The long non-coding RNA *Dreg1* is required for optimal ILC2 development

Sara Quon[1,2,6], Adelynn Tang[3,4,5,6], Nadia Iannarella[1,2], Kael Schoffer[1,2], Wing Fuk Chan[1,2], Timothy M. Johanson[1,2], Ajithkumar Vasanthakumar[3,4,5,7]*, Rhys S. Allan[1,2,7]*

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia.

[2]Department of Medical Biology, The University of Melbourne, Parkville, Victoria, 3010, Australia.

[3]Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria, Australia

[4]La Trobe University, Bundoora, Victoria, Australia

[5]Peter MacCallum Cancer Centre, Melbourne, VIC, 3000, Australia

[6]These authors contributed equally.

[7]These authors contributed equally.

***Corresponding authors:**

Ajithkumar Vasanthakumar (ajith.vasathakumar@petermac.org) and Rhys S. Allan (rallan@wehi.edu.au)

Running title: *Dreg1* tunes ILC2 development.

Keywords: *Dreg1*, *Gata3*, ILC2, *Tcf1*, haematopoiesis, Non-coding RNA

Abstract

Gata3 is an essential transcription factor for the development of several distinct immune cell lineages such as T cells, natural killer (NK) cells and innate lymphoid cells (ILC). As such, the levels and timing of *Gata3* expression are critical for directing lineage fate decisions. The *Gata3* locus has a complex and dynamic distal regulatory enhancer landscape. Recently we identified a non-coding RNA, *Dreg1*, located immediately upstream of the classic +280kb T/NK cell enhancer (Tce1). To test its function, we excised the *Dreg1* locus in mice and observed a selective reduction of group 2 ILCs (ILC2) across multiple tissues, but mature T, NK and other ILC lineages remained unchanged. In bone marrow, common innate lymphoid cell progenitors (ILCP) increased while ILC2 progenitors (ILC2P) decreased, with a modest reduction of *Gata3* in upstream progenitors consistent with an early developmental bottleneck. Chromatin profiling showed the Dreg1 locus is accessible in early lymphoid progenitors and became decorated with H3K27ac in ILCP in a Tcf1-dependent manner. Furthermore, Tcf1-deficient cells did not express *Dreg1* and showed alterations in the epigenetic landscape of the *Dreg1* locus. Finally, we discovered that potential homologues of *Dreg1* harboured in a syntenic enhancer of *GATA3* are also highly expressed in human ILC2. Taken together we conclude that *Dreg1* is a Tcf1-dependent non-coding RNA critical for fine tuning the high level of *Gata3* required for the optimal development of the ILC2 lineage.

Introduction

The immune system comprises a collection of functionally distinct cell types which act in concert to protect the body from infection and malignancy. These diverse cell populations develop from common haematopoietic progenitors in a highly controlled fashion driven by a network of transcription factors that control lineage-specific gene expression programs. GATA-binding protein 3 (Gata3) is one of the master transcription factors that regulates the development of immune cells such as T cells, Natural Killer (NK) cells, and innate lymphoid cells (ILCs) (1, 2). Gata3 is also critical for the development of $CD4^+$ T helper 2 (Th2) cells (3) which produce cytokines such as IL4, 5 and 13 that are critical for protection against extracellular pathogens. It has recently emerged that a population of ILCs known as group 2 ILCs (ILC2) with similar characteristics to Th2 cells reside in non-lymphoid tissues such as the small intestine, visceral adipose tissue and lung and have been shown to play a non-redundant role in immunity to helminths and drive allergic immune responses(4). ILC2s develop in the bone marrow via common ILC progenitor (CILP)(5), common helper ILC progenitor (CHILP)(6), innate lymphoid cell progenitor (ILCP) (7) and finally ILC2 progenitor (ILC2P)(8). Gata3 is critical to the development of ILCs (8, 9) and high levels are required for the development of ILC2P in the bone marrow (8, 10).

Dosage of Gata3 plays a critical role in T, NK and ILC development and peripheral immunity (8, 10, 11). To support this *Gata3* expression is spatiotemporally tuned by several distal enhancers that dictate *Gata3* activity in these individual lineages. Previous research identified a 7.1kb region ~280kb from the *Gata3* gene that is important for its expression in NK and T cell development, known as T cell enhancer 1 (Tce1**)** (12, 13). Subsequent work found that this region (also termed *Gata3* +278/285) also impacts ILC2 development but not Th2 cell

differentiation (14). ILC2-specific effects have been attributed to additional distal regulatory regions, such as *Gata3* +674/762 (14) and the Gata3 tandem super enhancers (+500/764)(15). In addition, a Th2 specific enhancer (*Gata3* +906/935) has recently been identified (16). Overall, this complex regulatory landscape mechanistically fit with a dosage-threshold model controlling *Gata3* expression.

Recently we discovered a long non-coding RNA gene that we named *Distal regulatory enhancer of Gata3 1 (Dreg1)* which lies ~0.6kb upstream of Tce1 enhancer (17, 18). Expression of *Dreg1* is high in early T cell progenitors and Th2 cells, mirroring *Gata3* expression (17). Overexpression studies indicated that it may be involved in the establishment but not maintenance of *Gata3* expression (17, 18), but it remains unclear how the loss of *Dreg1* affects immune system development. Here, we explore the role of *Dreg1* in the regulation of immune cell differentiation. We find that deletion of the *Dreg1* locus leads to a specific loss of ILC2 progenitors and their downstream progeny in non-lymphoid tissues but not T or NK cells. Chromatin profiling reveals that the regulatory regions around the *Dreg1* gene are open in haematopoietic progenitors and are bound by the transcription factor TCF1 in early innate lymphoid progenitors and TCF1-deficient mice lack *Dreg1* expression. Finally, we discovered potential homologues of *Dreg1* that highly expressed in human ILC2 emanating from a syntenic enhancer of GATA3.

Results

Deletion of the Dreg1 locus results in a specific reduction of ILC2

To understand how *Dreg1* contributes to the development of the immune system we generated *Dreg1*-deficient mice by CRISPR-Cas9-mediated excision of the complete 3kb of the *Dreg1* gene (Figure 1A). We confirmed that the *Dreg1* gene was removed from the germline and found that these mice reproduced at Mendelian frequencies and exhibited no overt developmental defects suggesting that *Dreg1* is not required for embryogenesis.

