Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity

Adam C. Wilkinson[1], Hiromitsu Nakauchi[1,2], Berthold Göttgens[3*]

[1]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, 265 Campus Drive, Stanford, CA 94305, USA

[2]Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[3]Department of Haematology, Cambridge Institute for Medical Research and Wellcome Trust and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 0XY UK

[*]Lead contact and corresponding author. Correspondence: bg200@cam.ac.uk

Abstract

Transcription factor (TF) networks are a key determinant of cell fate decisions in mammalian development and adult tissue homeostasis, and are frequently corrupted in disease. However, our inability to experimentally resolve and interrogate the complexity of mammalian TF networks has hampered the progress in this field. Recent technological advances, in particular large-scale genome-wide approaches, single-cell methodologies, live-cell imaging, and genome editing, are emerging as important technologies in TF network biology. Several recent studies even suggest a need to re-evaluate established models of mammalian TF networks. Here, we provide a brief overview of current and emerging methods to define mammalian TF networks. We also discuss how these emerging technologies facilitate new ways to interrogate complex TF networks, consider the current open questions in the field, and potential future directions and biomedical applications.

Introduction

During mammalian development, hundreds of unique cell types are specified in a complex spatio-temporal patterning process. In adults, stem and progenitor cell populations replenish mature cell types to maintain tissue homeostasis throughout life. Concerted gene expression programs are responsible for these fundamental biological processes and the underlying cell fate decisions. Transcription represents a major control point in gene expression (**Figure 1A**) and occurs within the context of chromatin. Precise spatial and temporal expression of combinations of a limited number of genes (~20,000 in humans) appears to be responsible for the intricate cellular processes of developmental specification and adult tissue homeostasis.

Sequence-specific transcription factors (TFs) are a large class of DNA binding protein that play central roles in regulating gene transcription, and account for almost 7% of genes (~1,400) in the human genome (Vaquerizas et al., 2009). TFs regulate gene promoter activity, but often act via interactions with other genomic locations that can be distant in primary DNA sequence. These are broadly defined as gene regulatory regions (Kellis et al., 2014), with an important subclass of positive regulatory regions being termed enhancers. Enhancers are composed of TF binding sites (TFBSs) or DNA motifs, which are are commonly short (4-12 nucleotides) (Jolma et al., 2013). Such motifs therefore frequently occur by chance in mammalian genomes and individual TF-DNA interactions can be weak. TF-DNA interactions must compete with histone-DNA interactions for stable and productive binding. Cooperativity in TF binding is therefore common, such as through protein-protein interactions with other TFs, co-activators, and/or co-repressors (Vaquerizas et al., 2009).

TFs can be thought of as "readers" of enhancers, with the combination (and spacing) of encoded TFBSs defining combinatorial binding capacity and stability. TF binding may directly activate or repress an enhancer and/or gene promoter, through recruitment of co-activators or co-repressors, or may act indirectly to influence gene expression such as through histone displacement (**Figure 1B**). The multi-protein complex Mediator is an important enhancer co-activator, which is thought to coordinate enhancer-promoter interactions and stimulate transcription (Malik and Roeder, 2010). TFs may also recruit other co-activators, such as histone methyltransferases, histone acetyltransferases, and chromatin-modifying complexes (Kouzarides, 2007). By contrast, enhancers and genes become repressed through TF recruitment of co-repressors such as histone demethylases (Whyte et al., 2012), histone deacetylases (HDACS), and polycomb complexes (Reynolds et al., 2013).

TFs have the ability to directly regulate their own expression through binding to enhancer(s) that control their own gene transcription. This can be thought of as a simple molecular circuit, a feedback loop. By understanding the concept that a TF can regulate its own expression, and expression of other TFs, it is possible to envisage the resulting TF circuits and networks that may be active within mammalian cells (Davidson, 2010). TF proteins, their genes and enhancers can be considered as the building blocks or constituents of a complex TF network (Alon, 2007). However, such a TF network is commonly not active in its entirety, but instead exists in various network states, comprising of active TF sub-networks. Of course, TFs not only regulate the transcription of TF genes, but also of genes involved in cellular structure/function. Therefore, the TF network state determines the global transcriptional program and the cell-type specific gene expression patterns that define cell identity and function.

Over the last thirty years, numerous experimental approaches have been used to define mammalian TF networks (**Box 1**). However, the inherent biological complexity (**Box 2**) has hampered these efforts. Recent technological innovations are now helping to define TF networks at unprecedented detail and accurately interrogate network logic and function, particularly in stem cell and cancer systems (**Box 3**). Several recent reports are even questioning established TF network models and suggest that some current paradigms may need updating. This Perspective aims to provide non-expert biologists with an overview of exciting recent developments in mammalian TF network biology, alongside a discussion of the open questions, the field's future directions, and its potential applications in biomedicine.

Recent insights from emerging technologies

New technologies remain major drivers in advancing our understanding of TF network biology. Most recently, single cell transcriptomics, live-cell imaging, and CRISPR/Cas9 genome editing have been applied to the field. These technologies are providing new angles from which to approach TF network biology, and alongside ongoing large-scale annotation approaches, including epigenome annotation, are helping to shed new light on mammalian TF networks.

Large-scale and high-throughput annotation

While TF ChIP-seq experiments have been possible for several years (Johnson et al., 2007) (**Box 1**), the reduction in next generation sequencing (NGS) costs are now allowing large-scale ChIP-seq studies. From assessing the binding of multiple TFs within the same and different cell types, it is possible to build a comprehensive annotation of TF network interactions. Several large-scale TF ChIP-seq studies have been undertaken, and have identified highly interconnected TF networks in several cell types (Tsankov et al., 2015, Wilson et al., 2016, Goode et al., 2016). Just as sequencing

of mammalian genomes has provided a blueprint to study mammalian genomics, these studies provide a central resource to investigate TF binding events within multiple cell types. By undertaking ChIP-seq experiments in similar cell types from different mammals, insights into evolutionary conservation of TF binding events and enhancer function are also being gained (Boyle et al., 2014, Cheng et al., 2014, Villar et al., 2015).

Recently large-scale efforts, including those by the Human Epigenome, ENCODE, and BLUEPRINT consortiums, have provided unprecedented resolution of the epigenetic state and conformation of chromatin in numerous cell types. For example, recent large-scale Hi-C analysis of almost 40 human cell types have helped to define topologically-associated domains (TADs) and assign enhancer-promoter interactions (both constitutive and cell-type specific) (Schmitt et al., 2016, Javierre et al., 2016). Such approaches are providing an important and data-rich annotation of mammalian epigenomes that is necessary for comprehensive TF network assignment. Interestingly, single cell DNase-seq, ATAC-seq, Hi-C and ChIP-seq protocols have also recently been published (Jin et al., 2015, Buenrostro et al., 2015, Cusanovich et al., 2015, Nagano et al., 2013, Rotem et al., 2015), suggesting single cell level chromatin structure and accessibility can also be employed to investigate TF network states in single cells.

