreased activation threshold phenotype, among these are Junb, Jund, Fos, Ifng, Il1b, Cebpb, Romo1, and Arhgap27. Junb (Jun B Proto-Oncogene) and Jund (Jun D proto-oncogene) are upregulated in response to activation, and protect cells from p53-dependent apoptosis. JunD oncogenic function is associated with human t-cell leukemia virus type 1. Fos (FBJ Murine Osteosarcoma Viral Oncogene Homolog) dimerizes with proteins of the JUN family, forming the transcription factor complex AP-1, and is involved in TGF-beta-mediated signaling activation. Ifng (Interferon gamma) and Il1b (Interleukin-1) are cytokines produced by activated lymphocytes and activated macrophages, respectively. IL1b stimulates thymocyte proliferation by inducing IL-2 release, and Cebpb (CCAAT/Enhancer Binding Protein (C/EBP), Beta) is a transcriptional activator that specifically binds to an IL-1 response element in the IL-6 gene. Romo1 (Reactive Oxygen Species Modulator 1) induces production of reactive oxygen species (ROS) that are necessary for cell proliferation. Arhgap27 is a Rho GTPase-activating protein that is associated with chronic lymphocytic leukemia. The most notable gene that stands out from group 5 (ΔF4 high at 30 min, 2 hrs) is Hes1 (Hairy And Enhancer Of Split 1). As previously mentioned, Hes1 is a Notch1 target gene, and is also a transcriptional repressor that influences the maintenance of certain stem cell genetic programs associated with pluripotency and proliferative expansion.

Due to the previously mentioned small populations for of each of these groups (group 1 = 34 genes, group 5 = 19 genes, group 3 = 38 genes), PSCAN and CSCAN analysis was less significant than for the larger gene pools, but recapitulated enrichment of several transcription factors

(Figure 8h). Group 1 showed PSCAN enrichment for three previously discussed "high ΔF4 specific" factors, RELA, Zfp423 and Ebf1 (A full PSCAN Z-score heatmap for Group 1 can be found in Supplementary Figure 8c). GO analysis for groups 1 and 3 produced a few significant biological function clusters (Supplementary Figure 8d). Group 1 functions included regulation of IL-6 production (Cebpb, Ifng, and Il1b), and basic-leucine zipper (bZIP) transcription factors (Cebpb, Junb, Jund, and Fos). Since there are only 19 genes in group 5, few significant p-values were generated for enriched factors using PSCAN, CSCAN, and no significant functions identified by GO analysis. Group 5 did show PSCAN enrichment for two other previously discussed "high ΔF4 specific" factors, NFkB and Hand::Tcfe2a. (A full PSCAN heatmap for Group 5 can be found in Supplementary Figure 8e). One novel PSCAN result worth mentioning is the appearance of RUNX1 enrichment for group 5. RUNX1 (runt-related transcription factor 1, also known as AML1, CBFA2) encodes core binding factor (CBF), which is involved in normal hematopoetic development. However, translocations of RUNX1 have been associated with several types of leukemia, and it might play a supporting role in malignant transformation.

Figure 8h: PSCAN and CSCAN output for over/under-represented TF binding sites (p-value <0.05) within promoters and cis-regulatory elements of genes from Group 1, 3, and 5 of Figure 8g

Group 1 (ΔF4 high, t=30m)

PSCAN			CSCAN					
TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE	TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 34 ENTRIES	OVER/ UNDER-REPRESENTED
SP1	3.07	1.07E-03	ZNF-MIZD-CP1_(ab65767)	MEL	14/34	3.90E-09	2	OVER
TBP	2.96	1.53E-03	CHD1_(NB100-60411)	CH12	13/34	9.93E-06	3	OVER
Zfp423	2.77	2.76E-03	ETS1	MEL	10/34	6.24E-05	2	OVER
CREB1	2.74	3.03E-03	COREST_(sc-30189)	CH12	17/34	1.03E-04	6	OVER
NFKB1	2.65	3.96E-03	BHLHE40_(NB100-1800)	CH12	22/34	3.86E-04	10	OVER
PLAG1	2.46	6.93E-03	CHD2_(AB68301)	CH12	21/34	4.00E-04	9	OVER
Mafb	2.21	1.35E-02	UBF_(sc-13125)	CH12	19/34	4.24E-04	8	OVER
RELA	2.16	1.52E-02	TCF3_(SC-349)	C2C12	9/34	1.74E-03	2	OVER
EBF1	2.08	1.87E-02	p300	MEL	7/34	6.24E-03	1	OVER
RREB1	2.00	2.29E-02	TBP	mESC	22/35	7.30E-03	12	OVER
INSM1	1.92	2.74E-02	E2F4	C2C12	14/34	9.06E-03	5	OVER
Klf4	1.80	3.61E-02	Max	MEL	14/34	1.53E-02	6	OVER
MZF1_1-4	1.76	3.95E-02	MafK_(ab50322)	CH12	8/34	1.67E-02	2	OVER
ZEB1	1.64	5.00E-02	STAT3	mESC	3/34	3.81E-02	0	OVER

Figure 8h continued:

Group 5 (ΔF4 high, t= 30m, 2Hr)

PSCAN

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
MZF1_1-4	2.60	4.69E-03
NF-kappaB	2.54	5.56E-03
RREB1	2.17	1.49E-02
RUNX1	2.08	1.89E-02
NFKB1	1.87	3.10E-02
Pax4	1.85	3.24E-02
REST	1.83	3.34E-02
Hand1::Tcfe2a	1.82	3.46E-02
Myf	1.77	3.84E-02
MZF1_5-13	1.76	3.88E-02
E2F1	1.72	4.22E-02
Arnt	1.68	4.61E-02

CSCAN

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 19 ENTRIES	OVER/ UNDER-REPRESENTED
FOSL1_(sc-605)	C2C12	2/19	3.08E-03	0	OVER
Med1	mESC	2/19	2.30E-02	0	OVER
Med12	MEFs	1/19	4.80E-02	0	OVER

Group 3 (WT high, t=30m)

PSCAN

TRANSCRIPTION FACTOR	Z-SCORE	P-VALUE
CTCF	3.86	5.52E-05
Klf4	3.47	2.57E-04
SP1	3.14	8.43E-04
Arnt::Ahr	2.98	1.44E-03
Mafb	2.92	1.74E-03
Pax2	2.81	2.44E-03
HIF1A::ARNT	2.78	2.73E-03
E2F1	2.76	2.85E-03
Egr1	2.59	4.78E-03
MIZF	2.57	5.07E-03
CREB1	2.41	7.88E-03
Myf	2.35	9.29E-03
NFYA	2.26	1.18E-02
ZNF354C	2.18	1.47E-02
Myc	2.08	1.88E-02
TFAP2A	1.98	2.37E-02
Arnt	1.97	2.45E-02
INSM1	1.85	3.23E-02
NFKB1	1.84	3.27E-02
PLAG1	1.84	3.31E-02
Mycn	1.72	4.23E-02

CSCAN

TRANSCRIPTION FACTOR	CELL LINE	HITS/ SAMPLE	P-VALUE	EXPECTED/ 38 ENTRIES	OVER/ UNDER-REPRESENTED
c-Jun	CH12	12/38	2.40E-08	1	OVER
SMC3_(ab9263)	CH12	19/38	6.12E-07	5	OVER
ETS1	CH12	21/38	6.61E-06	7	OVER
COREST_(sc-30189)	MEL	9/38	1.09E-05	1	OVER
p300_(SC-584)	CH12	12/38	1.08E-04	3	OVER
JunD	CH12	8/38	1.28E-04	1	OVER
Rad21	CH12	16/38	1.73E-04	5	OVER
Smc1a	MEFs	11/39	1.28E-03	3	OVER
CHD1_(NB100-60411)	CH12	11/38	2.10E-03	3	OVER
TCF3_(SC-349)	C2C12	9/38	5.06E-03	2	OVER
CHD2_(AB68301)	CH12	21/38	5.08E-03	10	OVER
GATA2	HSPC	2/39	2.61E-02	0	OVER
MafK_(ab50322)	MEL	3/38	3.15E-02	0	OVER
Max	CH12	24/38	4.96E-02	14	OVER

Chapter 8 endnote: all gene description summaries without specific reference numbers are from the Entrez gene database and the Human Gene Compendium; www.genecards.org.

Chapter 9: Future directions for understanding the role of Ikaros

Results discussed in Chapter 7 using CFSE to observe *in vitro* proliferation of T cells in whole splenocyte culture have shown a tenfold decrease in α-CD3/28 concentration for activation of *Ikzf1*$^{\Delta F4/\Delta F4}$ T cells versus WT. This is approximately the same decrease in activation threshold that was seen for *Ikzf1*$^{null/+}$ and *Ikzf1*$^{DN/+}$ mice[45]. At sub-threshold conditions for WT activation, the percentage of *Ikzf1*$^{\Delta F4/\Delta F4}$ cells that upregulate activation markers CD25 and CD69 is 4-5 fold higher than WT. Upregulation also appears to be proportional to number of WT copies of Ikaros, as *Ikzf1*$^{\Delta F4/+}$ heterozygotes show an intermediate level for both markers. Time course indicates that kinetics of upregulation for CD25 and CD69 occur normally in the *Ikzf1*$^{\Delta F4/\Delta F4}$ and *Ikzf1*$^{\Delta F4/+}$ cells despite the increase in expression. The difference in activation threshold is T cell autonomous, as it is maintained for MACS (and FACS) populations depleted of other cell types. However, cells from both genotypes (*Ikzf1*$^{\Delta F4/\Delta F4}$ and WT) need more highly concentrated stimulation by α-CD3/28 in the absence of other supporting cells and factors. Naïve *Ikzf1*$^{\Delta F4/\Delta F4}$ CD8$^+$ T cells isolated by FACS sorting maintain the activation threshold difference, therefore the source of cells with a lower activation threshold is not from altered populations of mature and memory cells. Many deregulated genes with potential contribution to decreased activation threshold and loss of tumor suppressor function were identified by RNA-seq analysis.

Transcription factors identified by RNAseq to have a potential role in driving transformation could be further investigated using RNAi knockdown in cultures of *Ikzf1*$^{\Delta F4/\Delta F4}$ T cells from mice that have progressed to lymphoma/leukemia.   If knockdown eliminates survival advantage of these cells, it would suggest that the given factor is involved in the malignant phenotype.

Another technique with the same culture setup would be to overexpress potentially repressed factors (ie. those that were more highly expressed in WT cells) and again observe the effect on cell survival. To determine specifically whether possible FLI1 translocations occur in the $Ikzf1^{\Delta F4/\Delta F4}$ T cells, PCR amplification of the putative junction could be used as a first step, and fluorescence microscopy with the same culture setup as above could be used to look for translocations t(11;22)(q24;q12) or t(4;11)(q21;q23).

RNA-seq analysis of a longer activation time course may also be informative for differential expression of late expressed genes that peak after 2 hours. A larger population of purified naïve CD8+ T cells can be obtained using the Stem Cell Technologies EasySep Naïve CD8+ T cell isolation kit, which can process 0.5 - 8.5 mL of sample (up to 8.5 x 10^8 cells).

Figure 9a: Purity testing of Stem Cell Technologies Naive CD8 T cell isolation kit
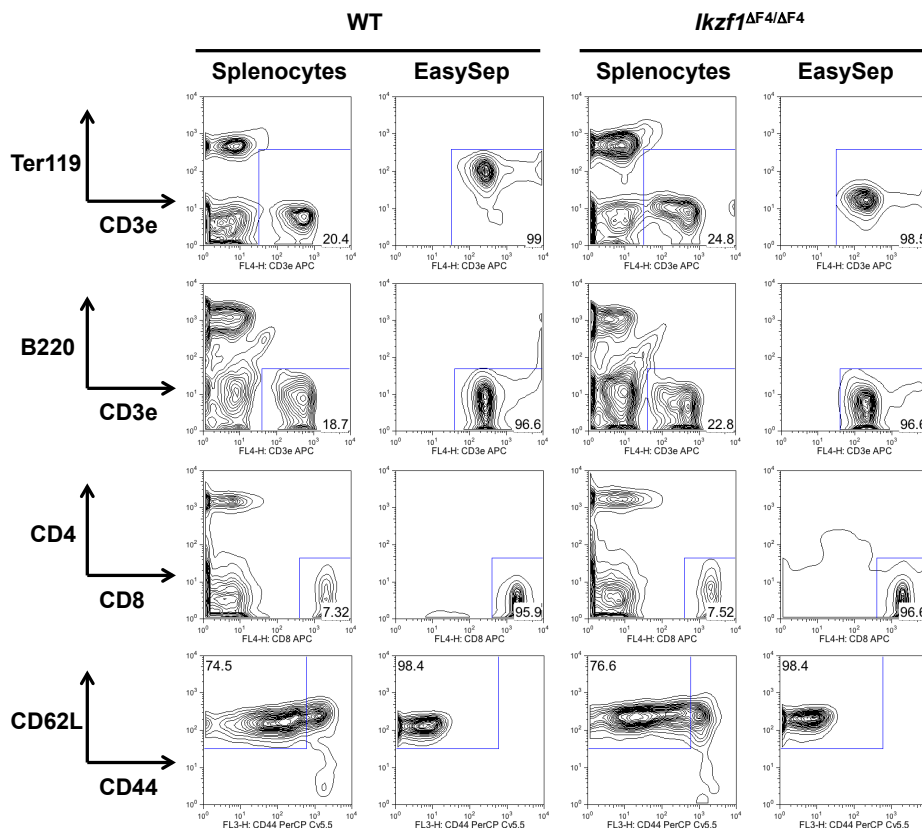
As shown in Figure 9a, purity from the EasySep kit is extremely high (~98% naïve CD8 T cells) and comparable to FACS sorting. This system has a nearly five fold higher yield than FACS sorting, presumably due to a reduced number of cells lost during fewer treatment steps and less death in the shorter time needed for isolation. Using this method it would be possible to purify enough cells from just two mice to collect sufficient RNA for at least 6 time points.

Other methods might also be used to test hypotheses for the mechanism of how Ikaros may be affecting activation threshold. One likely possibility is that $Ikzf1^{\Delta F4/\Delta F4}$ T cells enter the cell cycle more easily and rapidly than WT due to decreased production or activation of cell cycle inhibitors. Cell cycle analysis can be performed using propidium iodide (PI), as was used in previous characterization of $Ikzf1^{null/+}$ and $Ikzf1^{DN/+}$ mice, and is shown in Figure 9b[45].

Figure 9b: Accelerated entry into S phase[45]
Purified splenic T cells were cultured in the presence of 20 mg/ml plate-bound anti-CD3e and DNA content was determined by propidium iodide (PI) staining after 24, 36, and 60 hr. Histograms display PI intensity. Shown below are the percentages of cells in G0/G1, S, or G2/M.

	G0/G1	S	G2/M	G0/G1	S	G2/M	G0/G1	S	G2/M
Null+/-	77.2	17.0	5.8	39.4	46.3	14.3	35.0	51.3	13.7
DN+/-	80.2	17.8	2.0	40.1	41.4	10.5	39.4	44.4	16.2
+/+	93.8	5.6	0.6	47.1	42.7	10.2	37.3	50.7	12.0
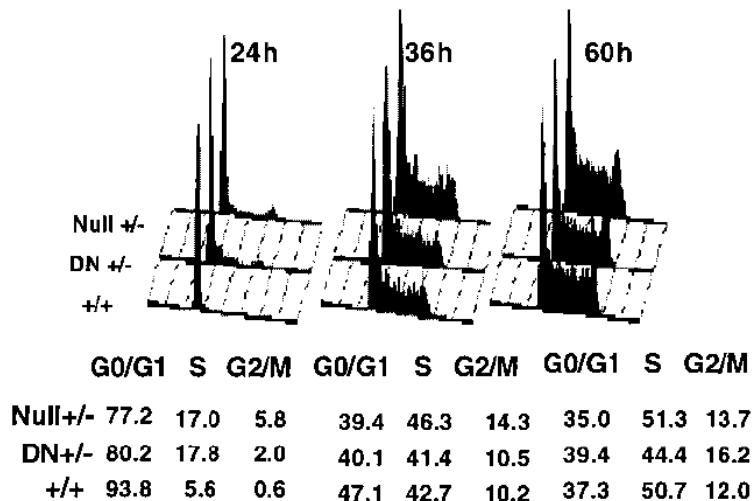
For this assay, cells are permeabilized following activation and stained with PI, which is a fluorescent dye that quantitatively stains DNA. The intensity of staining is analyzed by FACS and

directly correlates to the amount of DNA present. Cells in the G2/M phase have twice as much

DNA as cells in G0/G1 (before S phase), and those in S have an intermediate range. This can be

interpreted to determine the fraction of cells in each of the stages of cell division. Another

option for cell cycle analysis is to use BrdU (bromodeoxyuridine, 5-bromo-2'-deoxyuridine)

incorporation, which was previously attempted several times without success and put on hold to

focus on other priorities. Troubleshooting of this method could be resumed, starting with new

reagents and research in protocol modifications. BrdU is a thymidine analog that can be

incorporated into newly synthesized DNA of proliferating cells. Cells are treated with BrdU prior

to activation and after a time course and permeabilization, BrdU antibodies can be used for FACS

analysis of cell DNA content. If BrdU/PI cell cycle analysis shows that $Ikzf1^{\Delta F4/\Delta F4}$ T cells enter the

cell cycle more rapidly than WT, it would be important to examine whether levels of Ikaros

phosphorylation are higher than in WT T cells. Increased phosphorylation of the Ikaros

serine/threonine rich conserved region in exon 8 at the G1 to S transition would decrease the

ability of Ikaros to negatively regulate cell cycle progression[51] and allow $Ikzf1^{\Delta F4/\Delta F4}$ T cells to

more easily initiate proliferation.

Another potential mechanism for decreased activation threshold is increased production of

cytokines that support lymphocyte survival and proliferation, such as IL-2, IL-1b, Ifng, and IL-6.

Levels of cytokine production can be assessed using ELISA (enzyme-linked immunosorbent

assay). If cytokine production is affected, the activation phenotype likely isn't related to cell

cycle. It is also possible that $Ikzf1^{\Delta F4/\Delta F4}$ T cells have lost the requirement for sustained TCR

signaling or CD28 costimulation. Different signaling pathways require different lengths of

sustained TCR engagement, and it is possible that activation of $Ikzf1^{\Delta F4/\Delta F4}$ T cells does not

require one or more of the pathways activated by prolonged TCR signaling. This can be tested

using a time course in which the cells are removed from stimulation at different early time

points and downstream phosphorylation events (MAPKs, p38, Jnk) are evaluated by western blot

or phospho-FACS. NFkB signaling occurs with just α-CD3 stimulation, but the prevalent dogma is

that α-CD28 and resulting downstream activation of NFAT signaling is necessary to drive

proliferation instead of apoptosis. If the same activation phenotype is seen for $Ikzf1^{\Delta F4/\Delta F4}$ T cells

with stimulation using just α-CD3, these cells have potentially have lost requirement for

expression of NFAT target genes.

Section 3: Optimization of Ikaros chromatin immunoprecipitation for high-throughput sequencing (ChIP-seq)

Chapter 10: Goals and challenges of Ikaros ChIP

Ikaros target genes have remained largely elusive in part due to a general previous lack of

success attempting to use traditional chromatin immunoprecipitation (ChIP) methods. ChIP is a

technique used to identify genomic sites of *in-vivo* interactions between specific proteins and

DNA. Briefly, protein and associated chromatin in a whole cell suspension are crosslinked, which

is commonly achieved using formaldehyde. Cells are lysed to release nuclei, and chromatin is

sheared using sonication, typically to a range of 500-1000 bp. Antibodies to the protein of

interest are incubated with the sheared chromatin, and the DNA-protein complexes are

precipitated. For precipitation of antibody-bound complexes, most traditional protocols use

agarose conjugated to Protein A or G, which binds the heavy chain within the Fc (constant)

region of the antibody. After rigorous washing steps, crosslinking is reversed by heat or an

elution buffer and DNA fragments can be purified from the protein. Prior to advent of

affordable high-throughput sequencing, the main method of analyzing immunoprecipitated

fragments was either QPCR or a microchip (ChIP-chip) with putative target sequences. The

application of high-throughput sequencing to ChIP (Chip-seq) has made identification of novel

target sites more feasible.

Multiple antibodies to Ikaros-GST fusion proteins were developed by the Smale lab many years

ago[52], but Ikaros ChIP has remained challenging. These Ikaros ChIP difficulties are not an issue

with the antibodies, as was confirmed with troubleshooting performed by a previous lab

member. If purified Ikaros protein was used to do ChIP on a reporter gene containing multiple

Ikaros known consensus binding sites, there would be negligible enrichment. However, if a

purified Ikaros gal-4 fusion protein was used to do ChIP with plasmid DNA containing a Gal4

specific binding site, the same Ikaros antibodies would give strong enrichment signals.

Essentially, when Ikaros was bound to DNA through a Gal4 site, ChIP with an Ikaros antibody

worked well, but not when it was bound to DNA through its own domain.

In the past, a few have been able to see Ikaros signals using standard ChIP protocols (and the

antibodies purified by the Smale lab) with previously identified high affinity binding sites, but

results have been only semi-reproducible and enrichment relatively low[53,54,55,56].   Early in

research presented here, the ChIP method from Thompson 2007[54] was tested using primers

from Reynaud 2008[55], with results shown in Figure 10a. Although Ikaros shows greater

enrichment over input than the negative control (pre-immune serum from the rabbit used for α-

IK-C antibody production, prior to inoculation with Ikaros protein), it is clearly insignificant when

compared to the efficiency of a positive control, H3K4me3.

Figure 10a: Ikaros ChIP method from Thompson 2007[54] (with primers from Reynaud 2008[55]).
% Input shown for α-IK-C comparison to no antibody or pre-immune serum negative controls, and α-
HSK4me3 positive control
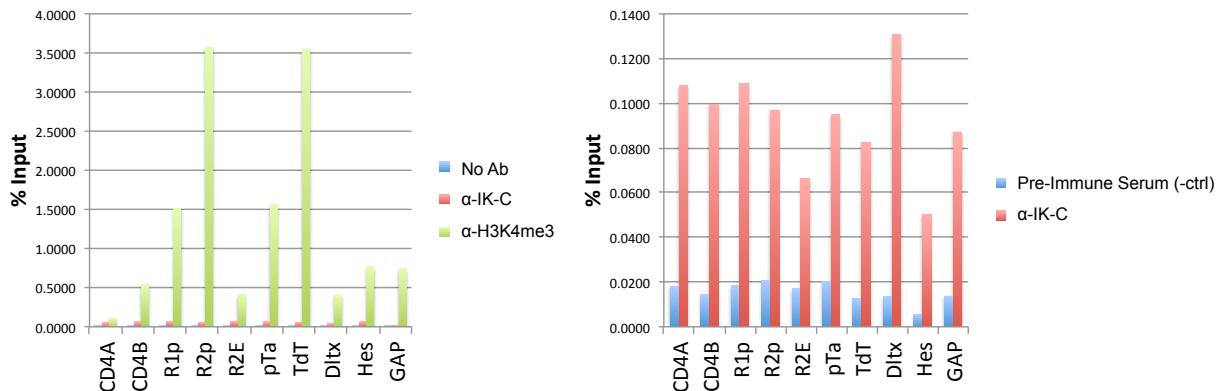
The ChIP method from Thompson 2007[54] was compared step-by-step to the standard Smale lab

protocol and both were performed in parallel with little difference in results (not shown).

Multiple conditions of crosslinking (length of time, media, crosslinking reagents) were tested

using an immortalized VL3 T cell line. Using a secondary crosslinking reagent, disuccinimidyl

glutarate (DSG), did increase enrichment, but it appeared ubiquitous and non-specific to Ikaros

target sequences (not shown). Troubleshooting ChIP was put on hold to pursue other topics, but

was resumed in 2013 as a higher priority, since recent publications have used different

techniques to improve upon Ikaros ChIP and successfully perform ChIP-seq analysis[29,57].

Chapter 11: Adapting ChIP techniques for the future

A postdoc from the lab was able to adapt the Farnham lab ChIP-seq protocol to see improved enrichment results with Ikaros[29]. However, signals were much weaker in the *Ikzf1* mutants, especially *Ikzf1*$^{\Delta F4/\Delta F4}$, due to their inherently lower binding affinity. Some generally low affinity sites had issues with reproducibility in WT as well. For these reasons, further testing of alternate protocols and overall optimization was pursued.

The UCLA Collaboratory offers workshops relating to high-throughput sequencing preparation and analysis, with applications for RNA- and ChIP-seq. These courses are taught by postdocs with expertise in each of these areas. The ChIP-seq workshop is led by a postdoc from the Kurdistani lab, which is known for techniques of ChIP optimization. Their current protocol uses an entirely different method of sonication and set of buffers than the Farnham lab. The primary difference between these two protocols is the type of chromatin sonication. The Farnham protocol uses a device called the Bioruptor, an entirely self-encompassed system, while the Kurdistani lab uses a Cup Horn accessory for a standard Misonex sonicator. Both use a chilled waterbath circulation system, and the main difference is in the intensity of the sonic disruption and size of the resulting DNA fragments. Figure 11a shows a comparison of testing both these methods on WT thymocytes.

Figure 11a: Comparison between sonication methods from Kurdistani and Farnham labs
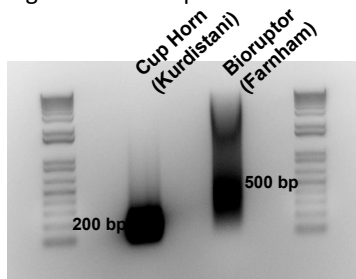
A standard sonication protocol typically suggests fragmenting chromatin to an average size of 500 bp. This is essentially what is seen for the Bioruptor method. The cup horn sonicator used by the Kurdistani lab achieves robust maximal sonication to a tight, consistent average size of 200 bp, corresponding to the length of DNA wrapped around a single nucleosome. This dramatically increases enrichment seen for ChIP, by reducing concentration of nonspecific DNA. It is also improves the accuracy of ChIP-seq, as it centers the predicted target site in a specific 200 bp window, regardless of which strand of the DNA was bound and/or sequenced in the library. Figure 11b shows the bimodal distribution of ChIP-seq enrichment peaks from library sequences of either strand, and relative distance from the putative binding site. Complementary strands will be sequenced from opposite ends, and knowing the specific length of the starting DNA fragment improves the accuracy of the binding site prediction algorithm.

