



Published in final edited form as:

*J Immunol.* 2013 April 1; 190(7): 3276–3288. doi:10.4049/jimmunol.1202970.

## Loss of T cell progenitor checkpoint control underlies leukemia initiation in *Rag1*-deficient NOD mice

Mary A. Yui<sup>\*</sup>, Ni Feng<sup>\*</sup>, Jingli A. Zhang<sup>\*</sup>, Chen Yee Liaw<sup>\*</sup>, Ellen V. Rothenberg<sup>\*</sup>, and Jeffrey A. Longmate<sup>†,1</sup>

<sup>\*</sup>Division of Biology, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

<sup>†</sup>Division of Biostatistics, City of Hope, Duarte, CA 91010

### Abstract

NOD mice exhibit major defects in the earliest stages of T cell development in the thymus. Genome-wide genetic and transcriptome analyses were used to investigate the origins and consequences of an early T cell developmental checkpoint breakthrough in *Rag1*-deficient NOD mice. QTL analysis mapped the presence of checkpoint breakthrough cells to several known NOD diabetes susceptibility regions, particularly *Idd9/11* on chromosome 4, suggesting common genetic origins for T cell defects affecting this trait and autoimmunity. Genome-wide RNA deep-sequencing of NOD and B6 *Rag1*-deficient thymocytes revealed the effects of genetic background prior to breakthrough, as well as the cellular consequences of the breakthrough. Transcriptome comparison between the two strains showed enrichment in differentially expressed signal transduction genes, prominently tyrosine kinase and actin-binding genes, in accord with their divergent sensitivities to activating signals. Emerging NOD breakthrough cells aberrantly expressed both stem cell-associated proto-oncogenes, such as *Lmo2*, *Hhex*, *Lyl1*, and *Kit* which are normally repressed at the commitment checkpoint, and post- $\beta$ -selection checkpoint genes including *Cd2* and *Cd5*. Co-expression of genes characteristic of multi-potent progenitors and more mature T cells persists in the expanding population of thymocytes, and in the thymic leukemias that emerge with age in these mice. These results show that *Rag1*-deficient NOD thymocytes have T cell defects that can collapse regulatory boundaries at two early T cell checkpoints, which may predispose them to both leukemia and autoimmunity.

### Introduction

All T cells arise from a small pool of multipotent progenitors, which undergo proliferation and tightly controlled developmental programming induced and sustained by interactions with the thymic environment rich in Notch ligands and cytokines (1, 2). Thymic precursors are CD4 and CD8 double negative (DN)<sup>2</sup>, and are defined by sequential changes in surface Kit (CD117), CD25, and CD44 into stages: DN1 (or early T cell progenitor, ETP; Kit<sup>hi</sup>CD44<sup>hi</sup>CD25<sup>-</sup>), DN2 (Kit<sup>hi</sup>CD44<sup>hi</sup>CD25<sup>+</sup>), DN3 (Kit<sup>lo</sup>CD44<sup>lo</sup>CD25<sup>+</sup>), and DN4 (Kit<sup>lo</sup>CD44<sup>lo</sup>CD25<sup>lo</sup>) (3, 4). A TCR-independent commitment checkpoint has recently been identified within the DN2 stage, marked by decreased Kit expression from the multipotent DN2a stage to the committed DN2b stage (5), and dependent on *Bcl11b* (6, 7). DN2a cells,

Address correspondence and reprint requests to Mary A. Yui: Phone: 626-395-4915; FAX: 626-449-0756; address: yui@caltech.edu.

<sup>1</sup>**Grant Support.** This work was supported by The National Institutes of Health (AI64590) (to M.A.Y.); Caltech Summer Undergraduate Research Fellowships (to C.Y.L.); the Albert Billings Ruddock Professorship (to E.V.R.); the Louis A. Garfinkle Memorial Laboratory Fund, and the Al Sherman Foundation.

<sup>2</sup>**Abbreviations used in this paper.** DN, double negative; DP, double positive; QTL, quantitative trait locus; RPKM, reads per kilobase per million reads, GeoMean, geometric mean; T-ALL, T cell acute lymphoblastic leukemia; chr, chromosome; WT, wild type

like more immature DN1/ETP cells, undergo proliferative expansion and express legacy stem/progenitor genes such as *Kit*, *Lyl1*, *Tal1*, *Gfi1b*, and *Sfp1* (PU.1). These features decline sharply in the DN2b/DN3 stages when T cell specification genes are strongly activated, proliferation slows, and efficient TCR gene rearrangement begins (5, 8). Thus, the T cell commitment checkpoint divides the TCR-negative stages of development into two phases: Phase I, wherein cells proliferate and retain alternative lineage potential, and Phase II, which prepares committed DN3 cells for the first TCR-dependent checkpoint,  $\beta$ -selection (9). Normally, only cells that successfully rearrange a TCR $\beta$  and assemble a signaling pre-TCR complex are permitted to pass through the  $\beta$ -selection checkpoint to DN4 and proliferate. These cells then become CD4<sup>+</sup>CD8<sup>+</sup> double positive (DP), express TCR $\alpha\beta$ , and undergo positive and negative selection (10, 11). *Rag*-deficient T cells are blocked at  $\beta$ -selection and do not generate DP cells.

The non-obese diabetic (NOD) mouse is a model of T cell mediated autoimmune Type 1 diabetes, with over 20 genetic regions associated with diabetes susceptibility, many of which specifically affect T cell activities (12–15). Most T lineages have been implicated in autoimmune susceptibility or resistance in NOD mice, including CD4, CD8, NKT, T<sub>reg</sub>, and  $\gamma\delta$ T cells (16), raising the possibility that all NOD T cells may share fundamental abnormalities that contribute to loss of self-tolerance and might be traceable to their common progenitors (8). We previously reported defects in the earliest stages of T cell development in both wild type (WT) and TCR-deficient NOD mice. Precursor cells from WT NOD mice exhibit impaired  $\alpha\beta$ T-cell development and enhanced  $\gamma\delta$ T-cell generation apparently arising from the ETP-DN2 stages (17). In addition, the NOD genetic background severely compromises  $\beta$ -selection checkpoint control: despite the absence of TCR expression in NOD.SCID and NOD.*Rag1*<sup>-/-</sup> (NOD.*Rag*) mice, aberrant breakthrough DP thymocytes spontaneously appear in all young adult animals (18). These TCR-negative DP cells do not fully mimic normal  $\beta$ -selected cells, as they still express receptors characteristic of earlier DN stages, like *Kit*, IL7R $\alpha$ , and CD25, which are normally turned off by the DP stage (18). Furthermore, older NOD.SCID and NOD.*Rag* mice develop thymic tumors at high frequency while mice of other strains with these immunodeficiencies do not (18–21). This suggests a possible link between early T cell checkpoint control and tumor suppression that may be jointly defective in the NOD genetic background.

We used genome-wide genetic and transcriptome analytical methods to investigate the source and consequences of the NOD.*Rag* thymocyte checkpoint defect. First, we found quantitative trait loci (QTLs) for this trait, all within several of known diabetes susceptibility regions mapped in WT NOD mice. A major QTL localized within the *Idd9/11* region of chromosome (chr)4 was confirmed using congenic mice. In addition, genome-wide transcriptome analyses revealed distinct differences in gene expression between thymocytes from NOD.*Rag* and B6.*Rag* control mice. The genes differentially expressed between the two strains were enriched for those encoding signaling proteins, suggesting aberrant signal transduction as a possible precondition for breakthrough. Furthermore, newly emergent NOD.*Rag* breakthrough cells fail to terminate gene expression programs from earlier stages: they co-express Phase I stem/progenitor genes along with T cell-specific genes characteristic of Phase II and post- $\beta$ -selection stages. This mixed gene expression profile foreshadows the phenotype of thymic tumors found in older mice of this strain, which share characteristics with classes of human early-type acute T cell lymphoblastic leukemia (T-ALL), suggesting that primary defects in early T cell checkpoint control underlie some forms of T-ALL.

## Materials and Methods

### Mice and crosses

B6.129S7-*Rag1<sup>tm1Mom</sup>/J* (B6.*Rag*) and NOD.Cg-*Rag1<sup>tm1Mom</sup> Prf1<sup>tm1Sdz</sup>/SzJ* (NOD.*Rag*) (The Jackson Laboratory) and NOD.B10<sup>*Idd9*</sup> Line 905 (14) (Taconic Farms) mice were bred and maintained in the Caltech Laboratory Animal Facility using autoclaved cages, food, and water. All animal protocols were reviewed and approved by the Animal Care and Use Committee of the California Institute of Technology.

### Genetic crosses, QTL analysis, and congenic mice

For the QTL analysis B6.*Rag* and NOD.*Rag* mice were crossed and intercrossed for F2 or backcrossed to NOD.*Rag* for N2 progeny. Thymocytes from 12–14 wk progeny were phenotyped by flow cytometric analysis. DNA was extracted from tail tips of 150 N2 and 30 F2 cross mice and 150 polymorphic SNPs were genotyped by Jackson Laboratories Genetic Services. QTL analysis was carried out using the R-qt1 program (22) and p-values were obtained from genome-wide significance test using 5,000 permutations (23). Congenic NOD.B10<sup>*Idd9*</sup> *Rag* mice were created by crossing NOD.*Rag* and NOD.B10<sup>*Idd9*</sup> mice and repeated backcrossing until the knockout *Rag1* gene and the B10<sup>*Idd9*</sup> region were homozygous, as determined by PCR analysis.

### Cell cultures and antibody staining

Freshly isolated thymocytes were either stained immediately for flow cytometric analysis or cultured on OP9-DL1 or -DL4 cells with 5 ng/ml of IL-7, as previously described (17). For cell stimulations, thymocytes were cultured for 1 h in RPMI supplemented with 10% fetal bovine serum (Gibco) before treatment with PMA. Cells were immediately fixed in 1.5% formaldehyde in PBS at 37°C and permeabilized by slow addition of ice-cold methanol to a final concentration of 90%. Cells were incubated on ice for 30 min, washed with PBS plus 0.5% BSA, and incubated with either phospho-p42/p44 (Erk1/2)-AlexaFluor 647 antibodies or isotype controls (Cell Signaling Technology, Danvers, MA) before washing and flow cytometric analysis.