Large-scale approaches have also been applied to understand cooperative TF binding. For example, Consecutive Affinity-Purification Systematic Evolution of Ligands by Exponential Enrichment (CAP-SELEX) (Jolma et al., 2015) has been developed to provide high-throughput TF pair binding site characterization, which have yielded important insights into cooperative TF binding events. Additionally, Fuxman Bass and co-workers recently provided a yeast one-hybrid based method to quantify human TF binding to enhancers, and probe the effects of genetic variation on TF

interactions (Fuxman Bass et al., 2015). Such methods afford high-throughput analysis from which general principles of mammalian TF cooperativity may be extracted.

High-throughput genome engineering

While CRISPR/Cas9 technologies can be used to make single genetic mutations, these methods are amenable to high-throughput studies, which allows for genome-wide coverage (Parnas et al., 2015) or saturation of a single genomic region (Canver et al., 2015). Several CRISPR gRNA libraries have recently been published (Sanjana et al., 2014, Tzelepis et al., 2016, Horlbeck et al., 2016), for both genetic deletion and transcriptional activation/repression. As with any screening method, it will be important to develop appropriate readouts and/or reporters for these assays. However, such tools are poised to provide significant insights into enhancer regulation and TF network interactions within mammalian cell types.

Canver et al. (2015) provided an elegant demonstration of the application of CRISPR/Cas9 for saturating mutagenesis of a single enhancer of *BCL11A*, allowing functional element mapping over a 12kb enhancer region (Canver et al., 2015). Importantly, such an approach allows for mutational analysis of endogenous enhancers, rather than the traditional reliance on heterologous enhancer reporter assays. Large-scale application of these CRISPR/Cas9 methods is likely to provide important fundamental insights into TF network architecture and principles.

Single cell transcriptomics

Fluorescent-Activated Cell Sorting (FACS) has long been used to purify single cells (Osawa et al., 1996), and has highlighted functional variability in highly purified cell populations (Yamamoto et al., 2013). In combination with microfluidics technologies, FACS has recently afforded single cell

gene expression analysis. Multiplexed qPCR initially allowed expression of 10-100 genes to be quantified in 100s of FACS-isolated single cells (Sanchez-Freire et al., 2012, White et al., 2011). Such methods have been particularly useful to investigate TF networks during embryogenesis and in adult stem cell populations, where limited cell numbers have often prevented population level analysis (Moignard et al., 2013, Moignard et al., 2015, Wilson et al., 2015). However, given the limited number of genes that could be simultaneously quantified by such methods, TF networks could not be comprehensively studied.

The advent of single-cell whole transcriptome RNA-seq methodologies (Tang et al., 2010, Picelli et al., 2013, Macosko et al., 2015) has provided important new opportunities. Several recent reports have demonstrated the potential of single cell RNA-seq to provide new resolution of TF network architecture (Kolodziejczyk et al., 2015, Olsson et al., 2016, Scialdone et al., 2016). It is of course important to remember the experimental caveats. First, mRNA levels do not always correspond to TF protein level (or activity). Second, single cell RNA-seq alone cannot distinguish indirect vs. direct TF interactions, although here, its combination with genetic deletion of specific TFs has yielded important resolution (Olsson et al., 2016, Scialdone et al., 2016). Third, current single cell RNA-seq approaches generally have lower sequencing coverage than bulk cell analyses, which may influence the transcript detection and/or observed intercellular heterogeneity.

For example, single cell RNA-seq has recently been used to re-evaluate the role of the TF Tal1 (also known as Scl) in mesoderm specification into hematopoietic and cardiac fates (Scialdone et al., 2016). Tal1 has been thought to autonomously activate (and stabilize) a hematopoietic TF network state while actively repressing a cardiac state, as its deletion blocks developmental blood cell formation and induces cardiomyocyte formation (Van Handel et al., 2012, Ismailoglu et al., 2008).

While single cell RNA-seq analysis confirmed that loss of Tal1 inhibited activation of blood-associated TFs, it failed to identity a corresponding immediate upregulation of a cardiac transcriptional program. These results suggest aberrant cardiac formation is likely a slower, and perhaps secondary consequence of *Tal1* deletion, rather than a direct lineage fate switch.

Recently, CRISPR/Cas9 screening has been combined with single cell RNA-seq in an approach that promises to provide detailed resolution of TF network circuitry in single cells (Dixit et al., 2016, Jaitin et al., 2016). Given that TF network states are heterogeneous at single cell level, the approaches by Dixit et al. (2016) and Jaitin et al. (2016) promise to provide a data-rich method to analyze CRISPR screening. Importantly, the approaches provide high-throughput and quantitative analysis of the direct transcriptomic consequences of genetic mutations. Through such methods, it should be possible to derive a comprehensive TF network map, and through profiling many single cells, infer fundamental principles of network state dynamics. However, to do so we will need to develop new bioinformatics methods to analyze and integrate these large and multidimensional datasets.

Live-cell imaging

While single cell transcriptomics determines the expression of many TFs, these technologies provide only a snapshot of gene expression. By lysing the cell for such gene expression studies, its future fate and potential cannot be simultaneously assessed (Hoppe et al., 2014). Live-cell imaging provides a powerful method to study the dynamics of TFs and allow TF network states to be correlated with cell fate decisions (within future generations of the cell). By tracking fluorescent reporters linked to TF expression (often directly fused to the TF of interest), live-cell imaging has provided important insights into the relationship between TF expression and the cell cycle (Kueh et

al., 2013), TF network plasticity (Filipczyk et al., 2015), TF antagonism (Hoppe et al., 2016), and extracellular signaling interactions (Kueh et al., 2016).

An early application of live-cell imaging was provided by Kueh et al. (2013), who focused on the transcription factor PU.1 and its interaction with the cell cycle. PU.1 is known to play a dose-dependent and context-dependent role in driving proliferation vs. terminal differentiation within hematopoietic cells (Mak et al., 2011). By tracking PU.1 expression over multiple cell divisions during differentiation into lymphoid and myeloid cell commitment, Kueh and colleagues were able to demonstrate that cell cycle kinetics could directly influence accumulation or loss of PU.1 over several generations, and thereby alter cell fate decisions. By regulating cell cycle progression, PU.1 could itself influence its own accumulation, outside of its regulation of positive auto-feedback at the transcriptional level (Kueh et al., 2013). One caveat of this study was that PU.1 levels were indirectly measured using a PU.1-IRES-GFP reporter, rather than a directly fused PU.1-fluorescent reporter. However, directly fused TF-fluorescent reporters hold their own caveats, such as potentially altering TF function and dynamics (TF-DNA or TF-TF interactions), as well as protein half-life.