Figure 11b: Bimodal distribution of predicted protein binding site location relative to starting end of DNA fragment sequenced[58]
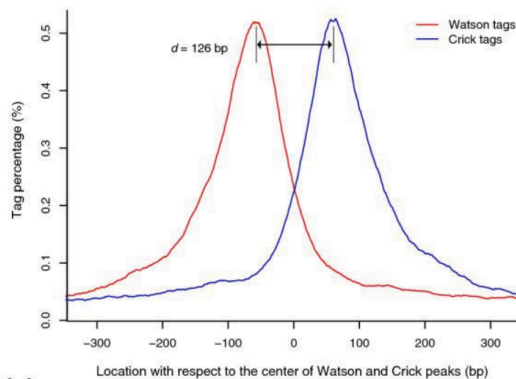
Figure 11c shows the QPCR results from testing both ChIP protocols in parallel on WT thymocytes, using both the Ikaros C-terminal antibody (α-IK-C) and a positive control, H3K4me3 (Histone H3 trimethyl Lys4). Primers from Reynaud 2008[55] for known Ikaros binding sites were used for QPCR analysis. Enrichment for H3K4me3 was approximately 5 fold greater using the Kurdistani method. However, it was unclear from these results whether intense sonication

down to a fragment size of 200 bp would be appropriate for a DNA-bound transcription factor as opposed to a histone modification such as H3K4me3. Histone modifications are contained in the nucleosome structure, while transcription factors may be bound outside of this range or have low affinity that is disrupted by the more robust sonication. Enrichment for Ikaros was approximately 10 fold lower than H3K4me3, and actually reduced by the cup horn sonication method in comparison to the Bioruptor. However, for both of these methods, crosslinking was performed with just formaldehyde, and so the Kurdistani lab protocol was tested again a second crosslinking reaction using DSG (disuccinimidyl glutarate).

Figure 11c: Direct enrichment comparison between ChIP protocols from Kurdistani and Farnham labs using α-IK-C and α-H3K4me3
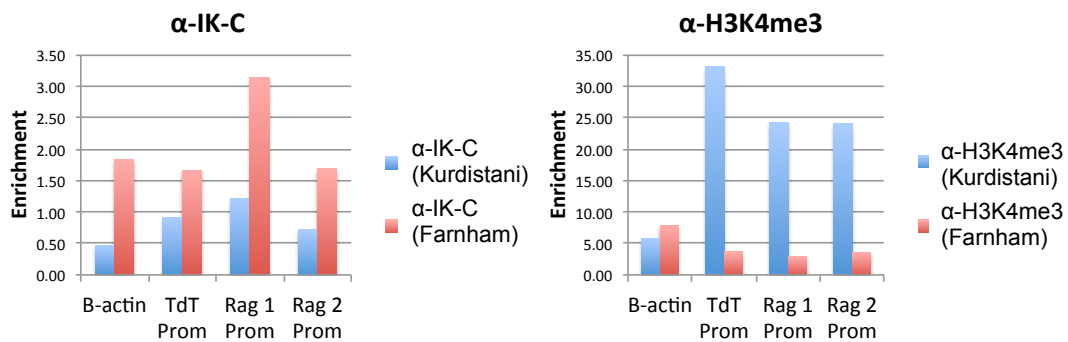
Addition of a second crosslinking step with DSG creates a more stable DNA-protein complex, and the DSG linker arm reaches farther to crosslink areas of the protein that are not is close proximity. Figure 11d shows that enrichment for Ikaros is also dramatically improved by the cup horn sonication method when the complex is more strongly attached. QPCR primers used in Figure 11d were designed from Ikaros target sites identified by previous ChIP-seq analysis[29]. Testing is also shown for another Ikaros antibody, α-IK-N, which recognizes the N-terminal end. This is the antibody that was used for previous ChIP-seq analysis[29]. Several repetitions (not

shown) concluded that α-IK-C produces higher enrichment signals. One reason for this may be better access to the recognition site due to distance from the N terminal DNA binding domain.

Figure 11d: Double-crosslinking modifications to Kurdistani ChIP protocol
Test of α-IK-C versus α-IK-N levels of enrichment, ΔF4 efficiency
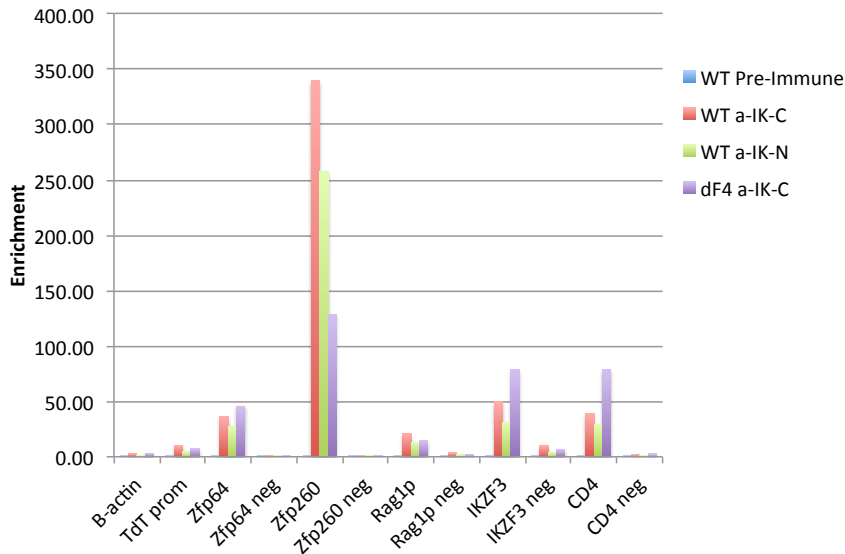
Figure 11d also shows that this protocol has similar enrichment for cells from *Ikzf1*$^{\Delta F4/\Delta F4}$ as from WT. This was one problem for the previously used protocol from the Farnham lab. One remaining issue was to troubleshoot obtaining enough DNA for ChIP-seq. The Illumina kits used previously required an input of 10 ng. The former protocol used DNA purification columns, which yielded an insufficient amount from a single thymus (a range of 5 to 10 ng maximum from *Ikzf1*$^{\Delta F4/\Delta F4}$), necessary for generating biological replicates. Phenol/chloroform extraction with ethanol precipitation was tested and found to reliably purify approximately twice as much DNA as the columns. This is enough to create sequencing libraries for *Ikzf1*$^{\Delta F4/\Delta F4}$ samples with the standard kit. Figure 11e shows QPCR testing of two ChIP replicates for WT and each mutant genotype, which are available for producing libraries if needed. The final Ikaros ChIP protocol adapted from the Kurdistani lab is included in Appendix I.

Figure 11e: QPCR test using Zfp260 primers to verify enrichment for ChIP-seq samples
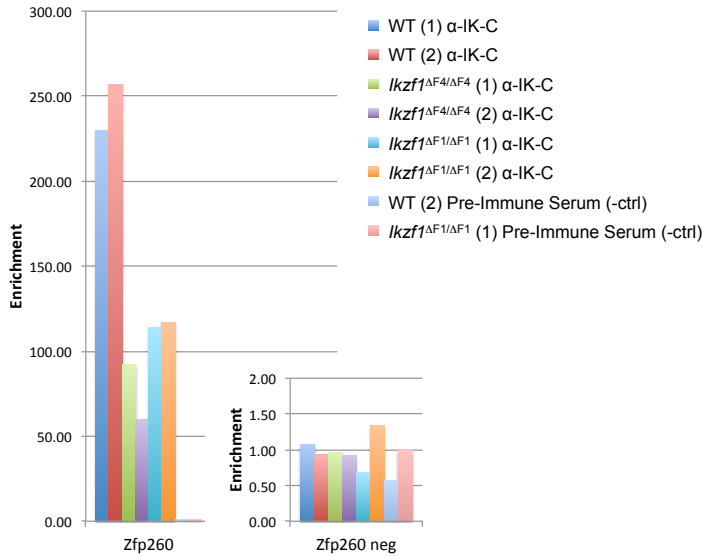
One future adaptation to this protocol would be the use of the Nugen Ovation Ultralow

Multiplex System, which allows for the use of as little as 1 ng input DNA. This would make it

possible to use ChIP-seq to analyze smaller cell populations, such as $Ikzf1^{\Delta F4/\Delta F4}$ samples from

lymphoma/leukemia, purified cells from individual stages of thymocyte development, and

activated T cells. In addition to ChIP for Ikaros, it would be helpful to look at factors known to be

associated with Ikaros (eg. Mi2, CtBP, Sin3, HDAC2) in any of these sample types. For T cell

activation in $Ikzf1^{\Delta F4/\Delta F4}$ mice, it would also be informative to use ChIP for putative cell

cycle/cytokine/receptor/NFAT/NF-KB target genes. In combination with RNA-seq, this data

could be used to validate functional targets and associate loss of Ikaros occupancy with

deregulation of gene expression.

Appendix I: Methods

Section 1:

Ikaros-GST fusion protein expression, purification and thrombin cleavage

Expression of protein and harvest of cells

Start overnight 10 mL LB cultures (Amp 1000X + Chlor 2000X, 25 mg/ml) from glycerol stocks or fresh colony

Next morning, inoculate fresh 500 mL culture 1:50 from ON culture

Grow 4 hours to an ~ OD of 0.6-0.8

Induce protein expression with 1:2000 stock IPTG (1M in water)

→ 250 μL IPTG (final conc. = 0.5 mM) +

→ 500 μL of 10 mM ZnCl2 (final conc. = 10 μM)

Induce ON at 37°C, shaking

Pellet cells by centrifugation at 6000 RPM for 10 min at 4°C

Remove LB supernatant and freeze pellet at -20°C (overnight or longer if needed)

Bacterial lysis, protein extraction and column purification

Resuspend bacterial pellet in 15 mL BPER extraction reagent with added:

1:100 10 mM ZnCl2 (final conc. 10 μM ZnCl2)
1:10 5 M NaCl (final conc. 500 mM NaCl)

1:1000 Aprotinin (2 mg/mL, 0.3 mM in dH2O)
1:1000 Bestatin (20 mg/mL, 65 mM in MeOH)
1:1000 E-64 (5 mg/mL, 14 mM in 1:1 dH2O:EtOH)
1:1000 Pepstatin (0.7 mg/mL, 1 mM in MeOH)
1:1000 Phosphoramidon (5 mg/mL, 9 mM in dH2O)

Note: PMSF inhibits thrombin, so it is not to be added until after cleavage step
1:100 PMSF stock (17 mg/mL, 1 mM in isopropanol)

Incubate/shake at RT for 10 mins to lyse cells

Pellet cell debris/insoluble proteins by centrifugation at 27,000 x g for 15 mins

Transfer supernatant containing soluble proteins into new tube and store on ice until adding to column

Running columns:

Equilibrate columns and buffers to RT (Best to get this started before the protein extraction –above)

Uncap the columns and allow sodium azide storage solution to drain

Prepare column bed by adding 2 x 5 mL of B-PER reagent to equilibrate

Apply up to 2 x 5 mL (10 mL) of sample to the column and collect flow through

Wash column with 2 x 3 mL (6 mL) wash buffer 1 (pH 7.5)

Wash column with 3 mL wash buffer 2 (pH 8.0)

Equilibrate columns for thrombin cleavage with 3 mL of thrombin cleavage buffer (TCB)

Thrombin cleavage buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 2.5 mM CaCl2
        Stock mixes: 25 mL 1M Tris pH 7.5, 15 mL 5M NaCl, 1.25 mL 1M CaCl2, up to 500 mL with dH20

Thrombin cleavage:

Add 1:1000 Bestatin, 1:1000 E-64, 1:1000 Pepstatin, 1:1000 Phosphoramidon protease inhibitors to a 25 mL aliquot of TCB.

Resuspend thrombin powder in vial by injecting TCB (with prot. inhib.) to give a final concentration of 25 UN/mL. Keep on ice.

Add 4 mL (100 UN) thrombin to each column, allow prior bed volume (~1 mL) to flow through into wash, cap columns, incubate at RT for 1 hr.

For the collection of cleaved protein, go one sample at a time and used a 10 mL plunger to push out 1 mL more precisely

Remove cap to allow flow through and collect 1 mL flow through on ice. Immediately cap column and incubate another 30 min at RT.

Repeat 30 min incubation 3X, adding 2 mL TCB (+prot inhib) after first 3 mL have gone through, to allow gravity flow for the last mL bed volume.

To each flow-through fraction: Add 100X PMSF stock, 10% glycerol, aliquot and freeze at -80C.

Wash columns 2X with 2 mL TCB, freeze at -80C.

Elution of remaining bound GST:

Prepare elution buffer by adding 12 mL of wash buffer 2 per vial of reduced glutathione (184 mg, makes 50 mM solution). Elute GST from each column with 4 x 3 mL and collect elution fractions. *

Store columns in wash buffer 2 with 0.02% sodium azide.

Electrophoretic Mobility Shift Assay (EMSA)

Annealing new oligos:

Resuspend in TE to 100 pmol/µL (0.1 nmol/µL)

	1X	
	2 µL	5 M NaCl (→ 50 mM)
	2 µL	1 M Tris, pH 7.5 (→ 10 mM)
	176 µL	ddH2O
	200 µL	
Each:	(Final 5 pmol/µL annealed oligos)	
	10 µL top	
	10 µL bottom	

Boil 5 min, remove block and let cool slowly to RT, store annealed oligos at -20C.

Labeling probes:

1X

22 µL	dH20
3 µL	10X PNK buffer
1 µL	PNK
1 µL	γ32P-ATP
Each:	
3 µL	ds oligo (5 pmol/µL)

37C for 1 hr
Add 20 µL dH20 or probe buffer, purify using G-50 columns
Transfer 1 µL probe to small eppendorf, place in glass vial and use scintillation counter to determine cpm

EMSA gel: 6% polyacrylamide, 0.25X TBE

1X

7.5 mL	40% Bis/Acrylamide
2.5 mL	5X TBE
39.5 mL	dH2O
0.5 mL	10X APS
16 µL	TEMED
50 mL	

EMSA Probe mix:

	1X
dH2O	8.5 µL
1 µg/µL BSA	1 µL
1 µg/µL dI/dC	1 µL
10 µM ZnCl2	1 µL
1 µL hot probe	1 µL
	12.5 µL total

Overall probe + protein mixes = 25 µL/reaction

Final salt concentration needs to be 50 mM for binding reaction--> make NE/protein mixes at 100 mM salt, adjust using HGED.0 and HGED.1 (0.15 M KCl) buffers:

	1 X
NE/protein	x μL
HGED.0	y μL
HGED.1	z μL
	12.5 μL total (0.1 M salt)

x(NE salt M) + z(.15) = 12.5 (0.1)

$x + y + z = 12.5$

- Aliquot protein mixes into tubes and leave on ice until probe mix is added
- Incubate (binding reaction) at RT for 30 min
- Pre-run gel in 0.25 TBE at 150 V for 30 min prior to loading EMSA samples

*Done simultaneously with RT 30 min sample incubations

- Load 1 lane with DNA dye (BpB + XC diluted to 2X in dH2O) in addition to samples
- Run at 150 V for ~1:30, avoid running probe out of gel
- Dry at 85 C for 30 min (higher % gel may need 45 min at 90 C)

Surface Plasmon Resonance (Biacore)

Running buffer:

(Thrombin cleavage buffer base + HBPS + $ZnCl_2$):

Final	Stock	Volume per 1L solution
50 mM Tris pH 7.5	1M	50 mL
150 mM NaCl	5M	30 mL
2.5 mM $CaCl_2$	1M	2.5 mL
1 μM $ZnCl_2$	10 mM	100 μL
50 mM Hepes pH 7.6	1M	50 mL
0.05% Tween-20	10% w/v	5 mL

Regeneration buffer:

(Running buffer +)

Final	Stock	Volume per 100 mL solution
1M NaCl	5M	20 mL
50 mM EDTA	0.5M	10 mL
0.1% SDS	10% w/v	1 mL

Labeling Biacore Streptavidin Chip:

Tools: Eject current chip

Insert new chip w/arrows first

Select: New Chip, Type SA, Name label, Dock

Connect running buffer to the left side (right side: small dH20 and large waste container)

       Running buffer = Thrombin cleavage buffer

(ie. 50 mM Tris pH 7.5, 150 mM NaCl, 2.5 mM CaCl$_2$)

Select Tools, Prime >2X

Insert into rack:	Conditioning buffer (1M NaCL, 50 mM NaOH)	Positions:	R1F1
	Regeneration buffer (Running buffer + 5 mM EDTA)		R1F2
	Probe for flow cell 2 (stock* diluted 1:1000, 1:100, 1:10)		A1,2,3
	Probe for flow cell 4 (1:1000, 1:100, 1:10)		B1,2,3
	Test protein IK 1-4 (5, 50, 500 uM)		C1,2,3

*annealed probe stocks at 5 pmol/ul

Select Run, Manual, flow rate of 30 ul/min (for conditioning and labeling)

Choose flowpath 1,2,3,4

Change Rack- select setup on right hand side

Save results to BioUsers/Smale/Sarah

Conditioning:

Command: Inject R1F1 (conditioning buffer), 60 sec, 3X

Stop run

Prime >2X

Labeling:

Select Run, Manual, flow rate of 30 ul/min (for conditioning and labeling)

Choose Flowpath 2-1, 4-3 and allow to run until response curve has reached a plateau

Select Flowpath 2 (or 4)

Commands: Inject, lowest probe concentration A1 (or B1), 45 sec

       View curve, Sensogram FC2 (or 4)

       Watch for an injection increase of ~500 response units

       Repeat and/or decrease dilution injection if needed

Repeat process for FC4

Stop run

Note: Best to take probes out of the rack at this time to avoid accidentally running through all 4 flow cells

Protein Test Binding:

Tools, Prime >2X

Manual run, Test

Flow rate 50 ul/min (response will increase)

Select Flowpath 2-1, 4-3 (FC1 and 3 are blank controls)

Commands: Inject protein (C1) 60 sec

Watch for response in sensogram curve FC 2-1 and FC 4-3

Regeneration:

Commands: Regeneration, F2, 60 sec

Repeat regeneration until the response returns to baseline levels

Finishing:

Stop run

Eject rack

Eject Chip *after removing chip, put in petri w/kimwipe soaked in running buffer, parafilm, 4C*

Insert maintenance chip, dock, switch left side buffer back to dH2O

       Select Reuse chip, maintenance

Set standby temp to 20C

Prime

Minimal Media (M9) for expression of proteins to analyze by NMR

1. Make M9 salts: aliquot 800ml H_2O and add
   a. 64g Na_2HPO_4-$7H_2O$
   b. 15g KH_2PO_4
   c. 2.5g NaCl
   d. 5.0g NH_4Cl (or $N^{15}H_4Cl$)
   e. Stir until dissolved
   f. Adjust to 1000ml with distilled H_2O
   g. Sterilize by autoclaving
2. Measure ~700ml of distilled H_2O (sterile)
3. Add 200ml of M9 salts
4. Add 2ml of 1M $MgSO_4$ (sterile)
5. Add 20 ml of 20% glucose (or other carbon source)
6. Add 100ul of 1M $CaCl_2$ (sterile)
7. Adjust to 1000ml with distilled H_2O

Section 2:

Splenic tissue preparation, CFSE staining and α-CD3/28 *in vitro* stimulation

1. Isolate spleen and store in sterile PBS (few mL in conical or petri) until made into single cell suspension: Mince with scissors, pulverize with plunger end from syringe, rinse through filter with PBS, collect in 50 mL conical, spin 5 min at 1200

2. Aspirate PBS, disrupt pellet and resuspend in 1 mL RBC lysis buffer per spleen, incubate 5 min and then chase with several mL PBS, spin 5 min at 1200, filter again into new tube, wash 2X more, resuspend in 2 mL 0.1% BSA/PBS (CFSE staining buffer)

3. Count cells, determine volume needed to obtain 2 x 10^7 cells (20 mil)

4. Transfer 2 x 10^7 cells for staining into a separate 15 mL conical, bring volume up to next whole mL

5. Prepare **2X CFSE staining solution:** (Enough to equal total volume of all samples)
   Stock solution = 10 mM   Final concentration 2-5 μM
   Make 2X final concentration, eg. 10 μM CFDA-SE (1:1000) in 0.1% BSA/PBS

6. Add equivalent volume of 2X CFSE staining solution to aliquot with 2 x 10^7 cells for staining, invert several times and incubate at 37°C for 10 min

7. Quench with 10 mL per sample pre-chilled DMEM 10% FBS, invert several times and spin down cells, wash 2 times with PBS

8. Resuspend cells in 500 μL complete media pre-warmed at 37°C

9. Count cells, determine total final volume of cell suspension needed for 1 x 10^6 cells/mL
   Bring up to final volume with complete media and distribute 200 μL/well (2 x 10^5 cells) in plate that has been precoated with α-CD3/28

10. Repeat #9 for unstained cell suspension needed for controls

Coating 96 U-bottom well plate with α-CD3/28:

1. Make α-CD3/28 dilutions in PBS:
   α-CD3/28 stock = 1 mg/mL
   Need 50 μL/well, determine total volume needed for each concentration
   Dilute to desired upper end concentration in PBS, serially dilute (eg 1:3) in PBS

2. Distribute 50 μL/well, leave at 37°C for 1 hr or 4°C overnight

3. Wash wells 3X with 150 μL/well (aspirate off from sides), leave 200 μL/well until used to keep moist

MACS negative selection for CD8+ T cell enrichment

1. Prepare splenic tissue as above

2. Save 5 x 10^6 unselected cells for FACS analysis, spin down the remainder and resuspend in 500 μL MACS buffer

   > MACS Buffer:
   > 2 mM EDTA
   > 0.5% BSA
   > PBS filtered

3. Make 2X mix of lineage-depleting biotin conjugated antibodies in MACS buffer, 500 μL of per sample:

	1X (500 μL)
B220	3 μL
CD4	2 μL
Gr-1	" μL
Ter119	"
CD11b	1 μL
CD11c	"
DX5	"

4. Add biotin antibody mix to sample, incubate on ice 20 min, wash 2X

5. Add 100 μL MACS buffer and 10 μL anti-biotin magnetic beads per 1 x 10^7 cells (or as specifically recommended), incubate on ice 10 min, wash 2X

6. Resuspend in 500 μL

7. Wet LS column w/ 3 mL MACS buffer

8. Add sample through filter onto column

9. Wash 3 x 3 mL into same collection tube and spin down cells from entire flow-through

10. Resuspend in 1 mL complete media for culture, count cells, resuspend in complete media at 1×10^6 cells/ml and plate 200 μL/well (2×10^5 cells)

Preparation of mRNA libraries for RNA-Seq with Illumina TruSeq Kit

Day 0:
Isolate RNA and measure RNA concentration. Freeze RNA at -70°C until use.
Day 1: Should NOT stop until finished Step 3: Plan to use approx 10 hours.
RNA sample date: _____
Decide on amount input RNA (e.g. 400 ng): _____
Dilute each sample to 50 ml with RNA-grade ddH$_2$O (can use from Qiagen RNaeasy Kit):
Add 50 ml **RNA-binding Beads** and mix by pipetting.
*While starting this, thaw **EFP and FSM-mix**

➔ 65 °C for 5 min ➔ cool to 4 °C
➔ RT for 5 min ➔ Magnet for 5 min

Follow protocol for wash, elute, re-binding w/ binding buffer, wash and last elute from **pages 44-46** in Illumina's "TruSeq RNA Sample Preparation Guide"

➔ **ELUTE in 19.5 ul EFP (Elute, Fragment, Prime) mix (pg 47)**:
➔ Magnet for 5 min ➔ remove 17 ul supernatant and proceed directly to next step.

Step 2: First Strand Synthesis:

RT master-mix	1x	x
FSM (First Strand Synthesis Mix):	7 ml	_____ ml
Home-made RT enzyme:	1 ml	_____ ml

            17 ml from magnet in Step 1 (EFP mix)
                  + 8 ml of RT-mix (above)
                  25 ml rxn volume

        PCR program (total time approx 1h 20 min):
Program "1ˢᵗ" under "Main" on Chem bench PCR machine
25 °C for 10 min
                  42 °C for 50 min
                  70 °C for 15 min
                  4 °C forever

Clean up with AmPure SPRI beads (2X volume): **+ 50 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **32.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ Remove 30 ml

Step 3: Strand-Specific 2nd Strand Synthesis
(Assuming Strand-specific library prep. If not, follow protocol on pages 47-50 in in Illumina's "TruSeq RNA Sample Preparation Guide".)

	1 x	x
sample	30 ml	_____
5 x First Strand Buffer (Invitrogen cat # P/N y02321): 4 ml	_____	
0.1 M DTT:	1 ml	_____
10 mM dUTP/dVTP:	1.5 ml	_____
10 x Nucleotide Free 2nd Strand Buffer (NEB, B6117S): 8 ml	_____	

2nd Strand Enzyme Mix (NEB, cat # E6111S):   4 ml   _____
ddH$_2$O: 31.5 ml   _____
Total volume: 80 ml   ____ x 50 ml
➔ 16 °C for 2 hours

Clean up with AmPure SPRI beads (2X volume): **+ 160 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **52.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ Remove **50 ml**

This is first safe stopping point. You may now either freeze sample or do next step (End Repair) and stop and freeze after purification of End Repair mix.

Step 4: End Repair:

Sample (from above):   50 ml
RSB: 10 ml
End Repair Mix: 40 ml
Total: 100 ml
➔ 30 °C for 30 min

Clean up with AmPure SPRI beads (1.6X volume): **+ 160 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **17.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ Remove **15 ml**

This is another safe stopping point.