### Genome-wide transcriptome analysis

CD25<sup>+</sup> DN thymocytes were FACS-sorted from NOD.*Rag* mice at 4 wks of age (pre-breakthrough) and 7 wks (at the time of first breakthrough), and age-matched B6.*Rag* mice for RNA extraction. mRNA purification and cDNA library building were performed as described (24). Sequencing was done using Illumina High Throughput Genome Analyzer IIx sequencers at Caltech's Jacobs Genetics and Genomics Laboratory and the data have been deposited in NCBI's Gene Expression Omnibus (25) and are accessible through GEO Series accession number GSE40688 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40688>). Computational analysis was carried out using software developed at Caltech (26, 27) as well as Mathematica and R. For analysis, 38 bp, single-read raw sequence reads were trimmed to 32 bp and mapped onto the mouse genome build NCBI37/mm9 using Bowtie (bowtie-0.12.1, <http://bowtie-bio.sourceforge.net/index.shtml>) with setting '-v 2 -k 11 -m 10 -t --best --strata'. The mappable data were then processed by the ERANGE v. 3.3 RNA-Seq analysis program to obtain the expression level in reads per kilobase per million reads (RPKM) for each gene (26). We obtained 13–17 × 10<sup>6</sup> mappable reads per sample for analysis. Based on a pair-wise comparison between samples, B4 values were divided by 1.27 to remove a scaling artifact. Genes were required to have at least one sample with 32 mappable reads (based on a significance level of 0.05, with a Bonferonni adjustment for 12000 comparisons) and have an annotated transcript name, leaving 11,098 named genes for further analysis. For each gene, comparisons of expression in the four

samples were calculated as differences on a  $\log_2$  scale, after replacing any zeros by 0.01 as an approximate limit of detection. The significance probabilities (p-values) were based on the assumption of Poisson distributions of counts, under the null hypothesis of equal rates, allowing for the difference in the number of Mb sequenced (and mapped). Three-fold change is typically the more stringent requirement and was used to determine the differentially expressed genes. The statistical significance requirement screens out some genes with low expression whose 3-fold differences might be attributable to sampling variation. The design does not permit assessing extra-Poisson variation, which would be more important at larger expression levels, where the 3-fold criterion rather than the statistical comparison is the dominant concern.

To profile the connection between different populations, hierarchical clustering was carried out for each selected subgroup of genes from B6.*Rag* and NOD.*Rag* samples or WT B6 ETP to DP populations (24). Individual mRNA data for the selected genes were first normalized by the corresponding geometric mean (GeoMean) of comparing populations, then, one-dimensional hierarchical clustering (along genes) was performed on the results after  $\log_2$  transformation. Euclidean distance and Ward linkage were used (MATLAB 7.10.0). Clusters were visualized as heat maps.

### Real-time QPCR analysis

QPCR was carried out based on selected RNA-Seq results using FACS-sorted thymocyte samples. RNA was extracted and reverse-transcribed and QPCRs were carried out using SyBr GreenER (Invitrogen) in a GeneAmp 7900HT Sequence Detection System (Applied Biosystems), as previously described (17). Relative expression was calculated for each gene by the  $\Delta C_T$  (change in cycling threshold) method, normalized to  $\beta$ -actin expression. Primer sequences were as previously published (6, 28) plus: *Zap70*: F-5' - TGTCCTCCTGAGATGTATGCAC-3', 5' - ATAGTTCCGCATACGTTGTTCC-3'; *Ptcra*: 5' - CTGGCTCCACCCATCACACT-3', R-5' - TGCCATTGCCAGCTGAGA-3'; *Notch1*: F-5' - CCACTGTGAACTGCCCTATGT-3', R-5' - TTGTTTCTGGACAGTCATCC-3'; *Heyl*: F-5' - AAGCTGGAGAAAGCTGAGGTC-3', R-5' - CCAATACTCCGGAAGTCAACA-3'; *Kit*: 5' - ACTTCGCCTGACCAGATTTAAA-3', R-5' - CGTACGTCAGGATTTCTGGTT-3'; *Hhex*: F-5' - ACTACACGCACGCCCTACTC-3', R-5' - GTCGTTGGAGAACCCTCACTTG-3'; *Epha2*: F-5' - CAGGAAGGCTACGAGAAGGTC-3', R-5' - CAGGGTATGCTCTGGACACTC-3'; *Milt4*: F-5' - GACTGGACAGTGACAGGGTGT-3', R-5' - CAGTATCAGTTCAGGCCAGT-3'; *Dapk1*: F-5' - CATCACCCCTGCATGAGGTCTA-3', R-5' - TGCCTCTCTCAGTCAGAGA-3'; *Vegfa*: F-5' - CTCCGAAACCATGAACCTTTCT-3', R-5' - ATGGGACTTCTGCTCTCCTTC-3'; *Tnfrsf9*: F-5' - GCTGGCCCTGATCTTCATTA -3', R-5' - ATCGGCAGCTACAAGCATCT -3'. The Tcf7 primers (28) yielded lower values than previously published (5), however, the relative efficiency was consistent between samples in these experiments.

## Results

### The NOD.*Rag* breakthrough trait maps to the *Idd9/11* region of chr4

We previously reported that aberrant Kit<sup>+</sup>DP breakthrough cells spontaneously appear in the thymuses of all male and female NOD.*Rag* mice by 8 weeks of age, while B6.*Rag* and (B6 X NOD)F1.*Rag* thymocytes arrest normally in DN3 at the  $\beta$ -selection checkpoint (18). To map the genetic basis of the T cell breakthrough, we tested progeny from a (B6 X NOD)F2.*Rag* intercross and an N2 backcross with NOD.*Rag* as the recurrent parent. Thymocytes from individual parental and cross progeny were phenotyped by flow cytometry at 12–14 wks of age for two traits: (1) the percentage of CD4<sup>+</sup> (including DP) cells, which is

indicative of illegitimate progression past the  $\beta$ -selection checkpoint, and (2) the percentage of Kit<sup>+</sup> cells, indicative of a failure to downregulate a key Phase I stem/progenitor gene (see Fig. 1E for a model of developmental stages). As shown in Figure 1A, B6.*Rag* and (B6 X NOD)F1.*Rag* mice at 12 wks of age exhibited very few CD4<sup>+</sup> or Kit<sup>+</sup> cells, whereas the majority of thymocytes from NOD.*Rag* mice were CD4<sup>+</sup> and Kit<sup>+</sup>. Results for individual progeny differed, e.g. #87 exhibited a non-breakthrough phenotype similar to B6.*Rag* and #86 showed Kit<sup>+</sup>CD4<sup>+</sup> breakthrough cells like NOD.*Rag* mice. The F2 intercross yielded a low percentage of progeny with elevated percentages of CD4<sup>+</sup> and/or Kit<sup>+</sup> cells (<10%), whereas approximately 35% of N2 backcross progeny exhibited T cell breakthrough. Despite indicating differing developmental defects, the traits are not independent, as shown in the strong correlation ( $R^2=0.82$ ) between the log-transformed percentages of CD4<sup>+</sup> and Kit<sup>+</sup> cells for individual progeny (Fig. 1B).

For the QTL analysis, 150 N2 and 30 F2 phenotyped progeny were genotyped for 150 polymorphic SNPs distributed over the 19 autosomal chromosomes. Genome-wide one-dimensional QTL scans were performed using R-qt1 on data from the N2 and F2 crosses alone and combined (22). Peak QTL LOD scores for the two phenotypes and the N2 and F2 crosses, individually and combined, are shown in Table 1, and genome-wide QTL scans for the combined cross data are shown in Fig. 1C. A highly significant QTL (LOD >12) was detected on distal chr4 for both %CD4<sup>+</sup> and %Kit<sup>+</sup> in both the N2 and the combined cross (Fig. 1C, D). Weaker QTLs were detected on chr17 for both phenotypes and N2 and combined crosses (LOD > 3 for %Kit) (Fig. 1C, D), and on chr13, in the N2 cross only (LOD>2.5). A scan using the F2 cross alone revealed only a suggestive locus on chr11. Remarkably, all four of these QTLs map to known NOD insulin dependent diabetes susceptibility (*Idd*) QTL regions. The chr4 QTL lies within the *Idd9/11* region (14, 29, 30), the chr17 QTL encompasses the region containing *Idd1*, *Idd16*, *Idd23*, and *Idd24* (31–33) (Fig. 1D), the chr13 QTL region overlaps with *Idd14* (34), and the chr11 QTL region overlaps with *Idd4* (29). These results allow the possibility that the same genes could be involved in the loss of checkpoint control and susceptibility to autoimmune diabetes in NOD mice.

Control of %CD4<sup>+</sup> and %Kit<sup>+</sup> mapped to the same genetic regions, therefore they arise from linked or identical genetic sources, and distinct from the *Cd4* (chr6) or *Kit* (chr5) genes themselves. The aberrant upregulation of CD4 occurs in cells that already express high levels of Kit (18), and results in Fig. 1B support this: no progeny were %CD4<sup>+</sup> high %Kit<sup>+</sup> low (right bottom quadrant), whereas a few progeny were %Kit<sup>+</sup> high %CD4 low (left top quadrant). Also, QTL LOD scores for %Kit<sup>+</sup> were typically higher than for %CD4 (Table 1, Fig. 1C, D), indicating that Kit may be a more sensitive measure of breakthrough. These cells express CD25 and IL7R $\alpha$ , as well as Kit (18), and their abnormal coexpression with CD4 (and CD8) suggests that the breakthrough arises from specific genes on the NOD genetic background that cause bypass of both the T cell commitment and  $\beta$ -selection checkpoints, as shown in Fig. 1E (and discussed further below).

### The congenic B10<sup>*Idd9*</sup> region of chr4 attenuates the NOD.*Rag* thymocyte breakthrough

To confirm the presence of a gene or genes in the NOD *Idd9/11* chr4 QTL region controlling the breakthrough trait, NOD.B10<sup>*Idd9*</sup> congenic mice (Line 905) (14) were bred to NOD.*Rag* to generate a NOD.B10<sup>*Idd9*</sup>*Rag* congenic mouse line. The congenic chr4 region contains at least three distinct diabetes resistance alleles from B10 mice, *Idd9.1*, *Idd9.2* and *Idd9.3* (Fig. 2A), as well as a possible *Idd11* locus (30). The NOD.B10<sup>*Idd9*</sup>*Rag* mice showed significantly lower percentages of CD4<sup>+</sup> and DP breakthrough cells than NOD.*Rag* mice (Fig. 2B, C), demonstrating that the chr4 congenic region includes at least one gene involved in susceptibility to thymocyte breakthrough. Other genetic regions also contribute, as

NOD.B10<sup>Idd9</sup>*Rag* congenic thymocytes exhibited more breakthrough cells than B6.*Rag* (Fig. 2B, C), in accord with the QTL analysis.

