Transcriptional and epigenetic regulation of innate-like T lymphocyte development
Mihalis Verykokakis\textsuperscript{1} and Barbara L Kee\textsuperscript{2}

Invariant Natural Killer T (iNKT) cells are a heterogeneous innate T cell population that recognizes lipid antigens. Despite the monospecific nature of their T cell receptor, iNKT cells differentiate into stable sublineages during thymic development, before foreign antigen encounter. How iNKT cell subsets acquire and maintain their functional programs is a central question in innate lymphocyte biology. Global transcriptional and epigenetic profiling of iNKT subsets has provided insights into the internal wiring of these subsets that defines their identity. Comparison of the iNKT transcriptional programs with those of other adaptive and innate lymphocyte lineages revealed common core regulatory circuits that may dictate effector functions. In this review, we summarize recent advances on the molecular mechanisms involved in iNKT cell development.

Addresses
\textsuperscript{1} Biomedical Sciences Research Center “Alexander Fleming”, 34 Fleming St, 16672 Vark, Greece
\textsuperscript{2} Department of Pathology and Committee on Immunology, The University of Chicago, Chicago, IL 60657, USA

Corresponding authors: Verykokakis, Mihalis (verykokakis@fleming.gr), Kee, Barbara L (bklee@bsd.uchicago.edu)

Introduction
Multiple evolutionarily conserved T cell populations with innate characteristics are present in mice and humans that differ in their antigen receptor expression and their functional properties [1,2]. In mice, invariant Natural Killer T cells (iNKT), which express an invariant V\textalpha{}14-J\textalpha{}18 T cell receptor alpha (TCR\textalpha{}) chain that can pair with a limited number of TCR\textbeta{} chains (V\textbeta{}2, V\textbeta{}7 and V\textbeta{}8.1–8.2), are the most abundant innate T cell population. These cells recognize glycolipid antigens presented by the non-classical, non-polymorphic Major Histocompatibility Complex (MHC) I family molecule CD1D. The majority of iNKT cells follow the same developmental trajectories in the thymus as conventional T cells; however, iNKT cells mature in the thymus and acquire a phenotype reminiscent of antigen-experienced T cells during development. The frequency and number of iNKT effector subsets varies between naïve mouse strains and human individuals and influences the fate of bystander immune cells, even at the steady state [3–5]. Due to their poised state, iNKT cells produce a vast amount of a variety of cytokines rapidly after activation, thereby influencing the functions of a number of innate and adaptive immune cells. iNKT cells have been implicated as positive or negative regulators of a wide range of diseased states, including microbial infection, cancer and autoimmunity and they are being targeted in immune therapies and vaccination strategies [6,7].

By contrast to adaptive T lymphocytes, which acquire their effector properties only after antigen and cytokine co-stimulation, innate T cell effector programs are ‘hard-wired’ during development, before foreign antigen exposure and are maintained throughout life [8]. Therefore, the preservation of these life-long effector states — despite the presence of an ever-changing microenvironment — is likely due to an imprinted genomic state, imposed and overseen by transcriptional and epigenetic networks. Recent studies that have shed light into the transcriptional and epigenetic wiring during iNKT cell development will be the focus of this review (Figure 1).