Day 2:

Step 5: A-Tailing:

Sample: 15 ml
RSB   + 2.5 ml
A-Tailing mix: + 12.5 ml
Total rxn volume =   30 ml

➔ 37 °C for 30 min
➔ Proceed directly to adapter ligation

Step 6: Adapter Ligation:

Sample: 30 ml
RSB   + 2.5 ml
Lig. mix + 2.5 ml
Adapter + 2.5 ml
Total rxn =   37.5 ml
➔ 30 °C for 10 min
➔ **+ 5 ml STOP Ligation**

93

➔ Purify with 2X clean-up with AmPure SPRI beads:

#1 Clean up with AmPure SPRI beads (1X volume): **+ 42.5 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **52.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ Remove **50 ml**

#2 Clean up with AmPure SPRI beads (1X volume): **+ 50 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **22.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ Remove **20 ml**

Step 7: For strand-specific Sequencing: USER (NEB, cat# M5505L) enzyme reaction

Sample: 20 ml
USER enzyme   +   1 ml
➔ 37 °C for 15 min
➔ 95 °C for 5 min (Heat inactivation of USER)

Step 8: PCR amplification:

	1x	x
Sample:	21 ml	
Primer cocktail:	5 ml	
PCR master mix:	25 ml	
Total volume =	51 ml	

PCR program: *Program "R-Seq-PC" on far side lab bench PCR machine* (41 min total)
        95 °C for 30 sec
        95 °C for 10 sec
        60 °C for 30 sec
        72 °C for 30 sec
Goto step 2 14 times (for 15 cycles)
        72 °C for 5 min
         4 °C forever

Clean up with AmPure SPRI beads (1X volume): **+ 51 ml Ampure SPRI Beads**
➔ 15 min @ RT ➔ 5 min magnet
➔ 2 x 80% EtOH (Fresh!) wash ***keep on magnet during washes*** ➔ 15 min air dry
Elute in **32.5 ml RSB** (Re-suspension Buffer):
Remove from magnet ➔ resuspend entire volume ➔ 2 min @ RT ➔ Magnet for 5 min
➔ **Remove 30 ml**

Measure DNA concentration by nanodrop and by Qubit:
Qubit directions: Add 2 ul sample to 198 ul working solution

Run 50-75 ng on a 2% gel (with small wells):
Shows average size of fragments and quality of library
Dilute individual libraries to 10 nM in EB w/ 0.1% Tween
Combine (from 10 nM dilutions) for multiplexing in one lane.

RNA-Seq Formatting and Analysis

Outline:

A. Format data

B. Align with Tophat via Galaxy

C. Count reads per exon by SeqMonk

D. Calculate rpkms in Excel

E. Uploading to UCSC Genome Browser

A. Download and format data

1. Compile read files and barcode files into one file using the following command:

    for readfile: gzcat s_3_1* > samplename_readfile.txt.gz

    for barcodefile: gzcat s_3_2* > samplename_barcodefile.txt.gz

2. Paste them together and delete the extra columns:

    paste samplename_readfile.txt samplename_barcodefile.txt | awk '{ $12 = $13 = $14 = $15 = $18 = $19 =

    $21 = $22 = ""; print }' > samplenamecodedreads.txt

3. Separate out the barcodes (first 6 bases).

    awk ' $14 ~ "TAGCTT"' filein.txt > fileout.txt

    gzip > label_qseq.txt

4. Upload to Galaxy by FileZilla

    →log in using Galaxy username and password

    →Select files from local site and drag to remote site to initiate uploading.

5. Run preset workflow "Doty qseq to fastq (imported from uploaded file)" (from Kevin Doty) on Galaxy:

    →Go to Galaxy website, log in and click "get data", then click "upload file".

    →Click "Workflow"→choose "Doty qseq to fastq (imported from uploaded file)"→choose "run"→Select

    the data to be run→run workflow→once started, the jobs will be shown on the right→it takes several

    hours to finish, and for each job, only the last fastq file will be kept→download it and rename it with the

    sample name, ended with .fq

B. Align with Tophat via Galaxy

Select "NGS: RNA Analysis" → Select "Tophat for Illumina"→choose the .fq file→"Use a built-in

index→Select a reference genome→"Single End"→Tophat setting to use→Full Parameter list→Library

Type choose "FR First Strand"→ Maximum number of alignments to be allowed select

1→"Execute"→Keep the last file ("Tophat for Illumina accepted hits"), and rename it with sample name

ending in .bam

C. Count reads per exon by SeqMonk

1. Obtain annotation file from Galaxy for use in mapping

"Get data"→UCSC main table→choose assembly→Track=refSeq→Group Genes & Gene

Predictions→Output=BED file→Click "get output"

For exon only mapping click "exons plus". This will give start and stop of each exon as opposed to each

gene.

Download this file from Galaxy

2. Add new project:

Open SeqMonk→choose New Project→Select Mouse NCBI37

3. Import Annotation:

File→Import Annotation→select annotation file from step 1

4. Import Data

File→Import Data→BAM/SAM→select the files→Import option, check "split spliced reads"

5. Count reads

Data→Define probe→Feature Probe Generator→select the annotation from step1

Read Count Quantitation→select opposite strand to probe→Quantitate

6. Export the reads

Report→ Create Annotated Probe Report→OK

This file(.txt) will be used to sum up number of reads in exons and give total number of reads for each

gene, and sum exons lengths to give gene length using "awk" command in Terminal.

→Open this .txt file in Excel→Delete everything else except Probe, start, end, reads→replace all "_exon*"

to get RefSeq ID only→Insert column next to End→Enter =(end-start), this will give the exon

length→Copy the column, paste as value→save file→upload to Galaxy and download again to be in the

right format

7. Sum up reads per exon

for example, if you have 14 columns in your file, then type in command line:

awk '{a[$1]+=$2; b[$1]+=$3; c[$1]+=$4; d[$1]+=$5; e[$1]+=$6; f[$1]+=$7; g[$1]+=$8; h[$1]+=$9;

x[$1]+=$10; j[$1]+=$11; k[$1]+=$12; l[$1]+=$13; m[$1]+=$14; n[$1]+=$15 } END {for (i in a) print

i,a[i],b[i],c[i],d[i],e[i],f[i],g[i],h[i],x[i],j[i],k[i],l[i],m[i],n[i]}' All_terminal_upstream_2.txt >

All_terminal_upstream_2_SumExon.txt

8. Open the file named "All_terminal_upstream_2_SumExon.txt" in Excel for RPKM calculation.

D. Calculate RPKM in Excel

1. Delete all those genes <200bp

2. Get gene names from file "refseq_to_genename" by VLOOKUP.

Example: =VLOOKUP(A2,G:H,2,FALSE)

A2: refseq ID

G:H: array in refseq_to_genename file

2: show the content in the second column

FALSE: look for exact the same as in A2

3. Calculate RPKM (add 1 to every data to get rid of /0 reads)

   rpkm=(10e9*C)/(N*L)

   C=mapped reads

   N=number of total mapped reads in that same sample/experiment

   L=length of gene (exon only, cDNA length) in bp

4. Set cutoff rpkm, usually it is 1 or 2.

   Sort the max rpkm in each group of experiment and delete those genes with max<1.

5. Calculate fold change and set cutoff.

6. Organize genes into clusters using Cluster3

   To format, add "UNIQID" as the name for Refseq column→add "NAME" as the name for gene name column→delete all other columns except UNIQID, NAME, reads→save as Text (Tab delimited) file to import into Cluster3

   →Open file in Cluster3→ check "log transform data"→check "center genes"→check "mean"→check "cluster"→check "Euclidean distance"→check "organize genes"→type in 10 numbers of clusters→click "Execute"→get two files: .CDT and .kgg files→Open CDT file in Excel and delete second row→Open kgg file and copy second column to CDT file→color format the data and adjust the row height to create heatmap

Command lines for running DESeq (with replicates) in R studio:

```
library(DESeq)
setwd("~/filename")
CountTable<-read.table(file="filename.txt", header=TRUE, row.names=1)
View(CountTable)
head(CountTable)
Design<-data.frame(
row.names = colnames(CountTable),
condition = c("condition1_replicate1","condition1_replicate2","condition2_replicate1","condition2_replicate2")
libtype = c("single-end", "single-end"))
Design
conds <- factor(c("condition1","condition2"))
cds<-newCountDataSet(CountTable,conds)
head(counts(cds))
cds<-estimateSizeFactors(cds)
sizeFactors(cds)
head(counts(cds,normalized=TRUE))
cds<- estimateDispersions(cds)
res <-nbinomTest(cds, "condition1","condition2")
write.table(res,file="newfilename.txt")
```

Command lines for running DESeq (without replicates) in R studio:

```
library(DESeq)
setwd("~/filename")
CountTable<-read.table(file="filename.txt", header=TRUE, row.names=1)
View(CountTable)
head(CountTable)
Design<-data.frame(
row.names = colnames(CountTable),
condition = c("condition1","condition2"),
libtype = c("single-end", "single-end"))
Design
conds <- factor(c("condition1","condition2"))
cds<-newCountDataSet(CountTable,conds)
head(counts(cds))
cds<-estimateSizeFactors(cds)
sizeFactors(cds)
head(counts(cds,normalized=TRUE))
cds<- estimateDispersions(cds, method="blind", sharingMode="fit-only")
res <-nbinomTest(cds, "condition1","condition2")
write.table(res,file="newfilename.txt")
```

Section 3:

Kurdistani lab ChIP-seq protocol adapted for Ikaros IP on murine thymocytes

<u>**Day 1:**</u>
Process thymus samples individually to generate a single cell suspension:
- Mince with scissors, pulverize with plunger end from syringe, rinse through filter with PBS, collect in 50 mL conical, spin 5 min at 1400 rcf
- Each thymus from a 4 week old wildtype mouse will yield approximately 2 x 10^8 cells, but *Ikzf1*$^{\Delta F4/\Delta F4}$ will be lower, closer to 1 x 10^8

<u>DSG (disuccinimidyl glutarate) crosslinking:</u>
Resuspend cells in 12.5 ml media in a 50 ml conical tube
- Prepare 0.5 M DSG (MW=326.26, 50 mg (Pierce), resuspended in 153 µl sterile DMSO)

Add 50 µl of 0.5M DSG (final concentration= 2 mM) and mix gently and immediately
Incubate 45 min on end-over-end rotator at room temp (RT)
Wash cells 3x with 10 ml PBS at RT
- Note: Spin at 1800 rcf=3000 rpm

<u>Formaldehye crosslinking:</u>
Resuspend cells in 40 ml media in a 50 ml conical tube
- Prepare 2.5 M Glycine: 3.753 g Glycine in 20 ml water (mix on rotator to resuspend)

Fix in 1% formaldehyde:
- Add 1112 **µ**l of 37% Formaldehyde (final 1%)

Incubate 10 min on end-over-end rotator at RT
Stop reaction (after 10 min) by adding 2.1 ml 2.5 M Glycine (final 0.125 M)
Mix at RT for 5 min and spin down cells at 4C (1800 rpm)
- Note: Discard media containing 1% formaldehyde as hazardous waste

Wash cells 3x with 10 ml PBS at RT, then transfer to eppendorf tube for last 1ml wash.
Spin down cells after last wash, remove supernatant, snap-freeze cell pellet in liquid nitrogen and store at -80 **°**C.

<u>Day 2:</u>
<u>Sonication:</u>
Note: Keep all chromatin on ice @ 4 °C at all times, unless otherwise noted.
Resuspend frozen cell pellet in 600 µl Lysis buffer (containing SDS) plus protease inhibitor cocktail (Roche complete)
Incubate 10 min on ice to release nuclei
Sonicate in Cup Horn sonicator (Qsonica)
- Notes: connected to Misonex 3000, Lauda E100 water bath circulation):

Transfer nuclei suspension to 5 mL polystyrene tube
Suspend with top of tube above water level and centered within cup horn ring just below edge
3 x 10 min cycles at 40 sec ON (amplitude= 80), 30 sec OFF
- Note: Transfer nuclei suspension to new 5 mL polystyrene tube after each cycle

Transfer to new eppendorf tube after last cycle, spin down at max speed to remove cell debris, and transfer to new eppendorf tube.

Test sonication on gel:
Remove 10 µl of sonicated chromatin sample
Dilute with 88 µl Dilution buffer and 12 µl 5 M NaCl
Incubate at 90 °C for 20 minutes
Add 1 µl DNase free RNase and incubate at 37 °C for 10 min

Purify DNA (from protein) with Qiagen PCR purification columns
Measure DNA concentration with nanodrop and run 1-3 **µ**g DNA on 1 % gel with ladder
Sonication fragments should appear in a tight concentrated band averaging 200 bp

Immunoprecipitation:
Measure chromatin concentration with nanodrop
Note: Remove 1-5% chromatin for input control separate from immunoprecipitation
Dilute sonicated chromatin (100 µg is enough) up to 5 ml in dilution buffer with fresh protease inhibitor cocktail in a 15 mL conical
  - Optional: Pre-clear with Protein A magnetic dynabeads
  - Wash beads 3 times with 1 ml dilution buffer using a magnetic rack and resuspend beads in the same volume they were in the original buffer.
  - For each IP add 50 µl washed beads and incubate the precleaning step for 1 hr at 4°C on an end-over-end rotator
  - Collect beads by adding 1.5 ml to eppendorf tube on magnetic rack. Transfer and combine (4 times) the pre-cleared chromatin to a new 15 ml conical and discard the beads.

Add 50 **µ**l α-IK-C antibody
Incubate at 4°C overnight (O/N) on an end-over-end rotator

Day 3:
IP with Protein A magnetic beads:
Pre-wash Protein A magnetic beads as described above
  - Add 100 µl beads per IP and incubate 2 hrs at 4°C on an end-over-end rotator

Collect beads by adding 1.5 ml to eppendorf tube on magnetic rack 4 times to combine

Washing beads:
Notes: After adding each wash buffer, vortex gently/briefly to resuspend pellet, invert several times to mix, pulse spin down, collect on magnet for 2 min and discard wash buffer. Remove wash buffer using a pipette, not aspirator, to avoid losing beads.
  - 2 x 1 ml Wash Buffer A
  - 2 x 1 ml Wash Buffer B
  - 2 x 1 ml LiCl Buffer
  - 2 x 1 ml TE
  - Note: TE buffer removal needs to be done carefully to avoid losing beads

Elution:
Add 2 x 100 µl Elution buffer, at 65°C for 10 min
  - Note: Vortex the beads every 5 min to avoid precipitation

Collect beads on magnetic rack. Transfer DNA elution to new eppendorf tube

Reverse crosslinking:
Add 24 µl 5 M NaCl to 200 µl DNA elution
Thaw input, add Elution buffer up to 200 µl + 24 µl 5 M NaCl
Incubate in 65°C oven O/N

Day 4:
RNAse treat and purify DNA:
Add 1 **µ**l of 10 **µ**g/**µ**l (1**µ**g) RNase
Incubate at 37°C for 30 min
Purify DNA using phenol/chloroform extraction:
- Add one volume of phenol:chloroform:isoamyl alcohol (25:24:1), and vortex or shake by hand thoroughly for approximately 20 seconds.
- Centrifuge at room temperature for 5 minutes at 16,000 × g. Carefully remove the upper aqueous phase, and transfer the layer to a fresh tube. Be sure not to carry over any phenol during pipetting.

Ethanol precipitation:
Add the following reagents to the aqueous phase, in the order listed:

Reagent	Volume
Glycogen (20 µg/µL)	1 µL (20 µg)
7.5 M NH$_4$OAc	0.5 × volume of sample
100% ethanol	2.5 × volume of sample +NH$_4$OAc

Place the tube at –20°C overnight (or at –80°C for at least 1 hour) to precipitate the DNA from the sample.
Centrifuge the sample at 4°C for 30 minutes at 16,000 × g to pellet the DNA.
Carefully remove the supernatant without disturbing the DNA pellet.
Add 150 µL of 70% ethanol.
- Centrifuge the sample at 4°C for 2 minutes at 16,000 × g and arefully remove the supernatant.
- Repeat once. Remove as much of the remaining ethanol as possible.

Dry the DNA pellet at room temperature for 5–10 minutes.
Resuspend the DNA pellet in desired amount of TE by pipetting up and down 30–40 times.

Buffers:
Lysis Buffer:1% SDS, 50 mM Tris-HCl pH 8, 20 mM EDTA, and freshly added protease inhibitors (Roche, one mini-complete tablet for 10 ml)

Dilution Buffer: 16.7 mM Tris-HCl, pH 8, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl

Wash Buffer A: 50 mM HEPES, pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 140 mM NaCl

Wash Buffer B: 50 mM HEPES, pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 500 mM NaCl

LiCl Buffer: 20 mM Tris-HCl, pH 8, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, 250 mM LiCl

Wash Buffer C: 5 mM Tris-HCl pH 8, 0.125% NP-40, 0.125% Dioxycholate, 0.25 mM EDTA, 62.5 mM LiCl

TE Buffer: 10 mM Tris-HCl, pH 8.0, 1 mM EDTA

Elution Buffer: 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 1% SDS

Appendix II: Supplementary Figures

Supplementary Figure 1a: Predicted DNA-recognition preferences by zinc finger amino acids at different positions in the α-helix (based on probability table[37])
Predicted consensus: Bases with a frequency 30-40% are lowercase, 40-60% are capitalized, and >60% are bold.

Predicted consensus:

C g C T T G G A A [g/t] [G/t] N T

α−Helix Positions:	-1	2	3	6
ZnF1:	CDICGIVCIGPNVLMVHKRSH			
ZnF2:	CNQCGASFTQKGNLLRHIKLH			
ZnF3:	CHLCNYACRRRDALTGHLRTH			
ZnF4:	CGYCGRSYKQRSGLEEHKERCH			

α−Helix Positions:	-1	2	3	6
ZnF1:	G	N	V	V
A	0.247	0.495	0.074	0.232
C	0.248	0.187	0.103	0.053
G	0.258	0.159	0.44	0.364
T	0.247	0.16	0.384	0.352
ZnF2:	Q	G	N	R
A	0.402	0.158	0.705	0.011
C	0.015	0.417	0.085	0.012
G	0.338	0.068	0.139	0.948
T	0.246	0.357	0.071	0.029
ZnF3:	R	D	A	G
A	0.044	0.049	0.048	0.297
C	0.023	0.928	0.085	0.125
G	0.878	0.011	0.083	0.125
T	0.055	0.012	0.784	0.453
ZnF4:	Q	S	S	E
A	0.218	0.132	0.242	0.146
C	0.685	0.326	0.201	0.566
G	0.031	0.422	0.328	0.23
T	0.066	0.12	0.228	0.057

Supplementary Figure 2a: pGEX-4T-1 cloning vector

pGEX-4T-1 (27-4580-01)

Thrombin
Leu Val Pro Arg↓Gly Ser Pro Glu Phe Pro Gly Arg Leu Glu Arg Pro His Arg Asp
CTG GTT CCG CGT GGA TCC CCG GAA TTC CCG GGT CGA CTC GAG CGG CCG CAT CGT GAC TGA
BamH I   EcoR I   Sma I   Sal I   Xho I   Not I   Stop codons
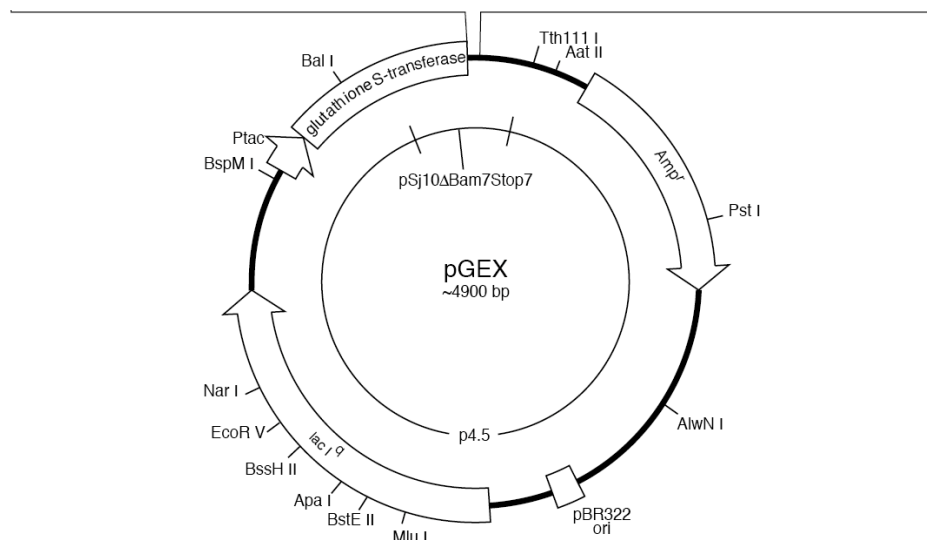
Supplementary Figure 3a: EMSA screen of PBM motif and mutant oligos
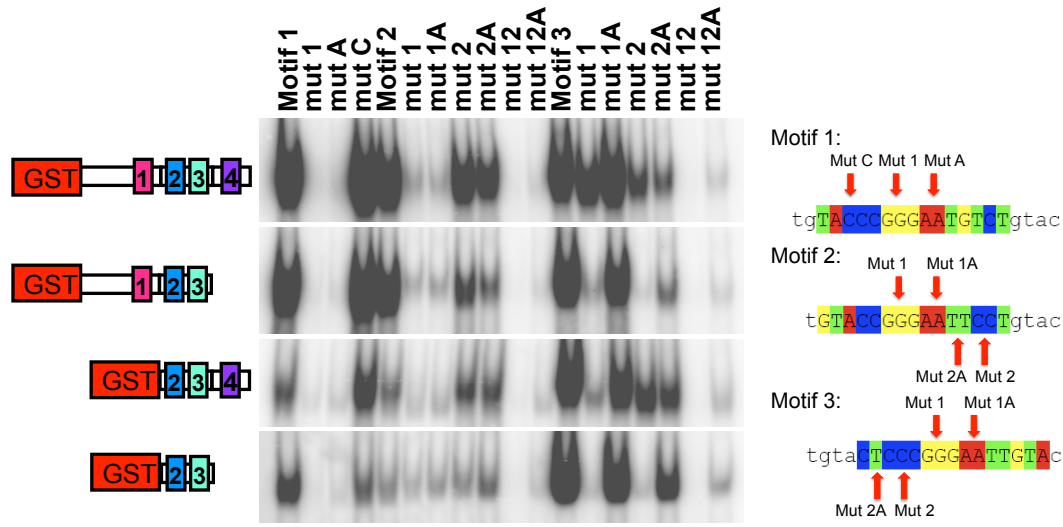
Supplementary Figure 3b: EMSA screen of single site isolation oligos
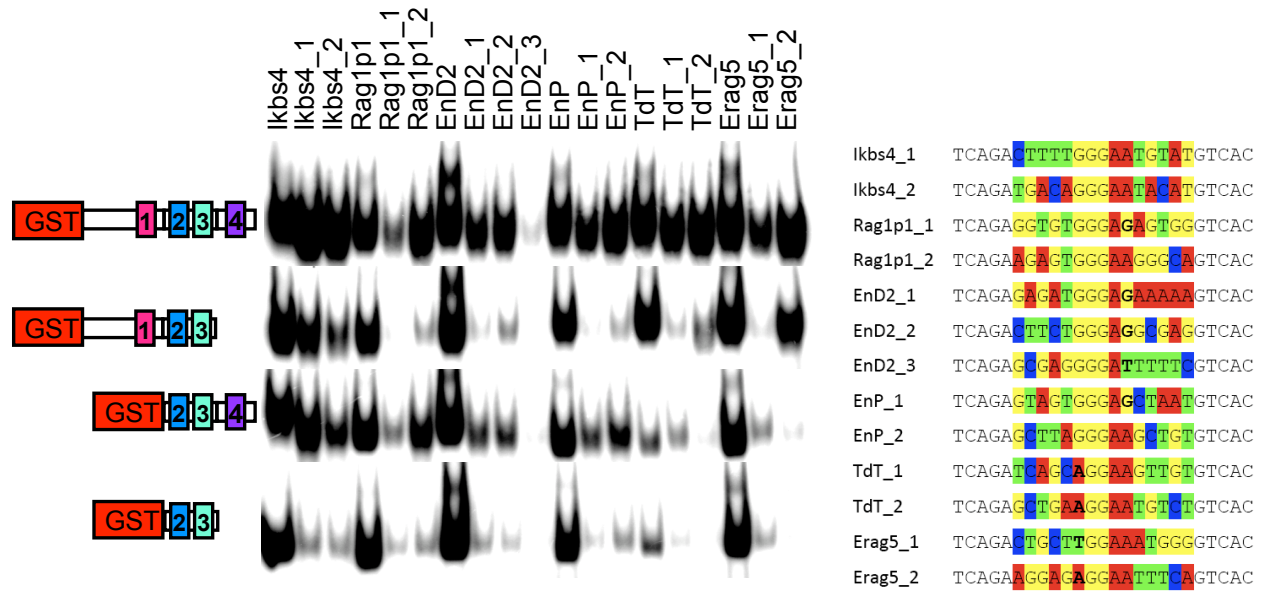
TABLE 1. Oligonucleotides used as probes in the DNA binding studies