### Differentially expressed genes between *Rag*-deficient NOD and B6 CD25<sup>+</sup>DN cells are enriched for actin-binding and signal transduction genes

Insights into the breakthrough program should emerge from both determining underlying strain-specific gene expression differences that precede the loss of checkpoint control, and investigating the earliest post-breakthrough changes. To first evaluate differences between NOD.*Rag* and B6.*Rag* thymocytes that restrict checkpoint control loss to the NOD background, we carried out a genome-wide RNA deep sequencing analysis (RNA-Seq) (26), comparing sorted CD4<sup>-</sup>CD25<sup>+</sup>DN thymocytes from NOD.*Rag* mice at 4 wks (N4, before breakthrough) and 7 wks (N7, early breakthrough), with corresponding non-breakthrough cells from B6.*Rag* mice at 4 (B4) and 7 (B7) wks. In each sample, the sorted cells are predominantly pre- $\beta$ -selection DN3 cells and they were phenotypically similar (Fig. 3A).

Of >11,000 expressed genes, 412 genes were differentially expressed by at least 3-fold between the two strains (excluding 10 genes more prominently differing between N7 and N4, see below) (Fig. 3B, C, and Supplemental Table 1). These genes are normally expressed in at least one stage of T cell development, from ETP to DP, based on our data from WT B6 mice (24), as shown in heatmaps in Fig. 3D. Genes with strain-specific differences in expression, at higher and lower levels in NOD cells, exhibit highly varied patterns of expression across these stages of T-cell development, in contrast to what we find with genes specific to N7 post-breakthrough cells (see below). Overall, 34% of the strain-dependent differentially expressed genes have peak expression in Phase I (23% ETP, 11% DN2a), 40% in Phase II (19% DN2b, 21% DN3), and 26% in post- $\beta$ -selection DP cells.

Gene ontology analysis of the differentially expressed genes using DAVID Bioinformatics Resources 6.7 (35) showed three types of genes were overrepresented: MHC (H2), signaling, and actin-binding protein genes (Table 2). Enrichment of H2 genes is not unexpected due to their expression diversity between mouse strains, although no role is known for them in early T cell development. Protein kinase genes, including tyrosine and serine/threonine kinases, were enriched, as were adherens junction protein genes, which include actin-binding, tyrosine kinase, and calmodulin-binding protein genes. All of these protein classes play critical roles in modulating signal transduction activity in T cells (36). Notably, all 11 of the differentially expressed tyrosine kinases were higher in NOD.*Rag* cells, suggesting that the two strains likely exhibit differences in signaling activity. This may not only contribute to checkpoint failure, but also promote the thymic lymphomas that develop in these mice, as tyrosine kinases are common oncogenes and therapeutic targets in leukemia and other malignancies (37).

Several genes differentially expressed between NOD.*Rag* and B6.*Rag* are located within and near the chr4 QTL peak region (Supplemental Table 1). RNA-Seq tracks for several of these genes are shown in Fig. 3E, including *Epha2*, an ephrin tyrosine kinase receptor strongly expressed in both NOD samples; *Padi3*, a peptidyl arginine deiminase; *Tnfrsf9* (CD137, 4-1BB, an *Idd9.3* candidate gene (14)); *Ctnnbip1* (ICAT), a Wnt pathway inhibitor; and *Hdac1*, a histone deacetylase, which is differentially spliced. These genes contrast with the stably expressed chr4 region gene, *Lck*. Furthermore, a number of poorly defined transcription units are differentially expressed between the two strains, including at least three in a region near the chr4 peak that is predicted to encode multiple KRAB-zinc finger proteins (unpublished observation). Differentially expressed genes of interest were also found within other QTL regions include *Vegfa*; *Notch4*; *Milt4* (Afadin), a Ras-associated adherens junction gene; and *Ddr1*, discoidin domain tyrosine kinase receptor, found in the chr17 QTL region, plus a tumor suppressor gene, death-associated protein kinase, *Dapk1*,

located within the chr13 QTL region and expressed in B6.*Rag* but not in NOD.*Rag* cells (Fig. 3F).

### Signaling and developmental differences between NOD.*Rag* and B6.*Rag* thymocytes

The enrichment of differentially expressed signal transduction pathway genes suggests that inappropriate signaling might underlie the NOD.*Rag* checkpoint breakthrough. Thymocytes from sex-matched 4–5 wk old pre-breakthrough NOD.*Rag* and B6.*Rag* mice were assayed for differences in responses to the protein kinase C activator PMA, in short- and long-term assays. Thymocytes from both strains used for these experiments were phenotypically similar and >90% were DN3 stage cells (18). Phosphorylation of Erk1/2, mediators of the Ras/MAPK pathway, was measured at by phospho-Erk1/2 (pErk) intracellular staining of thymocytes from mice stimulated with PMA (Fig. 4A). NOD.*Rag* populations exhibited cells with consistently lower levels of pErk than B6.*Rag* in response to PMA stimulation, over all PMA concentrations tested, as shown by plotting MFI ratios between pairs of matched cells (NOD:B6), over multiple experiments, which are almost uniformly below an expected equal value of 1 (red line) (Fig. 4B). These results are in agreement with a recent report showing differences in Erk1/2 phosphorylation between B6 and NOD TCR-transgenic T cells (38).

PMA with calcium ionophore can mimic  $\beta$ -selection-promoting signals *in vitro* to drive *Rag*-deficient thymocytes to generate DP cells in long-term cultures. Therefore, NOD.*Rag* and B6.*Rag* thymocytes were also tested for developmental responses to these signals in co-culture with OP9 stromal cells expressing Notch ligand, delta-like 1 (OP9-DL1). OP9-DL1 coculture supports T-cell development through the  $\beta$ -selection checkpoint in thymic progenitors from WT B6 and NOD mice (17, 39), but not from *Rag*-deficient mice unless an artificial TCR signal is delivered. Cells were treated with graded doses of PMA and calcium ionophore and analyzed after 7 days. NOD.*Rag* cells (dashed lines) were more sensitive to PMA than B6.*Rag* cells (solid lines) in this assay, as seen by the shift in dose response curves for both total cell numbers (Fig. 4C, left), and the percentage of cells that had progressed through the  $\beta$ -selection checkpoint, as measured by %CD4<sup>+</sup> cells (including DP) (Fig. 4C, right). These results demonstrate consistent differences between the two strains in the responses of their thymocytes to a specific stimulus.

### Breakthrough NOD.*Rag* DN cells upregulate Phase I stem/progenitor cell genes

NOD.*Rag* DP breakthrough cells only partially resemble post- $\beta$ -selection cells. They express surface CD4, CD8, and CD2 and initiate *TCR- $\alpha$*  and *Bcl- $X_L$*  transcription like normal or induced DP cells, while also retaining characteristics of pre- $\beta$ -selection DN cells such as surface CD25, KIT, and IL7 $\alpha$ , and *Spib* and *Hes1* transcription (18). To determine whether these were isolated cases or indicators of more general program derangement, the transcriptome data were analyzed for the earliest breakthrough-specific gene changes by comparing NOD.*Rag* cells (N7) with pre-breakthrough NOD.*Rag* (N4) and normal B6.*Rag* (B4, B7) cells. Sorted N7 cells excluded breakthrough cells expressing surface CD4 but included their precursors (Fig. 3A).

Very few genes exhibited a >3-fold difference between N7 and N4, and almost all were higher in N7 (Fig. 5A). This finding is consistent with the emergence of a subpopulation of abnormal cells, expressing novel genes; any down-regulated genes are likely to be masked by the dominant presence of conventional DN3 cells. Fifty-two breakthrough-specific differentially expressed genes were identified, including only those genes that distinguished N7 from all other samples, N4, B4 and B7, by at least 3-fold (Fig. 5B, Supplemental Table 2). To determine if the emergent population more closely resembles pre- or post- $\beta$ -selected cells, we used our RNA-Seq data for WT B6 DN1 to DP cells (24) as a reference for the

developmentally regulated expression of the N7-specific genes, and their patterns of expression are shown in a heatmap in Fig. 5C. Strikingly, although the breakthrough phenotype was originally identified by expression of post- $\beta$ -selection markers CD2, CD4 and CD8, most of the genes preferentially expressed by N7 are normal Phase I genes. Of the differentially expressed genes, 71% exhibited peak expression in Phase I (48% in DN1/ETP and 23% in DN2a), 21% in Phase II (15% in DN2b and 6% in DN3), and only 8% in DP. RNA-Seq tracks for some key Phase I genes, *Lmo2*, *Hhex*, *Lyl1*, *Kit*, and *Mef2c* (9), as well as a post- $\beta$ -selection gene, *Cd2*, are shown in Figure 5D. An additional Phase I gene, *Dlk1* (Delta-like-1; Pref1), was expressed only in N7, as a previously unreported truncated splice form (arrow). *Dlk1* supports proliferation in hematopoietic progenitors, maintains cells in an undifferentiated state, and is overexpressed in some leukemias (40).

The prominent expression of Phase I stem/progenitor genes suggests that the N7 emergent population either failed to silence or reestablished the early T-cell program, which normally occurs at the DN2 commitment checkpoint. At the same time, the emerging cells also appear to bypass the  $\beta$ -selection checkpoint, as shown by expression of *Cd2* and *Cd5*, which are normally turned on in DN4 ([www.immgen.org](http://www.immgen.org)). While the assayed cells were selected to be negative for surface CD4, *Cd4* gene expression was detectably upregulated (1.9X higher in N7 than in N4 cells). These results show that the emerging NOD.*Rag* breakthrough cells have greatly disordered developmental programming and fail to passage or arrest properly at the early T cell checkpoints (Fig. 1E).

### **NOD.*Rag* thymocyte numbers expand with age and *in vitro***

These features of the transcriptome analysis suggest a possible linkage between T cell breakthrough and the high incidence of thymic lymphomas that appear in NOD.*Rag* mice (18, 20, 21). NOD.*Rag* thymic cell numbers increase dramatically with age while B6.*Rag* thymocyte cell numbers change very little or decline over time (Fig. 6A). Beginning at 20 wks, NOD.*Rag* mice exhibiting thymic lymphomas appeared, as previously reported (18, 21). In addition, when DN thymocytes from young NOD.*Rag* and B6.*Rag* mice were cultured with Notch ligands and IL-7, both expanded but only NOD.*Rag* DN cells violated the  $\beta$ -selection checkpoint and upregulated CD4 (Fig. 6B). Furthermore, NOD.*Rag* cells isolated at 9–12 wks of age, after breakthrough but before lymphoma formation, expanded continuously in cultures in the presence of Notch ligand and IL7. While sorted NOD.*Rag* DP cells from 3 individuals expanded 6.7-fold  $\pm$  3.2 (SEM) over 5 days, and only generated more DP cells, sorted CD25<sup>+</sup> DN cells from the same mice expanded 3.1  $\pm$  1.0 and generated both DP cells and a small population of DN cells (Fig. 6C), suggesting that cells within the DN population may be a source of illegitimate differentiation. These thymocytes failed to survive when cultured in the absence of Notch ligands or IL-7 (data not shown), indicating that they are committed to the T lineage.