iNKT cell development
Early experiments based on cell fate-mapping and intrathymic cell transfer demonstrated that iNKT cells derive from the pool of CD4\textsuperscript{+}CD8\textsuperscript{+} (double positive, DP) cells in the thymus, through random T cell receptor alpha (T\textalpha{}) rearrangements that lead to the generation of the canonical V\textalpha{}14-J\textalpha{}18 TCR [9,10]. Following antigen receptor generation, iNKT cells are positively selected on the non-classical MHC molecule CD1D, which presents antigen on DP thymocytes, and is assisted by interactions between members of the Signaling Lymphocytic Activated Molecules (SLAM) family of receptors [11]. Although the exact nature of the selecting antigens is currently not clear, there is strong evidence that α-linked glycosylceramides represent one type of endogenous ligands for iNKT cells in mice [12]. This CD1D-mediated positive selection process results in the selective induction of the iNKT-signature transcription factor Promyelocytic Leukemia Zinc Finger (PLZF), which
iNKT cells differentiate into at least 3 stable lineages in the thymus (NKT1, NKT2 and NKT17), whereas there is evidence that NKT10 cells are also developing in the thymus before migration to the adipose tissue. Although the vast majority of NKT cells emerges from DP progenitors, a small NKT1 subpopulation with superior cytotoxic ability may derive from DN cells, bypassing the DP stage. The branching point between lineages is under investigation. It is possible that a common progenitor emerges after positive selection that expresses small amounts of a mixture of lineage-specific transcripts and cellular fate is resolved in subsequent stages. Alternatively, slight differences in TCR specificity during selection may direct the gene expression program of each lineage immediately as cells exit from positive selection. (b) The underlying chromatin state influences the transcriptional program of the iNKT lineage. The promoter of Zbtb16, although inactive, is in a bivalent state in DP cells, marked by both H3K4me3 (active transcription) and H3K27me3 (repressed transcription). Upon entry into the iNKT lineage, H3K27me3 is lifted by the demethylase UTX and Zbtb16 is expressed in an EGR2-dependent and ID-dependent manner. UTX can also interact with PLZF, to help activate a subset of PLZF target genes. Zbtb16 mRNA is regulated post-transcriptionally by the let-7 miRNA to promote the NKT1 lineage fate. In addition to histone methylation, a subset of iNKT lineage genes is controlled through DNA methylation by the TET proteins that catalyze the conversion of the
depends on the expression of the Early Growth Response protein-2 (EGR2) and the ID proteins [13,14].

Recently, an alternative developmental pathway for iNKT cells has been proposed that bypasses the DP stage and preferentially gives rise to T helper 1 (Th1)-like iNKT cells with strong cytotoxic potential [15*]. The authors used a novel mouse model of cell-fate mapping, based on the expression of Cre recombinase under the control of the Cd8 enhancer E8III (E8III-Cre) that is activated in DP cells, which in turn drove the expression of ROSA26-YFP. In this model, the authors identified a minor population of iNKT cells in the spleen and the thymus that was negative for YFP expression, suggesting that it did not derive from DP precursors. Conditional deletion of Rag2, using the E8III-Cre transgene resulted in the loss of the Vα14Jα18 transcripts at the DP stage, but not at the DN stage. Of note, 90% of the DN transcripts encoded in-frame, invariant Vα14Jα18 TCR products, suggesting that a subset of iNKT cells matured in the absence of recombination at the DP stage. However, it is still possible that some cells escape Rag2 deletion and develop normally into the iNKT cell lineage. Indeed, conventional CD4 and CD8 T cells are present in the thymus of E8III-Cre*Rag2Δ/Δ mice, albeit at lower numbers. Regardless, the iNKT cells that may develop through this alternative developmental pathway comprise just ~1% of the total iNKT cell pool and their biological relevance remains to be determined.

### iNKT cell subsets

Following positive selection, iNKT cells undergo a stepwise differentiation process that results in the generation of at least three distinct mature subsets in the thymus [16]. Early differentiation stages can be identified based on the expression of cell surface markers, including CD24 and CD44. CD24*CD44+ iNKT cells comprise a very rare population that corresponds to cells that have just undergone positive selection (stage 0, ST0). Stage 1 (ST1) cells downregulate CD24 expression and begin production of cytokines, particularly interleukin-4 (IL-4), followed by a subsequent upregulation of CD44 (Stage 2, ST2), a T cell activation marker. As iNKT cells develop, their differentiation programs diversify and can be distinguished by their signature cytokine and transcription factor expression pattern from the classification of conventional CD4 Th subsets [3,17,18]. NKT1 cells express TBET and produce predominantly interferon-γ (IFN-γ), whereas NKT2 cells express high levels of GATA3 and PLZF and secrete IL-4 and IL-13. A third subset expresses RORγt and IL-17 and is, thus, named NKT17. Although these three subsets represent stable iNKT sublineages [3], it is currently unclear at which point and how their effector programs unfold during development. A similar developmental pathway has been recently described for the Mucosal Associated Innate T (MAIT) cells, which are a distinct PLZF-expressing innate-like T lymphocyte subset that recognizes vitamin metabolites presented by the non-classical MHC molecule MR-1 [19**].