Probe	Sequence
IK-BS1	TCAGCTTTTGGGAATACCCTGTCA
IK-BS2	TCAGCTTTTGGGAATCTCCTGTCA
IK-BS3	TCAGCTTTTGGGAATTCCCTGTCA
IK-BS4	TCAGCTTTTGGGAATGTATTCCCTGTCA
IK-BS5	TCAGCTTTTGAGAATACCCTGTCA
IK-BS6	TCAGCTTTTGGGATTACCCTGTCA
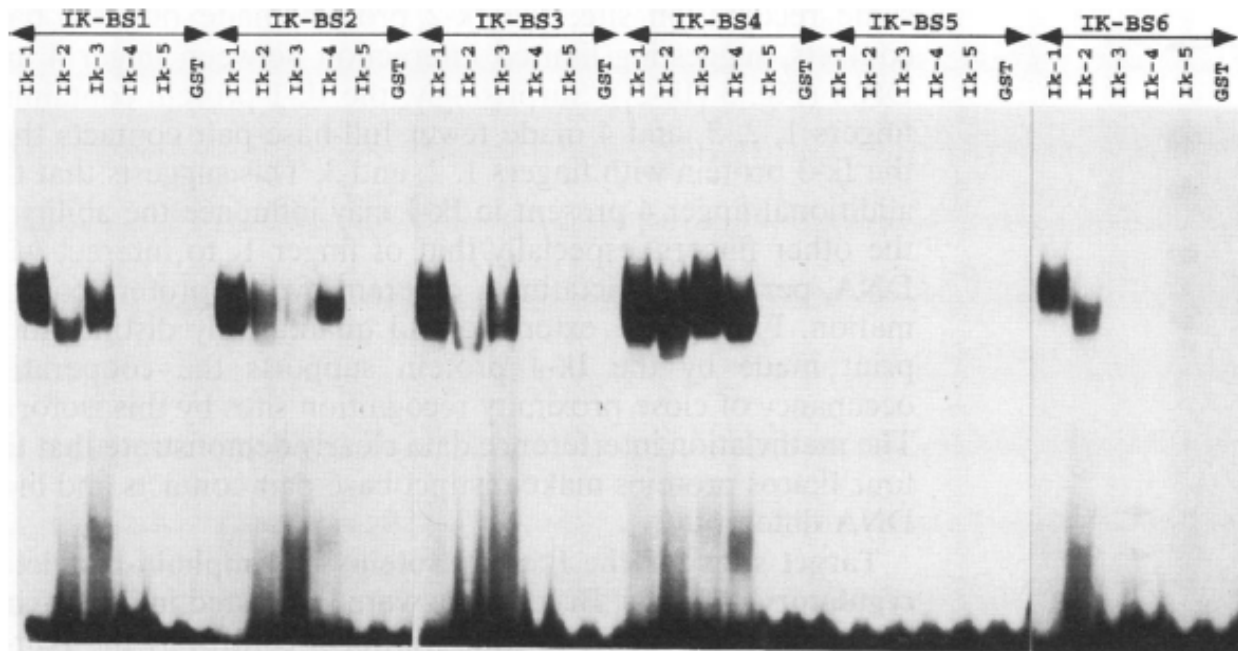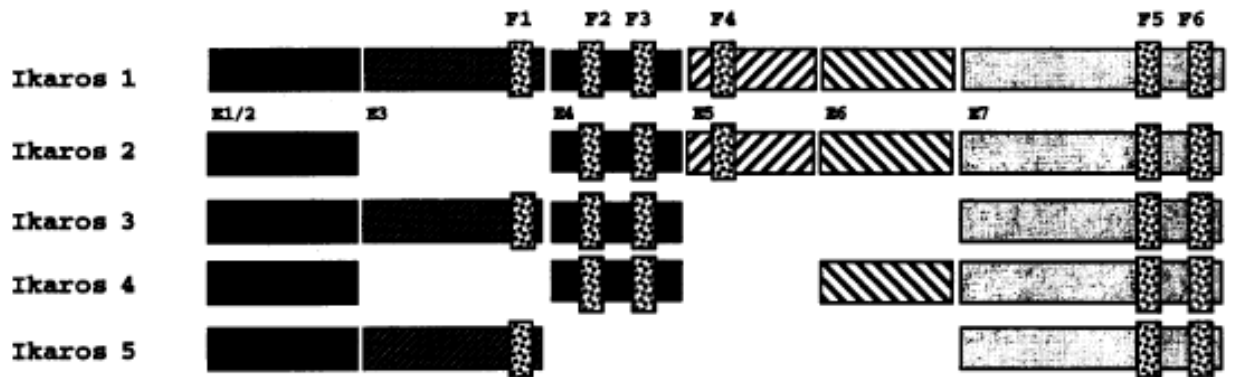
Supplementary Figure 3d: Affinity analysis for PWM calculation

Bind all:	[Note: Ikbs4 and Mot3 have 2 potential binding sites]	IK 1-4	IK 1-3	IK 2-4	IK 2-3
Ik bs 4	GCTTTTGGGAATGTATTCCCTGTCA TGACAGGGAATACATTCCCAAAAGC	++++	++++	++++	++++
Ik Motif 3	TGTACTCCCGGGAATTGTAC GTACAATTCCCGGGAGTACA	++++	++++	++++	++++
Ik Mot 3 mut 1A	TGTACGCCCGGGAATTGTAC	++++	++++	++	++++
Ik bs 2	TCAGCTTTTGGGAATCTCCTGTCA	++++	+++	+	++
Any 3 ZnF:					
Ik bs 1	TCAGCTTTTGGGAATACCCTGTCA	++++	++++	++	-
Ik Motif 1 mut C	TGTAACCGGGAATGTCTGTAC	++++	++++	++	-
Ik bs 3	TCAGCTTTTGGGAATTCCCTGTCA	++++	++++	++	+
Ikbs4_1	TCAGACTTTTGGGAATGTATGTCAC	++++	++	+++	+
Ikbs4_2	TCAGATGACAGGGAATACATGTCAC	++++	++	++	+
Hes1_3'	AGAAAGTTTGGGAAGTTTCACACGAG	+++	++	+++	+
Ik Motif 2 mut 2	TGTACCGGGAATTAATGTAC	+++	+	+	-
w/ ZnF1:					
Ik Motif 1	TGTACCCGGGAATGTCTGTAC	++++	++++	+	+
Ik Motif 2	TGTACCGGGAATTCCTGTAC	++++	++++	-	-
Erag5_2	TCAGAAGGAGAGGAATTTCAGTCAC	+++	+++	-	-
TdT_2	TCAGAGCTGAAGGAATGTCTGTCAC	+++	++	-	-
w/ZnF4:					
EnP_2	TCAGAGCTTAGGGAAGCTGTGTCAC	++++	+	+++	+
Rag1p1_2	TCAGAAGAGTGGGAAGGGCAGTCAC	+++	+	+++	+
EnD2_1	TCAGAGAGATGGGAGAAAAAGTCAC	+++	-	++	+
EnD2_2	TCAGACTTCTGGGAGGCGAGGTCAC	+++	+	++	-
EnP_1	TCAGAGTAGTGGGAGCTAATGTCAC	+++	-	++	-
Dltx1_1424	TCACTCTTGTGGGAAGTCGGACAAA	++	+	++	+
Hes1_5'	GAAAGTTACTGTGGGAAAGAAAGTTT	++	-	++	-
Rag1p1_1	TCAGAGGTGTGGGAGAGTGGGTCAC	++	-	+	-
Ik bs 6	TCAGCTTTTGGGATTACCCTGTCA	++	-	+	-
Ik Mot 2 mut 2A	TGTACCGGGAAGGCCTGTAC	++	-	+	-
Ik Motif 3 mut 2	GTACAATTAAAGGGAGTACA	+	-	+	-

Supplementary Figure 3e: ZnF4 affinity analysis with cleaved proteins
Variation in 5' ends for F4 selection; 3' ends continue to the right where not shown.
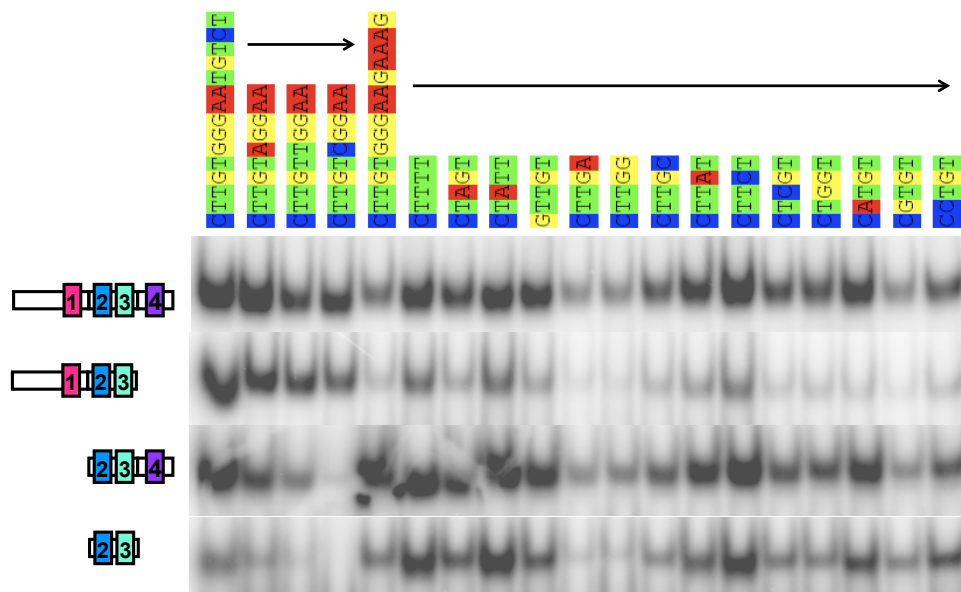
Supplementary Figure 4a-d: First round SPR
Second order reaction modeling for each protein titration binding to ZnF1+ZnF4+ sequence

a. IK 1-4

Name	k_a	k_d	K_D
1-4 62.5	1.565e5	3.164e-3	20.2212nM
1-4 125	1.715e5	4.741e-3	27.6512nM
1-4 250	2.94e5	6.842e-3	23.2736nM
1-4 500	2.075e5	8.471e-3	40.823nM
1-4 1	1.068e5	9.509e-3	89.0291nM

b. IK 1-3

Name	k_a	k_d	K_D
1-3 125	4(1)e5	0.1828	500(200)nM
1-3 250	4.6(7)e5	0.1936	420(60)nM
1-3 500	4.1(4)e5	0.2073	510(50)nM
1-3 1	2.2(2)e5	0.2169	1.00(9)uM

c. IK 2-4

Name	k_a	k_d	K_D
2-4 62.5	1(1)e5	0.01094	70(60)nM
2-4 125	5(4)e4	0.01057	200(200)nM
2-4 250	1.09(9)e5	0.01328	120(10)nM
2-4 500	3.5(2)e5	0.01637	47(2)nM
2-4 1	3.7(2)e5	0.01798	49(2)nM

d. IK 2-3

Name	k_a	k_d	K_D
2-3 250	4.054e5	0.38(2)	940.177nM
2-3 500	3.249e5	0.449(8)	1.38113nM

Supplementary Figure 4e-h: Second round SPR
Second order reaction modeling for each protein titration binding to ZnF1+ZnF4+ sequence

e. IK 1-4

t1/2□40.46□31.07□25.88□20.04□17.98
Avg t1/2 = 27.086

Name	k_a	k_d	K_D
1-4 62.5	18132.2	0.01713	944.735nM
1-4 125	65162.3	0.02231	342.344nM
1-4 250	3.481e5	0.02678	76.9451nM
1-4 500	1.38e6	0.03459	25.0586nM
1-4 1	4.057e6	0.03855	9.50276nM

f. IK 1-3

t1/2□1.46□1.391□1.199□0.9191□1.33□
Avg t1/2= 1.26

Name	k_a	k_d	K_D
1-3 62.5	9.865e6	0.4749	48.1416nM
1-3 125	8.705e6	0.4981	57.2236nM
1-3 250	7.866e6	0.5782	73.5028nM
1-3 500	6.76e6	0.7541	111.562nM
1-3 1	3.006e6	0.5211	173.362nM

g. IK 2-4

t1/2□11.06□11.22□10.42□9.173□8.766□
Avg t1/2 = 10.13

Name	k_a	k_d	K_D
2-4 62.5	41109.2	0.06269	1.5249uM
2-4 125	86225.8	0.06179	716.611nM
2-4 250	2.305e5	0.06655	288.759nM
2-4 500	7.498e6	0.07557	100.783nM
2-4 1	2.49e6	0.07907	31.7566nM

h. IK 2-3

t1/2□1.188□1.478□1.192□0.653□0.5892
Avg t1/2 = 1.02

Name	k_a	k_d	K_D
2-3 62.5	4(5)e4	0.58(4)	10(20)uM
2-3 125	1.078e5	0.469	4.34905uM
2-3 250	1.639e5	0.5815	3.54827uM
2-3 500	2.792e5	1.061	3.80184uM
2-3 1	1.845e7	1.176	63.7536nM

Supplementary Figure 7a: Representative FACS staining showing B cell, macrophage and erythrocyte lineage depletion by MACS
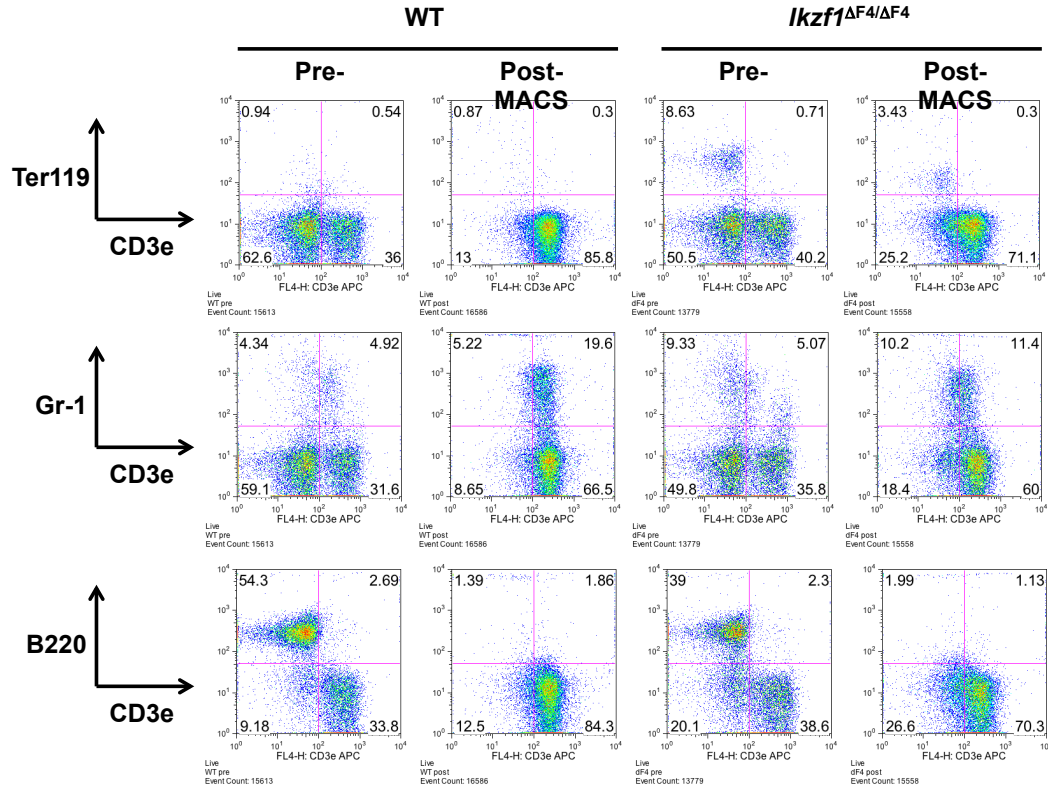
Supplementary Figure 8a: Gene ontology (GO) function analysis for genes in Groups 3, 5, and 9 from Figure 8c.

Group 5 (WT High, early)

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Exocytosis	2.91	1.11E-04	8
domain:EF-hand 2	1.83	1.02E-02	7
domain:C2 1	1.78	4.06E-03	5
domain:DEP	1.67	1.58E-02	3
Zinc finger region:Phorbol-ester/DAG-type	1.64	9.22E-03	4
domain:N-terminal Ras-GEF	1.48	1.58E-02	3
Rab/Ras GTPase-binding domain	1.47	6.44E-03	3
domain:IQ	1.27	2.26E-02	4
GPCR, family 2, extracellular region	1.21	3.10E-02	3
Positive regulation of protein modification process	1.17	4.26E-02	4

Group 3 (ΔF4 High, early)

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Purine ribonucleotide binding	2.71	1.51E-03	40
Positive regulation of cytokine production	2.28	2.77E-04	7
Hemopoiesis	2.15	3.80E-03	11
Protein kinase, ATP binding site	1.68	1.27E-02	14
Transcription factor, T-box	1.61	2.23E-02	3
Ig-like C1-type	1.26	6.92E-03	4
Neutrophil chemotaxis	1.24	2.49E-02	3
Regulation of cell migration	1.20	4.54E-02	5
Cellular ion homeostasis	1.19	3.76E-02	9
Regulation of phosphorylation	1.16	6.22E-02	9
Regulation of production of molecular mediator of immune response	1.15	7.04E-02	3
Regulation of adaptive immune response	1.15	4.81E-02	4
Glycosaminoglycan binding	1.07	6.81E-02	5
Serine/threonine protein kinase, active site	1.07	6.13E-02	10
Protein amino acid phosphorylation	1.06	2.50E-02	17
Programmed cell death	0.95	8.58E-02	12
T cell activation	0.90	2.73E-02	6
bZIP transcription factor, bZIP-1	0.89	6.36E-02	3
Positive regulation of protein secretion	0.87	2.49E-02	3
Transmembrane receptor protein tyrosine kinase activity	0.83	3.96E-02	4

Group 9 (ΔF4 High, All)

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Cadherin	12.31	7.14E-14	14
Membrane fraction	5.29	2.38E-06	16
Salivary gland morphogenesis	2.66	6.42E-04	4
Blood vessel morphogenesis	2.48	1.59E-03	7
Metal ion binding	1.90	1.07E-02	37
Coagulation	1.84	9.88E-03	4
Galectin	1.50	9.52E-04	3
FERM/acyl-CoA-binding protein, 3-helical bundle	1.41	3.01E-02	3
Regulation of apoptosis	1.20	6.07E-02	8
T cell activation	1.07	6.19E-03	3

Supplementary Figure 8a Continued:

REGULATION OF APOPTOSIS/PROGRAMMED CELL DEATH GENE CLUSTER	
Group 3 (ΔF4 High, early)	Group 9 (ΔF4 High, All)
BCL2-like 2	CD74 antigen (invariant polypeptide of MHC class II antigen-associated)
BCL2/adenovirus E1B 19kD interacting protein like	Kv channel interacting protein 3, calsenilin
N-acylsphingosine amidohydrolase 2	cadherin 1
NLR family, CARD domain containing 4	carbohydrate sulfotransferase 11
amyloid beta (A4) precursor protein	caspase 1; hypothetical protein LOC100044207
death associated protein kinase 1	caspase 4, apoptosis-related cysteine peptidase
engulfment and cell motility 3, ced-12 homolog (C. elegans)	discs, large homolog 5 (Drosophila)
growth arrest and DNA-damage-inducible 45 gamma	transglutaminase 2, C polypeptide
interferon gamma	
kit oncogene	
phorbol-12-myristate-13-acetate-induced protein 1	
programmed cell death 1	

T CELL ACTIVATION GENE CLUSTER	
Group 3 (ΔF4 High, early)	Group 9 (ΔF4 High, All)
RIKEN cDNA 5830411N06 gene	CD74 antigen (invariant polypeptide of MHC, class II antigen-associated)
eomesodermin homolog (Xenopus laevis)	histocompatibility 2, class II antigen A, alpha, class II antigen E, alpha
growth arrest and DNA-damage-inducible 45 gamma	lymphocyte-activation gene 3
histocompatibility 2, class II, locus DMa	
integrin alpha M	
misshapen-like kinase 1 (zebrafish)	

Supplementary Figure 8b: Gene Ontology (GO) function analysis for genes with >10 fold higher expression in dF4 versus WT at t=0 for both data sets

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Coagulation	2.09	6.78E-03	4
Serine protease	1.64	1.36E-02	5
domain:Ig-like C2-type 1	1.62	7.46E-03	5
Endopeptidase activity	1.55	8.85E-03	8
Laminin G, subdomain 2	1.54	1.99E-02	3
Cell migration	1.22	4.36E-02	5
Peptidase S1A, chymotrypsin	1.19	4.58E-02	4
Hemopoiesis	1.19	4.99E-02	5
Calcium-binding region:1	1.16	3.45E-02	4
LIM domain	1.11	6.32E-02	3
Regulation of apoptosis	1.06	8.45E-02	7

Supplementary Figure 8c: Full PSCAN Z-score heatmap for overrepresented TF binding sites (p-value <0.05) within promoters of individual genes from Group 1 of Figure 8g

TF NAME	SP1	TBP	Zfp423	CREB1	NFKB1	PLAG1	Mafb	RELA	EBF1	RREB1	INSM1	Klf4	MZF1_1-4	ZEB1
Z-SCORE	3.069	2.961	2.773	2.742	2.654	2.460	2.210	2.163	2.080	1.996	1.919	1.797	1.756	1.644
P-VALUE	1.07E-03	1.53E-03	2.76E-03	3.03E-03	3.96E-03	6.93E-03	1.35E-02	1.52E-02	1.87E-02	2.29E-02	2.74E-02	3.61E-02	3.95E-02	5.00E-02
INDIVIDUAL GENE PROMOTER Z-SCORES:														
Crlf2	1.451	0.556	-1.390	1.700	-0.381	1.087	-0.828	-0.413	-0.768	2.064	1.874	1.371	0.373	0.575
Bbc3	1.451	-2.092	1.787	2.018	3.064	-0.283	0.302	2.785	0.413	0.795	0.965	1.186	0.821	0.940
Tnfaip2	1.451	-0.754	0.531	-0.160	1.323	-0.712	1.597	1.974	-0.554	-0.374	0.495	1.186	0.821	0.575
Hmga1	1.451	1.363	0.321	1.168	1.103	0.913	2.190	0.271	0.023	0.664	0.215	1.186	0.821	0.940
Hmga1	1.451	1.363	0.321	1.168	1.103	0.913	2.190	0.271	0.682	0.664	0.215	1.186	0.821	0.940
Hmga1-rs1	1.451	1.363	0.321	1.168	1.103	0.913	2.190	0.271	0.682	0.664	0.215	1.186	0.821	0.575
Il1b	1.391	1.476	-0.011	0.624	-0.532	0.814	0.055	1.653	2.774	0.246	1.131	0.493	0.163	-0.648
Id3	1.391	0.910	1.724	1.700	-0.049	0.546	0.260	-0.942	0.430	-1.271	0.427	0.229	0.373	0.575
Junb	1.391	1.942	1.157	0.846	0.655	2.083	0.812	1.113	-0.351	-1.092	1.337	0.464	0.821	0.940
Morf4l1	1.391	0.310	0.138	2.018	-0.532	0.813	-0.381	-1.484	1.348	0.798	0.970	0.493	0.821	0.836
Fos	1.391	1.848	0.645	0.601	1.750	0.896	2.190	0.713	-0.151	0.545	0.585	-0.444	0.163	0.575
6430527G18Rik	1.391	0.209	-1.045	2.623	-0.299	1.886	1.654	-0.455	-0.266	0.716	-0.269	0.354	0.373	-1.281
Cebpb	1.000	1.519	0.091	2.018	1.777	1.428	-0.381	-0.135	0.932	-0.190	1.840	0.174	0.373	0.940
Uchl1	1.000	0.201	-0.249	-1.139	0.228	0.189	0.617	0.738	-1.374	2.081	1.912	0.523	0.163	-0.544
Phlda1	1.000	0.084	2.569	0.445	0.914	0.250	0.812	1.130	1.709	1.116	2.267	1.186	0.821	0.940
Prr7	1.000	-0.107	0.793	0.933	0.185	0.113	0.560	-0.291	-0.680	0.751	1.479	0.189	-1.257	0.940
Hnrnpa0	0.696	-0.993	0.113	0.123	0.934	-0.318	1.154	0.271	1.233	-1.018	-0.898	0.418	0.821	-1.029
H2-t9	0.696	0.560	-0.361	-0.835	-0.380	-0.063	0.055	-0.890	-1.054	-0.125	0.859	1.325	0.821	0.940
Arhgap27	0.647	0.169	0.798	-0.322	2.113	0.775	-0.761	0.912	0.783	-0.392	0.036	0.354	-1.257	-0.544
Dusp1	0.647	0.576	0.749	2.623	1.116	-0.071	-0.592	0.149	-0.730	2.011	-0.513	1.371	0.163	0.940
Ier5l	0.619	-0.930	1.817	-0.866	0.246	0.801	-0.735	-0.612	0.783	-1.142	0.506	1.186	0.821	0.940
Lyz2	0.206	0.408	-0.416	-0.932	-0.977	-0.940	0.644	0.149	0.769	-1.064	-1.498	-1.199	-1.257	-0.648
Acaa1b	0.178	0.008	-1.128	-1.014	-0.807	-0.576	-1.418	0.789	0.154	1.926	-0.398	-0.342	0.163	0.940
Rpl36	-0.036	0.343	0.623	-0.478	-0.187	0.169	-0.292	-0.846	1.695	-0.503	-0.923	-0.129	-0.104	-1.281
Rplp1	-0.219	0.238	1.069	-0.727	0.403	-0.303	0.302	1.397	0.052	1.603	-1.162	-0.041	0.163	0.836
Ifng	-0.236	1.363	0.447	-0.770	-0.063	0.233	0.617	0.373	-0.439	1.303	-0.034	0.354	0.821	0.940
H2-Q2	-0.246	0.630	0.708	-0.322	0.228	-0.133	-0.735	0.738	0.408	0.214	0.000	0.603	0.821	0.940
Id1	-0.338	1.329	1.438	-1.554	-0.780	-0.015	1.154	-1.252	-0.064	-0.777	1.117	-0.083	0.373	-0.909
Romo1	-0.367	0.305	0.725	0.445	0.620	0.189	-0.735	-0.030	1.824	-0.785	-0.220	-1.421	0.373	-1.029
Romo1	-0.367	0.305	0.725	-1.461	0.620	0.189	-0.735	-0.030	1.824	-0.785	-0.322	-1.421	0.373	-1.281
Romo1	-0.367	-0.282	0.725	0.445	0.620	0.189	-0.735	-0.187	1.824	-0.075	-0.220	-0.160	0.373	-1.029
Ier2	-0.810	0.786	-0.192	2.623	0.278	1.469	-0.592	2.220	0.797	0.592	-0.746	0.937	0.821	0.836
Scand1	-0.860	1.473	0.841	-0.704	0.329	-0.171	1.296	0.347	-0.615	1.011	-0.074	0.189	-0.104	0.940
Jund	-1.323	-0.653	-0.331	0.528	-0.353	-0.144	1.154	-0.708	-0.709	1.352	-0.617	-1.596	-1.906	0.836
Irg1	-1.402	1.700	0.362	1.700	0.339	1.429	0.201	2.819	-1.069	0.293	0.816	-0.672	0.821	0.575