### **Transcriptional regulatory relationship between NOD.*Rag* breakthrough thymocytes and thymic lymphoma cells**

If gene expression abnormalities in NOD.*Rag* breakthrough cells favored their differentiation into lymphoma cells, then at least some of their aberrant features should be preserved in the lymphomas that develop a few months after the breakthrough. To test this prediction, we used real time QPCR on populations of NOD.*Rag* cells to track changes in gene expression related to age and lymphoma status. We tested the persistence of representative genes from various developmental stages, as well as a few genes that showed differential expression between NOD.*Rag* and B6.*Rag* DN cells (Fig. 6D–G), using RNA from the following sorted thymocyte populations: CD25<sup>+</sup>DN cells from 4–10 wk old B6.*Rag* (arrested at the  $\beta$ -selection checkpoint, brown); CD25<sup>+</sup>DN cells from 4–6 wk old NOD.*Rag* mice (arrested at the  $\beta$ -selection checkpoint, before overt breakthrough, dark

green); DN, CD4<sup>+</sup>, and DP cells from post-breakthrough 10–16 wk old NOD.*Rag* mice (light green); unsorted thymic lymphoma cells from >4 mo old NOD.*Rag* mice (pink); and ETP, DN2a, DN2b, DN3a, and DP cells from WT NOD thymuses for comparison (blue).

NOD.*Rag* thymocytes obtained from all ages, including CD4<sup>+</sup> and DP breakthrough cells and thymic lymphomas, expressed critical T-cell genes, including *Bcl11b*, *Tcf7*, *CD3e*, and *Zap70* (Fig. 6D). *Bcl11b* expression in all NOD.*Rag* samples is particularly significant, as it is critical for Phase I gene repression and T-cell differentiation past the DN2a stage (6, 7). *Bcl11b* was reported to be a haploinsufficient tumor suppressor in human T-ALL (41, 42), and expression of *Bcl11b*, as well as *CD3e*, was indeed lower by 2–3x in breakthrough and lymphoma cells in comparison with B6.*Rag* and NOD WT DN3a/DP controls. A key TCR signaling gene, *Zap70*, was consistently expressed at higher levels in NOD.*Rag* cells than in B6.*Rag* cells, and *Ii7ra*, a critical early T-cell cytokine receptor, was sustained at high levels in almost all populations. Continuing expression of *Ii7ra* in all breakthrough cell stages and in thymic lymphoma cells supports the requirement for IL-7 for cell proliferation in early T cells, and this pathway has been reported to be critical in human early T-ALL (43).

Notch signaling is required for early but not later stages of T-cell development (44) and constitutive Notch activation is one of the most potent and common oncogenic factors in T-ALL (45). Expression of *Notch1*, and target genes *Dtx1*, *Hes1*, *Hey1*, and *Ptcr*, was sustained in all stages, including in lymphoma cells (Fig. 6E), implying sustained Notch signaling.

Among Phase I genes (Fig. 6F), *Kit* expression was sustained at high levels in NOD.*Rag* cells at all ages, and in lymphoma cells. *Lmo2* was upregulated >100-fold between 4 and 6 wks in NOD.*Rag* cells, declined in breakthrough CD4<sup>+</sup> and DP cells, but was still expressed in all lymphoma cells. *Hhex* was also elevated in most NOD.*Rag* cells including all lymphomas. *Lyl1* and *Sfp1* levels varied between different NOD.*Rag* cell populations, but, like *Lmo2* and *Hhex*, both genes were still expressed in all thymic lymphoma cells at levels higher than WT NOD DP cells. These Phase I genes are known proto-oncogenes and are likely to be major contributors to the thymocyte expansion and lymphomas found in NOD.*Rag* mice.

The lymphomas did not passively maintain gene expression patterns from DN and breakthrough cells, as shown in Fig. 6G for representative differentially expressed genes, each showing distinctive expression patterns. Of particular interest, the NOD-specific high-level expression of *Epha2* (in the chr4 QTL peak) began before breakthrough and was maintained in all stages including all lymphomas. QPCR analysis also showed *Epha2* to be differentially expressed between WT NOD and B6 cells, at all DN stages, and in DP and  $\gamma\delta$ T cells (Supplemental Fig. 1A). Higher NOD.*Rag* cell expression of *Milt4* (Afadin, in chr17 QTL peak) was also confirmed, but expression levels declined with age. Expression of the tumor suppressor kinase, *Dapk1* (chr13 QTL peak) was lower in NOD.*Rag* cells, although some expression persisted in the lymphomas. WT NOD cells also showed a lower expression of *Dapk1* in all stages of development in comparison with B6 cells, especially in the DN2a and DP stages where it is undetectable (Supplemental Fig. 1B). Expression of *Vegfa* and *Tnfrsf9* declined with age, however *Vegfa* was consistently expressed in the lymphomas while *Tnfrsf9* was variable. Overall, the lymphoma cells maintained many of the unique gene expression patterns found in NOD.*Rag* N7 cells, especially expression of Phase I proto-oncogenes, suggesting that origins of the transformed cells may lie in the early breakthrough cells which fail to shut off the genes that sustain the pleuripotency and proliferation of progenitor T cells (Fig. 1E).

## Discussion

Thymic lymphomas generated in mice with an NOD genetic background preserve many gene expression features of the early breakthrough cells, which themselves emerge as the result of an earlier dysfunction. From the transcriptome analysis, the dominant feature of breakthrough cells emerging spontaneously in the NOD.*Rag* thymus is the collapse of two regulatory boundaries: one that normally distinguishes pre-commitment ETP and DN2a stage cells from committed DN3 cells, the other operating at the  $\beta$ -selection checkpoint that requires a TCR signal for DN3 to DP progression (Fig. 1E). The breakthrough cells and the later-arising lymphoma cells share a highly abnormal co-expression of legacy progenitor cell genes, including proto-oncogenes, normally restricted to uncommitted ETP and DN2a cells, along with T cell genes characteristic of later stage cells, before and after TCR-dependent  $\beta$ -selection. Unlike normal *Rag*-deficient cells, breakthrough thymocytes undergo long-term expansion and aberrant differentiation *in vitro*, and thymic cell numbers increase with age *in vivo*, even before thymic lymphomas appear. Thus, the breakthrough cells may include a population of pre-leukemic cells, which can provide cellular targets of oncogenic transformation.

One striking feature of the T cell commitment checkpoint in the DN2 stage is the brief period of overlap in Phase I and T cell identity genes during the transition from dominant expression to silencing of legacy stem cell genes after the onset of expression of T cell specific genes (5, 9). The most notable feature of our current transcriptome and QPCR data is the apparent failure to terminate the Phase I ETP-DN2 progenitor/stem cell program in early breakthrough cells despite upregulation of T cell specific genes including *Bcl11b*, which is involved in their silencing (6), making the breakthrough cells appear to be retained in a DN2-like state. Expression of genes such as *Cd2*, *Cd5*, and *Cd4* indicate a bypass of the later  $\beta$ -selection checkpoint as well, possibly due to an aberrant pre-TCR-like signal. It is also possible that failure of the first checkpoint might prevent the normal establishment of the second checkpoint, making the DN2 stage the critical point of breakthrough vulnerability. Aberrant signal transduction at one or both checkpoints as a promoter of breakthrough is suggested by the transcriptome data showing a marked overrepresentation of signaling genes, especially actin-binding proteins and tyrosine kinases, among differentially expressed genes between NOD.*Rag* and B6.*Rag* thymocytes. Actin/cytoskeletal and signal transduction pathways are interrelated in TCR signaling (36), and alterations in expression of key genes may trigger inappropriate signaling at the checkpoints. Evidence that activated NOD CD4<sup>+</sup> cells exhibit a higher overall level of phospho-tyrosine than B6 cells (46) supports a potential role for at least some of the 11 differentially overexpressed tyrosine kinases in NOD.*Rag* cells and we show *in vitro* evidence for subtle but consistent strain-specific differences in responses to PMA between cells from the two strains. Furthermore, consistent differences in Erk1/2 phosphorylation have also been reported for immature and mature NOD and B6 TCR-transgenic T cells (38).

The NOD.*Rag* mixed gene expression signature of early breakthrough T cells and thymic lymphoma cells that emerge at high frequency after 4–5 mo, is shared with certain human acute T cell lymphocytic leukemias (T-ALL), especially of the early *LYL1-LMO2* T-ALL subgroup (47), as well as the recently described ETP-ALL, whose gene signatures share features with normal ETP and myeloid cells (48). ETP-ALL also shares similarities with both T cell and myeloid stem cells, and mutations in cytokine (*Flt3* and *IL7R*) and Ras signaling pathways are prominent (43). In common with ETP-ALL, NOD.*Rag* breakthrough cells express genes indicative of their early T cell programming, e.g., *Bcl11b*, *Ptcra* and *Ii7r*, in addition to stem/progenitor genes such as *Lmo2*, *Hhex*, *Lyl1*, *Kit*, *Sfpi1*, which also likely contribute to the myeloid potential of WT NOD and B6 ETP/DN2a cells (5). The prominent overexpression of *Lmo2* in the NOD.*Rag* breakthrough cells is noteworthy as DN cells

overexpressing transgenic *Lmo2* develop into self-renewing thymocytes, with striking gene expression similarities to NOD.*Rag* breakthrough cells, followed by T-ALL eight months later (49). The NOD.*Rag* breakthrough and thymic lymphoma development may be only partially attributable to the enhanced expression of *Lmo2*, as thymic tumors appear much more rapidly in NOD.*Rag* mice, beginning two months after breakthrough cells first appear, possibly due to the elevated expression of tyrosine kinases, which are commonly oncogenic (37).

Genes within the identified QTL regions control propensity for breakthrough in the NOD genetic background and they mapped to known autoimmune diabetes susceptibility regions. The strong chr4 QTL peak in the *Idd9/11* region (14) clearly includes at least one recessive NOD allele needed to promote efficient breakthrough, as shown by the reduced breakthrough in congenic NOD.B10<sup>*Idd9*</sup>*Rag* mice. Possible breakthrough-enhancing QTL regions were also detected in the *Idd16/23/1/24* region on chr17, *Idd14* on chr13, and *Idd4* on chr11. Although all of the QTL regions remain too large to determine if any of the same genes control both autoimmunity and the early T-cell breakthrough, the close mapping of the two different T-cell related phenotypes is highly suggestive and requires further analysis to pinpoint source genes. The B10 *Idd9* congenic region, which protects against the breakthrough trait, is reported to protect NOD mice against Type 1 diabetes after lymphocytic infiltration in the pancreas, suggesting control over the balance between pathogenic and protective T-cell responses (14). If the differentially expressed genes found within the QTL regions are also used in establishing self-tolerance, then the autoimmunity promoting T cell defects may be traceable to the earliest common T cell progenitors, even before TCR rearrangement occurs.