Given our previous finding of the relatively high levels of *Dreg1* expression in T cells (17), and of a critical role of the Tce1 (*Gata3* +278/285) enhancer (12, 14), we expected that the deletion of *Dreg1* would impact T cell and NK cell development. We therefore performed an analysis of the different populations of T cells. Mature $CD4^+$ and $CD8^+$ T cells in the spleen appeared unaltered in frequency (Supplementary Figure 1). Moreover, both populations had similar proportions of naïve, effector, and memory cells (Supplementary Figure 1B &C). This suggests that *Dreg1* is not required for conventional T cell development or maintenance. In addition, the loss of *Dreg1* did not affect the development of NK cells and γδ T cells as normal proportions were observed in the spleen (Supplementary Figure 1D). Overall, this analysis suggests that *Dreg1* is dispensable for NK and T cell development.

Gata3 is also critical for the development of all innate lymphoid cells (ILCs)(9) and the levels of Gata3 are critical in directing ILC2 lineage development (8, 19). Therefore, we next explored peripheral sites known to harbour substantial numbers of these cell types such as visceral adipose tissue (VAT), the small intestine lamina propria (SI LP) and the lung. While these tissues had

relatively normal proportions of NK, ILC1 and ILC3 (Figure 1B, C & E) we noted a specific reduction in the population of ILC2 in each of the tissues we examined (Figure 1D). Thus far the data obtained is from mice in which Dreg1 is deleted in all cells, hampering the ability to determine whether the loss of Dreg1 intrinsically regulates ILC2 development or the phenotype is driven by extrinsic signals. To demonstrate a cell-intrinsic role of Dreg1 in ILC2s we generated mixed bone marrow irradiation chimeras comprising a ratio of 50:50 of wild type (CD45.1^{+}) to Dreg1-deficient (CD45.2^{+}) cells. In these mice we observed a reduction in Dreg1-deficient ILC2 in comparison to their wild type ILC2 counterparts in all tissues examined (Figure 1F and Supplementary Figure 1E-G). In contrast we observed comparable proportions of wild type and Dreg1-deficient B cells from these same tissues (Supplementary Figure 1F-G). Taken together these results show that the loss of *Dreg1* results in the specific and intrinsic loss of ILC2 in different peripheral tissues.

Dreg1 deficiency results in the reduction of ILC2 progenitors

We next investigated the origins of the ILC2 defect in the *Dreg1*-deficient mice. It is known that expression of *Gata3* in the early ILC progenitor stage is required for development of mature ILC2s (8, 10, 19). Indeed, analysis of publicly available RNAseq data (20) demonstrated that Gata3 expression was very high in ILC2P and ILC2 compared to other populations, whereas the highest expression of Dreg1 was found in ILC2P (Figure 2A). We examined ILC development in the bone marrow (Figure 2B-G, gating strategy in Supplementary Figure 2A). We began by examining the common lymphoid (CLP) and lymph-myeloid primed progenitors (LMPP) and found no overt changes in proportions or total number of these populations (Figure 2B & C). However, we found increased CILP, CHILP and ILCP in the *Dreg1*-deficient mice (reaching

significance in the CILPs) (Figure 2D-F). We observed a substantial reduction of the ILC2P population suggesting that the loss of *Dreg1* leads to a block in ILC development (Figure 2G). This suggests that the loss of *Dreg1* results in a bottleneck at the early stages of ILC development. We next examined the levels of Gata3 in these populations (Figure 2H). We observed a reduction in Gata3 signal in the upstream CHILP and the ILCP from the *Dreg1*-deficient mice, but interestingly Gata3 levels were normal in the ILC2P even though this population is reduced in the *Dreg1*-deficient mice. This suggests that *Dreg1* acts to regulate Gata3 prior to the ILC2P stage which leads to a block in their development, however, if cells can overcome this perturbation, they express relatively normal levels of Gata3.

Dynamic epigenetic regulation of the Dreg1 locus in a Tcf1-dependent manner during ILC development

To gain a greater understanding of how the expression of *Dreg1* is regulated we next studied the chromatin state of the *Dreg1* locus during early ILC development via analysis of publicly available ATAC-Seq data generated on progenitors and their downstream ILC progeny (14)(Figure 3A). Of note, the original Refseq annotation of Dreg1 was missing the first exon and TSS which we identified (17) and show in Figure 3A. We observed three significant accessible chromatin “peaks”, with the predominant one at the promoter which was open at the CLP stage remained so in ILCP and ILC2P. One region downstream became more accessible in ILCP and ILC2P in line with that observed by others (14). Overall, we find the promoter of *Dreg1* is open in early lymphoid progenitors and the acquisition of chromatin accessibility downstream correlates with increased *Dreg1* expression in ILC2 progenitors.

To understand the transcription factors that might be involved in regulating *Dreg1* expression we examined the motifs at the *Dreg1* promoter and downstream accessible region. We observed a specific enrichment of motifs for Tcf1, Lef1 and Ets1 suggesting that these transcription factors may be involved in regulating *Dreg1* expression. As Tcf1 (encoded by the *Tcf7* gene) has been implicated in ILC2 development (21, 22) we focussed on its role in regulating *Dreg1* expression. Firstly, we examined publicly available Tcf1 CUT&Run data (23) and found that indeed Tcf1 was specifically bound to the accessible sites of the *Dreg1* locus in early innate lymphoid progenitors (EILP, which are a population that sit between CILP and CHILP) (Figure 3B). We next examined ChIC-Seq data (24) of active (H3K27ac) or repressive (H3K27me3) histone modifications at different stages of ILC development (Figure 3D). Firstly, we found no evidence of H3K27me3 in this region in any of the cell types examined. However, there was a modest accumulation of H3K27ac around the accessible regions in the *Dreg1* locus at the EILPs which was noticeably increased at the ILCP stage. Of note, Tcf1-deficient (via deletion of the *Tcf7* gene) EILP showed lower levels of H3K27ac suggesting that Tcf1 is likely involved in creating a permissive chromatin landscape at the *Dreg1* locus during ILC development. Indeed, the expression of *Dreg1* was dependent on the presence of Tcf1 as *Tcf7*-/- EILPs lacked *Dreg1* expression (23)(Figure 3E). Taken together this data suggests that the alterations to the chromatin state *Dreg1* locus and its subsequent transcriptional activity during ILC development is dependent on Tcf1.

Identification of a human GATA3 enhancer harbouring Dreg1 homologues that are active in ILC2s

Previously we identified a region syntenic to the mouse *Dreg1* locus on human chromosome 10 that is a similar distance from the *GATA3* gene and forms an 3D interaction in T cells but not B cells suggesting it may a represent an enhancer element (17). Interestingly, bidirectional transcription initiates from the sequence-conserved *cis* element to produce two potential human homologs of *Dreg1*, the lncRNAs CAT00000105356.1 (we named *DREG1.1*) and CAT00000117261.1 (*DREG1.2*). To explore whether these transcripts are also expressed in ILC populations we examined publicly available RNAseq data from healthy human blood (25)(Figure 4A). We found that GATA3 was highly expressed in human ILCP, ILC2 and TH2 cells. We also observed this pattern for *DREG1.1* and to an even greater extent for *DREG1.2* suggesting that the syntenic region harbours transcripts which reflect the same expression pattern as murine *Dreg1*.

