Duel of the fates: the role of transcriptional circuits and noise in CD4+ cells
Daniel Hebenstreit1, Andrew Deonarine1, M Madan Babu and Sarah A Teichmann

CD4+ T cells play key roles in orchestrating adaptive immune responses, and are a popular model for mammalian cell differentiation. While immune regulation would seem to require exactly adjusted mRNA and protein expression levels of key factors, there is little evidence that this is strictly the case. Stochastic gene expression and plasticity of cell types contrast the apparent need for precision. Recent work has provided insight into the magnitude of molecular noise, as well as the relationship between noise, transcriptional circuits and epigenetic modifications in a variety of cell types. These processes and their interplay will also govern gene expression patterns in the different CD4+ cell types, and the determination of their cellular fates.

Address
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

Corresponding author: Teichmann, Sarah A (sat@mrc-imb.cam.ac.uk)
1 These authors contributed equally to this work.

Introduction
Cells change phenotype when they specialize or adapt to different environments. There are several well-known systems that illustrate this phenomenon. Some examples include the recently reported reprogramming of differentiated cells into induced pluripotent stem cells (iPS) [1] or the differentiation of CD4+ T cells [2]. These changes or ‘switches’ in cellular phenotype are influenced by changes in expression levels of key genes [3], and are subject to molecular fluctuations [4]. Such noise can be classified as either intrinsic or extrinsic [5–8]. Intrinsic noise is caused by gene-specific stochastic fluctuations in abundances of molecules that are due to transcription, translation, and related biochemical mechanisms, while extrinsic noise originates from variations in the state or concentration of other components in the cell [5]. As a result, abundances of mRNAs and proteins of identical genes can vary widely from cell to cell, constituting ‘noise,’ even within a homogenous cell population.

Noise propagation is strongly influenced by topologies of regulatory networks [9,10], such as the transcriptional circuits controlling cell fate. Examples include small subnetworks, such as feedforward loops (FFLs) [11] and feedback loops which can buffer or amplify noise, depending on the parameters that influence the kinetics of gene expression [12]. Recent theoretical studies have tried to explore the structural requirements of network controllability [13,14]. This question is important in the context of ‘master regulators,’ which are essentially small numbers or even single transcription factors (TFs) that are able to ‘reprogram’ a cell type [15]. Besides the basic TF-target gene interactions, other factors such as miRNA [16], and epigenetic modifications [17,18] are part of broader regulatory circuits and contribute to stabilization or switching of phenotypes.

In this review we will focus on CD4+ T cells, which is a good biological system for illustrating the relationships between transcriptional circuits, expression level noise, and epigenetics. CD4+ T cells form an important branch of the adaptive immune system and function mainly by secreting certain cytokines that regulate the behavior of other cells, thus orchestrating the immune response [2]. Six major CD4+ T cell subtypes have been well characterized to date: the naïve, T-helper 1 (Th1), Th2 and Th17 cells, and induced and natural regulatory T cells (iTreg and nTreg). Most of the (‘mature’-) subtypes are derived from naïve CD4+ T cells, which enter a proliferation and differentiation process upon antigenic activation in the presence of cytokines (Figure 1a).

The phenotypic switches that CD4+ cells undergo upon differentiation are linked to changes in epigenetic modifications [19]. These switches are largely driven by TFs, and illustrate the importance of master regulators, such as Gata3 [20,21]. Recent studies on Th2 cells suggest that phenotypic switches are mirrored by transcriptomic switches [22]. Protein expression levels of cytokines are usually subject to large cell-to-cell variability, highlighting the importance of noise in protein and mRNA expression.

To understand the interplay between these factors, several questions need to be addressed: how do transcriptional circuits mediate phenotypic switching? What are
The roles of epigenetic modifications in this process? What role does noise play: is it detrimental, tolerated, or even beneficial? On the basis of the CD4+ T cell system, we explore possible answers.