To date, animal models of T-ALL have typically depended on artificial gene expression alterations such as ablation of tumor suppressor genes or transgenic overexpression of oncogenes (50). The results reported here suggest that the NOD.*Rag* mouse may be a useful unmanipulated model for studying oncogenic progression from a pre-leukemic state, which arises spontaneously and predictably from defects in early T cell checkpoint control. Moreover, our genetic analysis suggests a potential linkage to autoimmune diabetes in WT NOD mice, possibly related to an intrinsic defect in thymic  $\alpha\beta$ T-cell development, which also arises from ETP-DN2 stages (17). Human epidemiological studies have linked T-ALL and autoimmune Type 1 diabetes, both of which are predominantly pediatric diseases (51, 52). Furthermore, genes and molecular pathways involved in lymphocyte immune checkpoints and tumor suppression overlap, and the consequences of loss of checkpoint control, which sets thresholds for appropriate proliferation and survival, can be key in both autoimmunity and in lymphoid cancers (53). Thus, our results could extend alterations in T cell selection and functions back to the earliest stages of their development and raise the possibility that genes involved in an early commitment checkpoint breakthrough may contribute to defects in lymphocyte homeostasis and both T cell leukemia and autoimmune disease.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Brian Williams, Justine Chia, Avni Gandhi, Sagar Damle for technical and bioinformatics assistance; Donna Walls, Weidong Zhang (Jackson Labs) for genotyping and preliminary statistical analysis; J.C. Zuñiga-Pflücker (Univ. of Toronto) for providing OP9-DL1 and OP9-DL4 cells; L.S. Wicker (Cambridge University) for NOD.B10<sup>*Idd9*</sup> congenic mice; Rochelle Diamond, Diane Perez, Pat Koen from The Caltech Flow Cytometry &

Cell Sorting Facility; Scott Washburn, Natasha Bouey for animal care; Ali Mortazavi and members of the Rothenberg lab, for helpful suggestions.

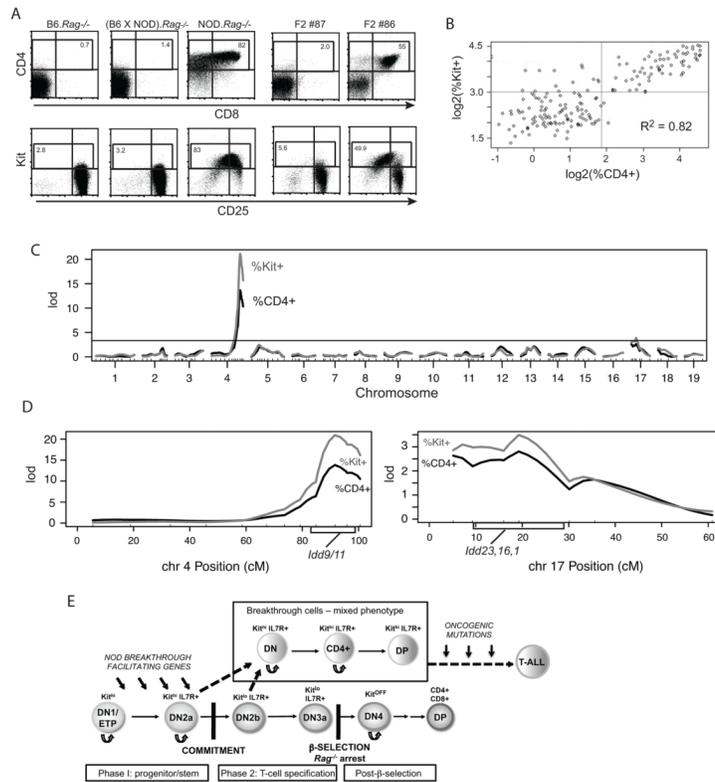
## References

1. Petrie HT, Zuniga-Pflucker JC. Zoned out: functional mapping of stromal signaling microenvironments in the thymus. *Annu Rev Immunol.* 2007; 25:649–679. [PubMed: 17291187]
2. Love PE, Bhandoola A. Signal integration and crosstalk during thymocyte migration and emigration. *Nat Rev Immunol.* 2011; 11:469–477. [PubMed: 21701522]
3. Godfrey DI, Kennedy J, Suda T, Zlotnik A. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8-triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *J Immunol.* 1993; 150:4244–4252. [PubMed: 8387091]
4. Porritt HE, Rumpf LL, Tabrizifard S, Schmitt TM, Zuniga-Pflucker JC, Petrie HT. Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity.* 2004; 20:735–745. [PubMed: 15189738]
5. Yui MA, Feng N, Rothenberg EV. Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol.* 2010; 185:284–293. [PubMed: 20543111]
6. Li L, Leid M, Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science.* 2010; 329:89–93. [PubMed: 20595614]
7. Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, Shibano-Satoh A, Kominami R, Katsura Y, Kawamoto H. An essential developmental checkpoint for production of the T cell lineage. *Science.* 2010; 329:93–96. [PubMed: 20595615]
8. Rothenberg EV, Moore JE, Yui MA. Launching the T-cell-lineage developmental programme. *Nat Rev Immunol.* 2008; 8:9–21. [PubMed: 18097446]
9. Rothenberg EV, Zhang J, Li L. Multilayered specification of the T-cell lineage fate. *Immunol Rev.* 2010; 238:150–168. [PubMed: 20969591]
10. Hoffman ES, Passoni L, Crompton T, Leu TM, Schatz DG, Koff A, Owen MJ, Hayday AC. Productive T-cell receptor beta-chain gene rearrangement: coincident regulation of cell cycle and clonality during development in vivo. *Genes Dev.* 1996; 10:948–962. [PubMed: 8608942]
11. Borowski C, Martin C, Gounari F, Haughn L, Aifantis I, Grassi F, von Boehmer H. On the brink of becoming a T cell. *Curr Opin Immunol.* 2002; 14:200–206. [PubMed: 11869893]
12. Anderson MS, Bluestone JA. The NOD mouse: a model of immune dysregulation. *Annu Rev Immunol.* 2005; 23:447–485. [PubMed: 15771578]
13. Delovitch TL, Singh B. The nonobese diabetic mouse as a model of autoimmune diabetes: immune dysregulation gets the NOD. *Immunity.* 1997; 7:727–738. [PubMed: 9430219]
14. Lyons PA, Hancock WW, Denny P, Lord CJ, Hill NJ, Armitage N, Siegmund T, Todd JA, Phillips MS, Hess JF, Chen SL, Fischer PA, Peterson LB, Wicker LS. The NOD Idd9 genetic interval influences the pathogenicity of insulinitis and contains molecular variants of Cd30, Tnfr2, and Cd137. *Immunity.* 2000; 13:107–115. [PubMed: 10933399]
15. Serreze DV, Leiter EH. Genes and cellular requirements for autoimmune diabetes susceptibility in nonobese diabetic mice. *Curr Dir Autoimmun.* 2001; 4:31–67. [PubMed: 11569409]
16. Todd JA, Wicker LS. Genetic protection from the inflammatory disease type 1 diabetes in humans and animal models. *Immunity.* 2001; 15:387–395. [PubMed: 11567629]
17. Feng N, Vegh P, Rothenberg EV, Yui MA. Lineage divergence at the first TCR dependent checkpoint: preferential  $\gamma\delta$  and impaired  $\alpha\beta$  T cell development in non-obese diabetic (NOD) mice. *J Immunol.* 2011:186.
18. Yui MA, Rothenberg EV. Deranged early T cell development in immunodeficient strains of nonobese diabetic mice. *J Immunol.* 2004; 173:5381–5391. [PubMed: 15494484]
19. Prochazka M, Gaskins HR, Shultz LD, Leiter EH. The nonobese diabetic scid mouse: model for spontaneous thymomagenesis associated with immunodeficiency. *Proc Natl Acad Sci U S A.* 1992; 89:3290–3294. [PubMed: 1373493]
20. Shultz LD, Lang PA, Christianson SW, Gott B, Lyons B, Umeda S, Leiter E, Hesselton R, Wagar EJ, Leif JH, Kollet O, Lapidot T, Greiner DL. NOD/LtSz-Rag1null mice: an immunodeficient and

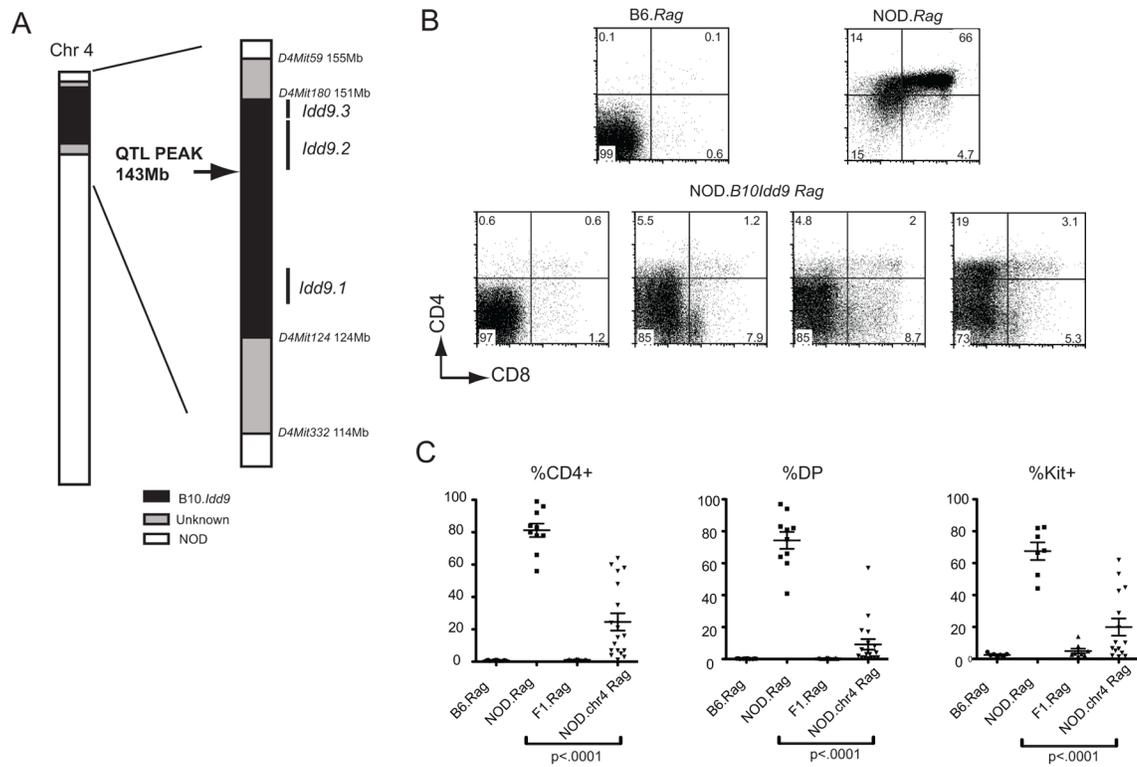
radioresistant model for engraftment of human hematolymphoid cells, HIV infection, and adoptive transfer of NOD mouse diabetogenic T cells. *J Immunol.* 2000; 164:2496–2507. [PubMed: 10679087]