Finally, we investigated the functional role of this potential enhancer of GATA3 by examining a recently published tiling CRISPR deletion screen searching for regulatory elements that control GATA3 expression in human Th2 cells (26). We focused on the syntenic region which was defined as functional sequence (FS) 23 (Figure 4B). Specifically, guides that deleted conserved sequences FS23-4 and FS23-5 led to a significant downregulation of GATA3 expression (26) suggesting that this region represents a distal enhancer of GATA3 (Figure 4B). ATACseq data from either cultured (pcILC2) or fresh ILC2 from healthy human blood (27) revealed a clear accessible region in these cells that overlaps the *DREG1.1/2* and FS23 (Figure 4C). Overall, we find that this syntenic region in humans represents an enhancer that is important for high levels of *GATA3* expression and contains two non-coding RNA genes that show a similar expression pattern to murine *Dreg1*.

Discussion

Here we show that germline excision of the *Dreg1* locus resulted in a specific reduction in ILC2 cells owing to a developmental bottleneck during ILC development in the bone marrow. These results are in line with the fact that very high levels of Gata3 are required for ILC2 development (8, 10). Given its proximity to the Tce1 (+280kb *Gata3*) enhancer, a surprising finding was that the deletion of *Dreg1* did not affect T or NK cells lineages. While it is tempting to speculate that the *Dreg1* locus represents a ILC2-specific enhancer element we favour the idea that *Dreg1* likely contributes to dosage tuning of Gata3 during ILC2 commitment and that ILC2s because of their higher Gata3 threshold are more sensitive to modest reductions, whereas T and NK cells are buffered by redundant modules.

Our analysis of publicly available data revealed that the Dreg1 locus was accessible in common lymphoid progenitors and subsequently gained H3K27ac during ILC differentiation coinciding

with increased Dreg1 expression. This region was enriched for Tcf1 motifs which was supported by analysis of data from Harly et al.(23), that showed indeed that Tcf1 bound to these regions in ILCP. Subsequent analysis revealed that H3K27ac of the Dreg1 locus and the expression of Dreg1 was reduced in Tcf1-/- EILP. Taken together this suggests that Tcf1, which is known to be critical for Gata3 expression during ILC development is upstream of, and may act in concert with Dreg1 to promote high levels of Gata3 expression in ILC2 development.

While the analysis of publicly available data was invaluable in providing insights into the regulation of Dreg1, further experiments will dig deeper into the mechanisms utilised by Dreg1 to control Gata3 expression. Indeed, how *Dreg1* is involved in regulating *Gata3* levels mechanistically is yet to be determined. *Dreg1* may act as an enhancer RNA that cooperates with chromatin modifying proteins (28) to reinforce local H3K27ac and enhancer-promoter communication. Future experiments that selectively suppress *Dreg1* transcription for example via antisense oligonucleotides or CRISPRi at the Dreg1 promoter will enable the discrimination of RNA-dependent from DNA element-dependent effects.

In humans, we identify two noncoding transcripts within the syntenic GATA3 distal enhancer that are highly expressed in ILCP, ILC2 and TH2 cells. Perturbations within this conserved block in an independent tiling-deletion screen performed by Chen et al.(26), resulted in reduced GATA3 expression in T cells. While these experiments were performed in T cells, we expect that this enhancer region will likely also play an important role in the regulation of ILC2 development. Given that ILC2 have been implicated in the development of allergic reactions (4)

the noncoding transcripts we identified may represent targets for the modulation of GATA3 levels and ILC2 function in disease.

Methods

Mice

The *Dreg1* knockout mouse line was generated at the Melbourne Advanced Genome Editing Centre (MAGEC) laboratory using CRISPR-Cas9 through microinjection into the pronucleus of one-cell stage C57BL/6 embryos with 2 sgRNAs (CTACTTGCTGACAAGTCGTC and TCTAGTAAGTCCAGTTGCTT) to target the 3,594 bp genomic region around *Dreg1*. The F0 offspring were validated for *Dreg1* deletion with forward primer GACCAGATATGGAGACGTGCA and reverse primer TCTTTGCCATCTTCTGTGTGC and were backcrossed with wild-type C57BL6/J mice for 2 or more generations.

Cells and tissues were obtained from littermate control mice aged at least 6 weeks unless otherwise specified. Due to sex differences in ILC populations (29), only female mice were used. All mice were maintained at the WEHI Animal Facility under specific pathogen-free conditions.

Bone marrow chimeras

Mixed bone marrow chimeras were generated by harvesting bone marrow from either wild type (CD45.1^{+}) mice or Dreg1-deficient (CD45.2^{+}) mice, mixing equally and injecting 5 million cells intravenously into lethally irradiated (2x550 Gy) F1 (CD45.1^{+}/CD45.2^{+}) mice. Mice were left to reconstitute for over 8 weeks.

Cell isolation:

Thymus and spleen

The spleen and thymus were made into single-cell suspension by mashing through a 70µm mesh. Splenic cells were treated with red cell lysis buffer.

Bone marrow

Both pairs of hips, femurs, and tibia bones were grounded in a mortar and pestle to collect the bone marrow. The cell suspension was filtered through a 70µm cell strainer and negatively enriched for ILC cells with the mouse lineage cell depletion kit (Miltenyi).

Extraction of lymphocytes from non-lymphoid tissues

Lymphocytes from the VAT and lung were finely chopped in 3-5mL digestion buffer (RPMI and 2% FCS with 2mg/mL Collagenase IV). The suspension was digested for 40 min at 37°C and shaken at 180 rpm. Digested VAT samples were topped up to 30ML with FACS buffer (1xPBS with 2% FCS) and spun at 800g for 15 min at 4°C. Lungs were mashed through a 70mM filter and topped to 30mL FACS buffer and pelleted at 1700rpm for 5min. Both VAT and lung pellets were treated with RBC lysis buffer.

Isolation of lymphocytes from small intestinal lamina propria

Mesentery and Peyer's patches were removed longitudinally, and faeces removed. Intestine was chopped into 0.5cm pieces into 1xPBS and vortexed. Intestinal pieces were placed into 20mLs of dissociation buffer (1xHBSS without Ca2+ and Mg2+ with 2% FCS, 10mM HEPES and 5 mM EDTA) and digested for 40min at 37°C at 180rpm. Digested pieces were washed with FACS buffer and then placed in 5-8mL of digestion buffer (1xHBSS with indicator containing 2% FCS, 10mM HEPES, 2mg/mL Collagenase IV and 2mg/mL DNAse I) and incubated for 30-40min at

37°C at 180rpm. Digested tissue was mashed through a 70 mM cell strainer and pelleted. A 40/80% percoll gradient was performed to enrich for lymphocytes.