Phenotypic switches in CD4+ T cells are mediated by transcriptional circuits

The difficulty of precisely defining cell types is well illustrated by the distinct yet closely related subtypes
of CD4+ T cells (Figure 1a). Though each of the six cell types illustrated in Figure 1a is associated with a particular master regulator and defined by a specific cytokine profile, there are many reports of intermediate cell types as well as cell-type switching after differentiation (i.e. plasticity), and reprogramming. For instance, mature Th2 cells can be induced to give rise to a mixed Th1/Th2 phenotype. Th17 cells can be transformed into Th1 cells, or iTreg cells can be differentiated towards Th17 cells [23] (Figure 1a).

The cytokines that are secreted by mature CD4+ T cells have crucial, often opposing roles in regulating the immune response. Thus, one would expect a tight regulation in maintaining lineage identity. The plasticity in the differentiation process as a whole is due to the complexity of the transcriptional regulatory circuits, which generally include many more components than just a single master regulator.

In Figure 2b, a small transcriptional circuit involving the master regulator Gata3 is illustrated. The circuit consists of overlapping FFLs which facilitate the differentiation of a CD4+ naïve cell to the Th2 type [24], and represents a small subnetwork in a much larger, more complex regulatory network involved in phenotypic switching and in stabilizing the cellular state.

Differentiation of CD4+ cells is strongly linked to cellular proliferation as many studies show a percentage increase of cytokine-secreting T cells in successive cell generations [25,26]. Notably, the percentage of differentiated cells added after each division is small and thus suggests the involvement of a stochastic mechanism. Where does this randomness come from and how can we reconcile it with the precision of immune responses?

Sources of stochasticity and their relation to circuit topology

T cell differentiation is regulated by TFs. It has become clear in recent years that transcriptional regulatory networks are subject to intrinsic noise in the expression of their components [27]. This has been shown by techniques that focus on single cells rather than population averages, such as single molecule RNA-FISH, which measures absolute numbers of mRNAs per cell [8].

One of the surprising findings from these assays was that the distributions of mRNA numbers per cell are usually non-Poissonian, meaning that the variances are higher than the mean levels of transcripts. This gives some mechanistic insight into transcription by RNA polymerase II (RNAPII) by hinting at a burst-like production of transcripts, where a gene is transcribed multiple times in short periods of time [28]. Our studies on Th2 cells [29] confirm such highly skewed, ‘noisy’ distributions of transcript numbers per cell for five genes classically associated with CD4+ cell biology, including Gata3 and CD4 itself (Figure 1b). Skewed distributions are observed for protein expression levels too (Figure 1b), suggesting that either the mRNA expression noise is not buffered completely or downstream processes serve as additional sources of noise. An interesting possibility is that cell-to-cell variation in the expression levels of genes such as Gata3 may simply reflect the heterogeneity in terms of their recent signaling experience. Thus some genes could be specifically programmed to be sensitive indicators of different levels of signaling molecules in each cell. Another contributor to the wide range of expression levels between individuals in a cell population could be asymmetric cell division. Polarized segregation of proteins has been observed upon antigen-induced T cell division, where activation and differentiation induced factors were preferentially found in the daughter cell close to the immunological synapse [30,31]. This included Tbx21 and may also affect other TFs and master regulators.

As a consequence of these sources of noise, there may not be a simple, discrete transcriptional ‘switch.’ This may vary from cell to cell depending on the number of copies of a set of signaling molecules, including TFs. As a result, statistical parameters such as the average burst size or the average waiting time may be altered upon transcriptional activation of RNAPII. Importantly, burst-like transcription demonstrates that transcription is an intrinsically noisy process per se, and in individual cells this phenomenon can result in large deviations in transcript numbers from the population average. In principle, random fluctuations give rise to different transcription rates in different cells over time, even if the cells started under identical conditions. For CD4+ cells, the large variance in the distributions shown in Figure 1b suggests that this may be relevant to phenotypic switching, as it means that a certain fraction of cells will have high numbers of mRNAs from genes that are lowly expressed, and vice versa.