Two additional subsets have been recently identified in the periphery. During activation with the iNKT cell ligand α-Galactosylceramide (α-GalCer), a subset of splenic iNKT cells upregulates the T follicular helper (Thf) transcription factor BCL6 and provides help to B cells in an IL-21 dependent manner (NKTTh) [20,21]. In addition, the recently identified NKT10 subset shows a distinct gene expression program and is characterized by a naïve phenotype (low expression of PLZF, CD44 and NK1.1), high E4BP4 expression, and secretes the immunosuppressive cytokine IL-10 [22,23]. NKT10 cells are resident in the adipose tissue of mice and help maintain an immunoregulatory microenvironment by controlling the fate of anti-inflammatory macrophages and the number of regulatory T cells (Treg). Importantly, mice that lack iNKT cells are predisposed to high fat diet induced obesity and iNKT cell numbers are reduced in obese mice and humans [24]. This reduction of iNKT cells is associated with a decrease in Treg cells and conversion of macrophages into a pro-inflammatory M1 state, thus perpetuating a state of chronic inflammation.

Although it is still unclear whether NKTTh cells originate from the thymus, recent evidence indicates that the transcriptional program of NKT10 cells is established during thymic development [25*]. Sant’Angelo and colleagues examined how the TCRα–TCRβ pairing may influence iNKT cell development by generating mice with a Tyrosine substitution of Phenylalanine 108 (F108Y) of the TCRβ chain, thus partially disrupting the interaction of the TCRβ with the TCRα chain. F108 lies within the αβ heterodimeric interface and is not directly involved in antigen recognition, consistent with results showing that antigen recognition, ligand binding avidity and downstream TCR signaling are unaffected in mutant iNKT cells. Surprisingly, F108Y mice had decreased numbers of iNKT cells in the thymus, spleen and liver accompanied with an accumulation of CD44−NK1.1− cells in the thymus. However, adipose iNKT cells were significantly enriched in F108Y mice, suggesting that F108Y iNKT cells preferentially home to the adipose tissue. In addition, PLZF expression was lower in F108Y thymic iNKT cells compared to F108 WT cells and E4BP4 and PD-1, two unique NKT10 markers, were already increased in the thymus of F108Y mice. Although it is not clear how the TCRα–TCRβ pairing influences NKT10 cell development,

(Figure 1 Legend Continued) repressive 5′-methylated cytosine mark (5mC) to 5′-hydroxymethylcytosin (5hmC). UTX binding on the DNA, H3K4me3 and 5hmC are correlated with increased accessibility of genomic regions in the vicinity of iNKT lineage genes.

these results suggest that possible conformational changes of the iNKT TCR and/or slight alterations in TCR specificities may determine distinct effector fates. Consistent with that, transgenic mice with various iNKT TCR specificities showed altered iNKT subset differentiation [26]. Regardless of the mechanism, iNKT cells with a phenotype associated with adipose tissue resident NKT10 cells are already present in the thymus, adding to the current thymic iNKT heterogeneity.

Transcriptional control of iNKT cell development

Several transcription factors are important for iNKT cell development and their functions have been recently reviewed [27,28]. The innate T cell lineage-specific transcription factor PLZF is induced in iNKT cells after positive selection, and directs key features of their innate program [29,30]. Recent data provided some evidence on the mechanism of PLZF function in iNKT cells. PLZF regulates the nuclear localization of the E3 ubiquitin ligase CULLIN3, which in turn modifies several proteins involved in gene regulation and chromatin organization and are essential for iNKT cell development [31,32]. In addition, PLZF directly binds and controls key genes encoding for homing, (Sel, Cd44) cytokine (Il12rb1, Il18r1, Ifngr1, Il21r) and chemokine (Ccr4, Icos) receptors, and signature transcription factors (Gata3, Rorc, c-Maf, Runx3) [33**]. Importantly, PLZF repressed the transcription factor Bach2, which maintains the naïve T cell state by repressing Th effector gene expression. Therefore, PLZF acts at multiple levels to impose an activated phenotype in iNKT cells and potentially in other innate-like T cell lineages. Interestingly, ChIP-seq experiments revealed that, rather than binding to its own specific DNA-binding motif, PLZF binds to genomic regulatory regions through interactions with other transcription factors, indicating a context-dependent recruitment of PLZF to target genes. However, how PLZF is recruited to target regulatory elements and how the underlying chromatin landscape may influence PLZF binding is not clear yet.