Flow cytometry and antibodies:

Fluorophore conjugated antibodies against mouse antigens were used for flow cytometry. Antibodies, clone names and manufactures: ***Biolegend***- CD4 (GK1.5) APCFire750, CD8a (53-6.7) BV570, NKp46 (29A1.4) PEcy7, ICOS (7E.17G9) BV421, Ki67 (11F6) BV650, CD62L (MEL-14) APC and T-bet (4B10) APC. ***BD***- CD19 (1D3) BV786, TCRb (H57-597) BV750, TCRgd (GL3) BV711, B220 (RA3-6B2) BV786, GATA3 (L50-823) BUV395, RORgt (Q31-378) BV421, RORgt (Q31-378) PerCPcy5.5, KLRG1 (2F1) BUV615, CD49a (Ha31/8), CD25 (PC61) BV480, NK1.1 (PK136) BUV805, NK1.1 (PK136) BB700, CD44 (IM7) BUV496. **Invitrogen/bioscience**- CD11c (N418) FITC, FOXP3 (FJK-16s) PE, IL-33R (ST2) (RMST2-2) PerCP-eFLuor710, Streptavidin BUV563.***WEHI***-CD45.2 (104) AF700. Live dead staining was performed using Fixable viability stain 440UV diluted in 1xPB at 1:1000 dilution. Surface staining of antibodies was for 30min on ice at 1/200, 1/300 or 1/400 in FACS buffer. Intracellular staining of transcription factors was overnight at 4°C using the Foxp3/Transcription factor staining buffer set from eBioscience overnight. Cell numbers for specific cell populations were calculated using the volume of the sample that was analyzed, the cell counts per sample, and the percentage of the specific cell population.

RNA-Seq Analysis:

Publicly available RNA-Seq data were downloaded from SRA and re-analysed. Briefly fastq files were analyzed with FASTQC (0.12.1) and adapters were removed using trimmomatic(30) (0.36). The reads were aligned to GRCm39 for mouse samples and GRCh38 for human samples using hisat2(31) (2.0.5). Gene counts were quantified using Rsubread(32) featureCounts (1.6.3) with annotation from gencode(33) (vM33 for mouse and v42 for human), which had the annotation for the *Dreg1* gene or human homologs of the *Dreg1* gene added, respectively. Samples were processed using edgeR(34) (3.42.4). Lowly expressed genes were removed, and samples were normalized according to library size. Gene expression was plotted as logCPM using ggplot2 (3.4.4).

ATAC-Seq Analysis:

Publicly available ATAC-Seq data were downloaded from SRA and re-analysed. Processing was based on the ENCODE ATAC-Seq pipeline. Briefly fastq files were analyzed with FASTQC and adapters were removed using trimmomatic (0.36). The reads were aligned to the GRCm39 gnome for mouse samples and GRCh38 genome for human samples using bowtie2 (2.4.4). Duplicate reads were marked using Picard tools (2.26.11) and reads were filtered using samtools(35) (1.9) with -F 1804 -f 2. Fold-change coverage tracks were created using macs2 (2.2.7.1) callpeak with a smooth window of 150 and a shift size of -75 and bdgcmp. The bedgraph was converted to a bigwig using ucsc tools (331) bedGraphToBigWig(36) and visualized using IGV(37) (2.17.4) with group scaling among samples from the same experiment.

Motif Scanning:

The regions of interest as identified through peak calling of ATAC-Seq data were scanned for motifs using FIMO(38) from the MEME suite (5.0.5) using the Hocomoco(39) (H12) core motif

database. The resulting list was filtered to remove factors that were not expressed in ILC2P cells as determined through bulk RNA-Seq (GSE77695) with a cutoff of >= 1 LogCPM.

ChIC and CUT&Run Analysis:

Publicly available ChIP-Seq or CUT&Run data were downloaded from SRA and re-analysed. Briefly fastq files were analyzed with FASTQC and adapters were removed using trimmomatic (0.36). The reads were aligned to the GRCm39 gnome for mouse samples using bowtie2 (2.4.4) with the options -q -5 0 -3 0. Duplicate reads were marked using Picard tools (2.26.11) and reads were filtered using samtools (1.9) with -F 1804 -f 2. Fold-change coverage tracks were created using macs2 (2.2.7.1) callpeak with a smooth window of 50 and a shift size of -25 for transcription factors and 150 and -75 for histone marks along with macs2 bdgcmp. The bedgraph was converted to a bigwig using ucsc tools (331) bedGraphToBigWig and visualized using IGV (2.17.4) with group scaling among samples from the same experiment.

CRISPR Guides

The guide counts were downloaded from GSE190860 and were used to create a bed file to identify the target regions. The bed file was visualized in IGV (2.17.4).

Publicly available datasets used in this study:

Dataset type	Accession Number
Bulk RNA-Seq (Mouse-ILC)(20)	GSE77695
Bulk RNA-Seq (Mouse-ILC w TCF1KO)(23)	GSE113767
Bulk RNA-Seq (Human)(25)	PRJEB35186

Bulk ATAC-Seq (ILCP)(14)	GSE169542
Bulk ATAC-Seq (HumanMouseILC2)(27)	GSE231999
Bulk ChIP-Seq (Histone)(40)	GSE85156
Bulk CUT&Run (EILP TCF1)(23)	GSE128483
Bulk ChIC-Seq (Histone Tcf1 KO)(24)	GSE142468
CRISPR Screen(26)	GSE190860

Acknowledgments

The authors thank the members of the Allan and Vasanthakumar labs for their input into this project. The authors also gratefully acknowledge the WEHI Flow Cytometry and Bioservices Facilities for their support and assistance in this work. We also thank the Melbourne Advanced Genome Editing Centre (MAGEC) laboratory for generation of the Dreg1$^{-/-}$ mice.

Funding

This work was supported by the NHMRC, Australia Ideas Grant 2001131 and a Stafford Fox Medical Foundation Research Grant to R.A and an NHMRC Investigator Grant 2009336 to A.V. This work was made possible through Victorian State Government Operational Support Program and the Australian Government NHMRC IRIISS.

Authorship contributions

Sara Quon: Conceptualization, Investigation, Formal Analysis, Writing-Original Draft, Review & Editing. **Adelynn Tang:** Conceptualization, Investigation, Formal Analysis, Writing-Review & Editing. **Nadia Iannarella**: Investigation. **Kael Schoffer:** Investigation. **Wing Fuk Chan:** Conceptualization, Methodology. **Timothy Johanson:** Investigation, Supervision. **Ajith Vasanthakumar:** Conceptualization, Supervision, Funding Acquisition, Writing-Review &

Editing. **Rhys Allan:** Conceptualization, Supervision, Funding Acquisition, Writing-Original Draft, Review & Editing.