The presence of large cell-to-cell variation may indicate that noise is somewhat inevitable. This might be due to the fact that noise suppression is a biologically expensive process: a recent study showed that suppression only occurs with the fourth root of the number of control factors in a transcriptional circuit [32]. However, many investigations in other cell types or systems have shown that stochastic transcriptional switches can even be advantageous over deterministic ones. For instance, stochastically switching phenotypes or noisy transcriptional circuits provide efficient mechanisms for adapting to fluctuating or unpredictable environments [33–35]. It has also been shown that FFL circuits that incorporate miRNA can dampen noise [36], while certain FFL architectures have specific noise profiles [37] and may be selected for these profiles, thereby facilitating the sampling of multiple phenotypes [38].

While it is uncertain whether such phenomena play a role in T cell biology, it is clear that the transcriptional
Switches in cellular fate as reflected in the transcriptome, regulatory network, and epigenetic state of the cell, each of which is influenced by noise (indicated by the Fano factor $\sigma^2/\mu$). (a) A density plot illustrating highly and lowly expressed transcripts (HE and LE, respectively), transcripts with no detectable expression (NE) and transcripts with intermediate levels (numbers of genes are in brackets). (b) Transcriptional feedforward circuits involved in the regulation of key cytokines that facilitate the ‘switch’ to Th2 from naïve cell [24]. Three feedforward loops are illustrated: Dec2-Gata3-II4, Dec2-Gata3-II5, and Dec2-Jun2-II4. (c) Histone modifications also play a role in mediating phenotypic switches, with activating histone medications (e.g. H3K4me3) being associated with an open chromatin state and HE transcripts, while repressive modifications (e.g. H3K27me3) are associated with LE transcripts. Bivalent chromatin is associated with transcripts that lie between HE and LE in expression level (intermediate).
regulatory networks of CD4+ differentiation are subject to intrinsic noise and the fundamental limit of noise suppression mentioned above.

**The role of epigenetic switches in CD4+ cells**

One mechanism that could have evolved in order to permit a cost-efficient control of such transcriptional noise could be the system of epigenetic modifications. Genes that become differentially expressed during Th differentiation are accompanied by changes in the chromatin landscape [19]. For instance, the chromatin at subtype-specific loci becomes relaxed, as illustrated by DNaseI hypersensitive regions. Other subtype-specific genes are marked with repressive histone modifications in cell types not expressing them [39]. Mariani et al. showed that in Th2 cells, stochastic IL4 expression patterns can be explained computationally using a two-state model, where promoter activation/inactivation states are explained in terms of chromatin dynamics (opening and closing due to activating or suppressing histone modifications). Here, different populations of cells are created stochastically over generations due to chromatin dynamics [6,40]. This work provides support for a link between noise in gene expression levels and histone modification levels.

Similarly, our recent work relates gene expression with histone modification levels, confirming such a two-state model. In this study, we identified groups of lowly expressed (LE) and highly expressed (HE) genes in Th2 cells based on the global distribution of gene expression levels measured by RNA-sequencing (RNA-seq) [29**], which is illustrated in Figure 2a. Here, we annotated another subset of genes (derived from the LE group) called not expressed (NE). NE genes are defined as genes with no reads mapping to that particular transcript, which is dependent on sequencing depth. This concept is similar to the three expression classes of mRNA previously identified by Hastie and Bishop [41]. While the HE genes featured the activating histone mark H3K9/14 acetylation at their promoters, the LE genes did not. The existence of the lowly expressed, tissue-specific LE group of genes suggests that the cell robustly tolerates a certain amount of background expression, representing a type of ‘transcriptional noise.’ This is often referred to as leaky expression and might be due to a tradeoff between the cost of (nearly) completely repressed transcription and the adverse consequences of stochastic expression [42,43]. The leaky expression reflects the inactive promoter state in the two-state model, while HE transcripts are more representative of the active promoter state.