21. Chiu PP, Ivakine E, Mortin-Toth S, Danska JS. Susceptibility to lymphoid neoplasia in immunodeficient strains of nonobese diabetic mice. *Cancer Res.* 2002; 62:5828–5834. [PubMed: 12384545]
22. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 2003; 19:889–890. [PubMed: 12724300]
23. Kruglyak L, Lander ES. A nonparametric approach for mapping quantitative trait loci. *Genetics.* 1995; 139:1421–1428. [PubMed: 7768449]
24. Zhang J, Mortazavi A, Williams B, Wold B, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell.* 2012; 149:467–482. [PubMed: 22500808]
25. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30:207–210. [PubMed: 11752295]
26. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628. [PubMed: 18516045]
27. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods.* 2009; 6:S22–32. [PubMed: 19844228]
28. David-Fung ES, Butler R, Buzi G, Yui MA, Diamond RA, Anderson MK, Rowen L, Rothenberg EV. Transcription factor expression dynamics of early T-lymphocyte specification and commitment. *Dev Biol.* 2009; 325:444–467. [PubMed: 19013443]
29. Ghosh S, Palmer SM, Rodrigues NR, Cordell HJ, Hearne CM, Cornall RJ, Prins JB, McShane P, Lathrop GM, Peterson LB, et al. Polygenic control of autoimmune diabetes in nonobese diabetic mice. *Nat Genet.* 1993; 4:404–409. [PubMed: 8401590]
30. Brodnicki TC, McClive P, Couper S, Morahan G. Localization of Idd11 using NOD congenic mouse strains: elimination of Slc9a1 as a candidate gene. *Immunogenetics.* 2000; 51:37–41. [PubMed: 10663560]
31. Inoue K, Ikegami H, Fujisawa T, Noso S, Nojima K, Babaya N, Itoi-Babaya M, Makimo S, Ogiwara T. Allelic variation in class I K gene as candidate for a second component of MHC-linked susceptibility to type 1 diabetes in non-obese diabetic mice. *Diabetologia.* 2004; 47:739–747. [PubMed: 15298352]
32. Boulard O, Damotte D, Deruytter N, Fluteau G, Carnaud C, Garchon HJ. An interval tightly linked to but distinct from the H2 complex controls both overt diabetes (Idd16) and chronic experimental autoimmune thyroiditis (Ceat1) in nonobese diabetic mice. *Diabetes.* 2002; 51:2141–2147. [PubMed: 12086944]
33. Hiromine Y, Fujisawa T, Noso S, Babaya N, Kawabata Y, Ikegami H. Congenic mapping of the MHC-linked susceptibility to type 1 diabetes in the NOD mouse: at least two genes contribute to the Idd16 effect. *Ann N Y Acad Sci.* 2008; 1150:90–92.
34. McAleer MA, Reifsnnyder P, Palmer SM, Prochazka M, Love JM, Copeman JB, Powell EE, Rodrigues NR, Prins JB, Serreze DV, et al. Crosses of NOD mice with the related NON strain. A polygenic model for IDDM. *Diabetes.* 1995; 44:1186–1195. [PubMed: 7556956]
35. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57. [PubMed: 19131956]
36. Burkhardt JK, Carrizosa E, Shaffer MH. The actin cytoskeleton in T cell activation. *Annu Rev Immunol.* 2008; 26:233–259. [PubMed: 18304005]
37. Wadleigh M, DeAngelo DJ, Griffin JD, Stone RM. After chronic myelogenous leukemia: tyrosine kinase inhibitors in other hematologic malignancies. *Blood.* 2005; 105:22–30. [PubMed: 15358622]
38. Mingueneau M, Jiang W, Feuerer M, Mathis D, Benoist C. Thymic negative selection is functional in NOD mice. *J Exp Med.* 2012; 209:623–637. [PubMed: 22329992]
39. Schmitt TM, Zuniga-Pflucker JC. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity.* 2002; 17:749–756. [PubMed: 12479821]

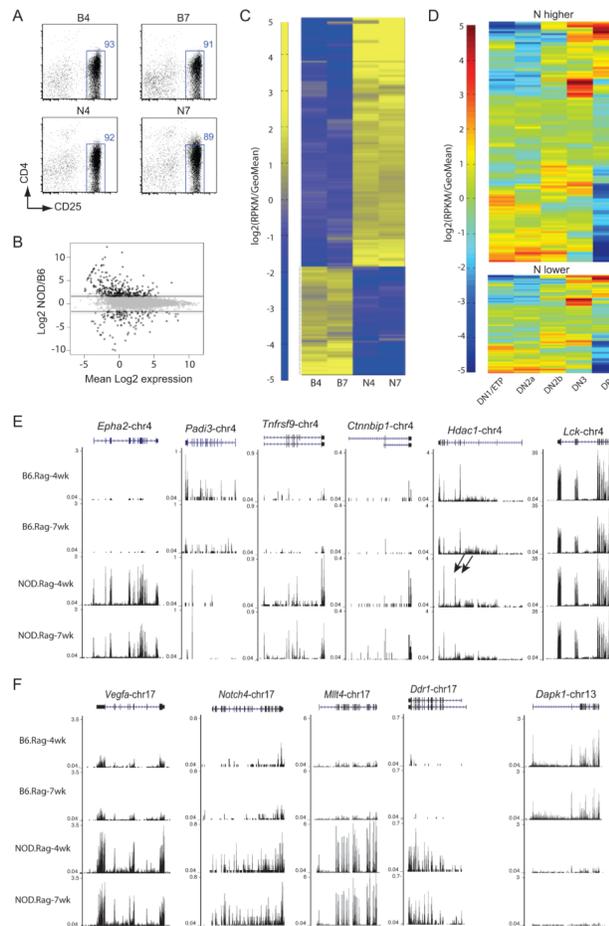
40. Moore KA, Pytowski B, Witte L, Hicklin D, Lemischka IR. Hematopoietic activity of a stromal cell transmembrane protein containing epidermal growth factor-like repeat motifs. *Proc Natl Acad Sci U S A*. 1997; 94:4011–4016. [PubMed: 9108096]
41. De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, Van Vlierberghe P, Barnes K, Castillo M, Sole X, Hadler M, Lenz J, Aplan PD, Kelliher M, Kee BL, Pandolfi PP, Kappes D, Gounari F, Petrie H, Van der Meulen J, Speleman F, Paietta E, Racevskis J, Wiernik PH, Rowe JM, Soulier J, Avran D, Cave H, Dastugue N, Raimondi S, Meijerink JP, Cordon-Cardo C, Califano A, Ferrando AA. The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat Med*. 2010; 16:1321–1327. [PubMed: 20972433]
42. Gutierrez A, Kentsis A, Sanda T, Holmfeldt L, Chen SC, Zhang J, Protopopov A, Chin L, Dahlberg SE, Neuberger DS, Silverman LB, Winter SS, Hunger SP, Sallan SE, Zha S, Alt FW, Downing JR, Mullighan CG, Look AT. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. *Blood*. 2011; 118:4169–4173. [PubMed: 21878675]
43. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfors J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK, Downing JR, Mullighan CG. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 2012; 481:157–163. [PubMed: 22237106]
44. Wolfer A, Bakker T, Wilson A, Nicolas M, Ioannidis V, Littman DR, Lee PP, Wilson CB, Held W, MacDonald HR, Radtke F. Inactivation of Notch 1 in immature thymocytes does not perturb CD4 or CD8T cell development. *Nat Immunol*. 2001; 2:235–241. [PubMed: 11224523]
45. Weng AP, Ferrando AA, Lee W, Morris JPT, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, Aster JC. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. 2004; 306:269–271. [PubMed: 15472075]
46. Iwai LK, Benoist C, Mathis D, White FM. Quantitative phosphoproteomic analysis of T cell receptor signaling in diabetes prone and resistant mice. *J Proteome Res*. 2010; 9:3135–3145. [PubMed: 20438120]
47. Ferrando AA, Neuberger DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, Lander ES, Golub TR, Look AT. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*. 2002; 1:75–87. [PubMed: 12086890]
48. Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D, Cheng C, Su X, Rubnitz JE, Basso G, Biondi A, Pui CH, Downing JR, Campana D. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol*. 2009; 10:147–156. [PubMed: 19147408]
49. McCormack MP, Young LF, Vasudevan S, de Graaf CA, Codrington R, Rabbitts TH, Jane SM, Curtis DJ. The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science*. 2010; 327:879–883. [PubMed: 20093438]
50. Aifantis I, Raetz E, Buonamici S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat Rev Immunol*. 2008; 8:380–390. [PubMed: 18421304]
51. Shu X, Ji J, Li X, Sundquist J, Sundquist K, Hemminki K. Cancer risk among patients hospitalized for Type 1 diabetes mellitus: a population-based cohort study in Sweden. *Diabet Med*. 2010; 27:791–797. [PubMed: 20636960]
52. Manda SO, Feltbower RG, Gilthorpe MS. Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes. *Eur J Epidemiol*. 2009; 24:743–752. [PubMed: 19784553]
53. Goodnow CC. Multistep pathogenesis of autoimmune disease. *Cell*. 2007; 130:25–35. [PubMed: 17632054]



**Figure 1.** Genome-wide QTL analysis of the *NOD.Rag* early T-cell breakthrough trait shows mapping to known diabetes susceptibility regions. (A) Flow cytometric data showing representative %CD4<sup>+</sup> vs. %CD8<sup>+</sup> and %Kit<sup>+</sup> vs. %CD25<sup>+</sup> phenotyping plots for thymocytes from control parental B6.*Rag*<sup>-/-</sup>, NOD.*Rag*<sup>-/-</sup>, and F1.*Rag*, as well as 2-(NOD X B6)F2.*Rag* cross mice at 14 wks of age. (B) Scatterplot showing the correlation between the two measured thymocyte phenotypic traits: %CD4<sup>+</sup> and %Kit<sup>+</sup> cells. (C) Genome-wide QTL scan for %CD4<sup>+</sup> and %Kit<sup>+</sup> thymocytes, using the combined N2 plus F2 crosses. LOD scores are plotted for %CD4<sup>+</sup> cells (black) and %Kit<sup>+</sup> cells (gray) for all 19 autosomal chromosomes. Black line indicates a LOD score of 3. (D) Expanded plots of LOD scores for the chromosomes with the highest peaks in the combined cross, chr4 and chr17, for the two phenotypes, %CD4<sup>+</sup> and %Kit<sup>+</sup> thymocytes. Approximate regions of known diabetes susceptibility *Idd* loci are also noted. (E) Diagram of key early T-cell developmental stages (based on (9)) and possible pathways (dashed and solid arrows) leading to pro-T-cell checkpoint breakthrough and T-cell lymphocytic leukemia (T-ALL) in NOD.*Rag* mice.