More recently, quantitative live-cell imaging has been further used to investigate PU.1 within the context of PU.1-Gata1 antagonism in the erythroid vs. myeloid cell fate decision during hematopoietic stem cell (HSC) differentiation (Hoppe et al., 2016). Bulk cell analyses have implicated a direct cross-antagonism of Gata1 and PU.1 proteins, and this is widely used as an example of mammalian TF circuit interactions (Burda et al., 2010). However, live-cell imaging of Gata1-mCherry and PU.1-eYFP expression during the differentiation of individual HSCs questioned these long-held assumptions as few cells co-express Gata1 and PU.1 simultaneously, a requirement

for cross-antagonism. Instead of being responsible for the cell fate decision, upregulation of Gata1 or PU.1 appears to only occur following cell fate decisions, and therefore more likely acts to reinforce a pre-determined cell fate decision. While it is currently unclear what is upstream, and thereby responsible for the erythroid vs. myeloid fate decision, these data question the simplistic TF network models for mammalian cell fate decisions.

An alternative to tagging TFs with fluorescent proteins is the use of HaloTag ligand-based technology to tag TFs with organic dyes, which even affords subcellular TF localization in live cells using light-sheet microscopy. For example, Liu et al (2014) used this powerful approach to track Sox2-DNA binding in human ESCs and 3D diffusion within the nucleus (Liu et al., 2014). However, it is worth noting that live cell imaging methods are currently limited in the number of TFs that can be simultaneous detected (due to fluorescent spectral overlap). However, it is clear such powerful approaches are providing important quantitative insights into TF network biology.

Open questions and future directions

As described above, recent technical advances are helping to drive forward our characterization of mammalian TF regulatory networks. However, many of these technologies are yet to reach their full potential. Several key questions remain open. We hope that these technologies and others will be able to provide answers in the future.

Moving towards protein level quantification

Cell fate decisions are made at the single cell level. Important biological understanding is therefore being yielded from the single cell approaches described above. Single cell RNA-seq technologies certainly provide unprecedented single cell resolution of transcriptional programs, but we must

remember that TF proteins are the functional products that determine TF network activity. While current technical limitations prevent single cell proteomics, new technologies are moving towards this goal, including mass cytometry (Bendall et al., 2011) and single cell protein expression using microfluidic systems and protein-PCR based quantification (Macaulay and Voet, 2014). Ultimately however, we need to develop *in vivo* live cell imaging systems to measure multiple endogenous proteins simultaneously.

Dynamic modeling of a comprehensive TF network

While several *in silico* computational modeling methods have been used to study TF networks, these have so far largely failed to accurately predict biological systems. More complex modeling approaches are needed to better predict and extract the biological logic of TF networks. Here, it will be important to move from static models to dynamic models of TF regulatory networks, which better reflect biological complexity. *In silico* models of mammalian TF networks have also so far tended to investigate limited sub-networks within the larger TF network. By moving towards larger-scale comprehensive analysis of enhancer activity and TF network interactions, such as methods used in other organisms (Arnold et al., 2013, MacNeil et al., 2015), modeling of entire mammalian TF regulatory networks should be possible. For example, Arnold et al. recently developed a powerful genome-wide assay called STARR-seq to quantitiate enhancer activity genome-wide in Drosophila (Arnold et al., 2013). Additionally, by integrating data from various experimental sources, greater predictive power is possible. For example, by integrating data from ChIP-seq and enhancer assays into dynamic Bayesian computational modeling, accurate TF sub-network modeling has been possible (Schütte et al., 2016).

General principles in mammalian TF network biology

While many general principles that govern mammalian gene expression, enhancer activity, and TF interactions, have been described, we still have few general principles that hold for explaining mammalian TF networks. We hope that through integrating the above technologies, along with others, we will soon start to be able to develop meaningful principles that govern this key biological decision-making process.

Applications of TF network biology

Understanding the human TF regulatory network has important implications for health and disease. These include improving disease diagnosis and developing new therapeutic strategies, as outlined below.

Disease diagnosis

Central to the realization of personalized medicine is the ability to discriminate whether a patient will respond to a particular treatment or develop resistance, will relapse or go into remission, etc. Accurate biomarkers are key to this predict power. TF network components are often mutated in cancers (see **Box 3**), but genetic mutations alone often fail to accurately predict disease progression. Understanding the TF network states associated with a certain disease, and unique output (e.g. gene expression profile), may help to accurately predict clinical response and outcome. For example, regression analysis has recently been applied to large gene expression datasets from leukemia patients to identify a highly predictive and clinically informative 17-gene signature for therapy resistance (Ng et al., 2016). The application of our understanding of TF networks associated with (and specific to) human diseases also has significant potential in identifying novel therapeutic targets.

Cellular engineering

Reprogramming, forward programming, and trans-differentiation approaches hold important promise for regenerative medicine (Graf and Enver, 2009). However, the generation of transplantation-grade cells is a major hurdle for the clinical application of these approaches. TFs are most commonly used in these approaches, with the best combinations of TFs being identified from experimental screening. However, such approaches are often expensive, time-consuming and fail to generate fully functional cell types. Several network biology-based bioinformatics tools have been developed to predict TF combinations optimal for reprogramming and trans-differentiation, such as the CellNet platform (Cahan et al., 2014, Morris et al., 2014). Recently, a more comprehensive bioinformatics platform has been developed, called Mogrify, which combines gene expression datasets from over 300 cell types with TF network information to predict TFs for cellular reprogramming and trans-differentiation (Rackham et al., 2016). Given the importance of TF network state for cellular function and potential, such approaches and *in silico* modeling methodologies will likely play an increasingly important role in future translational research efforts.

Conclusion

TF regulatory network biology is an inter-disciplinary research field, with molecular, cellular, genetic, genomic, and computational approaches currently driving forward our understanding in a range of mammalian cell lineages and developmental stages. It is particularly exciting that human TF regulatory network biology is becoming an increasing research focus within the field, with its significance in understanding human health and disease. Using recent technological advances, we can now quantify global gene expression at single cell resolution, track TF dynamics within live cells, and investigate the impact of multiple mutations on cellular (and molecular) function. These new approaches are questioning some of the long-held assumptions in TF network biology. We

expect that future advances will further drive forwards our understanding of mammalian TF networks.

Display items:

Glossary

Sequence-specific transcription factor (TF): A sequence-specific DNA binding protein that regulates gene expression. These differ from other classes of transcription factors, such as general transcription factors, that regulate gene expression but do not bind DNA with sequence specificity.

TF network: The entire collection of TFs, their genes, and enhancer regions, within a cell that directly or indirectly interact to form a complex, interconnected molecular circuit.

Network state: The subset of the TF network, including the TFs and genes/enhancers that are expressed/active at a given moment.

Enhancer region: A genomic region (often composed of multiple DNA elements) that positively regulates (enhances) transcription from a gene promoter. An enhancer may be proximal to or distal from its target sequence in primary DNA sequence.