Supplementary Figure 8d: Gene Ontology (GO) function analysis for genes in Group 1 of Figure 8g

Group 1 (ΔF4 high, t=30)

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Regulation of interleukin-6 biosynthetic process	2.95	2.11E-04	3
Basic-leucine zipper (bZIP) transcription factor	2.81	7.38E-05	4
Positive regulation of macromolecule biosynthetic process	1.85	1.12E-02	5
Positive regulation of cytokine production	1.75	4.87E-03	3
Positive regulation of transcription	1.30	4.48E-02	4
Positive regulation of apoptosis	0.90	6.53E-02	3
Cytokine	0.76	2.34E-02	3

Group 3 (WT high, t=30)

CLUSTER FUNCTION	ENRICHMENT	P-VALUE	# GENES
Basic-leucine zipper (bZIP) transcription factor	3.62	5.14E-05	4
Regulation of transcription, DNA-dependent	3.13	2.58E-04	10
Transcription regulation	1.61	1.16E-02	7
Positive regulation of transcription from RNA polymerase II promoter	1.37	1.93E-02	4

Supplementary Figure 8e: Full PSCAN Z-score heatmap for TF binding sites (p-value <0.05) within promoters of genes from Group 5 of Figure 8g

TF NAME	MZF1_1-4	NF-kappaB	RREB1	RUNX1	NFKB1	Pax4	REST	Hand1::Tcfe2a	Myf	MZF1_5-13	E2F1	Arnt
Z-SCORE	2.598	2.538	2.173	2.075	1.865	1.846	1.833	1.817	1.769	1.765	1.725	1.684
P-VALUE	4.69E-03	5.56E-03	1.49E-02	1.89E-02	3.10E-02	3.24E-02	3.34E-02	3.46E-02	3.84E-02	3.88E-02	4.22E-02	4.61E-02
INDIVIDUAL GENE PROMOTER Z-SCORES:												
Sema4c	0.821	2.665	1.983	1.735	2.779	-0.489	0.760	-0.157	-0.151	-0.012	1.214	-0.150
Lmna	0.821	0.984	-0.388	1.853	-0.381	0.494	1.390	0.300	0.153	-0.116	-1.405	2.085
Lmna	0.821	1.351	-0.193	0.050	0.185	-0.263	-0.719	0.309	0.943	-0.019	0.846	-1.149
Marcksl1	0.821	-0.301	1.313	-0.012	0.549	0.254	0.968	0.604	0.428	0.780	0.846	0.582
Epha2	0.821	-0.035	1.291	0.487	-0.119	0.226	0.177	-0.551	-0.301	-0.302	-0.006	-0.417
Hk2	0.821	0.361	-0.552	1.137	0.475	-0.677	0.612	1.268	0.216	-0.544	0.846	0.582
Lgals7	0.821	-0.393	2.263	-0.389	0.293	0.427	1.899	0.062	1.619	0.926	-0.077	1.353
Plagl1	0.821	0.741	1.370	0.869	1.116	-0.091	0.514	0.652	0.269	1.237	2.439	2.085
Irf4	0.821	1.053	-0.445	-1.654	0.105	-0.119	0.107	-0.244	1.266	0.532	0.188	-0.417
Kcnk5	0.821	-0.452	0.740	1.503	-0.381	0.830	-0.730	0.830	0.817	-0.731	0.846	2.085
Tnfsf11	0.821	1.244	1.199	1.799	-0.049	0.762	0.186	0.392	-1.307	0.936	-0.184	-0.298
Trim36	0.821	0.417	1.148	-1.466	0.644	-0.268	0.679	0.656	-0.163	1.875	2.903	-0.150
Klf9	0.821	-0.572	-0.921	-0.842	0.208	0.867	0.718	0.513	0.653	0.653	0.188	2.085
Emp1	0.373	0.880	0.748	1.524	0.528	1.639	-0.429	1.650	2.225	1.041	-0.843	1.353
Slc6a8	0.163	-1.222	0.554	-1.338	-0.999	1.027	0.415	0.194	1.078	1.875	1.214	-0.417
Rnd1	0.163	2.845	-0.367	0.345	2.658	0.461	0.226	0.022	0.023	0.144	-1.599	-0.150
Trim36	0.163	-0.382	0.673	-0.582	-0.723	1.571	1.444	0.172	0.481	-0.587	-0.077	-1.149
Lphn2	-0.104	1.259	0.543	3.313	-0.597	1.218	-1.932	0.490	-0.851	0.560	-0.006	-1.149
Hes1	-0.104	0.625	-1.484	0.721	1.843	0.180	1.707	0.756	0.316	-0.557	0.188	0.582

nature immunology

Selective regulation of lymphopoiesis and leukemogenesis by individual zinc fingers of Ikaros

Hilde Schjerven[1,2,8], Jami McLaughlin[1], Teresita L Arenzana[1,2], Seth Frietze[3,8], Donghui Cheng[1,4], Sarah E Wadsworth[1,2], Gregory W Lawson[5,6], Steven J Bensinger[2,6], Peggy J Farnham[3], Owen N Witte[1,2,4,7] & Stephen T Smale[1,2,7]

C2H2 zinc fingers are found in several key transcriptional regulators in the immune system. However, these proteins usually contain more fingers than are needed for sequence-specific DNA binding, which suggests that different fingers regulate different genes and functions. Here we found that mice lacking finger 1 or finger 4 of Ikaros exhibited distinct subsets of the hematological defects of Ikaros-null mice. Most notably, the two fingers controlled different stages of lymphopoiesis, and finger 4 was selectively required for tumor suppression. The distinct defects support the hypothesis that only a small number of genes that are targets of Ikaros are critical for each of its biological functions. The subcategorization of functions and target genes by mutagenesis of individual zinc fingers will facilitate efforts to understand how zinc-finger transcription factors regulate development, immunity and disease.

In the post-genomics era, a critical goal is to define the target genes and mechanisms of action of transcription factors that contribute to the development of multicellular organisms, disease and the response to environmental cues. Although progress in this area has been rapid, a major hurdle is that many factors are involved in several biological pathways and can contribute to multiple steps in a single pathway, which makes it difficult to study key regulatory events in isolation. Ikaros is one transcription factor that regulates numerous biological events. Ikaros-null (*Ikzf1*[null]) mice lack B lineage cells, natural killer cells, peripheral lymph nodes and fetal T cells, and they exhibit many other abnormalities[1–3]. B cells are absent from adult *Ikzf1*[null] mice because of an inability of lymphoid-primed multipotent progenitor cells to develop into common lymphoid progenitor cells[4], but Ikaros also regulates later stages of B cell development[5–9]. Moreover, mice with mutations in the gene encoding Ikaros (*Ikzf1*) develop thymic lymphoma with high penetrance[10,11].

In humans, mutations at the *IKZF1* locus have been observed in a high percentage of samples of BCR-ABL[+] B cell acute lymphoblastic leukemia (B-ALL) and high-risk progenitor B-ALL[12,13]. *IKZF1* was the only gene for which mutations were found to be useful in predicting a poor response to therapy[13], yet the mechanisms responsible for tumor suppression by Ikaros remain poorly understood. Ikaros is thought to contribute to both the activation and repression of transcription, with the Mi-2–NuRD complex as a predominant interacting partner[14–16]. Notably, the target genes responsible for most of the key phenotypes of Ikaros-mutant mice remain unknown[7,17–19]. The widespread deregulation of gene expression in Ikaros-mutant cells and the finding that Ikaros binds several thousand genomic sites have increased the challenge of identifying the mechanisms by which Ikaros regulates lymphocyte development and leukemogenesis[16,20].

Members of the Ikaros family have a conserved DNA-binding domain near the amino (N) terminus, usually with four C2H2 zinc fingers[21]. Ikaros proteins also have two carboxy (C)-terminal C2H2 zinc fingers dedicated to dimerization and multimerization[22–24]. C2H2 zinc fingers are found in many ubiquitous transcription factors and in numerous proteins involved in tissue-specific and developmental stage–specific transcription[25–27], including PLZF, Bcl-6, Gfi1, Blimp-1 and ThPOK of the immune system.

Stable binding of DNA *in vitro* by a C2H2 zinc-finger protein usually requires two or three tandem fingers[27,28]. In Ikaros, fingers 2 and 3 are sufficient for stable binding and recognize the core consensus sequence GGGAA. In contrast, fingers 1 and 4 seem to modulate binding to specific sites[29–31]. In addition to expressing full-length Ikaros (Ik-1), most or all hematopoietic cells have abundant expression of a smaller isoform that lacks finger 1 (Ik-2). Ikaros isoforms that lack other fingers because of alternative splicing of pre-mRNA have been observed, but these isoforms are generally much less abundant except in transformed cell lines[29,32,33].

Given that two or three fingers are generally sufficient for stable binding of DNA[28], it is striking that most zinc-finger DNA-binding proteins contain more than three tandem fingers. The reasons for this phenomenon have been of interest since the early studies of the nine-finger protein TFIIIA[27,34,35]. All of the tandem fingers may be required for stable binding to key targets *in vivo*, with the multiple fingers recognizing extended DNA regions with high specificity. Alternatively, defined subsets of fingers may recognize distinct target genes. Initial support for the latter hypothesis emerged from *in vitro* protein-DNA–interaction studies showing that different combinations of fingers from TFIIIA, Ikaros, CTCF and other proteins can bind different DNA sequences[29,30,35–39]. In addition, studies of ectopically expressed mutant CTCF by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) have revealed that CTCF-binding sites can be separated into distinct classes on the basis of binding by different finger subsets[40].

To determine in a physiological setting whether multifinger DNA-binding proteins use different fingers to regulate different genes and biological functions, or whether all functions require the full complement of fingers, we generated two new Ikaros-mutant mouse strains with germline deletion of *Ikzf1* exons encoding fingers 1 and 4. Phenotypic analysis suggested that the two fingers regulate distinct biological events. Transcriptome profiling confirmed that the two fingers regulate distinct sets of genes, and DNA-binding analyses confirmed that fingers 1 and 4 can modulate binding to different genomic sites. Finally, *in vitro* and *in vivo* models of BCR-ABL[+] B-ALL established parallels between tumor suppression in thymocytes and that in B-ALL. Together these results suggest that multifinger DNA-binding domains exist at least in part to allow different fingers to regulate different genes. Moreover, the notably different phenotypes of the mutant strains suggest that each developmental step and biological function regulated by Ikaros is dependent on the proper regulation of only a small number of genes that are targets of Ikaros.

RESULTS

Generation of mutant mouse strains

We generated mice lacking Ikaros zinc-finger 1 (*Ikzf1*[ΔF1/ΔF1]) or zinc-finger 4 (*Ikzf1*[ΔF4/ΔF4]) by germline deletion of exon 4 or exon 6, respectively (**Supplementary Fig. 1a**). Each deletion generated a mutant protein lacking a zinc finger and a small number of adjacent

residues. Exon 4 encodes 87 residues, including the 23 residues of finger 1, 63 residues preceding the finger, and one residue following the finger. Exon 6 encodes only 41 residues. The 24-residue finger 4 is preceded by four residues that constitute the linker between fingers 3 and 4; these residues closely resembles the linker consensus sequence found in most zinc-finger proteins[28]. After the finger-encoding sequence, exon 6 encodes 13 additional residues; two prolines near the end of this region may separate the DNA-binding domain from an exon 7–encoded domain.

We confirmed correct targeting of the *Ikzf1* exons by Southern blot analysis, PCR (data not shown) and immunoblot analysis (**Fig. 1a**). *Ikzf1*[ΔF1/ΔF1] thymocytes lacked the full-length Ikaros isoform (Ik-1) but expressed Ik-2, which in wild-type mice is generated by alternative splicing of pre-mRNA[2]. In the *Ikzf1*[ΔF4/ΔF4] strain, deletion of exon 6 reduced the size of both Ik-1 and Ik-2 (**Fig. 1a**). Analysis of thymocyte mRNA by high-throughput sequencing technology (RNA-Seq) further confirmed correct targeting, as 'reads' from exons 4 and 6 were absent from the *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] samples, respectively (**Fig. 1b**).

Initial analysis of mutant strains

We initiated examination of the *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] phenotypes soon after the mice were generated but confirmed all phenotypes after backcrossing those mice onto the C57BL/6 background through nine or more generations. Efforts to backcross *Ikzf1*[null] mice have been unsuccessful (K. Georgopoulos and S. Winandy, personal communication), presumably because the null mutation results in embryonic death on the C57BL/6 background. The successful backcrossing of the *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] strains provided initial evidence that neither finger 1 nor finger 4 was required for all functions of Ikaros.

Initial studies revealed that the development of CD11b[+] myeloid cells was unperturbed in each strain, similar to the phenotype of *Ikzf1*[null] mice (**Fig. 1c** and **Supplementary Fig. 1b**). Conventional B cells were also present in the bone marrow and spleen of each mutant strain, albeit in smaller amounts than those in wild-type mice (**Fig. 1d** and **Supplementary Fig. 1c,d**). The presence of B cells was in contrast to the absence of all cells of the B lineage in *Ikzf1*[null] mice. These results suggested that DNA-binding fingers 2 and 3 were sufficient for the successful progression of a fraction of hematopoietic progenitors through B cell development.

Figure 1 Adult B cells are present in mice with germline deletion of *Ikzf1* exons encoding zinc-finger 1 or 4. (**a**) Immunoblot analysis of extracts of wild-type (WT), *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] thymocytes; *, unidentified form of Ikaros, most probably a post-transcriptionally modified form of Ik-1 (and Ik-2 after deletion of exon 4). SNP70 serves as a loading control. (**b**) RNA-Seq analysis of mRNA from wild-type, *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] thymocytes (*n* = 4 mice per strain); red outlined areas indicate targeted exons 4 (*Ikzf1*[ΔF1/ΔF1]) and 6 (*Ikzf1*[ΔF4/ΔF4]). (**c**) Expression of CD11b by myeloid cells in bone marrow from 6-week-old wild-type mice (*n* = 21), *Ikzf1*[ΔF1/ΔF1] mice (*n* = 10) and *Ikzf1*[ΔF4/ΔF4] mice (*n* = 14). Numbers above bracketed lines indicate percent CD11b[+] myeloid cells (average ± s.d.). (**d**) CD19[+] cells (left) and CD19[+]B220[+] cells (right) of the B lineage in the bone marrow and spleen of 6-week-old wild-type mice (*n* = 12–20), *Ikzf1*[ΔF1/ΔF1] mice (*n* = 9–18) and *Ikzf1*[ΔF4/ΔF4] mice (*n* = 6–17), presented as total cells (left and middle) and frequency of live cells (right). Each symbol (right) represents an individual mouse; small horizontal lines indicate the mean (± s.e.m.). Data are representative of three experiments (**a**), four experiments, (**b**) fourteen experiments (**c**) or five to fifteen experiments (**d**; average and s.e.m., left and middle).

Figure 2 B cell development is disrupted at different stages in $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice. (**a**) B cell development in whole bone marrow from 6- to 8-week-old mice: top row, total bone marrow; second row, pre-pro B cells to large pre-BII cells (B220+CD43+); third row, Lin−B220+CD43+CD19− cells (for analysis of pre-pro-B cells); bottom row, small pre-BII cells to mature B cells (Mat B; B220+CD43−). Lineage markers included CD3, Ter-119, Gr-1 and CD11b. Imm B, immature B cells. (**b**) Absolute number of c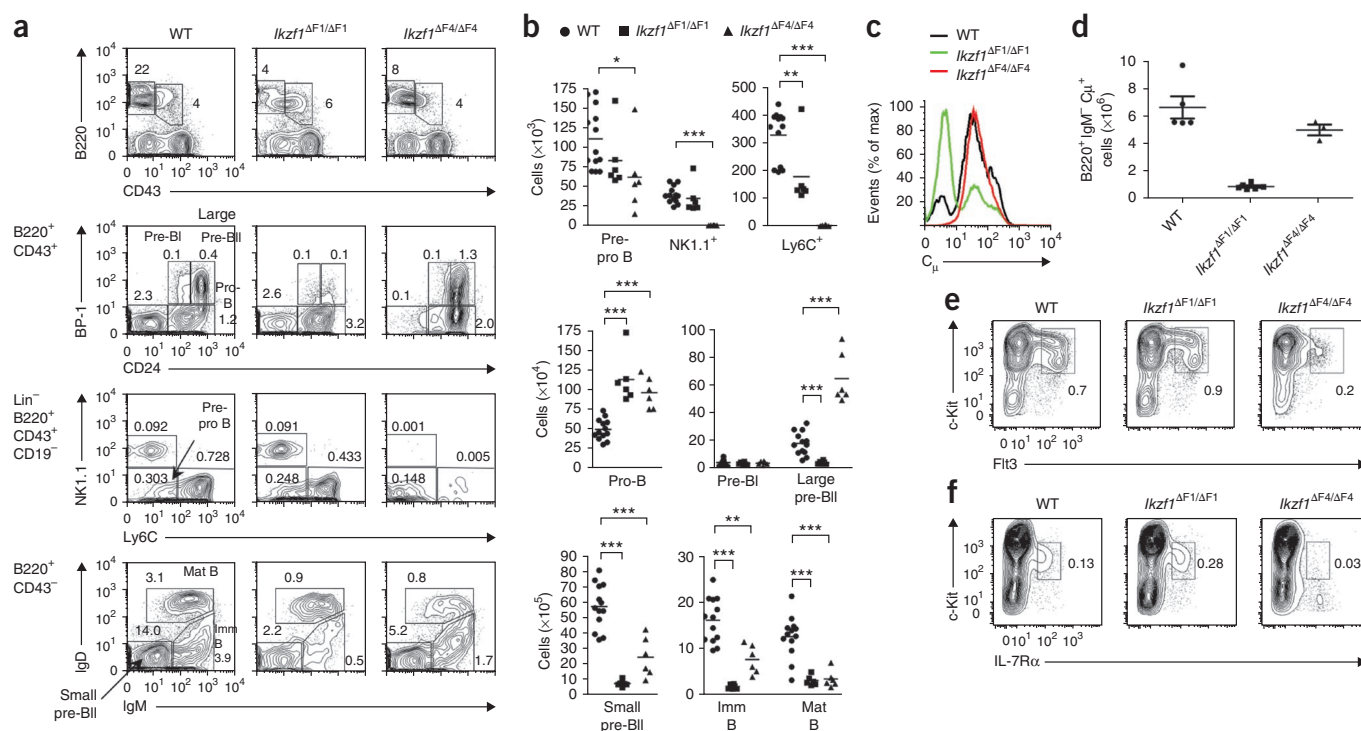ells of various populations (horizontal axes) in total bone marrow from wild-type mice ($n = 13$), $Ikzf1^{\Delta F1/\Delta F1}$ mice ($n = 6$) and $Ikzf1^{\Delta F4/\Delta F4}$ mice ($n = 6$–8). *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$ (unpaired, two-tailed Student's t-test, 95% confidence interval). (**c**) Cytoplasmic expression of immunoglobulin-μ (C_μ) in bone marrow cells gated on B220+IgM− cells. (**d**) Absolute number of B220+IgM− immunoglobulin-μ–positive (C_μ^+) cells per femur from wild-type mice ($n = 5$), $Ikzf1^{\Delta F1/\Delta F1}$ mice ($n = 6$) and $Ikzf1^{\Delta F4/\Delta F4}$ mice ($n = 3$). (**e,f**) Expression of Flt3 (**e**) and IL-7Rα (**f**) on c-Kit+ bone marrow cells from wild-type mice ($n = 12$–14), $Ikzf1^{\Delta F1/\Delta F1}$ mice ($n = 3$–4) and $Ikzf1^{\Delta F4/\Delta F4}$ mice ($n = 8$–9), gated on Lin^{neg-lo} cells. Numbers adjacent to outlined areas (**a,e,f**) indicate percent cells in each among total live bone marrow cells. Each symbol (**b,d**) represents an individual mouse; small horizontal lines indicate the mean (± s.e.m. in **d**). Data are from one experiment representative of five independent experiments (**a**; average), four independent experiments (**c**) or two to seven experiments (**e,f**; average) or are representative of five experiments (**b**) or three experiments (**d**).

B cell development phenotypes

Detailed examination of B cell development[41,42] revealed striking differences among $Ikzf1^{\Delta F1/\Delta F1}$, $Ikzf1^{\Delta F4/\Delta F4}$ and $Ikzf1^{null}$ mice (**Supplementary Fig. 2**). First, in contrast to the absence of pre-pro-B cells in $Ikzf1^{null}$ mice due to a block in the maturation of lymphoid-primed multipotent progenitor cells[4], $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice had relatively normal numbers of these cells (**Fig. 2a,b**). The B220+CD43+CD24−BP-1− population includes precursors of natural killer cells and plasmacytoid dendritic cells, in addition to pre-pro-B cells; when initially analyzed, $Ikzf1^{\Delta F4/\Delta F4}$ mice had a lower abundance of these cells than did wild-type mice (**Fig. 2a**). However, natural killer cells (NK1.1+) and plasmacytoid dendritic cells (Ly6C+) were selectively absent from $Ikzf1^{\Delta F4/\Delta F4}$ mice (similar to their absence in $Ikzf1^{null}$ mice), with only slightly reduced numbers of pre-pro-B cells (**Fig. 2a,b**). Both $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice also had relatively normal percentages and numbers of pro-B cells and pre-BI cells (**Fig. 2a,b**).

In contrast to the similar numbers of early progenitor cells in the two strains, $Ikzf1^{\Delta F1/\Delta F1}$ mice had substantially decreased numbers of large pre-BII cells, yet $Ikzf1^{\Delta F4/\Delta F4}$ mice had increased numbers of these cells, relative to their abundance in wild-type mice (**Fig. 2a,b**). Staining for cytoplasmic immunoglobulin-μ, which is expressed in most large pre-BII cells, confirmed the developmental block in $Ikzf1^{\Delta F1/\Delta F1}$ mice, as only a small number of B220+ cells that lacked surface expression of immunoglobulin M (IgM−) but expressed cytoplasmic immunoglobulin-μ

were present (**Fig. 2c,d**). Notably, an essential role for Ikaros at this developmental stage has been suggested[7].

Consistent with the deficiency in large pre-BII cells in $Ikzf1^{\Delta F1/\Delta F1}$ mice, we observed considerably fewer small pre-BII cells, immature B cells and mature recirculating B cells in these mice than in wild-type mice (**Fig. 2a,b**). In $Ikzf1^{\Delta F4/\Delta F4}$ mice, the number of cells in these populations was reduced to a lesser degree (**Fig. 2a,b**). Notably, although much of B cell development seemed to be largely intact in $Ikzf1^{\Delta F4/\Delta F4}$ mice, c-Kit+ early progenitor cells with low to negative expression of lineage markers (Lin^{neg-lo}) had lower expression of the receptor tyrosine kinase Flt3 and of interleukin 7 receptor α-chain (IL-7Rα) than did their wild-type counterparts (**Fig. 2e,f**). This phenotype was reminiscent of that of $Ikzf1^{null}$ mice, although in the $Ikzf1^{\Delta F4/\Delta F4}$ mice, expression of these proteins apparently was not reduced to a sufficient extent to block development.