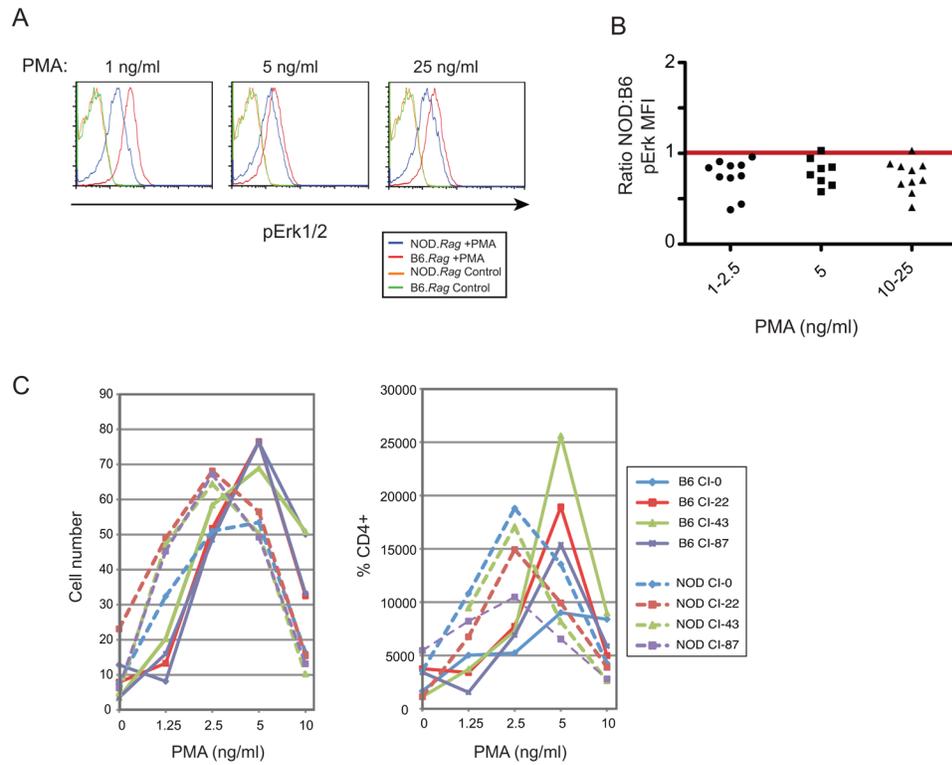


**Figure 2.** NOD.B10<sup>Idd9</sup>Rag congenic mouse strain analysis of the chr4 *Idd9* QTL peak region. (A) Diagram of the chr4 congenic region from NOD.B10<sup>Idd9</sup> (L. Wicker, Cambridge, UK), which was crossed onto NOD.Rag congenic mice. Indicated are the genomic regions derived from B10<sup>Idd9</sup> (black) and NOD (white), and the recombination region (gray), as well as the Mappair markers used in the congenic cross and the QTL peak (arrow). Approximate locations for *Idd9.1*, *9.2* and *9.3* are shown (from T1Dbase.org). (B) Flow cytometric analysis of B6.Rag, NOD.Rag and four NOD.B10<sup>Idd9</sup>Rag congenic mice at 12 weeks of age. (C) Summary of %CD4+, %DP, and %Kit<sup>+</sup> cells, for individual NOD.B10<sup>Idd9</sup>Rag (NOD.chr4Rag) congenic mice in comparison with parental B6.Rag, NOD.Rag, F1.Rag mice, all at 12–16 wks old. p-values are from t-tests comparing data from NOD.Rag and NOD.B10<sup>Idd9</sup>Rag mice.

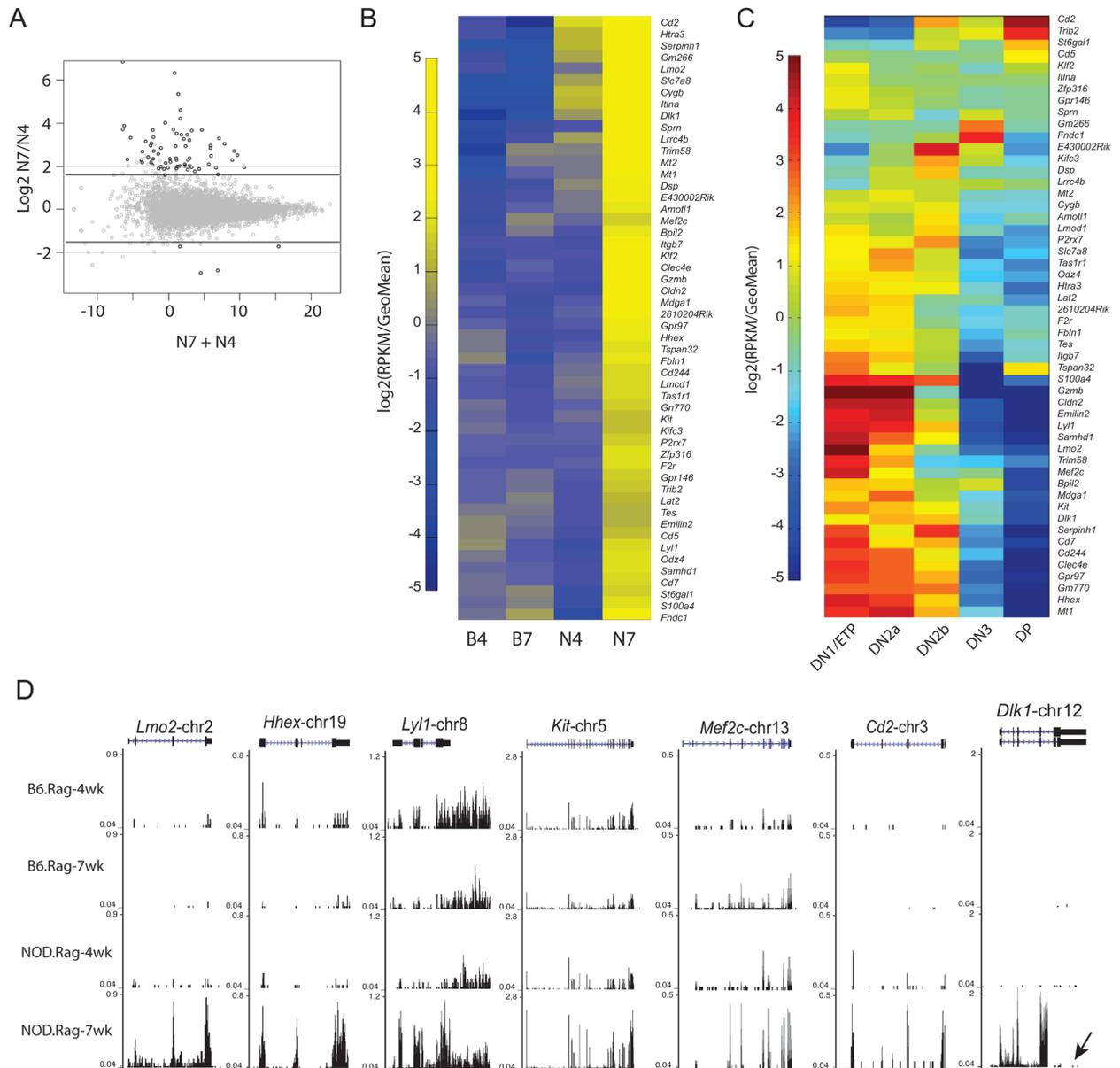


**Figure 3.**

RNA-Seq transcriptome analysis comparing CD25<sup>+</sup>DN thymocytes from B6.*Rag* and NOD.*Rag* mice. (A) Gates used for sorting phenotypically similar CD4<sup>+</sup>CD25<sup>+</sup> cells for RNA purification: B6.*Rag* at 4 and 7 wk (B4, B7) and NOD.*Rag* at 4 and 7 wk (N4, N7). (B) Difference versus mean of log<sub>2</sub> expression of N4/N7 and B4/B7, including 11,098 genes expressed at >32 reads in at least one sample. Horizontal gray lines show 3- and 4-fold differences. Genes differing by >3-fold are shown in dark circles, <3-fold by gray circles. (C) Heat map for hierarchical clustering of 412 genes differing by >3X between N(4 and 7) vs. B(4 and 7). Data are shown as log<sub>2</sub>(RPKM/GeoMean). (D) Hierarchical cluster heat maps showing the normal (WT B6) developmentally regulated expression (log<sub>2</sub>(RPKM/GeoMean)) of differentially expressed genes that are higher (top) and lower (bottom) in N(4 and 7) compared with B(4 and 7) using our previously published data (24). (E–F) RNA-Seq tracks showing sequence read histograms for expressed genes of interest in QTL regions from B6.*Rag* and NOD.*Rag* CD25<sup>+</sup> DN cells at 4 and 7 wks of age. Gene tracks are shown using the UCSC browser and the data are mapped onto the mouse genome build NCBI37/mm9. (E) Genes located in the chr4 major QTL region, including differentially expressed genes, *Epha2*, *Padi3*, *Tnfrsf9* (4–1BB), *Cttnbip1* (*Icat*), a differentially spliced gene, *Hdac1* (differentially spliced exons indicated with arrows), and *Lck*, a non-differentially expressed gene for reference. (F) Differentially expressed genes in the chr17 QTL region, including *Vegfa*, *Notch4*, *Mllt4* (*Afadin*), and *Ddr1*, and in the chr13 QTL region, *Dapk1*.