Transcription factor binding site (TFBS): A specific (usually short) DNA sequence or motif to which a TF binds, a functional DNA element within an enhancer region.

Box 1: Defining TF networks in mammals

To define mammalian TF networks, we must be able to identify its constituents and understand its underlying network logic. Numerous technological advances have helped us to define enhancer location and gene targets, measure enhancer activity, determine TF activity and function, and develop TF network models to predict logic. These are briefly summarized below, but further details can be found elsewhere (Brent, 2016, Blais and Dynlacht, 2005).

Defining enhancer location and gene targets: Putative enhancers were initially identified from conservation in non-coding elements within the genome, although we now know that sequence conservation is often a poor method to identify TFBSs (Villar et al., 2015). Numerous approaches have since been developed to identify putative enhancers, including DNase I hypersensitivity site (DHS) assays and chromatin immunoprecipitation

(ChIP) (Noonan and McCallion, 2010), and more recently ATAC-seq (Buenrostro et al., 2013). Additionally, chromatin conformation capture (3C) methods (Dostie et al., 2006) and particularly the genome-wide adaptation of this method (Hi-C) are helping to better define enhancer-promoter interactions as well as spatial chromatin structure information including topologically-associated domains (Pombo and Dillon, 2015). Certain histone modifications, particularly histone 3 lysine 4 mono-methylation (H3K4me1) (Heintzman et al., 2007) and H3K27 acetylation, are commonly used to identify putative enhancers (Creyghton et al., 2010, Rada-Iglesias et al., 2011). Large (~10 kb) regions of H3K27Ac enrichment within the genome have recently been defined as "super-enhancers", which appear to often contain numerous TFBSs (and perhaps multiple individual enhancers) (Hnisz et al., 2013, Whyte et al., 2013). However, the functional relevance of this classification is still unclear (Hay et al., 2016). While histone modification, TF enrichment, or open chromatin are often fairly good indicators of active enhancers, we do not yet have a universal active enhancer "mark" (Dogan et al., 2015).

Measuring enhancer activity: Enhancers are classically defined as functional DNA sequences with the ability to activate (enhance) the rate of transcription from a heterologous promoter, independent of location and orientation (Maniatis et al., 1987, Kim and Shiekhattar, 2015). This is the basis of the widely used enhancer assay: a putative enhancer is cloned up or downstream of a minimal promoter that drives expression of a reporter gene (Noonan and McCallion, 2010) (**Figure 2A**). However, recent technological advances are now allowing endogenous enhancers to be functionally interrogated (see the **Emerging Technologies section**). Synthetic TFs such as Zinc Fingers, Transcription Activator-Like Effectors (TALEs), and Cas9, can also be used to perturb endogenous enhancer activity and study TF networks. This is achieved by fusing genome-specific synthetic DNA binding domains to transcriptional effector domains such as the VP64 (activator) and KRAB (repressor) domains (Gao et al., 2013, Wilkinson et al., 2014). These approaches provide an opportunity to inducibly activate or silence enhancers, or synthetically engineer more complex transcriptional circuitry.

Determining TF function: The genome-editing revolution associated with CRISPR/Cas9 technologies (Cho et al., 2013, Cong et al., 2013, Mali et al., 2013) now allows TF genes, as well as enhancer regions, to be deleted and mutated easily. Importantly, these methods also allow for the first time, large-scale analysis within human cells, where efficiencies of homologous recombination are normally too low for traditional targeted genetic manipulation (see **Emerging Technologies section**).

Defining TFBSs: TF DNA-binding specificity can be determined by *in vitro* assays such as electrophoretic mobility shift assays (EMSAs) or systematic evolution of ligands by exponential enrichment (SELEX) (**Figure 2B**). DNase-seq is also being used to describe the binding patterns of TF-DNA interactions through deep sequencing, which allows TF binding motifs or "footprints" within DHSs to be resolved (Hesselberth et al., 2009). Although such DNase footprinting only generates candidate assignments for a class of TF (not a specific TF), this methodology has provided fundamental insights into TF network topology, and its conservation within mammals (Stergachis et al., 2014, Boyle et al., 2014).

Developing TF network models: By combining our knowledge of TF binding events within TF gene loci with enhancer assays and functional analysis, we can build models of TF networks to predict biological behavior *in silico* (**Figure 2C**). To date, most network models are relatively simple, often consisting of simple diagrams of TF sub-networks annotated with nodes and edges. However, Boolean (Xu et al., 2014, Dunn et al., 2014, Moignard et al., 2015) and Bayesian (Dowell et al., 2014) modeling approaches have been built from such information to provide dynamic and executable network models. PetriNets, a mathematical modeling approach to graphically model networks, have also been successfully used to computationally encode TF networks (Bonzanni et al., 2013). Alternative methods have been used to "reverse engineer" networks from gene expression data, such as by using mutual information or partial correlation analysis including the ARACNE algorithm (Margolin et al., 2006), and have been recently applied to single cell gene expression data (Wilkinson et al., 2014, Moignard et al., 2015). We refer readers to recent reviews of TF network modeling for further information (Woodhouse et al., 2016, Le Novère, 2015).

Box 2: Emerging concepts in the regulation of TF networks

TF network states are not static but dynamic and often unstable. It is well understood that phenotypically different cells will contain different TF network states. However, phenotypically identical cells also have considerable functional heterogeneity, even with highly purified cell populations (Yamamoto et al., 2013), and differ in their response to extracellular signals (Satija and Shalek, 2014). Such functional heterogeneity highlights that TF network states can also differ between individual phenotypically similar (identical) cells. Extracellular signaling plays a major role in influencing the TF network state in mammalian cells, although numerous intrinsic mechanisms also influence TF network state stability and transitions between them, including intrinsic dynamics within the TF network (and its constituents), indirect TF interactions, cell cycle progression, and metabolic state (**Figure 3**). It is worth highlighting that different TF sub-networks (regulating particular biological functions such as cell cycle or response to stress) may be overlapping or independent of each other. Recent examples of these mechanisms of TF network regulation are briefly summarized below, and can be found in more detail elsewhere (Long et al., 2016, Davidson, 2010).

Extracellular signaling: All multicellular organisms require intercellular signaling pathways to allow the coordinated formation and maintenance of complex tissues. Signaling pathways often have multiple functions at different stages of development and in different cell types (Massagué, 2012). This can at least partially be explained by the transcriptional response to a stimulus being dependent on the TF network state (Trompouki et al., 2011, Mullen et al., 2011). Many signaling pathway effectors are themselves TFs, and therefore directly integrate with the TF network. The downstream transcriptional targets of these signaling effectors also often include TF genes (Kageyama et al., 2007), which go on to influence future TF network states.