Conflicts of interest

The authors have no conflicts of interest.

References

1. Harly, C., M. Cam, J. Kaye, and A. Bhandoola. 2018. Development and differentiation of early innate lymphoid progenitors. *J Exp Med* 215: 249-262.
2. Ho, I. C., T. S. Tai, and S. Y. Pai. 2009. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol* 9: 125-135.
3. Zheng, W., and R. A. Flavell. 1997. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell* 89: 587-596.
4. Jarick, K. J., P. M. Topczewska, M. O. Jakob, H. Yano, M. Arifuzzaman, X. Gao, S. Boulekou, V. Stokic-Trtica, P. S. Leclere, A. Preusser, Z. A. Rompe, A. Stamm, A. M. Tsou, C. Chu, F. R. Heinrich, G. M. Guerra, P. Durek, A. Ivanov, D. Beule, S. Helfrich, C. U. Duerr, A. A. Kuhl, C. Stehle, C. Romagnani, M. F. Mashreghi, A. Diefenbach, D. Artis, and C. S. N. Klose. 2022. Non-redundant functions of group 2 innate lymphoid cells. *Nature* 611: 794-800.
5. Yu, X., Y. Wang, M. Deng, Y. Li, K. A. Ruhn, C. C. Zhang, and L. V. Hooper. 2014. The basic leucine zipper transcription factor NFIL3 directs the development of a common innate lymphoid cell precursor. *Elife* 3.
6. Klose, C. S. N., M. Flach, L. Mohle, L. Rogell, T. Hoyler, K. Ebert, C. Fabiunke, D. Pfeifer, V. Sexl, D. Fonseca-Pereira, R. G. Domingues, H. Veiga-Fernandes, S. J. Arnold, M. Busslinger, I. R. Dunay, Y. Tanriver, and A. Diefenbach. 2014. Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell* 157: 340-356.
7. Constantinides, M. G., B. D. McDonald, P. A. Verhoef, and A. Bendelac. 2014. A committed precursor to innate lymphoid cells. *Nature* 508: 397-401.
8. Hoyler, T., C. S. Klose, A. Souabni, A. Turqueti-Neves, D. Pfeifer, E. L. Rawlins, D. Voehringer, M. Busslinger, and A. Diefenbach. 2012. The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. *Immunity* 37: 634-648.
9. Yagi, R., C. Zhong, D. L. Northrup, F. Yu, N. Bouladoux, S. Spencer, G. Hu, L. Barron, S. Sharma, T. Nakayama, Y. Belkaid, K. Zhao, and J. Zhu. 2014. The transcription factor GATA3 is critical for the development of all IL-7Ralpha-expressing innate lymphoid cells. *Immunity* 40: 378-388.
10. Klein Wolterink, R. G., N. Serafini, M. van Nimwegen, C. A. Vosshenrich, M. J. de Bruijn, D. Fonseca Pereira, H. Veiga Fernandes, R. W. Hendriks, and J. P. Di Santo. 2013. Essential, dose-dependent role for the transcription factor Gata3 in the development of

IL-5+ and IL-13+ type 2 innate lymphoid cells. *Proc Natl Acad Sci U S A* 110: 10240-10245.
11. Scripture-Adams, D. D., S. S. Damle, L. Li, K. J. Elihu, S. Qin, A. M. Arias, R. R. Butler, 3rd, A. Champhekar, J. A. Zhang, and E. V. Rothenberg. 2014. GATA-3 dose-dependent checkpoints in early T cell commitment. *J Immunol* 193: 3470-3491.
12. Hosoya-Ohmura, S., Y. H. Lin, M. Herrmann, T. Kuroha, A. Rao, T. Moriguchi, K. C. Lim, T. Hosoya, and J. D. Engel. 2011. An NK and T cell enhancer lies 280 kilobase pairs 3' to the gata3 structural gene. *Mol Cell Biol* 31: 1894-1904.
13. Ohmura, S., S. Mizuno, H. Oishi, C. J. Ku, M. Hermann, T. Hosoya, S. Takahashi, and J. D. Engel. 2016. Lineage-affiliated transcription factors bind the Gata3 Tce1 enhancer to mediate lineage-specific programs. *J Clin Invest* 126: 865-878.
14. Kasal, D. N., Z. Liang, M. K. Hollinger, C. Y. O'Leary, W. Lisicka, A. I. Sperling, and A. Bendelac. 2021. A Gata3 enhancer necessary for ILC2 development and function. *Proc Natl Acad Sci U S A* 118.
15. Furuya, H., Y. Toda, A. Iwata, M. Kanai, K. Kato, T. Kumagai, T. Kageyama, S. Tanaka, L. Fujimura, A. Sakamoto, M. Hatano, A. Suto, K. Suzuki, and H. Nakajima. 2024. Stage-specific GATA3 induction promotes ILC2 development after lineage commitment. *Nat Commun* 15: 5610.
16. Kumagai, T., A. Iwata, H. Furuya, K. Kato, A. Okabe, Y. Toda, M. Kanai, L. Fujimura, A. Sakamoto, T. Kageyama, S. Tanaka, A. Suto, M. Hatano, A. Kaneda, and H. Nakajima. 2024. A distal enhancer of GATA3 regulates Th2 differentiation and allergic inflammation. *Proc Natl Acad Sci U S A* 121: e2320727121.
17. Chan, W. F., H. D. Coughlan, N. Iannarella, G. K. Smyth, T. M. Johanson, C. R. Keenan, and R. S. Allan. 2021. Identification and characterization of the long noncoding RNA Dreg1 as a novel regulator of Gata3. *Immunol Cell Biol* 99: 323-332.
18. Chan, W. F., H. D. Coughlan, Y. Chen, C. R. Keenan, G. K. Smyth, A. C. Perkins, T. M. Johanson, and R. S. Allan. 2022. Activation of stably silenced genes by recruitment of a synthetic de-methylating module. *Nat Commun* 13: 5582.
19. Zhong, C., M. Zheng, K. Cui, A. J. Martins, G. Hu, D. Li, L. Tessarollo, S. Kozlov, J. R. Keller, J. S. Tsang, K. Zhao, and J. Zhu. 2020. Differential Expression of the Transcription Factor GATA3 Specifies Lineage and Functions of Innate Lymphoid Cells. *Immunity* 52: 83-95 e84.
20. Shih, H. Y., G. Sciume, Y. Mikami, L. Guo, H. W. Sun, S. R. Brooks, J. F. Urban, Jr., F. P. Davis, Y. Kanno, and J. J. O'Shea. 2016. Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell* 165: 1120-1133.
21. Mielke, L. A., J. R. Groom, L. C. Rankin, C. Seillet, F. Masson, T. Putoczki, and G. T. Belz. 2013. TCF-1 controls ILC2 and NKp46+RORgammat+ innate lymphocyte differentiation and protection in intestinal inflammation. *J Immunol* 191: 4383-4391.
22. Yang, Q., F. Li, C. Harly, S. Xing, L. Ye, X. Xia, H. Wang, X. Wang, S. Yu, X. Zhou, M. Cam, H. H. Xue, and A. Bhandoola. 2015. TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat Immunol* 16: 1044-1050.
23. Harly, C., D. Kenney, G. Ren, B. Lai, T. Raabe, Q. Yang, M. C. Cam, H. H. Xue, K. Zhao, and A. Bhandoola. 2019. The transcription factor TCF-1 enforces commitment to the innate lymphoid cell lineage. *Nat Immunol* 20: 1150-1160.
24. Ren, G., B. Lai, C. Harly, S. Baek, Y. Ding, M. Zheng, Y. Cao, K. Cui, Y. Yang, J. Zhu, G. L. Hager, A. Bhandoola, and K. Zhao. 2022. Transcription factors TCF-1 and GATA3 are