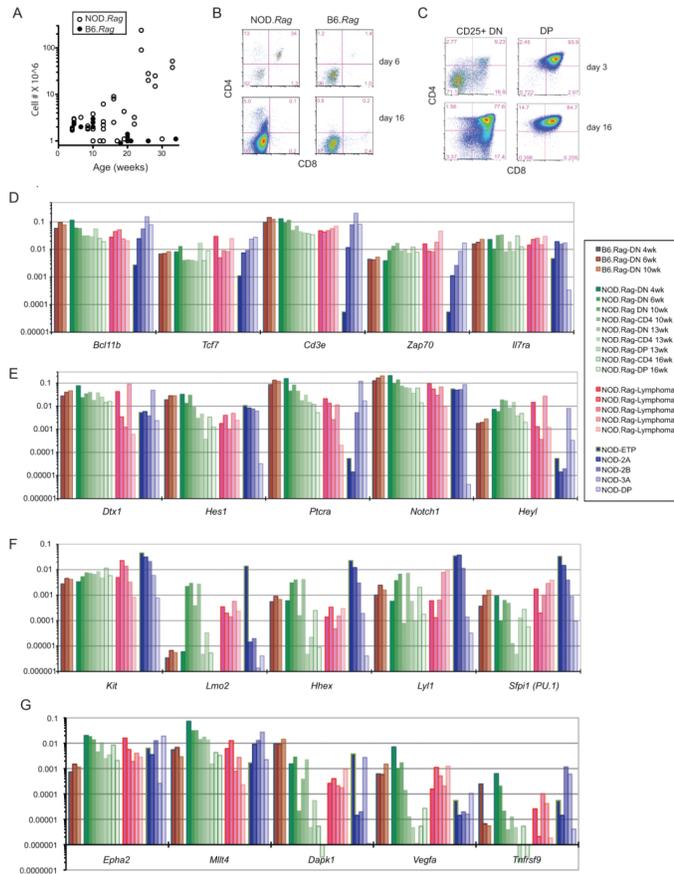
**Figure 4.**

NOD.*Rag* and B6.*Rag* thymocytes exhibit differential responses to PMA stimulation. (A) Representative flow cytometric histogram showing intracellular Erk1/2 phosphorylation (pErk1/2) in freshly isolated thymocytes (>90% DN3 cells) from pre-breakthrough NOD.*Rag* and B6.*Rag* mice treated for 2 min. with PMA. (B) Summary plot of the NOD:B6 ratios of pErk1/2 (using geometric mean fluorescence intensity (MFI) from intracellular staining) for thymocytes from age- and sex-matched NOD.*Rag* and B6.*Rag* mice treated with 1 to 25 ng/ml PMA. The MFI was consistently lower for NOD.*Rag* cells than for B6.*Rag* cells as shown by ratios lower than 1 (equal expression, indicated with a red line). Data included are combined from 5 independent experiments. (C) Responses of thymocytes from 6 wk old NOD.*Rag* (dashed lines) and B6.*Rag* (solid lines), placed in co-culture with OP9-DL1 cells and stimulated with graded doses of PMA, and calcium ionophore A23187 (CI: doses of 0, 22, 43, and 87 nM as indicated), and assayed at day 7 for cell number (left graph) and %CD4<sup>+</sup> cells (right graph). Representative of 3 independent experiments.

**Figure 5.**

RNA-Seq transcriptome analysis of NOD.*Rag* breakthrough cells. (A) Difference versus mean of  $\log_2$  expression of NOD.*Rag* 7 wk (N7, at breakthrough) vs. 4 wk (N4, pre-breakthrough). Horizontal gray lines show 3- and 4-fold differences. Genes differing by >3-fold are shown in dark circles, <3-fold by gray circles. (B) Heat map for hierarchical clustering of 52 genes differing by >3X between N7 vs. all other samples. Data are shown as  $\log_2(\text{reads per kilobase per million reads (RPKM)}/\text{Geometric mean (GeoMean)})$ . (C) Hierarchical cluster heat map showing the WT B6 expression profile ( $\log_2(\text{RPKM}/\text{GeoMean})$ ) of 52 genes differentially expressed between N7 and other samples using data from ETP to DP developmental stages (24). (D) RNA-Seq gene tracks from B6.*Rag* and NOD.*Rag* CD25<sup>+</sup> DN cells at 4 and 7 wks of age showing a selection of genes of interest expressed >3-fold higher in N7 in comparison to N4, B4 and B7, including Phase I genes, *Lmo2*, *Hhex*, *Lyl1* *Kit*, and *Mef2c*, post- $\beta$ -selection gene, *Cd2*, and a differentially

expressed and spliced Notch-related gene, *Dlk1* (lack of expression of 3' exon(s) is indicated with an arrow).



**Figure 6.**

Changes in thymocyte cell number, *in vitro* potential, and gene expression with age and lymphoma. (A) Plot of thymocyte cell numbers with age showing the increase in cell numbers with age in NOD.*Rag* but not B6.*Rag* mice. (B) DN cells from 6 wk old pre-breakthrough NOD.*Rag* mice cultured with OP9-DL1 stroma plus IL-7 exhibit upregulation of CD4 while B6.*Rag* cells do not express CD4. (C) Sorted CD25<sup>+</sup> DN and DP cells from a 12 wk old NOD.*Rag* mouse proliferate and differentiate when cultured on OP9-DL4 stroma supplemented with 5 ng/ml IL-7 for 3 and 16 days before FACS analysis for expression of CD4 and CD8. Data are representative of 2 experiments. (D–G) Real time QPCR expression of some developmentally regulated genes in sorted B6.*Rag* CD25<sup>+</sup>DN cells, from 4, 6, and 10 wks of age (brown bars), sorted NOD.*Rag* pre-lymphoma CD25<sup>+</sup>DN, CD4<sup>+</sup> and DP cells, from 4, 6, 10, 13, and 16 wks of age (green bars), unsorted cells from 5 independent NOD.*Rag* lymphomas (pink bars), and representative WT NOD DN1/ETP, DN2a, DN2b, DN3a, and DP cells for reference (blue bars). Relative gene expression levels are shown for representatives of (D) key T-cell genes, (E) Notch and target genes, (F) Phase I (ETP/DN2a) genes, and (G) differentially expressed genes from the transcriptome analysis.

Table 1

Peak regions from a genome-wide QTL scan of N2, F2 and combined N2 plus F2 crosses for %CD4<sup>+</sup> and %Kit<sup>+</sup> thymocytes

Cross	Trait	chr <sup>1</sup>	Position (cM) <sup>2</sup>	Peak (Mb) <sup>3</sup>	LOD <sup>4</sup>	p-value <sup>5</sup>
N2	%CD4 <sup>+</sup>	4	91.6	142.8	12.49	<<0.01
		13	40.3	76.2	2.85	<0.05
		17	19.2	40.2	2.69	<0.1
	%Kit <sup>+</sup>	4	91.6	142.8	18.97	<<0.01
		13	40.3	76.2	2.59	<0.1
		17	19.2	40.2	3.2	<0.05
F2	%CD4 <sup>+</sup>	11	42.3	70.3	2.85	<0.5
		11	31	60	3.46	<0.2
		4	91.6	142.8	13.67	<<0.01
	%CD4 <sup>+</sup>	17	19.2	40.2	2.6	<0.025
		4	91.6	142.8	21.13	<<0.01
		17	19.2	40.2	3.87	<0.01
Combined N2 + F2	%CD4 <sup>+</sup>	4	91.6	142.8	21.13	<<0.01
		17	19.2	40.2	3.87	<0.01
		4	91.6	142.8	21.13	<<0.01
	%Kit <sup>+</sup>	17	19.2	40.2	3.87	<0.01
		4	91.6	142.8	21.13	<<0.01
		17	19.2	40.2	3.87	<0.01

<sup>1</sup> chromosome

<sup>2</sup> Genetic map position in centimorgans (cM), calculated from cross genotypes

<sup>3</sup> Peak locations in NCBI37/mm9 mouse assembly (Mb) obtained from peak SNP location

<sup>4</sup> log<sub>10</sub> likelihood ratio (LOD) calculated with R-ql using a simple model for the N2 and F2 crosses, and adding cross to the model for the combined N2 + F2 cross data; only LOD scores >2.5 are shown

<sup>5</sup> p-value based on genome-wide significance tests for LOD thresholds (5,000 permutations) for each cross and phenotype

Table 2

Gene ontology analysis on >3-fold differentially expressed genes between NOD.*Rag* and B6.*Rag* CD25<sup>+</sup> DN cells showing major enriched categories of genes.

GOTERM CATEGORY	GENES <sup>1</sup>	FOLD ENRICHMENT	p-value	Benjamini
MHC protein complex	<i>H2-K1</i> , <i>H2-Q2</i> , <i>H2-M6-PS</i> , <i>H2-Q10</i> , <i>C920025E04rik</i> , <i>H2-OA</i> , <i>H2-BL</i> , <i>H2-Q1</i> , <i>H2-T10</i> , <i>H2-AB1</i> , <i>H2-Q7</i> , <i>H2-T3</i> (all chr17) <sup>2</sup>	9.8	6.5E-07	1.6E-04
Actin binding	<i>Lima1</i> , <i>Ccdc88a</i> , <i>Mybpc2</i> , <i>Tln2</i> , <i>Inpp1l1</i> , <i>Myo7a</i> , <i>Evl</i> , <i>Spire2</i> , <i>Tpm3</i> , <i>Coro2b</i> , <i>Fmn2</i> , <i>Syne2</i> , <i>Mtap1a</i> , <i>Capg</i> , <i>Tmod4</i> , <i>Eps8l1</i> , <i>Mylk</i> , <i>Myh10</i> (chr11)	3.5	1.8E-05	7.9E-03
Protein tyrosine kinase activity	<i>Tyro3</i> , <i>Ddr1</i> (chr17), <i>Fgfr1</i> , <i>Ptk2</i> , <i>Ltk</i> , <i>Ptpn3</i> , <i>Ptk2b</i> , <i>Hck</i> , <i>Txk</i> , <i>Ephb4</i> , <i>Epha2</i> (chr4)	3.7	7.7E-04	1.1E-01
Adherens junction	<i>Ptk2</i> , <i>Lima1</i> , <i>Itga6</i> , <i>Ptk2b</i> , <i>Tln2</i> , <i>Lmo7</i> , <i>Evl</i> , <i>Mllt4</i> (chr17), <i>Arhgap26</i>	4.3	1.1E-03	6.8E-02
Protein kinase activity	<i>Tyro3</i> , <i>Fgfr1</i> , <i>Alpk1</i> , <i>Ltk</i> , <i>Ptpn3</i> , <i>Hck</i> , <i>Ulk4</i> , <i>Ephb4</i> , <i>Epha2</i> (chr4), <i>Dapk1</i> (chr13), <i>Rps6kl1</i> , <i>Ddr1</i> (chr17), <i>Ptk2</i> , <i>Ptk2b</i> , <i>Camk1</i> , <i>Stk39</i> , <i>Camk2B</i> , <i>Txk</i> , <i>Grk5</i> , <i>Camk2a</i> , <i>Mylk</i>	2.0	4.3E-03	1.9E-01
Calmodulin binding	<i>Adcy1</i> , <i>Myo7a</i> , <i>Camk1</i> , <i>Camk2b</i> , <i>Camk2a</i> , <i>Mylk</i> , <i>Myh10</i> (chr11), <i>Dapk1</i> (chr13)	3.9	4.5E-03	1.8E-01

<sup>1</sup> Genes expressed at higher amounts in NOD.*Rag* cells are indicated in bold

<sup>2</sup> Genes located within a mapped QTL region in this study are noted (chr)