In Figure 3, we analyzed the chromosomal organization of NE, LE, and HE genes. As illustrated in Figure 3a, the physical distances between adjacent HE and LE genes were found to be less than between HE and NE genes. This suggests that expression in LE genes may be due to their proximity to HE genes. Similarly, as depicted in Figure 3b, both HE and NE genes tend to form homogeneous multigene clusters more often than expected by chance. In contrast, LE genes usually occur alone. These findings support a genomic organization illustrated in Figure 3c, with LE genes close to clusters of HE genes, and clusters of NE genes further away from HE gene clusters. In this model, expression from neighboring HE genes leads to leaky expression of neighboring LE genes via some form of chromatin state spreading.

An interesting aspect of histone modifications is the existence of ‘bivalent’ genes, which feature both activating and repressive marks. In mammalian cells, there is increasing evidence from sequential ChIP and other experiments that these marks are present simultaneously at the same allele [44–48]. In *Drosophila* and *Xenopus*, it is unclear whether this is the case [49–51]. In mouse T-helper cells, Wei et al. observed such bivalent marks at master regulator loci [22]. The authors argued that such bivalent marks could poison them for expression. This could explain the plasticity of these cell types as well as why perfectly ‘pure’ cultures of single CD4+ subtypes are difficult to obtain. Using CD4+ microarray expression and H3K4me3/H3K27me3 histone modification data (analyzed with EpiChip [52]) from Wei et al., we found that genes with bivalent marks have a higher probability of differential expression upon maturation than genes with the H3K27me3 mark or no modification (see Figure 3d). They are also enriched in TFs relative to the genomic background. Whether these bivalent genes also have a specific noise profile is yet to be elucidated.

**The functional relevance of noise in transcriptional regulatory networks**

The question remains concerning how much of the observed stochasticity in CD4+ gene expression is tolerated and how much is beneficial. With regard to the beneficial aspects of stochastic expression, Guo et al. argued that probabilistic expression of the Th2 cytokine IL4 by a fraction of cells is a more flexible mechanism of immune regulation than fine-tuning of the exact IL4 amounts secreted by a homogenous cell population [40*,53].

Another recently published study on CD8+ cells, closely related to CD4+ cells, demonstrates how a certain amount of ‘controlled’ randomness in protein expression generates useful variation in the antigen responsiveness [54*] of the cell. This system is based on a combination of positive and negative feedback loops and demonstrates that the architecture of the regulatory circuits can have a profound impact on the existence of stable states and/or extent of noise. In line with this, a recent analytical study showed that feedback systems provided by cytokine secretion and stimulation, including downstream transcriptional interactions, are sufficient for yielding mathematically stable states that resemble cell types among CD4+ cells. The
The role of highly and lowly expressed transcripts. (a) The physical distances (defined as the shortest genomic distance between genes in nucleotides) between adjacent pairs of highly expressed (HE) and not expressed (NE)/lowly expressed (LE) genes in mouse Th2 cells. RNA-seq derived gene expression levels from Hebenstreit et al. [29**] were classified as LE or HE; Tophat [63] was also used for junction mapping. In general, LE genes are found closer to HE genes than NE genes. The difference between the LE–HE versus NE–HE distributions is statistically significant (Wilcoxon rank sum test, \( P < 10^{-4} \)). This observation may support the concept of leaky expression, or the ripple effect [42,43]. (b) The percentage of all NE, LE, and HE genes (with intermediate genes between LE and HE being classified as HE) in the genome that occur in a particular cluster size, compared to the cluster sizes obtained by averaging values across 10,000 randomized sequences of NE, LE, and HE genes. HE and NE genes are significantly more likely to form larger clusters than expected by chance (\( P < 10^{-4} \) estimated from simulation). (c) A model describing the organization of NE, LE, and HE genes according to their distances and clustering. (d) Data from Wei et al. [22*] were normalized using MAS5 [64], and differential expression between naïve and Th1, Th2, Th17, and iTreg was calculated using Limma [65] (an adjusted \( P \)-value < 0.05 was used to determine significance). More than half of the genes with bivalent histone marks are differentially expressed upon CD4+ cell maturation. As well, genes with bivalent modifications were enriched in transcription factors (annotated by DBD [66]) relative to the whole mouse genome (Fisher’s exact test, \( P < 10^{-4} \)).