key factors for the epigenetic priming of early innate lymphoid progenitors toward distinct cell fates. *Immunity* 55: 1402-1413 e1404.
25. Ercolano, G., T. Wyss, B. Salome, P. Romero, S. Trabanelli, and C. Jandus. 2020. Distinct and shared gene expression for human innate versus adaptive helper lymphoid cells. *J Leukoc Biol* 108: 723-737.
26. Chen, H. V., M. H. Lorenzini, S. N. Lavalle, K. Sajeev, A. Fonseca, P. C. Fiaux, A. Sen, I. Luthra, A. J. Ho, A. R. Chen, K. Guruvayurappan, C. O'Connor, and G. McVicker. 2023. Deletion mapping of regulatory elements for GATA3 in T cells reveals a distal enhancer involved in allergic diseases. *Am J Hum Genet* 110: 703-714.
27. Laurie, S. J., J. P. Foster, 2nd, D. W. Bruce, H. Bommiasamy, O. V. Kolupaev, M. Yazdimamaghani, S. G. Pattenden, N. J. Chao, S. Sarantopoulos, J. S. Parker, I. J. Davis, and J. S. Serody. 2024. Type II innate lymphoid cell plasticity contributes to impaired reconstitution after allogeneic hematopoietic stem cell transplantation. *Nat Commun* 15: 6000.
28. Statello, L., C. J. Guo, L. L. Chen, and M. Huarte. 2021. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22: 96-118.
29. Kadel, S., E. Ainsua-Enrich, I. Hatipoglu, S. Turner, S. Singh, S. Khan, and S. Kovats. 2018. A Major Population of Functional KLRG1(-) ILC2s in Female Lungs Contributes to a Sex Bias in ILC2 Numbers. *Immunohorizons* 2: 74-86.
30. Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
31. Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907-915.
32. Liao, Y., G. K. Smyth, and W. Shi. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 47: e47.
33. Mudge, J. M., S. Carbonell-Sala, M. Diekhans, J. G. Martinez, T. Hunt, I. Jungreis, J. E. Loveland, C. Arnan, I. Barnes, R. Bennett, A. Berry, A. Bignell, D. Cerdan-Velez, K. Cochran, L. T. Cortes, C. Davidson, S. Donaldson, C. Dursun, R. Fatima, M. Hardy, P. Hebbar, Z. Hollis, B. T. James, Y. Jiang, R. Johnson, G. Kaur, M. Kay, R. J. Mangan, M. Maquedano, L. M. Gomez, N. Mathlouthi, R. Merritt, P. Ni, E. Palumbo, T. Perteghella, F. Pozo, S. Raj, C. Sisu, E. Steed, D. Sumathipala, M. M. Suner, B. Uszczynska-Ratajczak, E. Wass, Y. T. Yang, D. Zhang, R. D. Finn, M. Gerstein, R. Guigo, T. J. P. Hubbard, M. Kellis, A. Kundaje, B. Paten, M. L. Tress, E. Birney, F. J. Martin, and A. Frankish. 2025. GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* 53: D966-D975.
34. Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.
35. Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10.
36. Kent, W. J., A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26: 2204-2207.

37. Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178-192.
38. Grant, C. E., T. L. Bailey, and W. S. Noble. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017-1018.
39. Kulakovskiy, I. V., I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, and V. J. Makeev. 2018. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* 46: D252-D259.
40. Gury-BenAri, M., C. A. Thaiss, N. Serafini, D. R. Winter, A. Giladi, D. Lara-Astiaso, M. Levy, T. M. Salame, A. Weiner, E. David, H. Shapiro, M. Dori-Bachash, M. Pevsner-Fischer, E. Lorenzo-Vivas, H. Keren-Shaul, F. Paul, A. Harmelin, G. Eberl, S. Itzkovitz, A. Tanay, J. P. Di Santo, E. Elinav, and I. Amit. 2016. The Spectrum and Regulatory Landscape of Intestinal Innate Lymphoid Cells Are Shaped by the Microbiome. *Cell* 166: 1231-1246 e1213.

Figure Legends

Figure 1: *Dreg1* deletion results in a specific reduction in peripheral ILC2 cells.

(A) Deletion of Dreg1 in mice using CRISPR-Cas9. (B-E) Representative FACS plots and % of $CD45^+$, $TCR\beta^-$, $CD19^-$, $CD11c^-$ NK and ILC populations from the visceral adipose tissue (VAT), small intestinal lamina propria (SI LP), and lung from wild type (WT) or Dreg1-/- mice. Shown is one representative experiment of two with n=4 mice/group. (F) SI LP from mixed bone marrow chimeras were examined for the proportion of ILC2 from the wildtype (CD45.1) or Dreg1-deficient (CD45.2) compartment. Data is from pooled from 3 independent experiments. Mean and SEM together with individual data points are shown. Data were statistically analysed by Student *t* test.

Figure 2: *Dreg1* deletion results in a bottleneck in ILC development and a reduction in ILC2P in the bone marrow.

(A) Boxplots showing expression of *Gata3* and *Dreg1* from RNA-Seq (GSE77695). (B-G) Quantification of ILC progenitor subsets in the bone marrow with percentages (upper) and total numbers (below). (H) Mean fluorescence intensity (MFI) of Gata3 measured on CHILP, ILCP and ILC2P. Shown is one representative experiment of two with n=3-4 mice/group. Mean and SEM together with individual data points are shown. Data were statistically analysed by Student *t* test.