Previous gene expression analysis demonstrated that iNKT cells share an extensive transcriptional program with NK, γδ T and activated CD8 T cells, indicating that a core transcriptional network operates within innate and activated lymphocytes [34]. However, in that study the bulk of mature iNKT cells was analyzed without considering the heterogeneity resulting from the distinct iNKT cells subsets. Indeed, recent studies that compared the transcriptional profile of NKT1/NKT2/ NKT17 separately with other lymphoid populations showed that only NKT1 cells are transcriptionally similar to NK, ILC1 and Th1 cells, corroborating previous results [35], whereas NKT2 and NKT17 cells are more closely related to their ILC and γδ T cell counterparts than the respective Th lineages [36**]. In addition, this and other studies [37**,38**] revealed that the iNKT subsets are characterized by highly diverse gene expression programs. Pairwise comparisons between thymic subsets showed that 300–900 genes were differentially expressed in mature subsets, indicating that the functional characteristics of iNKT sublineages rely on distinct transcriptional programs. A significant fraction of these genes encodes for chemokine receptors, which are implicated in cell migration and tissue homing, consistent with the fact that distinct iNKT subsets are preferentially localized in different tissues. Not surprisingly, the iNKT subsets differ in their expression of cytokines and cytokine receptors, confirming their distinct effector properties. In addition, expression of genes involved in TCR signaling varies among iNKT subsets [38**], possibly implying that TCR signal perception may be different for each iNKT subset. Alternatively, it is possible that, as iNKT cells exit from positive selection, TCR signaling may dictate iNKT sublineage fate. Although this remains to be proven, elimination of various molecules just downstream of TCR — leading to altered TCR signaling — favors NKT2 cell development at the expense of NKT1 cells [4,5,39–41], consistent with the hypothesis that the iNKT cell effector fate is driven by differences in TCR signaling.

iNKT cell progenitors show a unique gene expression profile compared to that of mature iNKT cells. This early progenitor program was significantly enriched in genes encoding for proteins involved in cell proliferation and metabolism, tissue homing and motility [36**,37**]. Single-cell RNA sequencing analysis revealed that there are cells within the NKT2 cell pool (sorted as Tetramer<sup>+</sup>TCRβ<sup>+</sup>CD24<sup>−</sup>NK1.1<sup>+</sup>CD27<sup>+</sup>CD4<sup>+</sup>), that show simultaneous expression of NKT1 signature genes (e.g. Tbx21, Ccr5) and cell cycle-related genes [37**]. These results confirm, at the global level, that the NKT2 population, as currently defined, is highly heterogeneous and contains both mature and transitioning cells.

Epigenetic regulation of iNKT cell development

Although several studies have addressed the transcriptional requirements for iNKT cell development, less information is available on the epigenetic and chromatin regulation of gene expression during this process. During iNKT cell positive selection, strong TCR signals induce transcription from the Zbb16 locus [13]. It has been demonstrated that the proximal regulatory region of Zbb16 is decorated with both repressive (trimethylation of Lysine 27 on histone 3, H3K27me3) and activated (trimethylation of Lysine 4 on histone 3, H3K4me3) chromatin marks in DP cells [42], indicating that Zbb16 is poised for expression. Upon iNKT positive selection, the repressive marks are lost, thus enabling activation of the Zbb16 promoter. Loss of the histone methyltransferase EZH2 (which mediates the repressive H3K27me3) in
DP cells led to an increase in total iNKT cells and in PLZF+ T cells with innate-like characteristics, independently of their TCR specificity, indicating that H3K27me3 links TCR specificity with PLZF expression. By contrast, sustained H3K27me3 severely impairs iNKT cell development, as shown in mice deficient for the histone demethylase UTX, by repressing iNKT lineage-specific genes including Tbx21 and Zbtb16 and the associated gene expression programs [43**].

iNKT cell subsets show distinct patterns of PLZF expression. Although PLZF is high in immature iNKT cells and NKT2 cells and intermediate in NKT17s, it is down-regulated to almost baseline expression in NKT1 cells. This dynamic expression pattern of PLZF is regulated post-transcriptionally by the microRNA (miRNA) let-7 [44*]. let-7 is upregulated during iNKT cell differentiation in an IL15-dependent manner, and targets Zbtb16 transcripts to inhibit PLZF expression, thus enabling differentiation towards the NKT1 sublineage. Therefore, PLZF expression is regulated both at the chromatin, transcriptional and post-transcriptional level to control iNKT cell lineage commitment and differentiation.