The model also allows for plasticity and gives rise to intermediate cell types [55].

Experimental studies have indeed identified various circuits involved in CD4+ cell differentiation, which probably have important roles in regulating transcriptional and thus phenotypic noise. Djuretic et al. noted that in Th1 cells, the TF Tbx21 induces another TF Runx3, and together these molecules repress Il4 by binding to its silencer [56] and activate Ifng by binding.
to its promoter. Hence, Tbx21 and Runx3 form FFLs with Il4 and Ifng. In Treg cells, it has been shown by Bruno et al. that the TF Runx3 can promote the expression of the master regulator Foxp3 [57], and both of these TFs regulate target cytokines such as Il2 thereby forming a coherent FFL.

FFL circuits also play an important role in phenotype determination in biological systems other than CD4+ cells. Foxp3 is known to suppress the oncogene Satb1. Additionally, it was recently found that Foxp3 also forms coherent FFLs involving miRNAs that bind to the 3′-UTR of Satb1 (miR-7 and miR-155) [58] in breast tissue. Satb1 also plays an important role in CD4+ cell biology, particularly cytokine regulation. As noted previously, the Th2-specific cytokines Il4 and Il15 are regulated by FFL circuits (see Figure 2) [24]. At the epigenetic level, Satb1 remodels chromatin to create a 200 kb transcriptionally active region associated with Stat6 and Gata3. These TFs physically bridge this region with the Th2 ‘cytokine locus,’ which contains the genes Il4, Il5, and Il13. Cai et al. showed that this Satb1 mediated remodeling, in addition to the presence of Stat6 and Gata3, is essential to cytokine induction [59].

Finally, the regulation of cytokine receptors plays an important role in the circuits that stabilize Th lineages. This is supported by the aforementioned modeling approach [55]. A well-known biological example includes the IL-4 receptor α-chain (Il4ra), which is upregulated by IL-4 via Stat6 [60,61] and thus constitutes a Th2-reinforcing positive feedback loop.

Conclusions
Studies on CD4+ cells and other systems have highlighted the importance of transcriptional circuits and other mechanisms in regulating noise and mediating phenotypic switches. The strongly skewed mRNA distributions suggest that the sharpness of lineage identities is subject to fundamental limitations. Some examples indicate that there can be benefits from tolerating a certain amount of stochasticity, but in general the cell employs multiple mechanisms to control noise. These include the topologies of transcriptional networks and circuits, and epigenetic mechanisms, as exemplified by the discovery of the LE/HE transcript expression classes. Future studies will provide insight into the role of transcriptional circuits and epigenetics in broader regulatory networks, how these networks affect noise, and the probabilistic nature of phenotypic switches and lineage stabilities. This may eventually lead to more accurate, probabilistic definitions of these switches and stabilities, paving the way for new pharmaceutical strategies [62], and improving how we confront the ‘dueling fates’ of pathological versus normal cellular phenotypes.

Acknowledgements
We thank Guilhem Chalancon, Jesper Nissen and Valentina Proserpio for helpful suggestions concerning this manuscript. The authors acknowledge the Lister Institute (UK), the Medical Research Council (UK) grant reference nos. U105161047 and U105185889, and the Clinic Investigator Program at the University of British Columbia (Canada) for their support.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


29. Hebenstreit D, Fang M, Gu M, Chaoensasawat V, van Oudenaarden A, Teichmann SA: RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol Syst Biol* 2011, 7:437. This article is the first to contain RNA-seq and single molecule RNA-FISH data for mouse Th2 cells. It describes two classes of mRNA expression levels which are distinguished by histone marks at the promoters of the genes.


54. Feinerman O, Veiga J, Dorfman JR, Germain RN, Altmann-Bonnet G: Variability and robustness in T cell activation from
This article shows how ‘controlled stochasticity’ can benefit the cell by generating useful variation during T cell activation.