Figure 3: *Dreg1* locus is dynamically regulated in a Tcf1-dependent manner during ILC2 development.

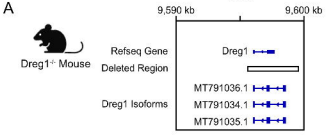
(A) Chromatin accessibility around Dreg1 in ILC progenitors (GSE169542). (B) Motifs enriched in accessible regions around Dreg1. (C) Tcf1 binding around Dreg1 locus in EILP (GSE128483). (D) Histone marks around Dreg1 at different stages of ILC development and in Tcf7-/- EILP (GSE142468). (E) *Gata3* and *Dreg1* expression quantified from RNA-Seq data in ILC progenitors and Tcf7-/- EILP (GSE113767).

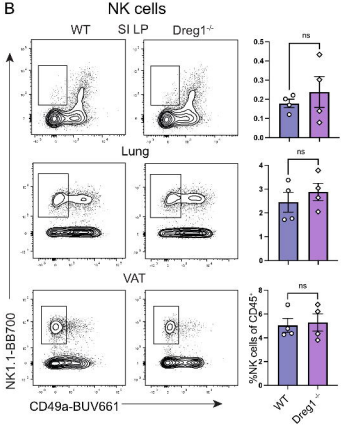
Figure 4: A syntenic region in humans represents a *GATA3* enhancer element with a transcriptional profile akin to murine *Dreg1*. (A) *GATA3* and *DREG1.1 and DREG1.2* expression quantified from RNA-Seq data in ILC and T helper populations from isolated from healthy human blood (PRJEB35186). (B) Alignment of Functional Sequence 23 (FS23) that overlaps the syntenic region revealed as regulating GATA3 expression in a CRISPR deletion screen of human TH2 cells (26). (C) Chromatin accessibility around FS23 guides from the CRISPR deletion screen (26) (GSE231999).

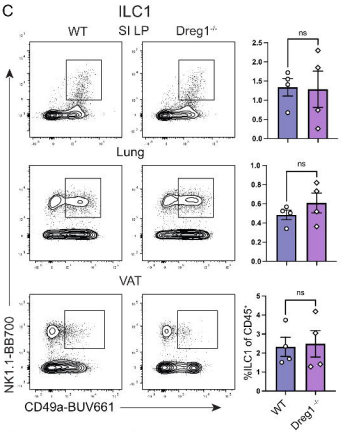
Supplemental Figure 1: *Dreg1* deletion in does not affect T or NK cells. (A-C) Representative FACS plots with quantification (right) showing expression of (A) CD8α and CD4 in splenic T cells. (B) CD44 and CD62L in splenic $CD8^+$ T cells (C) CD44 and CD62L in splenic $CD4^+$ T cells. (D) Quantification of splenic γδ T cells and NK cells. Shown is one representative experiment of two with n=3 mice/group. Mixed bone marrow chimeras were examined for the proportion of SI LP B cells (E), VAT ILC2 or B cells (F) or Lung ILC2 or B cells from the wildtype (CD45.1) or Dreg1-deficient (CD45.2) compartment. Data is from pooled from 3 independent experiments. Mean and SEM together with individual data points are shown. Data were statistically analysed by Student *t* test.

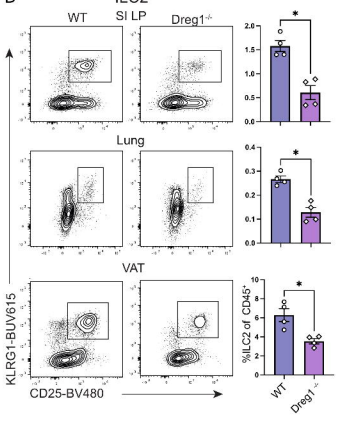
Supplemental Figure 2: Gating of ILC progenitors in bone marrow. (A) Flow cytometric analysis of bone marrow cells was used to quantify the changes between wild type and Dreg1-deficient haematopoietic and ILC progenitors in Figure 2. Viable, lineage negative, CD45.2+ bone marrow cells were gated by the expression of cell surface markers into the delineated progenitor populations, and this was used to quantify their proportions and numbers.

Figure 1


A
Chr 2
9,590 kb
9,600 kb
Dreg1-/- Mouse
Refseq Gene
Dreg1
Deleted Region
Dreg1 Isoforms
MT791036.1
MT791034.1
MT791035.1


B
NK cells
WT
SI LP
Dreg1-/-
Lung
VAT
NK1.1-BB700
CD49a-BUV661
ns
%NK cells of CD45+
WT
Dreg1-/-


C
ILC1
WT
SI LP
Dreg1-/-
Lung
VAT
NK1.1-BB700
CD49a-BUV661
ns
%ILC1 of CD45+
WT
Dreg1-/-


D
ILC2
WT
SI LP
Dreg1-/-
Lung
VAT
KLRG1-BUV615
CD25-BV480
*
%ILC2 of CD45+
WT
Dreg1-/-

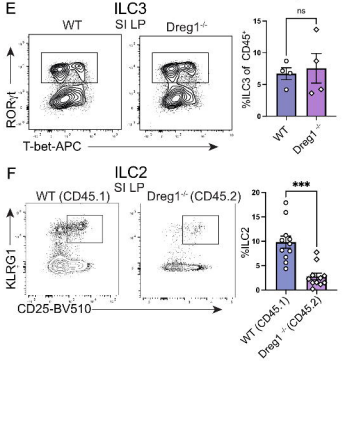

E
ILC3
WT
SI LP
Dreg1-/-
RORγt
T-bet-APC
ns
%ILC3 of CD45+
WT
Dreg1-/-
F
ILC2
SI LP
WT (CD45.1)
Dreg1-/- (CD45.2)
KLRG1
CD25-BV510
***
%ILC2
WT (CD45.1)
Dreg1-/- (CD45.2)

Figure 2

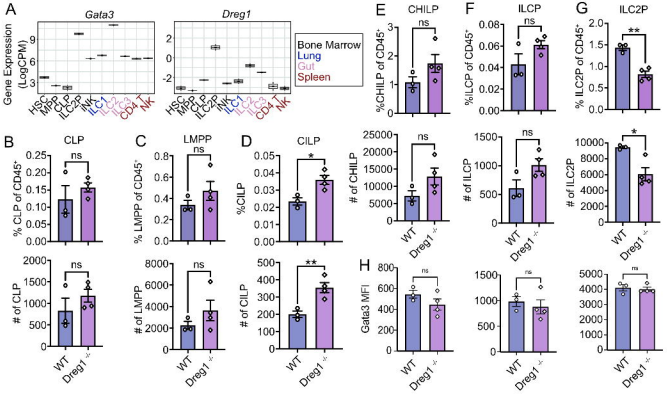
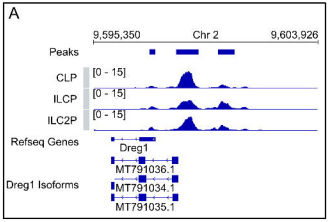
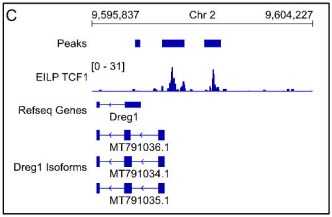