The epigenetic regulation of iNKT cell development through the miRNA pathway has long been recognized, as shown in Dicer-deficient mice [45,46], which completely lack iNKT cells. Consistent with that, specific miRNA molecules fine-tune several aspects of iNKT cell development, including agonist selection [47,48], metabolic fitness [49], and subset differentiation and maturation [50,51], through modulation of TCR, PTEN/PI3K and TGFβ signaling pathways.

Recent studies have shed light on the underlying enhancer landscape that directs chromatin accessibility of transcription factors, thus influencing the gene expression program of iNKT cells. Based on the deposition of histone 3 lysine 27 acetylation (H3K27ac) peaks, Winau and colleagues identified 396 super enhancers (SE) in bulk iNKT cells, around genes encoding for known regulators of the iNKT lineage, including Tbx21 and Zbtb16 [43**]. Using ATAC-seq experiments, they showed that SEs are characterized by an open chromatin environment. Importantly, elimination of UTX resulted in accumulation of repressive H3K27me3 marks accompanied by reduced H3K4me3, diminished transcription and loss of chromatin accessibility. In addition to histone modifications, DNA modifications such as cytosine hydroxymethylation (5hmC), regulated by the Ten-Eleven Translocation (TET) dioxygenases, correlate with increased chromatin accessibility around lineage-specific genes and enhanced transcription [52**]. Moreover, alterations in the 5hmC pattern, caused by simultaneous genetic deletion of Tet1 and Tet2, lead to a proliferative expansion of iNKT cells and a skewing of the iNKT cell differentiation program towards the NKT17 lineage at the expense of NKT1 cells [52**]. Although these data demonstrate for the first time how chromatin and DNA modifications affect chromatin accessibility and the gene expression program of iNKT cells, it is currently unclear how all these modifications interact with each other and what is the hierarchy of events. Furthermore, how the chromatin landscape differs among mature subsets and in developing iNKT cells is not well understood. Indeed, genome-wide data showed that more than 10,000 regions were differentially enriched in H3K27Ac between NKT1, NKT2 and NKT17 cells [37**], suggesting that the reported functional heterogeneity observed among iNKT cells has a chromatin basis that potentially determines their distinct transcriptional programs.

Conclusions

Although significant progress has been made regarding the identification of the individual transcription factors that control iNKT cell development, how these are coordinated with each other and what is the chromatin background that enables proper gene expression is currently unclear. Recent studies have started illuminating the nuclear events in iNKT cells that mediate the functional properties of mature iNKT cell subsets and revealed the gene expression programs that establish and maintain iNKT cell identity. The advance of techniques that allow global gene expression analysis at the single-cell level in combination with powerful genetic mouse models will contribute to the delineation of the developmental trajectories of iNKT cells and the corresponding gene regulatory networks.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

Related work from the authors lab was supported by the National Institute of Health R01 AI125360 to BLK, the European Commission H2020-MSCA- IF 655271 (MV) and the Stavros Niarchos Foundation Start-Up grant (MV).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

5. Verykokakis M, Boos MD, Bendelac A, Kee BL: SAP protein-dependent natural killer T-like cells regulate the development


Demonstrates the presence of DN-derived iNKT cells suggesting the existence of a potential alternative developmental pathway for iNKT cells that bypasses the DP stage.


Demonstrated that MAIT cells follow a similar developmental trajectory as iNKT cells.


Demonstrated for the first time that adipose tissue resident NKT10 may originate from the thymus.


Identified for the first time direct target of PLZF in iNKT cells.


See annotation to Ref. [38*]


See annotation to Ref. [38*]


These three studies analyzed the global gene expression profiles of thymic NKT1, NKT2 and NKT17 cells and their enhancer landscape and revealed that the three subsets are transcriptionally distinct.


Identified that the histone demethylase UTX controls the epigenetic program and chromatin accessibility in iNKT cells.


Identified that the miRNA let-7 downregulates the Zbtb16 mRNA post-transcriptionally to promote NKT1 differentiation


Demonstrated that DNA 5'-hydroxymethylation of cytosine regulates NKT subset diversification and chromatin accessibility.