TF network states are dynamic and heterogeneous: Positive feedback loops help to reinforce TF expression, and can thereby stabilize a TF network state. By contrast, TF antagonism causes inherent instability and appears to play important roles in cell fate decisions. One of the best-described examples is the antagonism

between the TFs Gata1 and PU.1 in erythroid-myeloid lineage specification during hematopoietic differentiation (Burda et al., 2010) (although this TF antagonism has recently been questioned – see the **Emerging Technologies** section). Of course, TF antagonism can occur by various mechanisms, for example, competition for binding to the same TFBS (Bresnick et al., 2010).

TF stability and dose-dependence: The stability of the TF network as a whole depends on the stability of its constituents, particularly protein stability, rates of transcription and mRNA stability. While the DNA encoding TF genes and enhancers is permanent (although copy number can change, as in polyploidy and aneuploidy), relative accessibility of these genomic regions can be modulated by epigenetic modification. TF protein concentration appears to be particularly important for their function, with several TFs having dosage-dependent functions in development and homeostasis (Sigvardsson, 2012).

Indirect interactions and regulation of TF activity: Considering only TFs, enhancers and gene loci as the TF regulatory networks oversimplifies the biological complexity. Numerous levels of regulation overlay each other to "fine tune" gene expression. For example, numerous microRNAs post-transcriptionally regulate TF genes and mediate indirect TF network interactions (Martinez and Walhout, 2009). The function of other noncoding RNAs such as enhancer RNAs (eRNAs), RNA transcripts that originate from active enhancers, are yet to be fully understood (Kim and Shiekhattar, 2015). It is important to also remember that TFs are frequently post-translationally modified, which can profoundly influence TF activity and localization (Filtz et al., 2014).

Cell cycle: A major cause of intrinsic destabilization is the cell cycle. Cell cycle progression and division directly impacts on DNA accessibility (Ma et al., 2015) as well as TF protein (and mRNA) concentration. It is important to remember that the TF network plays a key role in regulating cell cycle progression (Müller, 1995), while cell cycle stage itself influences rates of gene expression (Bertoli et al., 2013) and cell fate decisions (Pauklin and Vallier, 2013).

Metabolic status and intracellular signaling: A cell must be able to adapt to its intracellular and extracellular metabolic status. There are several highly conserved metabolic signaling pathways that regulate such cellular adaption (Efeyan et al., 2015). While these pathways, in particular the Integrated Stress Response (ISR) pathway and mTOR signaling, can alter TF networks through altering rates of global translation, these pathways also more directly influence transcriptional states (de Nadal et al., 2011). For example, activation of the ISR pathway suppresses global translation, it acts to increase translation of certain TFs, notably its canonical effector ATF4 (Wek et al., 2006).

Box 3: Using stem cells and cancer models to understand TF network regulation and dysregulation

While TF networks have been investigated in numerous mammalian cell types, such networks have been most intensively studied in pluripotent stem cells (Ng and Surani, 2011, Orkin et al., 2008), adult muscle stem cells (Buckingham and Rigby, 2014, Tapscott, 2005), and adult hematopoietic stem cells (Göttgens, 2015) (**Figure 4**). Additionally, TF network dysregulation is a common theme in cancer, particularly leukemia (Sive and Gottgens, 2014).

Pluripotent stem cells: Pluripotent stem cells (PSCs) have the capacity to form any embryonic type (Figure 4B) (Murry and Keller, 2008). *In vitro* PSC self-renewal and differentiation provides an important and widely used tractable model of early developmental cell fate decisions. Induced pluripotent stem (iPS) cell reprogramming experiments have highlighted the importance of TFs (all four Yamanaka factors are TFs) in the acquisition and maintenance of the pluripotent state (Takahashi and Yamanaka, 2006). Live cell imaging and single cell RNA-seq methods have recently revealed unexpected heterogeneity of the TF network associated with pluripotency (Filipczyk et al., 2015, Kolodziejczyk et al., 2015), suggesting we still do not fully understand this TF network state. Consistent with this, several pluripotency TF network models recently built from detailed knowledge of the key TFs regulating pluripotency (Xu et al., 2014, Dunn et al., 2014, Dowell et al., 2014) were unable to fully predict cellular behavior.

Adult stem cells: Adult stem cells are thought to provide life-long homeostasis of several adult mammalian tissues. TF interactions regulating the generation, self-renewal, and differentiation of unipotent muscle stem cells and multipotent hematopoietic stem cells (HSCs) are arguably the best understood (**Figure 4B**). The TF MyoD is a central regulator of muscle formation; it is upregulated in differentiating muscle stem cells and its overexpression in a number of cell types can induce trans-differentiation to muscle (Tapscott et al., 1988, Davis et al., 1987). Numerous TFs have been found to regulate HSC function, and can be found reviewed elsewhere (Wilkinson and Gottgens, 2013). Several laboratories have recently developed methods to reprogram, trans-differentiate and forward program cells into hematopoietic stem and progenitor cells (Riddell et al., 2014, Xie et al., 2004, Sandler et al., 2014). All the methods published so far have used TF overexpression (usually in multi-TF combinations), highlighting the instructive role of TFs, and their combinatorial interactions, in initiating and maintaining hematopoietic cell identity.

Dysregulation in cancer: A diverse set of molecular mechanisms has so far been described to interfere with normal TF network logic in cancer including the mutation of TFs, co-activators/co-repressors, and enhancer regions. Two novel mechanism by which enhancer regions are mutated are particularly noteworthy. First, a chromosomal inversion event has been shown to cause spatial rearrangement of a *GATA2* enhancer to be proximal to the *EVI1*. This alteration in the TF regulatory network logic results in oncogenic *EVI1* overexpression, which results in leukemogenesis (Groschel et al., 2014, Yamazaki et al., 2014). Second, somatic mutation of an enhancer upstream of the *TAL1* gene has been shown to introduce a novel Myb DNA motif, which drives oncogenic *TAL1* overexpression and acute lymphoblastic leukaemia (Mansour et al., 2014).

Figure legends:

Figure 1: Central dogma of molecular biology and functions of transcription factors

(A) Gene expression is the process of gene transcription into messenger (m)RNA followed by translation into protein. Genes are encoded within genomic DNA and packaged within the nucleus as chromatin. Genomic sequencing has allowed protein-coding genes to be identified and annotated. A range of techniques have been developed to investigate chromatin structure, including DNase I hypersensitivity assays (such as DNase-seq), chromatin immunoprecipitation (such as ChIP-seq for histone modifications and TF enrichment) and chromatin conformation capture (3C) methods. Gene products can be measured at both RNA and protein levels by a range of techniques.