Additional finger-specific functions in hematopoiesis

The developmental abnormalities of T cells were far more severe in $Ikzf1^{\Delta F4/\Delta F4}$ mice than in $Ikzf1^{\Delta F1/\Delta F1}$ mice. Both strains exhibited slightly reduced thymic cellularity, with a greater reduction in $Ikzf1^{\Delta F4/\Delta F4}$ mice, relative to that of their wild-type littermates (**Fig. 3a**). During the CD4−CD8− double-negative stages 1–4 (DN1–DN4) of thymopoiesis, reduced cell numbers were readily apparent in $Ikzf1^{\Delta F4/\Delta F4}$ mice but not in $Ikzf1^{\Delta F1/\Delta F1}$ mice, relative to their abundance in wild-type mice (**Fig. 3b** and **Supplementary Fig. 3a–d**).
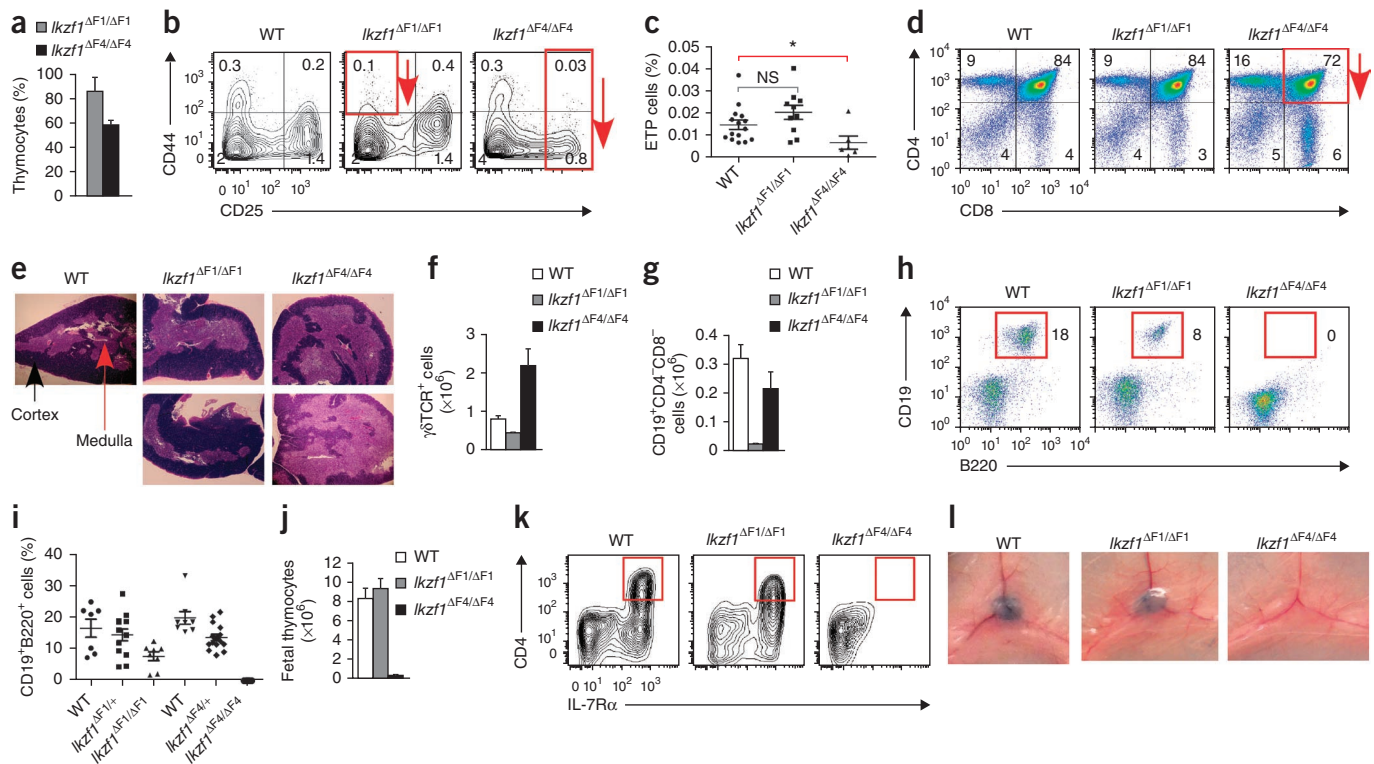
Figure 3 Selective thymocyte and fetal phenotypes in *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] mice. (**a**) Thymic cellularity in 4-week-old *Ikzf1*[ΔF1/ΔF1] mice ($n = 8$) and *Ikzf1*[ΔF4/ΔF4] mice ($n = 9$), presented relative to that of their wild-type littermates (controls). (**b**) Early stages of thymocyte development in 4-week-old wild-type mice ($n = 16$), *Ikzf1*[ΔF1/ΔF1] mice ($n = 8$) and mice *Ikzf1*[ΔF4/ΔF4] ($n = 8$), among cells gated on CD4$^-$CD8$^-$ (DN) cells. Numbers in quadrants indicate percent CD44$^+$CD25$^-$ (DN1) cells (top left), CD44$^+$CD25$^+$ (DN2) cells (top right), CD44$^-$CD25$^+$ (DN3) cells (bottom right) and CD44$^-$CD25$^-$ (DN4) cells (bottom left). (**c**) Analysis of the early thymic progenitor (ETP) population (Lin$^-$Kit$^+$CD44$^+$CD25$^-$) in wild-type mice ($n = 15$–22), *Ikzf1*[ΔF1/ΔF1] mice ($n = 10$–11) and *Ikzf1*[ΔF4/ΔF4] mice ($n = 6$–11). *$P < 0.05$ (unpaired, two-tailed Student's *t*-test, 95% confidence interval). NS, not significant. (**d**) Expression of CD4 and CD8 in the thymus (ages and n values as in **b**). (**e**) Histological analysis of thymi from 4-week-old mice, stained with hematoxylin and eosin. Original magnification, ×4. (**f,g**) Quantification of thymic cells that express the γδ T cell antigen receptor (γδTCR$^+$; **f**) and thymic B cells (CD19$^+$CD4$^-$CD8$^-$; **g**) in 4-week-old wild-type mice ($n = 12$–17), *Ikzf1*[ΔF1/ΔF1] mice ($n = 8$–9) and *Ikzf1*[ΔF4/ΔF4] mice ($n = 8$–11). (**h,i**) Quantification of fetal B cells (CD45$^+$CD19$^+$B220$^+$) in wild-type mice ($n = 15$), *Ikzf1*[ΔF1/ΔF1] mice ($n = 8$) and *Ikzf1*[ΔF4/ΔF4] mice ($n = 10$) at embryonic day 18.5 (E18.5). Numbers adjacent to red outlined area (**h**) indicate percent CD45$^+$CD19$^+$B220$^+$ B cells. Each symbol (**i**) represents an individual mouse; small horizontal lines indicate the mean (± s.e.m.). (**j**) Quantification of fetal thymocytes in wild-type mice ($n = 17$), *Ikzf1*[ΔF1/ΔF1] mice ($n = 6$) and *Ikzf1*[ΔF4/ΔF4] mice ($n = 10$) at E18.5. (**k**) Flow cytometry to detect lymphoid tissue–inducer cells (red outlined areas; CD45$^+$CD3ε$^-$CD4$^+$IL-7Rα$^+$) in the fetal mesentery of wild-type, *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] mice at E18.5. (**l**) Inguinal lymph nodes in 4-week-old adult wild-type, *Ikzf1*[ΔF1/ΔF1] and *Ikzf1*[ΔF4/ΔF4] mice, visualized with Evan's blue dye (images of other peripheral lymph nodes, not shown). Data are representative of eight experiments (**a**; error bars, s.e.m.), eight to ten experiments (**f,g**; error bars, s.e.m.), four to five experiments (**h,i**) or three to four experiments (**j**; error bars, s.e.m.) or are from one experiment representative of ten to eighteen experiments (**b,d**; average) or two or more experiments (**e,k,l**).

The DN1 population (CD44$^+$CD25$^-$)[43] seemed to be smaller in *Ikzf1*[ΔF1/ΔF1] mice than in wild-type mice (**Fig. 3b** and **Supplementary Fig. 3a**). However, that reduction was due to greatly reduced numbers of other cell types that contaminated the DN1 population, including thymic B cells (discussed below). Consistent with that interpretation, the percentage of early thymic progenitor cells was normal in *Ikzf1*[ΔF1/ΔF1] mice but reduced in *Ikzf1*[ΔF4/ΔF4] mice (**Fig. 3c**). The frequency of DN2 (CD44$^+$CD25$^+$) and DN3 (CD44$^-$CD25$^+$) cells was also greatly reduced in *Ikzf1*[ΔF4/ΔF4] mice but normal in *Ikzf1*[ΔF1/ΔF1] mice, relative to their abundance in wild-type mice (**Fig. 3b** and **Supplementary Fig. 3b,c**). DN4 cells (CD44$^-$CD25$^-$) recovered in the *Ikzf1*[ΔF4/ΔF4] strain to percentages and numbers similar to those observed in wild-type mice, presumably by homeostatic proliferation (**Fig. 3b** and **Supplementary Fig. 3d**). The ratio of CD4$^+$CD8$^+$ double-positive (DP) thymocytes to CD4$^+$ or CD8$^+$ single-positive (SP) thymocytes in *Ikzf1*[ΔF1/ΔF1] mice was similar to that of wild-type mice but was significantly reduced in *Ikzf1*[ΔF4/ΔF4] mice (**Fig. 3d** and **Supplementary Fig. 3e**). The reduced number of DP thymocytes in *Ikzf1*[ΔF4/ΔF4] mice correlated

with a reduction in the size of the thymic cortex (**Fig. 3e**). Neither *Ikzf1*[ΔF1/ΔF1] nor *Ikzf1*[ΔF4/ΔF4] mice exhibited the extreme skewing toward the CD4$^+$ lineage that characterizes *Ikzf1*[null] mice[1] (**Fig. 3d** and **Supplementary Fig. 3f**).

Ikzf1[null] mice have reduced numbers of thymic γδ T cells[1]. These cells were largely unperturbed in number in *Ikzf1*[ΔF1/ΔF1] mice and increased in number in *Ikzf1*[ΔF4/ΔF4] mice, relative to their abundance in wild-type mice (**Fig. 3f**). Thymocyte analysis also revealed a severe reduction in the number of thymic B cells (CD19$^+$CD4$^-$CD8$^-$) in *Ikzf1*[ΔF1/ΔF1] mice relative to that of wild-type mice (**Fig. 3g**). This selective and consistent phenotype is difficult to explain, given the similar number of conventional B-2 cells in the spleen of the two mutant strains. Furthermore, examination of the peritoneal cavity revealed reduced numbers of the B-1a population in each mutant strain and a reduced number of B-1b cells only in the *Ikzf1*[ΔF1/ΔF1] strain, relative to that of wild-type mice (**Supplementary Fig. 3g,h**). Further analysis of these phenotypes may provide insight into the developmental origin, homing mechanism and/or function of thymic B cells and peritoneal B-1 cells.
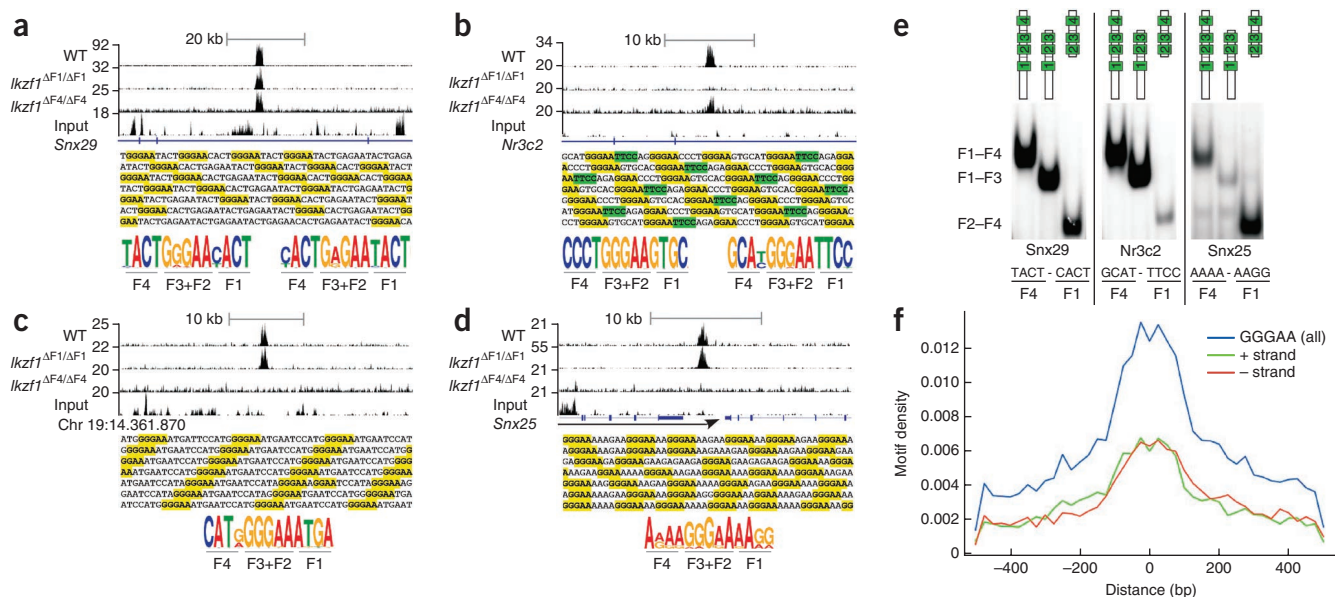
Figure 4 Differences in the binding of Ikaros to DNA in *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes. (**a**–**d**) ChIP-seq analysis of Ikaros binding in thymocytes from 4-week-old wild-type, *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ mice, presented as UCSC Genome Browser tracks for four genomic regions with repetitive Ikaros-recognition motifs (which yields unusually strong ChIP-seq peaks), showing interactions that do not depend on finger 1 or finger 4 (*Snx29*) or that depend on finger 1 (*Nr3c2*) or finger 4 (chromosome (Chr) 19 and *Snx25*). Input, sequencing of input genomic DNA (negative control). Below, repetitive genomic sequences, with putative Ikaros-recognition sites (yellow highlight, GGGAA or GGAA; green highlight, TTCC); bottom, position weight matrices of the putative Ikaros-recognition sequences that were repeated, with the zinc fingers (F1–F4) predicted to contact or lie in close proximity to the nucleotides. (**e**) EMSA with oligonucleotide probes for three repetitive sequences and recombinant proteins containing Ikaros fingers 1–4, 1–3 or 2–4 (above). Below, zinc fingers (F1 or F4) predicted to contact or lie in close proximity to the nucleotides flanking the GGGAA core recognition sequence. (**f**) Bioinformatics analysis of a wild-type Ikaros ChIP-seq data set, showing the localization of the core GGGAA sequence relative to the center of each ChIP-seq peak. Data are representative of three experiments (**a**–**d**), three or more experiments (**e**) or two experiments (**f**).

Although adult mutant mice had substantial numbers of mature B cells and T cells, the *Ikzf1*$^{\Delta F4/\Delta F4}$ strain lacked cells of both the B lineage and T lineage in the fetus (**Fig. 3h**–**j** and **Supplementary Fig. 4a**–**c**). In contrast, fetal T cell numbers were normal and fetal B cells were only moderately reduced in the *Ikzf1*$^{\Delta F1/\Delta F1}$ strain (**Fig. 3h**–**j** and **Supplementary Fig. 4a**–**c**). Lymphoid tissue–inducer cells (CD45+CD3ε−CD4+IL-7Rα+) were also selectively absent from the fetal intestinal mesentery of *Ikzf1*$^{\Delta F4/\Delta F4}$ mice (**Fig. 3k**). This phenotype correlated with the selective absence of lymph nodes (inguinal and lumbar) and Peyer's patches in adult *Ikzf1*$^{\Delta F4/\Delta F4}$ mice (**Fig. 3l** and **Supplementary Fig. 4d**), consistent with the important role of fetal lymphoid tissue–inducer cells in the development of these lymphoid structures[1,44]. Nasal-associated lymphoid tissue, which develops postnatally[44], was intact (**Supplementary Fig. 4e**). These developmental defects were reminiscent of those described for *Ikzf1*^{null} mice[1], which suggested that zinc-finger 4 was required for the proper regulation of genes encoding molecules involved in early stages of fetal lymphopoiesis. Notably, zinc-finger 1 and all other residues encoded by exon 4 seemed to be dispensable for these developmental events.

Finger-specific DNA binding *in vivo*

The generation of the *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ strains was inspired by evidence that fingers 1 and 4 support the binding to different DNA sequences *in vitro*[29–31]. To determine whether the two fingers also participated in binding to different DNA sequences *in vivo*, we used ChIP-seq analysis of total thymocytes from 4-week-old wild-type, *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ mice. The most prominent ChIP-seq peaks were present at simple repetitive elements that contained multiple copies of the core Ikaros consensus sequence of GGGAA (**Fig. 4a**–**d**). Some of these simple repeats were located in close proximity to protein-encoding

genes, but these repeats have not been conserved through evolution and therefore may not have important functions. Nevertheless, the ChIP-seq profiles confirmed that fingers 1 and 4 contributed to differences in DNA binding *in vivo*. For example, a repeat region in *Snx29* (which encodes a sorting nexin) exhibited a prominent ChIP-seq peak in wild-type thymocytes and in thymocytes from both mutant strains (**Fig. 4a**), which suggested that fingers 2 and 3 were sufficient for binding to this region. In contrast, a prominent ChIP-seq peak in *Nr3c2* (which encodes the mineralocorticoid receptor Mlr) was present in wild-type and *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes but not in *Ikzf1*$^{\Delta F1/\Delta F1}$ thymocytes (**Fig. 4b**). At a repeat region adjacent to *Snx25* (which encodes a sorting nexin) (**Fig. 4d**), and at a repeat in chromosome 19 that is not linked to an annotated gene (**Fig. 4c**), ChIP-seq peaks were present in wild-type and *Ikzf1*$^{\Delta F1/\Delta F1}$ thymocytes but not in *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes.

Inspection of the repeats revealed regions containing the Ikaros consensus sequence GGGAA but with different flanking nucleotides. Motif analysis of these regions in *Snx29* and *Nr3c2* revealed two repetitive sequences that may support the binding of Ikaros, while one repetitive sequence that may support the binding of Ikaros was present in the regions in *Snx25* and chromosome 19 (**Fig. 4**). *In vitro* binding-site-selection experiments have suggested that finger 1 recognizes a pyrimidine-rich sequence downstream of the Ikaros core[29] (S.E.W. and S.T.S., data not shown). On the basis of those studies, finger 1 would be predicted to bind the TCC sequence in one of the *Nr3c2* repeats (**Fig. 4b**, bottom right). The other *Nr3c2* consensus (**Fig. 4b**, bottom left) and the two *Snx29* consensus sequences (**Fig. 4a**) may also contact finger 1, but perhaps more weakly because the nucleotides that would be contacted by finger 1 include a purine. In contrast, we predict that the repeats in *Snx25* and chromosome 19 (**Fig. 4c,d**) have little ability to interact with finger 1 because the GGGAA core is followed by purine residues.

Figure 5 Deregulation of distinct sets of genes in *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ DP thymocytes. (**a**) RNA-Seq analysis of mRNA from DP thymocytes sorted by flow cytometry from 4-week-old mice, presented as a Venn diagram of genes upregulated (Up) or downregulated (Down) at least threefold ($P \leq 0.001$) in either *Ikzf1*$^{\Delta F1/\Delta F1}$ or *Ikzf1*$^{\Delta F4/\Delta F4}$ mice relative to their expression in wild-type mice (for genes with an RPKM value of ≥ 4 in at least one of the six samples). (**b**) Genes in **a** with an increased (Up) or decreased (Down) mRNA abundance of at least tenfold ($P \leq 0.001$) in one of the mutant strains. (**c**) Gene-ontology analysis of genes upregulated more than threefold ($P \leq 0.001$) in *Ikzf1*$^{\Delta F4/\Delta F4}$ DP thymocytes. (**d**) Distribution of RPKM values (mRNA abundance) for all annotated genes (left) and all genes upregulated by at least threefold ($P \leq 0.001$) in *Ikzf1*$^{\Delta F1/\Delta F1}$ mice (middle) or *Ikzf1*$^{\Delta F4/\Delta F4}$ mice (right). Data are representative of two experiments (duplicates).
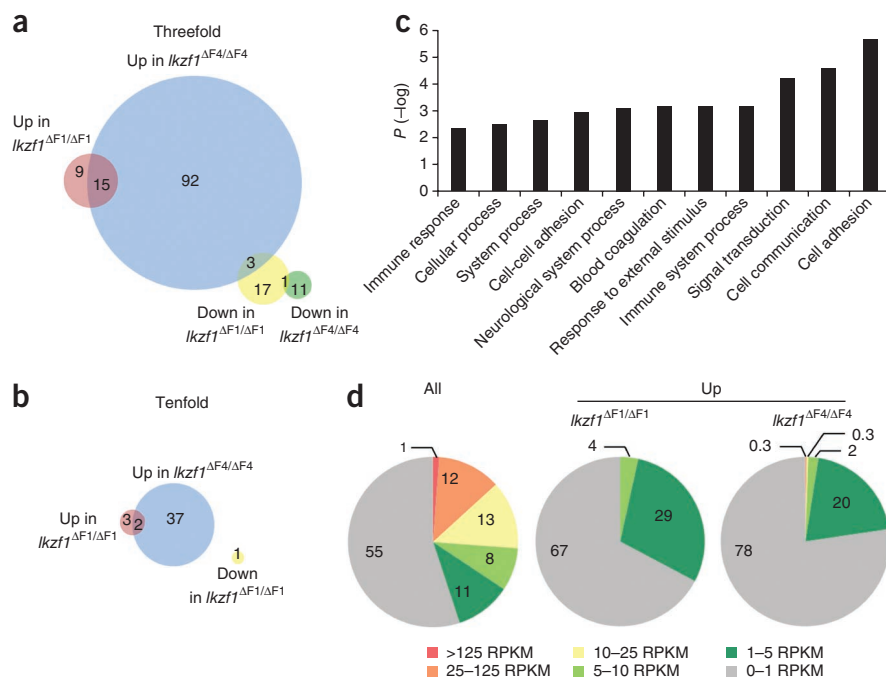
The binding site–selection data[29] and additional findings (S.E.W. and S.T.S., data not shown) failed to reveal a strong DNA sequence 'preference' for finger 4. Instead, finger 4 seemed to stabilize binding in a DNA sequence–independent manner, with its DNA contacts particularly important at sites not stably bound by fingers 2 and 3 alone (S.E.W. and S.T.S., data not shown). We hypothesize that finger 4 may be important for the repeats in *Snx25* and chromosome 19 because the DNA-binding energy of fingers 2 and 3 is weakened by the presence of a purine just downstream of the GGGAA core; we have found that a pyrimidine at this position is important for stable protein-DNA interactions by fingers 2 and 3 (S.E.W. and S.T.S., data not shown).

To determine whether the difference in binding observed by ChIP-seq was due to intrinsic DNA-binding 'preferences', we did electrophoretic mobility-shift assays (EMSAs) with recombinant proteins containing different combinations of zinc fingers. Consistent with the ChIP-seq results, we observed substantial binding to one of the *Snx29* sequences with a protein containing all four fingers, as well as with proteins containing only fingers 1–3 or 2–4 (**Fig. 4e**). The second *Snx29* repeat (with a GaGAA core instead of the GGGAA core, where the lowercase 'a' indicates the difference from the consensus core sequence) was unable to bind any of the proteins (**Supplementary Fig. 5a**). We observed substantial binding to one of the *Nr3c2* sequences with proteins containing fingers 1–4 or 1–3 but not with a protein containing fingers 2–4 (**Fig. 4e**), which demonstrated finger 1–dependent binding. Finally, we observed substantial binding to the *Snx25* repetitive sequence with proteins containing fingers 1–4 or fingers 2–4 but not with a protein containing fingers 1–3 (**Fig. 4e**). Thus, the intrinsic *in vitro* binding capacity mirrored the *in vivo* ChIP-seq results.

In addition to identifying the prominent interactions noted above (**Fig. 4**), the ChIP-seq results revealed thousands of other Ikaros-binding sites throughout the genome. More than 60% of the ChIP-seq peaks coincided with the Ikaros core sequence of GGGAA (**Fig. 4f** and **Supplementary Fig. 5b**). In experiments with wild-type or mutant thymocytes, the number of 'called peaks' varied from about 1,000 to 17,000. Those numbers are in the same range as the 7,000 Ikaros peaks in thymocytes reported before[16], with extensive overlap observed between the peaks in our experiments and those reported before. *Zfp64*, *Zfp260*, *Cd4*, *Notch1*, *Hdac7*, and *Bcl11b* are examples of genes exhibiting the same Ikaros peaks in our experiments and in published reports[16,18,19] (**Supplementary Fig. 5c**). Motif-enrichment analysis revealed that the core Ikaros recognition sequence of GGGAA was highly prevalent at peaks that overlapped in the wild-type data set and in the data sets from both mutant strains (**Supplementary Fig. 5d**). Furthermore, the sequence GGGAAAAGGGAA was prevalent at peaks selectively absent from the *Ikzf1*$^{\Delta F4/\Delta F4}$ sample (**Supplementary Fig. 5d**). That sequence is very similar to the sequences noted above (**Fig. 4c,d**) that exhibited finger 4 dependence, which suggested that finger 4 would be broadly important for binding to such sequences throughout the genome. Thus, the ChIP-seq data supported a model in which the different functions of fingers 1 and 4 are due at least in part to their ability to facilitate binding to distinct genomic sites (discussed below).

Selective misregulation of gene expression

Although thousands of Ikaros-binding sites were identified by ChIP-seq, the highly selective phenotypes of the *Ikzf1*$^{\Delta F1/\Delta F1}$ and *Ikzf1*$^{\Delta F4/\Delta F4}$ strains suggested that DNA-binding fingers 1 and 4 may contribute to the regulation of distinct sets of genes. To examine the roles of fingers 1 and 4 in transcriptional control, we obtained mRNA from CD4$^+$CD8$^+$ thymocytes sorted from 4-week-old wild-type, *Ikzf1*$^{\Delta F1/\Delta F1}$, and *Ikzf1*$^{\Delta F4/\Delta F4}$ mice and analyzed the mRNA by RNA-Seq. The results revealed misregulation of small number of genes in each of the mutant strains, relative to their expression in wild-type mice. In *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes, only 110 genes and 12 genes (with RPKM (reads per kilobase per million mapped reads) values of ≥ 4) were upregulated and downregulated, respectively, by more than threefold, with only 24 and 21 genes upregulated and downregulated, respectively, by this magnitude in *Ikzf1*$^{\Delta F1/\Delta F1}$ thymocytes (**Fig. 5a** and **Supplementary Fig. 6**). Notably, only 15 genes were upregulated in both mutant strains and 1 gene was downregulated in both strains. Three genes upregulated in *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes were downregulated in *Ikzf1*$^{\Delta F1/\Delta F1}$ thymocytes. Notably, much smaller numbers of genes were upregulated or downregulated by tenfold or more (**Fig. 5b** and **Supplementary Fig. 6**). Although it is difficult to accurately compare RNA-Seq data sets to microarray data sets, many of the genes found by RNA-Seq to be misregulated in *Ikzf1*$^{\Delta F1/\Delta F1}$ or *Ikzf1*$^{\Delta F4/\Delta F4}$

Figure 6 $Ikzf1^{\Delta F4/\Delta F4}$ mice develop spontaneous thymic lymphoma but $Ikzf1^{\Delta F1/\Delta F1}$ mice do not. (**a**) Survival of wild-type, $Ikzf1^{\Delta F1/+}$ and $Ikzf1^{\Delta F1/\Delta F1}$ mice (left) and wild-type, $Ikzf1^{\Delta F4/+}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice (right), presented as Kaplan-Meier curves. (**b**) Histochemical analysis of hematoxylin- and eosin-stained heart tissue from an $Ikzf1^{\Delta F4/\Delta F4}$ mouse (right; $n = 3$) and a wild-type control mouse (left), showing a thymic lymphoma at right. (**c**) Histochemical analysis of hematoxylin- and eosin-stained lung tissue from an $Ikzf1^{\Delta F4/\Delta F4}$ mouse (right) and a wild-type mouse (left), showing an infiltrating thymic lymphoma at right. (**d**) Expression of CD4 and CD8 in lymphomas that had metastasized to the spleen in $Ikzf1^{\Delta F4/\Delta F4}$ mice ($n = 3$) and a wild-type mouse (control). Data are from one experiment with a large cohort of mice (**a**) or with one representative mouse of three or more mice (**b–d**).

thymocytes have been found by microarray to be misregulated in $Ikzf1^{null}$ thymocytes[16]. Thus, misregulated genes in $Ikzf1^{null}$ thymocytes seem to represent a composite of the genes misregulated in $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes.