A
Gata3
Dreg1
Gene Expression (Log2CPM)
HSC
MPP
CLP
ILC2P
INK
ILC1
ILC2
ILC3
CD4
NK
Bone Marrow
Lung
Gut
Spleen
B
CLP
% CLP of CD45+
# of CLP
C
LMPP
% LMPP of CD45+
# of LMPP
D
CILP
%CILP
# of CILP
E
CHILP
%CHILP of CD45+
# of CHILP
F
ILCP
%ILCP of CD45+
# of ILCP
G
ILC2P
% ILC2P of CD45+
# of ILC2P
H
Gata3 MFI
ns
*
**
WT
Dreg1 -/-

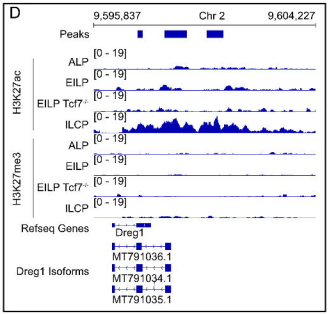
Figure 3



B

ID	Motif Match Score	q-Value
Tcf1	12.1089	0.036
Lef1	11.4158	0.031
Ets1	12.2131	0.039





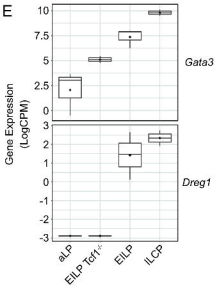
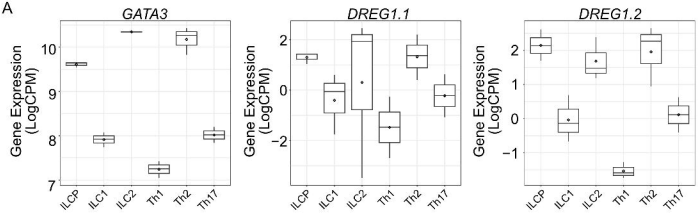
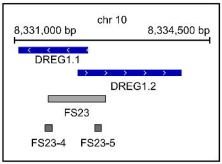


Figure 4

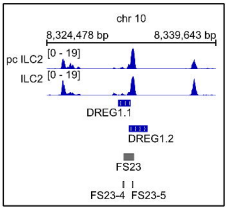
A


GATA3
DREG1.1
DREG1.2
Gene Expression (LogCPM)
ILCP
ILC1
ILC2
Th1
Th2
Th17

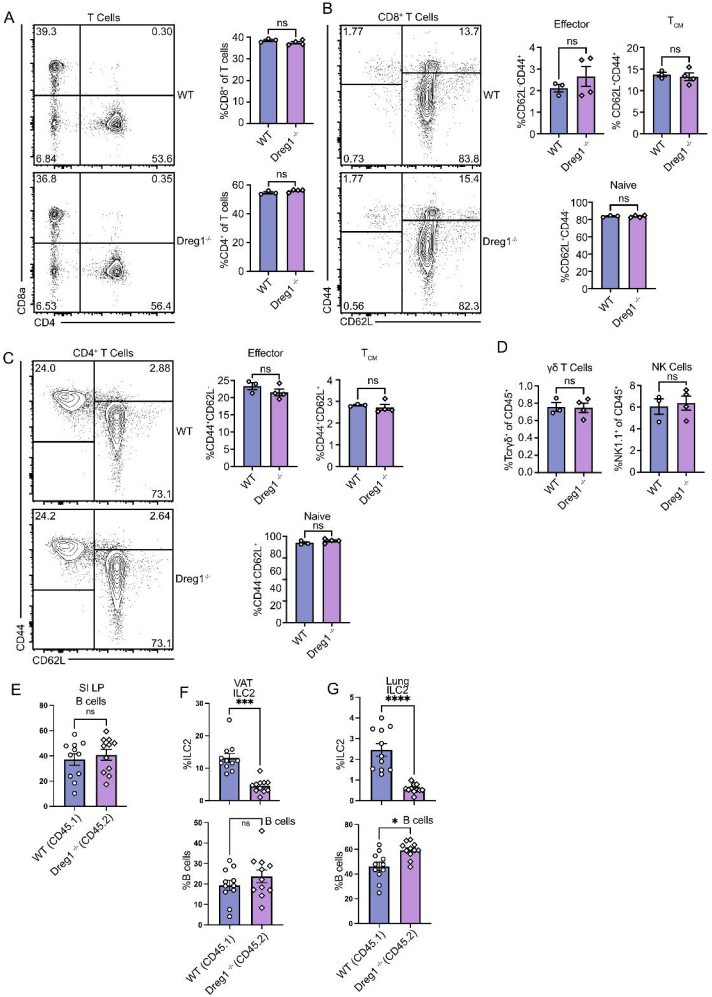
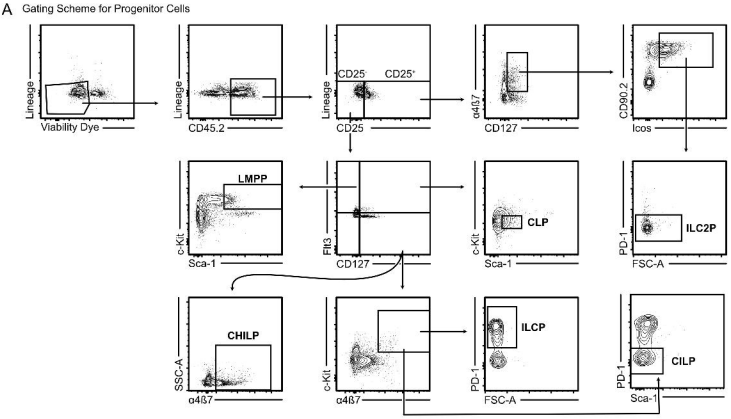
B


chr 10
8,331,000 bp
8,334,500 bp
DREG1.1
DREG1.2
FS23
FS23-4
FS23-5

C


chr 10
8,324,478 bp
8,339,643 bp
pc ILC2
[0 - 19]
ILC2
[0 - 19]
DREG1.1
DREG1.2
FS23
FS23-4
FS23-5

Supplmentary Figure 1



Supplementary Figure 2