(B) Regulation of TF expression, activity and function. TFs are regulated at transcriptional, post-transcriptional and post-translational levels. TFs (green) can function by multiple mechanisms including: (i) recruitment of co-activators (yellow) that may add activating histone modifications (H3K4me or H3K27Ac; denoted as orange histones) or recruit RNA pol II to promote gene transcription; (ii) recruitment of co-repressors (red) that apply repressive histone modifications (such as H3K29me; denoted by black histones) to promote histone compaction and gene silencing; or (iii) DNA binding that results in histone displacement, which allows other TFs (blue) to bind;. TFs usually bind cooperatively and regulation of TF expression levels (and post-translational modifications) may influence TF function and activities.

Figure 2: Approaches to build TF regulatory network models

(A) Enhancers. Putative enhancers can be identified by a number features including DNase I hypersensitivity sites (DHSs), histone modifications (such as H3K4me), TF enrichment and DNA looping (measured by chromatin conformation capture methods such as Hi-C). Enhancer activity can be assessed using *in vivo* or *in vitro* enhancer assays, and the function of the TFBSs (DNA motifs) identified within such enhancers can assessed by mutational analysis.

(B) Transcription factors. TFs can be identified by their DNA binding domains. TFs also contain effector domains, which are responsible for protein-protein interactions. A range of methods including electrophoretic

mobility shift assays (EMSAs) and systematic evolution of ligands by exponential enrichment (SELEX) have been used to determine individual and cooperative TF DNA binding specificities.

(C) Building TF network models. Methods in (A) and (B) can be combined with functional assays (such as enhancer mutagenesis), gene expression analysis and/or TF pertubation analysis to build and train TF regulatory network models that can be "executed" *in silico*. These models provide important insights into the biological logic underpinning mammalian cell fate decisions, which feedback into experimental research.

Figure 3: Mechanism of TF network regulation

(A) A summary of mechanisms that influence TF network state stability. Numerous extrinsic and intrinsic mechanisms regulate TF network stability. These mechanisms are also often influenced by TF network state). TF network stabilization results in maintenance of a cell identity/function, such as stem cell self-renewal, while TF network destabilization induces TF network state transitions can lead to changing cellular identity/function and cellular differentiation.

(B) A schematic of how different signalling pathways activate certain sub-networks or states of a TF network. Depending on the logic of TF interactions and signalling inputs, states may be (i) stabilized or (ii) destabilized (resulting in state transitions). The set of TFs expressed determines the selection of genes regulated/expressed, which influences cellular identity and function. This review focuses on the TF networks, rather than upstream signaling inputs or downstream regulated genes/patterns of expression. For simplicity, the TF protein, its enhancer(s) and gene are represented as a single circle (A-G). As described in the main text, TF proteins regulate the activity of enhancers of other TFs, to activate or repress gene transcription (and may also be involved in auto-feedback regulation).

Figure 4: Commonly used mammalian systems to study TF networks

Mammalian TF networks have been commonly investigated in the context of pluripotency, muscle formation and hematopoiesis. (i) Pluripotent stem cells (embryonic stem cells or induced pluripotent stem cells) can self-renew or differentiate into any embryonic cell type through commitment to mesoderm, endoderm or

ectoderm germ layers. (ii) Muscle stem cells (or satellite cells) can self-renew or differentiate into muscle cells. (iii) Hematopoietic stem cells (HSCs) have the ability to self-renew or differentiate into any mature blood cell type, through increasingly lineage-restricted haematopoietic progenitor cells. Red blood cells (RBCs), megakaryocytes (MKs), myeloid cells and lymphoid cells can be specified from HSCs.

ACKNOWLEDGEMENTS

We apologize to those authors whose work could not be cited due to space constraints. We thank Dr. Hannah Long and Dr. Craig Mak for helpful discussion and feedback on the manuscript. ACW is funded by a Bloodwise Visiting Fellowship. HN is funded by the Japan Science and Technology Agency, the California Institute of Regenerative Medicine and Ludwig Foundation. BG is funded by Bloodwise, Cancer Research UK, the Wellcome Trust, the MRC, NIH-NIDDK and core funding from the Wellcome Trust to the Cambridge Stem Cell Institute.

REFERENCES

Alon, U. (2007) 'Network motifs: theory and experimental approaches', *Nat Rev Genet*, 8(6), pp. 450-61.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł., Rath, M. and Stark, A. (2013) 'Genome-wide quantitative enhancer activity maps identified by STARR-seq', *Science*, 339(6123), pp. 1074-7.

Bendall, S. C., Simonds, E. F., Qiu, P., Amir, e.-A., Krutzik, P. O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O. I., Balderas, R. S., Plevritis, S. K., Sachs, K., Pe'er, D., Tanner, S. D. and Nolan, G. P. (2011) 'Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum', *Science*, 332(6030), pp. 687-96.

Bertoli, C., Skotheim, J. M. and de Bruin, R. A. (2013) 'Control of cell cycle transcription during G1 and S phases', *Nat Rev Mol Cell Biol*, 14(8), pp. 518-28.

Blais, A. and Dynlacht, B. D. (2005) 'Constructing transcriptional regulatory networks', *Genes Dev*, 19(13), pp. 1499-511.

Bonzanni, N., Garg, A., Feenstra, K. A., Schütte, J., Kinston, S., Miranda-Saavedra, D., Heringa, J., Xenarios, I. and Göttgens, B. (2013) 'Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model', *Bioinformatics*, 29(13), pp. i80-8.

Boyle, A. P., Araya, C. L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L. W., Janette, J., Jiang, L., Kasper, D., Kawli, T., Kheradpour, P., Kundaje, A., Li, J. J., Ma, L., Niu, W., Rehm, E. J., Rozowsky, J., Slattery, M., Spokony, R., Terrell, R., Vafeados, D., Wang, D., Weisdepp, P., Wu, Y. C., Xie, D., Yan, K. K., Feingold, E. A., Good, P. J., Pazin, M. J., Huang, H., Bickel, P. J., Brenner, S. E., Reinke, V., Waterston, R. H., Gerstein, M., White, K. P., Kellis, M. and Snyder, M. (2014) 'Comparative analysis of regulatory information and circuits across distant species', *Nature*, 512(7515), pp. 453-6.

Brent, M. R. (2016) 'Past Roadblocks and New Opportunities in Transcription Factor Network Mapping', *Trends Genet*, 32(11), pp. 736-750.

Bresnick, E. H., Lee, H. Y., Fujiwara, T., Johnson, K. D. and Keles, S. (2010) 'GATA switches as developmental drivers', *J Biol Chem*, 285(41), pp. 31087-93.

Buckingham, M. and Rigby, P. W. (2014) 'Gene regulatory networks and transcriptional mechanisms that control myogenesis', *Dev Cell*, 28(3), pp. 225-38.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nat Methods*, 10(12), pp. 1213-8.

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. and Greenleaf, W. J. (2015) 'Single-cell chromatin accessibility reveals principles of regulatory variation', *Nature,* 523(7561), pp. 486-90.

Burda, P., Laslo, P. and Stopka, T. (2010) 'The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis', *Leukemia,* 24(7), pp. 1249-57.