Gene-ontology analysis of genes upregulated threefold or more in $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes revealed enrichment for genes encoding molecules involved in cell adhesion, cell communication and signal transduction (**Fig. 5c**). Consistent with the invasive properties of the thymic lymphomas that arose in older $Ikzf1^{\Delta F4/\Delta F4}$ mice (reported below), several genes encoding molecules involved in tumor invasion and metastasis were among the genes upregulated the most, including *Mmp14* (which encodes a matrix metalloprotease), *Ctnnd1* (which encodes δ-catenin) and *Dock1* (which encodes a guanine nucleotide–exchange factor)[45–48] (**Supplementary Fig. 6e,g**).

Notably, genes that were upregulated in either $Ikzf1^{\Delta F1/\Delta F1}$ or $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes generally had very low expression in wild-type thymocytes (**Fig. 5d**). We assessed the distribution by the expression (RPKM) in thymocytes of all annotated genes and the RPKM distribution in wild-type cells of genes upregulated by at least threefold in $Ikzf1^{\Delta F1/\Delta F1}$ or $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes (**Fig. 5d**). As the vast majority (>96%) of upregulated genes were upregulated from very low expression (RPKM < 5), these results were consistent with evidence that Ikaros often functions as a transcriptional repressor.

Zinc-finger requirements for tumor suppression
Published Ikaros-mutant strains develop thymic lymphoma[10,11]. We never observed lymphoma in $Ikzf1^{\Delta F1/\Delta F1}$ mice, but $Ikzf1^{\Delta F4/\Delta F4}$ mice developed thymic lymphoma with a penetrance similar to that reported for other Ikaros-mutant strains (**Fig. 6a**). Also similar to other Ikaros-mutant strains, the lymphomas in $Ikzf1^{\Delta F4/\Delta F4}$ mice were aggressive (highly invasive and metastatic) (**Fig. 6b,c**) and they displayed variable expression of the coreceptors CD4 and CD8 (**Fig. 6d**), clonal rearrangement of the gene encoding the T cell antigen receptor β-chain, and aberrant expression of the intracellular domain of the signaling receptor Notch (data not shown).

Ikzf1 mutations in a mouse model of BCR-ABL⁺ B-ALL
Although *Ikzf1* mutations invariably give rise to T cell malignancies in mice, human malignancies of T cell origin rarely have *IKZF1* mutations[49]. Instead, human *IKZF1* mutations are frequently associated with BCR-ABL⁺ B-ALL and other progenitor–B cell malignances[12,13]. To investigate this species difference, and with the additional goal of developing models with which to study progenitor–B cell malignancies associated with *IKZF1* mutations, we first made use of a well-established *in vitro* culture assay for BCR-ABL⁺ B-ALL[50]. We obtained bone marrow cells from wild-type, $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice and transduced those cells with a retrovirus
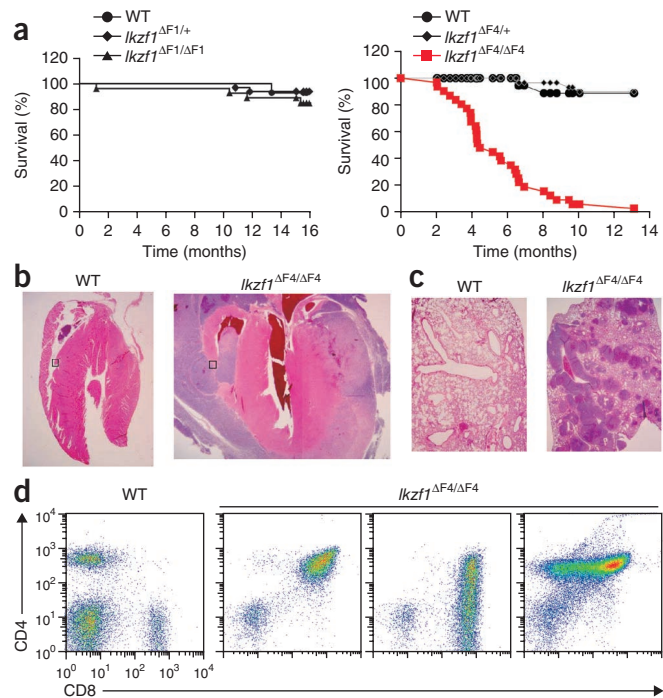
expressing the oncogenic tyrosine kinase BCR-ABL, then monitored cell numbers in the progenitor B cell cultures over the course of approximately 3 weeks. Transduced $Ikzf1^{\Delta F4/\Delta F4}$ cells proliferated much more rapidly than did their wild-type counterparts (**Fig. 7a**). Notably, transduced $Ikzf1^{\Delta F1/\Delta F1}$ cells proliferated more slowly than did their wild-type counterparts and eventually stopped growing (**Fig. 7a** and data not shown).

We next used an *in vivo* model of BCR-ABL⁺ B-ALL in which we transplanted bone marrow cells transduced with the retrovirus expressing BCR-ABL into irradiated wild-type C57BL/6 recipient mice. In this assay, transduced $Ikzf1^{\Delta F4/\Delta F4}$ cells yielded much more aggressive malignancies than did their wild-type or $Ikzf1^{\Delta F1/\Delta F1}$ counterparts (**Fig. 7b**). These results demonstrated that an *Ikzf1* mutation was able to contribute to B cell malignancy in mice when combined with BCR-ABL expression. Furthermore, these results established an important parallel between the tumor-suppressor function of Ikaros in thymocytes and progenitor cells–B cells; in both cell types, finger 4 was essential for tumor suppression, whereas finger 1 was dispensable.

To approach an understanding of the tumor-suppressor function of Ikaros, we transduced cells with retrovirus expressing BCR-ABL, obtained mRNA from the cells after 21 or 28 d of culture and analyzed the mRNA by RNA-Seq. We also analyzed sorted pro-B cells and pre-BI cells plus large pre-BII cells, which are thought to most closely represent the developmental stages that expand following BCR-ABL transformation. Cluster analysis revealed extensive differences between the normal cells and transformed cells in their gene expression (**Fig. 7c**). However, most of the changes in gene expression were unaffected by the *Ikzf1* mutations. Clusters 1–3, for example, showed differences between all of the data sets from untransformed cells and all of the data sets from transformed cells (**Fig. 7c**). Gene-ontology analysis of the two largest clusters (1 and 2) revealed, as expected, that several genes encoding molecules involved in the cell cycle and control of cell growth were upregulated after transformation with BCR-ABL, whereas genes encoding molecules involved in leukocyte differentiation and activation were downregulated (**Supplementary Fig. 7**).

Figure 7 Selective synergy between BCR-ABL and the $Ikzf1^{\Delta F4/\Delta F4}$ mutation *in vitro* and *in vivo*. (**a**) *In vitro* growth of bone marrow cells from wild-type, $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice transduced with a BCR-ABL-expressing retrovirus and grown under B-ALL culture conditions. (**b**) Survival of irradiated recipient mice given transplantation of 1×10^6 BCR-ABL-transduced bone marrow cells from wild-type mice ($n = 15$), $Ikzf1^{\Delta F1/\Delta F1}$ mice ($n = 8$) or $Ikzf1^{\Delta F4/\Delta F4}$ mice ($n = 14$), presented as Kaplan-Meier curves. (**c**) RNA-Seq analysis of mRNA from sorted pro-B cells (pro-B) and pre-BI plus large pre-BII cells (pre-BI-II$_L$), as well as from BCR-ABL-transformed cells at day 21 or 28 of culture, for genes whose mRNA abundance differed by threefold or more between any two samples among the 12 samples analyzed (presented as k-means clustering). (**d**) Genes of the two clusters with genes selectively upregulated or downregulated in BCR-ABL-transformed $Ikzf1^{\Delta F4/\Delta F4}$ cells, identified by k-means clustering of the data sets from BCR-ABL-transformed cultures; expression data from pro-B cells and pre-BI cells plus large pre-BII cells were aligned after the cluster analysis was completed. (**e**) Flow cytometry of the *in vitro* cultures in **a**, gated on cells expressing yellow fluorescent protein as a marker of BCR-ABL transduction. (**f**) Expression of *Il2ra* mRNA (encoding CD25) and *Kit* mRNA (encoding c-Kit), from the RNA-Seq data in **c** ($n = 2$ samples per genotype). Data are from one experiment representative of three experiments (**a**) or three or more experiments (**e**) or are from three (**b**) or two (**d**–**f**) combined experiments (error bars (**f**), s.e.m.).

The remaining three clusters (4–6) included genes with higher expression in BCR-ABL-transformed $Ikzf1^{\Delta F1/\Delta F1}$ cells than in their wild-type or $Ikzf1^{\Delta F4/\Delta F4}$ counterparts (**Fig. 7c**). Those same genes also had higher expression in untransformed $Ikzf1^{\Delta F1/\Delta F1}$ pre-BI cells plus large pre-BII cells (**Fig. 7c**), which suggested that the $Ikzf1^{\Delta F1/\Delta F1}$ cells that proliferated in culture following BCR-ABL transformation were maintained at the pre-BI cell–to–large pre-BII cell developmental stage, a stage at which many genes were misregulated in $Ikzf1^{\Delta F1/\Delta F1}$ mice. Notably, most genes in all six clusters had similar expression in wild-type and $Ikzf1^{\Delta F4/\Delta F4}$ cells (**Fig. 7c**), despite the greatly enhanced proliferation of BCR-ABL-transformed $Ikzf1^{\Delta F4/\Delta F4}$ cells.

To identify genes selectively misregulated in BCR-ABL-transformed $Ikzf1^{\Delta F4/\Delta F4}$ cells, we did an additional cluster analysis with the six data sets obtained with transformed cells; we then aligned the expression of the respective genes in untransformed cells after we defined the clusters. This analysis revealed 155 genes that were selectively upregulated and 133 genes that were selectively downregulated in $Ikzf1^{\Delta F4/\Delta F4}$ cultures relative to their expression in wild-type and $Ikzf1^{\Delta F1/\Delta F1}$ cultures, at both day 21 and day 28 (**Fig. 7d**). Among the genes with higher expression in $Ikzf1^{\Delta F4/\Delta F4}$ cells was *Kit* (which encodes the stem cell factor receptor c-Kit) (**Fig. 7d**–**f**), which is known to be silenced during B cell maturation. Among the genes with lower expression were *Il2ra* (which encodes the IL-2 receptor α-chain CD25) and *Enpep* (which encodes the cell surface marker BP-1), which are activated during B cell maturation (**Fig. 7d**–**f**). Flow cytometry confirmed misregulation of the proteins encoded by *Kit* (c-Kit) and *Il2ra* (CD25) (**Fig. 7e**).
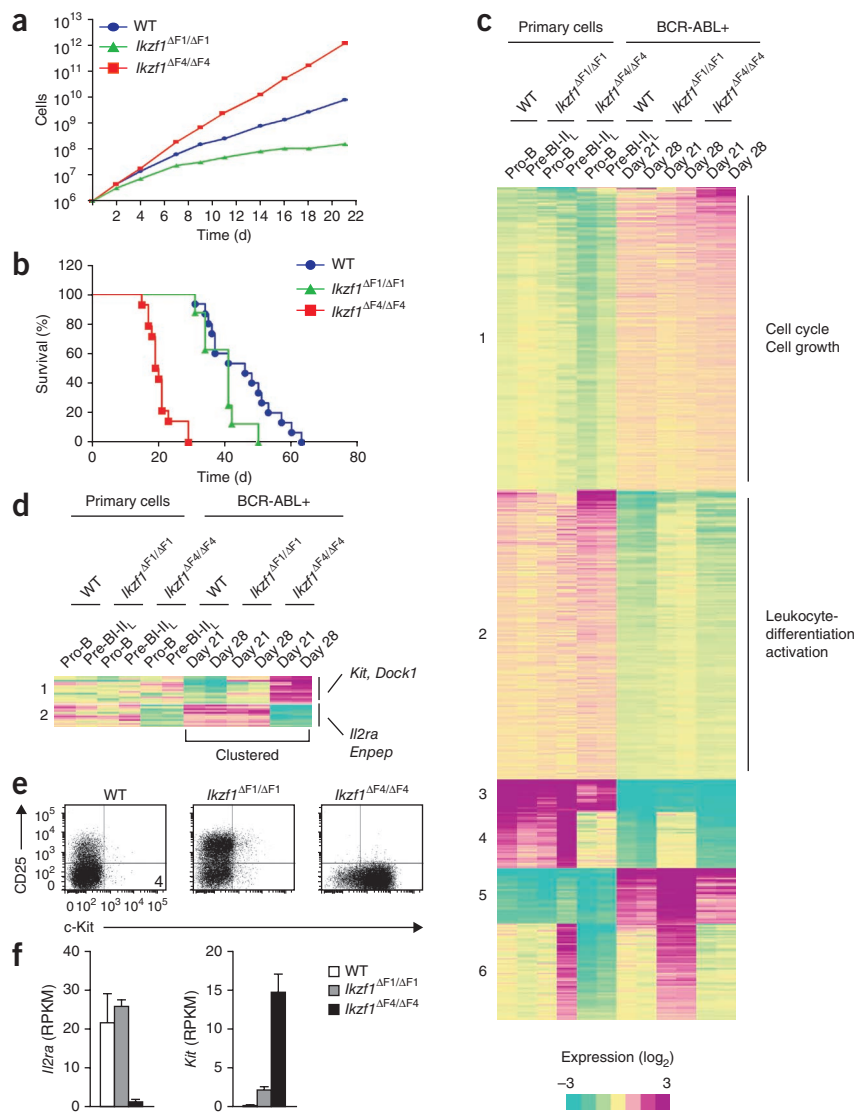
These findings suggested that the $Ikzf1^{\Delta F4/\Delta F4}$ mutation may influence the developmental stage of cells transformed by BCR-ABL. However, most of the other genes that were selectively misregulated in BCR-ABL-transformed $Ikzf1^{\Delta F4/\Delta F4}$ cells are not developmentally regulated, and several, such as *Dock1* and *Ctnnd1*, correspond to genes that were also misregulated in $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes. This identification of a limited set of genes selectively misregulated in transformed cells in the context of the $Ikzf1^{\Delta F4/\Delta F4}$ mutation provides an important step toward delineating the elusive mechanisms responsible for the tumor-suppressor function of Ikaros.

DISCUSSION

We have created two new mouse strains in which exons encoding individual zinc fingers of Ikaros were deleted. Our results have demonstrated that different zinc fingers in a single DNA-binding domain can participate in the regulation of distinct sets of genes and contribute to distinct biological functions. The disruption of individual fingers may be a generally valuable strategy for delineating the complex biological functions and mechanisms of action of zinc-finger transcription factors.

When designing the mutant strains, we considered three different strategies: deletion of the exons encoding fingers 1 and 4;

deletion of DNA sequences encoding fingers 1 and 4, with retention of the remaining sequences of exons 4 and 6; or mutagenesis of specific nucleotides encoding key residues of fingers 1 and 4 involved in DNA binding. Each strategy had notable advantages and limitations. An advantage of the second and third strategies is that it would be possible to exclude the possibility of involvement of non-finger residues in the resulting phenotypes. However, a major limitation of those strategies is that the mutant proteins do not normally exist in mouse cells, which would raise the possibility of dominant-negative or gain-of-function activities. Consistent with that, a point mutation in the *Ikzf1* region encoding finger 3 has been found to have much stronger dominant-negative properties in mice than a deletion mutant that results in complete removal of exons encoding fingers 1, 2 and 3 (ref. 51). Furthermore, preliminary data have suggested that exon 4–encoded residues immediately upstream of finger 1 autoregulate the DNA-binding activity of finger 1 (S.E.W. and S.T.S., data not shown); retention of those residues with deletion of finger 1 would have an uncertain outcome.

Because of the limitations noted above, we deleted exons 4 and 6, which encode fingers 1 and 4, respectively, as well as the small number of additional residues encoded by these exons. With this strategy, it remains possible that some of the hematological defects are due to the loss of activities other than the DNA-binding activity of the deleted finger. However, we were able to conclude with confidence that finger 1 and finger 4 were fully and differently dispensable for several biological functions of Ikaros, and our data suggested that at least a subset of the defects of each strain was due to loss of finger-dependent DNA interactions. Furthermore, the $Ikzf1^{\Delta F1/\Delta F1}$ strain revealed the specific functions of full-length Ikaros (Ik-1), as this strain lacked the Ik-1 isoform while retaining the naturally occurring Ik-2 isoform. Moreover, since the long-term goal is to delineate the highly selective phenotypes of the mutant strains for the purpose of understanding how Ikaros regulates lymphopoiesis and leukemogenesis, we can proceed with relatively little concern about the possibility of aberrant gain-of-function activities.

We unexpectedly found that the two fingers regulated different biological functions and even different steps in the same developmental pathway. We originally anticipated that Ikaros would regulate numerous genes involved in each developmental step in which it participates, such that each mutant strain would exhibit phenotypes similar to that of $Ikzf1^{null}$ mice. That expectation was based on evidence that gene expression is substantially altered in $Ikzf1^{null}$ cells[16,20] and that Ikaros binds to several thousand genomic sites, as shown by ChIP-seq experiments[16]. Instead, we found that finger 1 was largely or fully dispensable for many biological functions of Ikaros, with finger 4 being dispensable for several other functions.

That selectivity provides support for a hypothesis in which only one target gene or a small number of target genes is (are) essential for each biological event in which Ikaros participates. It is difficult to envision that Ikaros would be a critical regulator of hundreds of genes required for each biological event, with finger 1 being dispensable for the regulation of all target genes involved in some events and finger 4 being dispensable for the regulation of all genes involved in other events. Instead, it seems much more likely that each event requires the proper regulation of only one or a few genes that are targets of Ikaros, with either finger 1 or finger 4 being critical for the regulation of those genes. Although we favor a model in which Ikaros is an essential regulator of only a few key genes involved in each biological event, an alternative possibility is that fingers 2 and 3 are sufficient for the regulation of almost all critical target genes involved in each of its biological functions, with fingers 1 or 4 essential for the regulation of only a few key genes.

A longstanding challenge in the transcription field has been the identification of direct targets of transcription factors that can explain their biological functions. Efforts to identify such targets have benefited from the development of gene-expression profiling methods and methods for examining DNA binding *in vivo* on a genome-wide scale (such as ChIP-seq). By merging ChIP-seq data with the gene-expression profiles of wild-type and mutant cells, insights into the direct targets of a transcription factor can be obtained. However, a major limitation of this approach is that many transcription factors may bind genomic sites at which they do not function[52]. Moreover, because some transcription factors may 'preferentially' bind sequences associated with the relatively open chromatin found at active promoters and enhancers, it can be difficult to conclude with confidence that the merging of gene-expression and ChIP data sets has successfully identified the direct functional targets.

In our RNA-Seq studies, we found that most misregulated genes in $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes had low expression in wild-type cells, whereas Ikaros-binding sites defined by ChIP-seq were distributed among genes at all levels of expression (data not shown). Furthermore, the number of ChIP-seq peaks greatly exceeded the number of misregulated genes. One possible and perhaps likely interpretation of these results is that most Ikaros-binding sites identified by ChIP-seq are not functionally relevant. However, many other possibilities must be considered. For example, Ikaros may contribute to the proper regulation of a much larger number of genes, but its absence may alter the level of expression of most target genes by a magnitude that fails to reach the threefold cutoff used for our analyses. Another possibility is that Ikaros acts redundantly with other members of the Ikaros family at many of its target genes.

We originally were hopeful that ChIP-seq analysis of our mutant strains would reveal a distinct loss of Ikaros binding to a well-defined subset of sites, which would lead to improved correlation between binding and transcriptional misregulation of nearby genes. However, although we observed reduced numbers of ChIP-seq peaks for $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ thymocytes (data not shown), we have been unable to do a meaningful analysis of the binding events that require or are enhanced by these fingers. A chief reason for this difficulty is that the ChIP-seq peaks were not clearly polarized, in that we did not observe a distinct subset of peaks that were entirely dependent on finger 4 and another subset of peaks that were entirely dependent on finger 1. Instead, we observed a continuum of effects, with different yet highly variable degrees of dependence on finger 1 or finger 4.

The potential value of targeting individual zinc fingers can be summarized as follows. First, the $Ikzf1^{\Delta F1/\Delta F1}$ and $Ikzf1^{\Delta F4/\Delta F4}$ mice exhibited select subsets of the hematopoietic defects of $Ikzf1^{null}$ mice. Second, when we observed a phenotype, it was as robust or nearly as robust as that of $Ikzf1^{null}$ mice, which makes the phenotype amenable to further analysis. Third, despite the robust phenotypes, gene-expression changes were quite limited, which narrowed the list of potential direct and indirect target genes responsible for the phenotype. These findings pave the way for detailed analyses of each phenotype of the mutant strains.

Our results raise the question of why Ikaros and perhaps other zinc-finger transcription factors acquired the ability to bind DNA through different subsets of their fingers. One possibility is that this strategy allows Ikaros to recognize a larger number of DNA sequences, so that each target gene does not need to have a sequence that matches a single, rigid consensus. Although that simple scenario is possible, we favor a model in which recognition of DNA through different combinations of fingers supports different functions. One possibility

is that binding to DNA through different combinations of zinc fingers leads to conformational differences in the bound protein that may influence coregulatory interactions. The zinc-finger DNA-binding protein CTCF has similarly been suggested to carry out different functions when bound to DNA through different subsets of its fingers[40], yet this hypothesis has not yet been tested through the disruption of individual CTCF fingers.

Finally, it can be argued that the most important function of Ikaros to understand is its tumor-suppressor function. The correlation between *IKZF1* mutations and the therapeutic response of high-risk progenitor B-ALL suggests that understanding of the tumor-suppressor function of Ikaros may be critical for the development of new therapies[12,13]. The establishment of a mouse model for BCR-ABL[+] B-ALL through the use the *Ikzf1*$^{\Delta F4/\Delta F4}$ strain confirms a published report that B cell malignancies can be observed in an Ikaros-mutant strain when coupled to BCR-ABL expression[53]. The finding that finger 4 was selectively required for tumor suppression both in thymocytes and in the BCR-ABL[+] B-ALL model suggests that tumor suppression in the two cell types may involve similar molecular mechanisms. That hypothesis is further supported by the finding that many of the same genes were misregulated in *Ikzf1*$^{\Delta F4/\Delta F4}$ thymocytes and in BCR-ABL-transformed progenitor B cells from *Ikzf1*$^{\Delta F4/\Delta F4}$ mice. The loss of the tumor-suppressor function of Ikaros in the *Ikzf1*$^{\Delta F4/\Delta F4}$ strain will make this strain particularly valuable for future studies, as it exhibits only a subset of the hematopoietic defects of *Ikzf1*^{null} mice. This feature will allow studies of the tumor-suppressor function on a background in which many target genes of Ikaros involved in its other functions are relatively unperturbed.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: RNA-Seq and ChIP-seq data, GSE33693.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

H.S., J.M., T.L.A., S.F., D.C., S.E.W. and G.W.L. designed and did experiments and analyzed data; S.J.B. provided intellectual input and experimental advice; P.J.F., O.N.W. and S.T.S. supervised research and analyzed data; and H.S. and S.T.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wang, J.H. *et al.* Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity* **5**, 537–549 (1996).
2. Georgopoulos, K. Haematopoietic cell-fate decisions, chromatin regulation and ikaros. *Nat. Rev. Immunol.* **2**, 162–174 (2002).
3. Yoshida, T., Ng, S.Y. & Georgopoulos, K. Awakening lineage potential by Ikaros-mediated transcriptional priming. *Curr. Opin. Immunol.* **22**, 154–160 (2010).
4. Yoshida, T., Ng, S.Y., Zuniga-Pflucker, J.C. & Georgopoulos, K. Early hematopoietic lineage restrictions directed by Ikaros. *Nat. Immunol.* **7**, 382–391 (2006).
5. Kirstetter, P., Thomas, M., Dierich, A., Kastner, P. & Chan, S. Ikaros is critical for B cell differentiation and function. *Eur. J. Immunol.* **32**, 720–730 (2002).
6. Thompson, E.C. *et al.* Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits. *Immunity* **26**, 335–344 (2007).
7. Reynaud, D. *et al.* Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat. Immunol.* **9**, 927–936 (2008).
8. Trageser, D. *et al.* Pre-B cell receptor-mediated cell cycle arrest in Philadelphia chromosome-positive acute lymphoblastic leukemia requires IKAROS function. *J. Exp. Med.* **206**, 1739–1753 (2009).
9. Ma, S. *et al.* Ikaros and Aiolos inhibit pre-B-cell proliferation by directly suppressing c-Myc expression. *Mol. Cell Biol.* **30**, 4149–4158 (2010).
10. Winandy, S., Wu, P. & Georgopoulos, K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell* **83**, 289–299 (1995).
11. Dumortier, A. *et al.* Notch activation is an early and critical event during T-cell leukemogenesis in Ikaros-deficient mice. *Mol. Cell Biol.* **26**, 209–220 (2006).
12. Mullighan, C.G. *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110–114 (2008).
13. Mullighan, C.G. *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* **360**, 470–480 (2009).
14. Kim, J. *et al.* Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10**, 345–355 (1999).
15. Sridharan, R. & Smale, S.T. Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. *J. Biol. Chem.* **282**, 30227–30238 (2007).
16. Zhang, J. *et al.* Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat. Immunol.* **13**, 86–94 (2011).
17. Harker, N. *et al.* The CD8α gene locus is regulated by the Ikaros family of proteins. *Mol. Cell* **10**, 1403–1415 (2002).
18. Naito, T., Gomez-Del Arco, P., Williams, C.J. & Georgopoulos, K. Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2β determine silencer activity and Cd4 gene expression. *Immunity* **27**, 723–734 (2007).
19. Gomez-del Arco, P. *et al.* Alternative promoter usage at the Notch1 locus supports ligand-independent signaling in T cell development and leukemogenesis. *Immunity* **33**, 685–698 (2010).
20. Ng, S.Y., Yoshida, T., Zhang, J. & Georgopoulos, K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity* **30**, 493–507 (2009).
21. John, L.B. & Ward, A.C. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. *Mol. Immunol.* **48**, 1272–1278 (2011).
22. Sun, L., Liu, A. & Georgopoulos, K. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* **15**, 5358–5369 (1996).
23. Trinh, L.A. *et al.* Down-regulation of TDT transcription in CD4+CD8+ thymocytes by Ikaros proteins in direct competition with an Ets activator. *Genes Dev.* **15**, 1817–1832 (2001).
24. McCarty, A.S., Kleiger, G., Eisenberg, D. & Smale, S.T. Selective dimerization of a C2H2 zinc finger subfamily. *Mol. Cell* **11**, 459–470 (2003).
25. Tupler, R., Perini, G. & Green, M.R. Expressing the human genome. *Nature* **409**, 832–833 (2001).
26. Ravasi, T. *et al.* Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. *Genome Res.* **13**, 1430–1442 (2003).
27. Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* **79**, 213–231 (2010).
28. Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
29. Molnar, A. & Georgopoulos, K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol. Cell Biol.* **14**, 8292–8303 (1994).
30. Cobb, B.S. *et al.* Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev.* **14**, 2146–2160 (2000).
31. Koipally, J., Heller, E.J., Seavitt, J.R. & Georgopoulos, K. Unconventional potentiation of gene expression by Ikaros. *J. Biol. Chem.* **277**, 13007–13015 (2002).
32. Hahm, K. *et al.* Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. *Genes Dev.* **12**, 782–796 (1998).
33. Payne, K.J. *et al.* Ikaros isoform x is selectively expressed in myeloid differentiation. *J. Immunol.* **170**, 3091–3098 (2003).
34. Miller, J., McLachlan, A.D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1614 (1985).
35. Shastry, B.S. Transcription factor IIIA (TFIIIA) in the second decade. *J. Cell Sci.* **109**, 535–539 (1996).
36. Filippova, G.N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell Biol.* **16**, 2802–2813 (1996).
37. Ohlsson, R., Renkawitz, R. & Lobanenkov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).