Cahan, P., Li, H., Morris, S. A., Lummertz da Rocha, E., Daley, G. Q. and Collins, J. J. (2014) 'CellNet: network biology applied to stem cell engineering', *Cell,* 158(4), pp. 903-15.

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G. C., Zhang, F., Orkin, S. H. and Bauer, D. E. (2015) 'BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis', *Nature,* 527(7577), pp. 192-7.

Cheng, Y., Ma, Z., Kim, B. H., Wu, W., Cayting, P., Boyle, A. P., Sundaram, V., Xing, X., Dogan, N., Li, J., Euskirchen, G., Lin, S., Lin, Y., Visel, A., Kawli, T., Yang, X., Patacsil, D., Keller, C. A., Giardine, B., Kundaje, A., Wang, T., Pennacchio, L. A., Weng, Z., Hardison, R. C., Snyder, M. P. and Consortium, M. E. (2014) 'Principles of regulatory information conservation between mouse and human', *Nature,* 515(7527), pp. 371-5.

Cho, S. W., Kim, S., Kim, J. M. and Kim, J. S. (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nat Biotechnol,* 31(3), pp. 230-2.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. and Zhang, F. (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science,* 339(6121), pp. 819-23.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. and Jaenisch, R. (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proc Natl Acad Sci U S A,* 107(50), pp. 21931-6.

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. and Shendure, J. (2015) 'Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing', *Science,* 348(6237), pp. 910-4.

Davidson, E. H. (2010) 'Emerging properties of animal gene regulatory networks', *Nature,* 468(7326), pp. 911-20.

Davis, R. L., Weintraub, H. and Lassar, A. B. (1987) 'Expression of a single transfected cDNA converts fibroblasts to myoblasts', *Cell,* 51(6), pp. 987-1000.

de Nadal, E., Ammerer, G. and Posas, F. (2011) 'Controlling gene expression in response to stress', *Nat Rev Genet,* 12(12), pp. 833-45.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S., Friedman, N. and Regev, A. (2016) 'Perturb-Seq: Dissecting Molecular Circuits with

Scalable Single-Cell RNA Profiling of Pooled Genetic Screens', *Cell,* 167(7), pp. 1853-1866.e17.

Dogan, N., Wu, W., Morrissey, C. S., Chen, K. B., Stonestrom, A., Long, M., Keller, C. A., Cheng, Y., Jain, D., Visel, A., Pennacchio, L. A., Weiss, M. J., Blobel, G. A. and Hardison, R. C. (2015) 'Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility', *Epigenetics Chromatin,* 8, pp. 16.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D. and Dekker, J. (2006) 'Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements', *Genome Res,* 16(10), pp. 1299-309.

Dowell, K. G., Simons, A. K., Bai, H., Kell, B., Wang, Z. Z., Yun, K. and Hibbs, M. A. (2014) 'Novel insights into embryonic stem cell self-renewal revealed through comparative human and mouse systems biology networks', *Stem Cells,* 32(5), pp. 1161-72.

Dunn, S. J., Martello, G., Yordanov, B., Emmott, S. and Smith, A. G. (2014) 'Defining an essential transcription factor program for naïve pluripotency', *Science,* 344(6188), pp. 1156-60.

Efeyan, A., Comb, W. C. and Sabatini, D. M. (2015) 'Nutrient-sensing mechanisms and pathways', *Nature,* 517(7534), pp. 302-10.

Filipczyk, A., Marr, C., Hastreiter, S., Feigelman, J., Schwarzfischer, M., Hoppe, P. S., Loeffler, D., Kokkaliaris, K. D., Endele, M., Schauberger, B., Hilsenbeck, O., Skylaki, S., Hasenauer, J., Anastassiadis, K., Theis, F. J. and Schroeder, T. (2015) 'Network plasticity of pluripotency transcription factors in embryonic stem cells', *Nat Cell Biol,* 17(10), pp. 1235-46.

Filtz, T. M., Vogel, W. K. and Leid, M. (2014) 'Regulation of transcription factor activity by interconnected post-translational modifications', *Trends Pharmacol Sci,* 35(2), pp. 76-85.

Fuxman Bass, J. I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D. E., Vidal, M. and Walhout, A. J. (2015) 'Human gene-centered transcription factor networks for enhancers and disease variants', *Cell,* 161(3), pp. 661-73.

Gao, X., Yang, J., Tsang, J. C., Ooi, J., Wu, D. and Liu, P. (2013) 'Reprogramming to Pluripotency Using Designer TALE Transcription Factors Targeting Enhancers', *Stem Cell Reports,* 1(2), pp. 183-97.

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., Challinor, M., Wallace, K., Gilmour, J., Assi, S. A., Cauchy, P., Hoogenkamp, M., Westhead, D. R., Lacaud, G., Kouskoff, V., Göttgens, B. and Bonifer, C. (2016) 'Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation', *Dev Cell,* 36(5), pp. 572-87.

Graf, T. and Enver, T. (2009) 'Forcing cells to change lineages', *Nature,* 462(7273), pp. 587-94.

Groschel, S., Sanders, M. A., Hoogenboezem, R., de Wit, E., Bouwman, B. A., Erpelinck, C., van der Velden, V. H., Havermans, M., Avellino, R., van Lom, K., Rombouts, E. J., van Duin, M., Dohner, K., Beverloo, H. B., Bradner, J. E., Dohner, H., Lowenberg, B., Valk, P. J., Bindels, E. M., de Laat, W. and Delwel, R. (2014) 'A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia', *Cell*, 157(2), pp. 369-81.

Göttgens, B. (2015) 'Regulatory network control of blood stem cells', *Blood*, 125(17), pp. 2614-20.

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O., Graham, B. J., Hanssen, L. L., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., Telenius, J., Williams, R., Rode, C., Li, P. S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S., Sauka-Spengler, T., Gibbons, R. J., Smith, A. J., Wood, W. G. and Higgs, D. R. (2016) 'Genetic dissection of the α-globin super-enhancer in vivo', *Nat Genet*, 48(8), pp. 895-903.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. and Ren, B. (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nat Genet*, 39(3), pp. 311-8.

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., Fields, S. and Stamatoyannopoulos, J. A. (2009) 'Global mapping of protein-DNA interactions in vivo by digital genomic footprinting', *Nat Methods*, 6(4), pp. 283-9.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke, H. A. and Young, R. A. (2013) 'Super-enhancers in the control of cell identity and disease', *Cell*, 155(4), pp. 934-47.

Hoppe, P. S., Coutu, D. L. and Schroeder, T. (2014) 'Single-cell technologies sharpen up mammalian stem cell research', *Nat Cell Biol*, 16(10), pp. 919-27.