38. Renda, M. *et al.* Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* **282**, 33336–33345 (2007).
39. Nurmemmedov, E., Yengo, R.K., Uysal, H., Karlsson, R. & Thunnissen, M.M. New insights into DNA-binding behavior of Wilms tumor protein (WT1)–a dual study. *Biophys. Chem.* **145**, 116–125 (2009).
40. Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678–1689 (2013).
41. Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173**, 1213–1225 (1991).
42. Rolink, A., Grawunder, U., Winkler, T.H., Karasuyama, H. & Melchers, F. IL-2 receptor α chain (CD25, TAC) expression defines a crucial stage in pre-B cell development. *Int. Immunol.* **6**, 1257–1264 (1994).
43. Rothenberg, E.V., Moore, J.E. & Yui, M.A. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* **8**, 9–21 (2008).
44. Randall, T.D., Carragher, D.M. & Rangel-Moreno, J. Development of secondary lymphoid organs. *Annu. Rev. Immunol.* **26**, 627–650 (2008).
45. Juric, D. *et al.* Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. *J. Clin. Oncol.* **25**, 1341–1349 (2007).
46. Lu, Q. delta-Catenin dysregulation in cancer: interactions with E-cadherin and beyond. *J. Pathol.* **222**, 119–123 (2010).
47. Wang, H. *et al.* The role of Crk/Dock180/Rac1 pathway in the malignant behavior of human ovarian cancer cell SKOV3. *Tumour Biol.* **31**, 59–67 (2010).
48. Perentes, J.Y. *et al.* Cancer cell-associated MT1-MMP promotes blood vessel invasion and distant metastasis in triple-negative mammary tumors. *Cancer Res.* **71**, 4527–4538 (2011).
49. Marcais, A. *et al.* Genetic inactivation of Ikaros is a rare event in human T-ALL. *Leuk. Res.* **34**, 426–429 (2010).
50. Wong, S. *et al.* Sole BCR-ABL inhibition is insufficient to eliminate all myeloproliferative disorder cell populations. *Proc. Natl. Acad. Sci. USA* **101**, 17456–17461 (2004).
51. Papathanasiou, P. *et al.* Widespread failure of hematolymphoid differentiation caused by a recessive niche-filling allele of the Ikaros transcription factor. *Immunity* **19**, 131–144 (2003).
52. Lickwar, C.R., Mueller, F., Hanlon, S.W., McNally, J.G. & Lieb, J.D. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2012).
53. Virely, C. *et al.* Haploinsufficiency of the IKZF1 (IKAROS) tumor suppressor gene cooperates with BCR-ABL in a transgenic model of acute lymphoblastic leukemia. *Leukemia* **24**, 1200–1204 (2010).

ONLINE METHODS

Mice. Germline-targeted deletion of exon 4 (encoding zinc finger 1) or exon 6 (encoding zinc finger 4) of *Ikzf1* was achieved by homologous recombination in embryonic stem cells. Embryonic stem cells were transfected by electroporation (Molecular Genetic Technology Center of the University of California, Los Angeles (UCLA)) and were screened for correct targeting by Southern blot analysis. Mutant embryonic stem cell lines were injected into 129 blastocysts by the UCLA Molecular Genetic Technology Center. After germline transmission was achieved, the *lox*P-flanked neomycin-resistance cassette introduced into the genome during homologous recombination was removed by crossing of the mice with *EIIa*-Cre mice on a C57BL/6 background. Germline transmission of the deletions was confirmed after backcrossing with wild-type C57BL/6 mice. The mutant strains were subsequently backcrossed through more than ten generations with wild-type C57BL/6. Mice were excluded from analysis of hematopoietic development if they were runted or if they had developed thymic lymphoma (*Ikzf1*$^{\Delta F4/\Delta F4}$ strain). No randomization or 'blinding' was used for the animal studies. Mice of both sexes were used, at E18.5 (fetal hematopoiesis), 4 weeks (thymic development) and 6–8 weeks of age (peripheral lymph nodes, peritoneal B cells, spleen and bone marrow). All phenotypes described were analyzed with a minimum sample size of $n = 5$, with wild-type littermates as controls and are represented by at least three separate experiments with mice from different litters. Animals were housed in the vivaria of the UCLA Division of Laboratory Animal Medicine. All experiments were approved by the UCLA Animal Research Committee and were done according to guidelines of the UCLA Institutional Animal Care and Use Committee. Wild-type C57BL/6 and *EIIa*-Cre mice were from The Jackson Laboratory.

Cell preparation and flow cytometry. Cell suspensions were prepared from adult hematopoietic tissues and were filtered through 70-μm nylon cell strainers (BD Biosciences). Fetal hematopoietic cells were prepared as described from mesentery[54] and fetal liver[55]. All antibodies for flow cytometry were from BD Biosciences or eBioscience, except for the antibody to cytoplasmic immunoglobulin-μ, which was from Southern Biotech (**Supplementary Table 1**). All antibodies were initially used at a dilution of 1:200 and were individually 'titrated' when necessary. Intracellular staining of cytoplasmic immunoglobulin-μ was done with reagents from the Foxp3 Staining Buffer set (eBioscience). A FACSCalibur, BD LSR I, or FACSAria II SOP (BD Biosciences) was used for flow cytometry or sorting, and data were analyzed with FlowJo 7.5 and FACSDiva 6.1.1 software.

Histological analysis and visualization of lymph nodes and Peyer's patches. Dissected tissues were fixed for >72 h in 10% formalin (4% formaldehyde) (Fisher Scientific), decalcified when needed, embedded in paraffin, sectioned and stained with hematoxylin and eosin by the Translational Pathology Core Laboratory at UCLA. Inguinal and lumbar lymph nodes and Peyer's patches were visualized as described[54,56].

Immunoblot analysis. Whole-cell extracts were prepared by resuspension of cells in one volume water in the presence of protease inhibitor 'cocktail' (Roche) and by immediate direct lysis by the addition of one volume of 2× SDS sample buffer. Samples were separated by SDS-PAGE and transferred to nitrocellulose membrane, and blots were probed with an antibody raised against the N-terminal region of Ikaros (residues 1–80)[32] or antibody to SNP70 (loading control)[57].

Retroviral transduction, cell culture and bone marrow transplantation. Retroviral supernatants for expression of BCR-ABL were produced with the plasmid pMSCV-YFP-IRES-p185, and bone marrow cells were transduced as described[50], except that whole bone marrow from untreated mice was used for transduction. For *in vitro* cell culture, 5×10^6 transduced cells were plated on top of pre-established feeder cells from wild-type C57BL/6 bone marrow stroma. Cells were split and counted and then reseeded at a density of 1×10^5 to 2×10^5 cells per ml in a volume of 5 ml every 2–3 d for growth analysis. For *in vivo* experiments, 1×10^6 transduced whole bone marrow cells were injected intravenously into irradiated wild-type C57BL/6 recipient mice. Mice were monitored for development of B-ALL, with the endpoint of complete paralysis of hind legs or a moribund condition.

RNA purification and RNA-Seq. RNA was prepared with TRI Reagent (Molecular Research Center), followed by purification with an RNeasy kit (Qiagen), with on-column treatment with RNase-free DNase I. For RNA-Seq analysis, MicroPoly(A)Purist Kit (Ambion) or the Tru-Seq RNA Sample Prep Kit (Illumina) was used to isolate mRNA according to the manufacturer's protocol. cDNA was prepared for the Illumina Sequencing platform by the fragmented double-stranded cDNA protocol[58] or an Illumina TruSeq kit. Samples were sequenced on an Illumina HiSeq instrument with 50–base pair single-end reads, at the UCLA Broad Stem Cell Research Center High Throughput Sequencing Core. Raw data were uploaded to the Galaxy website (Pennsylvania State University) and were mapped (with the TopHat fast splice junction mapper for RNA-Seq) to the July 2007 annotation of the mouse (*Mus musculus*) genome (the mm9 assembly of the mouse genome of the National Center for Biotechnology Information), with filtering for uniquely mapped 'reads'[59–63]. Genome coverage visualization files were created with the BEDtools flexible software suite of utilities for the comparison of genomic features and were uploaded to University of California at Santa Cruz (UCSC) Genome Browser[64]. Relative mRNA expression (RPKM)[65] was calculated on the basis of exonic 'reads' with SeqMonk software (Babraham Bioinformatics) and reference genome annotations from the National Center for Biotechnology Information (the mm9 assembly of the mouse genome). For RNA-Seq analysis of DP thymocytes, exonic 'reads' were analyzed by the Bioconductor DESeq program[66] for the identification of genes with statistically significant different expression in two populations. Clustering analysis (k-means) was done on $\log_2$-transformed, mean centered RPKM. values, with Cluster 3.0 software[67,68] and visualized with Java Treeview software[69]. Gene ontology analysis was performed using Panther software[70].

ChIP and ChIP-seq analysis. Primary thymocytes were fixed for 10 min at 25 °C in 1% formaldehyde, then were washed in PBS and 'snap frozen'. ChIP of Ikaros with an antibody directed against the N-terminal region of Ikaros (residues 1–80)[32] was followed by PCR analysis of enrichment at published Ikaros-binding sites[18,19] and unpublished (Zfp260) Ikaros-binding sites identified in initial ChIP-seq experiments. Input DNA was prepared from wild-type thymocytes as a negative control. Libraries were prepared for sequencing on the Illumina platform as described[71] and were sequenced at the University of Southern California Epigenome Data Production Facility and the UCLA Broad Stem Cell Research Center High Throughput Sequencing Core. Raw sequences were mapped (with Bowtie software for the alignment of short DNA sequences) to the July 2007 annotation of the mm9 assembly, and significant peaks over background were 'called' with Sole-Search software[72,73]. Genome coverage visualization files were created with BEDtools and were uploaded to the UCSC Genome Browser[64]. Homer software[74] was used for *de novo* motif analysis, and MEME software[75] was used for position weight matrix analysis.

Recombinant protein preparation and EMSA. *Ikzf1* sequences encoding DNA-binding zinc fingers 1–4 were amplified by PCR from cDNA and were subcloned into the bacterial expression plasmid pGEX-4T-1 (GE Healthcare) by standard methods. Recombinant proteins were expressed in Rosetta (DE3) competent cells (Novagen), were induced with isopropyl β-D-thiogalactopyranoside, and purified using B-PER GST Fusion Protein Purification Kit (Thermo Scientific). The glutathione S-transferase tags were removed and proteins were purified by on-column thrombin cleavage. EMSA of Ikaros was done as described[30], except that bacterially expressed recombinant Ikaros proteins were used instead of nuclear extracts. Sequences of probes are in **Supplementary Table 2**.

Statistical analysis. GraphPad Prism software was used for statistical analyses (unpaired, two-tailed, Student's *t* test, 95% confidence intervals).

54. Sun, Z. *et al.* Requirement for RORgamma in thymocyte survival and lymphoid organ development. *Science* **288**, 2369–2373 (2000).

55. Montecino-Rodriguez, E., Leathers, H. & Dorshkind, K. Identification of a B-1 B cell-specified progenitor. *Nat. Immunol.* **7**, 293–301 (2006).

56. Aliahmad, P., de la Torre, B. & Kaye, J. Shared dependence on the DNA-binding factor TOX for the development of lymphoid tissue-inducer cell and NK cell lineages. *Nat. Immunol.* **11**, 945–952 (2010).

57. Pandya-Jones, A. & Black, D.L. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**, 1896–1908 (2009).

58. Nagalakshmi, U., Waern, K. & Snyder, M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol.* **89**, 4.11.1–4.11.13 (2010).

59. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).

60. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).

61. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **89**, 19.10.1–19–10–21 (2010).

62. Fujita, P.A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2010).

63. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).

64. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

65. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

66. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

67. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).

68. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).

69. Saldana, A.J. Java Treeview - extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).

70. Thomas, P.D. *et al.* Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* **34**, W645–W650 (2006).

71. O'Geen, H., Frietze, S. & Farnham, P.J. Using ChIP-seq technology to identify targets of zinc finger transcription factors. *Methods Mol. Biol.* **649**, 437–455 (2010).

72. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

73. Blahnik, K.R. *et al.* Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Res.* **38**, e13 (2010).

74. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

75. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

References

[1] 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(9):R137.

[2] Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A. The Ikaros gene is required for the development of all lymphoid lineages. Cell. 1994 Oct 7;79(1):143-56.

[3] Georgopoulos K, Moore DD, Derfler B. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science. 1992 Oct 30;258(5083):808-12.

[4] Trinh LA, Ferrini R, Cobb BS, Weinmann AS, Hahm K, Ernst P, Garraway IP, Merkenschlager M, Smale ST. Down-regulation of TDT transcription in CD4(+)CD8(+) thymocytes by Ikaros proteins in direct competition with an Ets activator. Genes Dev. 2001 Jul 15;15(14):1817-32.

[5] Koipally J, Heller EJ, Seavitt JR, Georgopoulos K. Unconventional potentiation of gene expression by Ikaros. J Biol Chem. 2002 Apr 12;277(15):13007-15.

[6] Harker N, Naito T, Cortes M, Hostert A, Hirschberg S, Tolaini M, Roderick K, Georgopoulos K, Kioussis D. The CD8 alpha gene locus is regulated by the Ikaros family of proteins. Mol Cell. 2002 Dec;10(6):1403-15.

[7] Thompson EC, Cobb BS, Sabbattini P, Meixlsperger S, Parelho V, Liberg D, Taylor B, Dillon N, Georgopoulos K, Jumaa H, Smale ST, Fisher AG, Merkenschlager M. Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits. Immunity. 2007 Mar;26(3):335-44.

[8] Naito T, Gómez-Del Arco P, Williams CJ, Georgopoulos K. Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression. Immunity. 2007 Nov;27(5):723-34.

[9] Reynaud D, Demarco IA, Reddy KL, Schjerven H, Bertolino E, Chen Z, Smale ST, Winandy S, Singh H. Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. Nat Immunol. 2008 Aug;9(8):927-36.

[10] Ng SY, Yoshida T, Zhang J, Georgopoulos K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity. 2009 Apr;30(4):493-507

[11] Kim J, Sif S, Jones B, Jackson A, Koipally J, Heller E, Winandy S, Viel A, Sawyer A, Ikeda T, Kingston R, Georgopoulos K. Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity. 1999 Mar;10(3):345-55.

[12] Koipally J, Renold A, Kim J, Georgopoulos K. Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. EMBO J. 1999 Jun 1;18(11):3090-100.

[13] Koipally J, Georgopoulos K. Ikaros interactions with CtBP reveal a repression mechanism that is independent of histone deacetylase activity. J Biol Chem. 2000 Jun 30;275(26):19594-602.

[14] Koipally J, Georgopoulos K. Ikaros-CtIP interactions do not require C-terminal binding protein and participate in a deacetylase-independent mode of repression. J Biol Chem. 2002 Jun 28;277(26):23143-9.

[15] Koipally J, Georgopoulos K. A molecular dissection of the repression circuitry of Ikaros. J Biol Chem. 2002 Aug 2;277(31):27697-705.

[16] Sridharan R, Smale ST. Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. J Biol Chem. 2007 Oct 12;282(41):30227-38.

[17] Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell. 1997 Dec 12;91(6):845-54.

[18] Klug CA, Morrison SJ, Masek M, Hahm K, Smale ST, Weissman IL. Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. Proc Natl Acad Sci U S A. 1998 Jan 20;95(2):657-62.

[19] Cobb BS, Morales-Alcelay S, Kleiger G, Brown KE, Fisher AG, Smale ST. Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. Genes Dev. 2000 Sep 1;14(17):2146-60.

[20] Hahm K, Ernst P, Lo K, Kim GS, Turck C, Smale ST. The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. Mol Cell Biol. 1994 Nov;14(11):7111-23.

[21] Molnár A, Georgopoulos K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Mol Cell Biol. 1994 Dec;14(12):8292-303.

[22] Wang JH, Nichogiannopoulou A, Wu L, Sun L, Sharpe AH, Bigby M, Georgopoulos K. Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. Immunity. 1996 Dec;5(6):537-49.

[23] Winandy S, Wu P, Georgopoulos K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. Cell. 1995 Oct 20;83(2):289-99.

[24] Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui CH, Relling MV, Shurtleff SA, Downing JR. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008 May 1;453(7191):110-4.

[25] Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, Ma J, Liu W, Cheng C, Schulman BA, Harvey RC, Chen IM, Clifford RJ, Carroll WL, Reaman G, Bowman WP, Devidas M, Gerhard DS, Yang W, Relling MV, Shurtleff SA, Campana D, Borowitz MJ, Pui CH, Smith M, Hunger SP, Willman CL, Downing JR. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009 Jan 29;360(5):470-80.

[26] Beverly LJ, Capobianco AJ. Perturbation of Ikaros isoform selection by MLV integration is a cooperative event in Notch(IC)-induced T cell leukemogenesis. Cancer Cell. 2003 Jun;3(6):551-64.

[27] Dumortier A, Jeannet R, Kirstetter P, Kleinmann E, Sellars M, dos Santos NR, Thibault C, Barths J, Ghysdael J, Punt JA, Kastner P, Chan S. Notch activation is an early and critical event during T-Cell leukemogenesis in Ikaros-deficient mice. Mol Cell Biol. 2006 Jan;26(1):209-20.

[28] Chari S, Winandy S. Ikaros regulates Notch target gene expression in developing thymocytes. J Immunol. 2008 Nov 1;181(9):6265-74.

[29] Kleinmann E, Geimer Le Lay AS, Sellars M, Kastner P, Chan S. Ikaros represses the transcriptional response to Notch signaling in T-cell development. Mol Cell Biol. 2008 Dec;28(24):7465-75.

[30] Schjerven H, McLaughlin J, Arenzana TL, Frietze S, Cheng D, Wadsworth SE, Lawson GW, Bensinger SJ, Farnham PJ, Witte ON, Smale ST. Selective regulation of lymphopoiesis and leukemogenesis by individual zinc fingers of Ikaros. Nat Immunol. 2013 Oct;14(10):1073-83.

[31] John LB, Ward AC. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. Mol Immunol. 2011 May;48(9-10):1272-8.

[32] Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu. Rev. Biochem. 79, 213–231 (2010).

[33] Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. 29, 183–212 (2000).

[34] Ravasi T, Huber T, Zavolan M, Forrest A, Gaasterland T, Grimmond S, Hume DA; RIKEN GER Group; GSL Members. Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Res. 2003 Jun;13(6B):1430-42.

[35] Shastry, B.S. Transcription factor IIIA (TFIIIA) in the second decade. J. Cell Sci. 109, 535–539 (1996).

[36] Filippova, G.N. et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell Biol. 16, 2802–2813 (1996).

[37] Nakahashi, H. et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 3, 1678–1689 (2013).

[38] Kaplan T, Friedman N, Margalit H. Ab initio prediction of transcription factor targets using structural knowledge. PLoS Comput Biol. 2005 Jun;1(1):e1.

[39] Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat Genet. 2004 Dec;36(12):1331-9.

[40] Stoll R, Lee BM, Debler EW, Laity JH, Wilson IA, Dyson HJ, Wright PE. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. J Mol Biol. 2007 Oct 5;372(5):1227-45.

[41] Schaufler LE, Klevit RE. Mechanism of DNA binding by the ADR1 zinc finger transcription factor as determined by SPR. J Mol Biol. 2003 Jun 20;329(5):931-9.

[42] Cooper MA. Optical biosensors in drug discovery. Nat Rev Drug Discov. 2002 Jul;1(7):515-28.

[43] Nurmemmedov E, Yengo RK, Uysal H, Karlsson R, Thunnissen MM. New insights into DNA-binding behavior of Wilms tumor protein (WT1)--a dual study. Biophys Chem. 2009 Dec;145(2-3):116-25.

44 Sakkhachornphop S, Jiranusornkul S, Kodchakorn K, Nangola S, Sirisanthana T, Tayapiwatana C. Designed zinc finger protein interacting with the HIV-1 integrase recognition sequence at 2-LTR-circle junctions. Protein Sci. 2009 Nov;18(11):2219-30.

45 Pio F, Assa-Munt N, Yguerabide J, Maki RA. Mutants of ETS domain PU.1 and GGAA/T recognition: free energies and kinetics. Protein Sci. 1999 Oct;8(10):2098-109.

46 Avitahl N, Winandy S, Friedrich C, Jones B, Ge Y, Georgopoulos K. Ikaros sets thresholds for T cell activation and regulates chromosome propagation. Immunity. 1999 Mar;10(3):333-43.

47 Bonetti P, Testoni M, Scandurra M, Ponzoni M, Piva R, Mensah AA, Rinaldi A, Kwee I, Tibiletti MG, Iqbal J, Greiner TC, Chan WC, Gaidano G, Piris MA, Cavalli F, Zucca E, Inghirami G, Bertoni F. Deregulation of ETS1 and FLI1 contributes to the pathogenesis of diffuse large B-cell lymphoma. Blood. 2013 Sep 26;122(13):2233-41.

48 Y Gu, G Cimino, H Alder, T Nakamura, R Prasad, O Canaani, D T Moir, C Jones, P C Nowell, C M Croce. The (4;11)(q21;q23) chromosome translocations in acute leukemias involve the VDJ recombinase. Proc Natl Acad Sci U S A. 1992 November 1; 89(21): 10464–10468.

49 Kornblau SM, Qiu YH, Zhang N, Singh N, Faderl S, Ferrajoli A, York H, Qutub AA, Coombes KR, Watson DK. Abnormal expression of FLI1 protein is an adverse prognostic factor in acute myeloid leukemia. Blood. 2011 Nov 17;118(20):5604-12.

50 Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013 Sep;8(9):1765-86.

51 Gómez-del Arco P, Maki K, Georgopoulos K. Phosphorylation controls Ikaros's ability to negatively regulate the G(1)-S transition. Mol Cell Biol. 2004 Apr;24(7):2797-807.

52 Hahm K, Ernst P, Lo K, Kim GS, Turck C, Smale ST. The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. Mol Cell Biol. 1994 Nov;14(11):7111-23.

53 Harker et al., 2002 N. Harker, T. Naito, M. Cortes, A. Hostert, S. Hirschberg, M. Tolaini, K. Roderick, K. Georgopoulos and D. Kioussis, The CD8α gene locus is regulated by the Ikaros family of proteins, Mol. Cell 10 (2002), pp. 1403–1415.

54 Naito T, Gómez-Del Arco P, Williams CJ, Georgopoulos K. Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression. Immunity. 2007 Nov;27(5):723-34.

55 Thompson EC, Cobb BS, Sabbattini P, Meixlsperger S, Parelho V, Liberg D, Taylor B, Dillon N, Georgopoulos K, Jumaa H, Smale ST, Fisher AG, Merkenschlager M. Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits. Immunity. 2007 Mar;26(3):335-44.

56 Reynaud D, Demarco IA, Reddy KL, Schjerven H, Bertolino E, Chen Z, Smale ST, Winandy S, Singh H. Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. Nat Immunol. 2008 Aug;9(8):927-36.

57 Ferreirós-Vidal I, Carroll T, Taylor B, Terry A, Liang Z, Bruno L, Dharmalingam G, Khadayate S, Cobb BS, Smale ST, Spivakov M, Srivastava P, Petretto E, Fisher AG, Merkenschlager M. Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation.Blood. 2013 Mar 7;121(10):1769-82.

58 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(9):R137.