Hoppe, P. S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K. D., Hilsenbeck, O., Moritz, N., Endele, M., Filipczyk, A., Gambardella, A., Ahmed, N., Etzrodt, M., Coutu, D. L., Rieger, M. A., Marr, C., Strasser, M. K., Schauberger, B., Burtscher, I., Ermakova, O., Bürger, A., Lickert, H., Nerlov, C., Theis, F. J. and Schroeder, T. (2016) 'Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios', *Nature*, 535(7611), pp. 299-302.

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M. and Weissman, J. S. (2016) 'Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation', *Elife*, 5.

Ismailoglu, I., Yeamans, G., Daley, G. Q., Perlingeiro, R. C. R. and Kyba, M. (2008) 'Mesodermal patterning activity of SCL', *Experimental Hematology*, 36(12), pp. 1593-1603.

Jaitin, D. A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T. M., Tanay, A., van Oudenaarden, A. and Amit, I. (2016) 'Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq', *Cell*, 167(7), pp. 1883-1896.e15.

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., Stunnenberg, H. G., Todd, J. A., Zerbino, D. R., Stegle, O., Ouwehand, W. H., Frontini, M., Wallace, C., Spivakov, M., Fraser, P. and Consortium, B. (2016) 'Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters', *Cell*, 167(5), pp. 1369-1384.e19.

Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T. M., Childs, R., Levens, D. and Zhao, K. (2015) 'Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples', *Nature*, 528(7580), pp. 142-6.

Johnson, D. S., Mortazavi, A., Myers, R. M. and Wold, B. (2007) 'Genome-wide mapping of in vivo protein-DNA interactions', *Science*, 316(5830), pp. 1497-502.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J. M., Vincentelli, R., Luscombe, N. M., Hughes, T. R., Lemaire, P., Ukkonen, E., Kivioja, T. and Taipale, J. (2013) 'DNA-binding specificities of human transcription factors', *Cell*, 152(1-2), pp. 327-39.

Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E. and Taipale, J. (2015) 'DNA-dependent formation of transcription factor pairs alters their binding specificity', *Nature*, 527(7578), pp. 384-8.

Kageyama, R., Ohtsuka, T. and Kobayashi, T. (2007) 'The Hes gene family: repressors and oscillators that orchestrate embryogenesis', *Development*, 134(7), pp. 1243-51.

Kellis, M., Wold, B., Snyder, M. P., Bernstein, B. E., Kundaje, A., Marinov, G. K., Ward, L. D., Birney, E., Crawford, G. E., Dekker, J., Dunham, I., Elnitski, L. L., Farnham, P. J., Feingold, E. A., Gerstein, M., Giddings, M. C., Gilbert, D. M., Gingeras, T. R., Green, E. D., Guigo, R., Hubbard, T., Kent, J., Lieb, J. D., Myers, R. M., Pazin, M. J., Ren, B., Stamatoyannopoulos, J. A., Weng, Z., White, K. P. and Hardison, R. C. (2014) 'Defining functional DNA elements in the human genome', *Proc Natl Acad Sci U S A*, 111(17), pp. 6131-8.

Kim, T. K. and Shiekhattar, R. (2015) 'Architectural and Functional Commonalities between Enhancers and Promoters', *Cell*, 162(5), pp. 948-59.

Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C., Ilicic, T., Henriksson, J., Natarajan, K. N., Tuck, A. C., Gao, X., Bühler, M., Liu, P., Marioni, J. C. and Teichmann, S. A. (2015) 'Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation', *Cell Stem Cell*, 17(4), pp. 471-85.

Kouzarides, T. (2007) 'Chromatin modifications and their function', *Cell*, 128(4), pp. 693-705.

Kueh, H. Y., Champhekhar, A., Nutt, S. L., Elowitz, M. B. and Rothenberg, E. V. (2013) 'Positive feedback between PU.1 and the cell cycle controls myeloid differentiation', *Science*, 341(6146), pp. 670-3.

Kueh, H. Y., Yui, M. A., Ng, K. K., Pease, S. S., Zhang, J. A., Damle, S. S., Freedman, G., Siu, S., Bernstein, I. D., Elowitz, M. B. and Rothenberg, E. V. (2016) 'Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment', *Nat Immunol,* 17(8), pp. 956-65.

Le Novère, N. (2015) 'Quantitative and logic modelling of molecular and gene networks', *Nat Rev Genet,* 16(3), pp. 146-58.

Liu, Z., Legant, W. R., Chen, B. C., Li, L., Grimm, J. B., Lavis, L. D., Betzig, E. and Tjian, R. (2014) '3D imaging of Sox2 enhancer clusters in embryonic stem cells', *Elife,* 3, pp. e04236.

Long, H. K., Prescott, S. L. and Wysocka, J. (2016) 'Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution', *Cell,* 167(5), pp. 1170-1187.

Ma, Y., Kanakousaki, K. and Buttitta, L. (2015) 'How the cell cycle impacts chromatin architecture and influences cell fate', *Front Genet,* 6, pp. 19.

Macaulay, I. C. and Voet, T. (2014) 'Single cell genomics: advances and future perspectives', *PLoS Genet,* 10(1), pp. e1004126.

MacNeil, L. T., Pons, C., Arda, H. E., Giese, G. E., Myers, C. L. and Walhout, A. J. (2015) 'Transcription Factor Activity Mapping of a Tissue-Specific in vivo Gene Regulatory Network', *Cell Syst,* 1(2), pp. 152-162.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A. and McCarroll, S. A. (2015) 'Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets', *Cell,* 161(5), pp. 1202-14.

Mak, K. S., Funnell, A. P., Pearson, R. C. and Crossley, M. (2011) 'PU.1 and Haematopoietic Cell Fate: Dosage Matters', *Int J Cell Biol,* 2011, pp. 808524.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) 'RNA-guided human genome engineering via Cas9', *Science,* 339(6121), pp. 823-6.

Malik, S. and Roeder, R. G. (2010) 'The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation', *Nat Rev Genet,* 11(11), pp. 761-72.

Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987) 'Regulation of inducible and tissue-specific gene expression', *Science,* 236(4806), pp. 1237-45.

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., Etchin, J., Lawton, L., Sallan, S. E., Silverman, L. B., Loh, M. L., Hunger, S. P., Sanda, T., Young, R. A. and Look, A. T. (2014) 'Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element', *Science,* 346(6215), pp. 1373-7.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R. and Califano, A. (2006) 'ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context', *BMC Bioinformatics,* 7 Suppl 1, pp. S7.

Martinez, N. J. and Walhout, A. J. (2009) 'The interplay between transcription factors and microRNAs in genome-scale regulatory networks', *Bioessays,* 31(4), pp. 435-45.

Massagué, J. (2012) 'TGFβ signalling in context', *Nat Rev Mol Cell Biol,* 13(10), pp. 616-30.

Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schutte, J., Calero-Nieto, F. J., Kinston, S., Joshi, A., Hannah, R., Theis, F. J., Jacobsen, S. E., de Bruijn, M. F. and Gottgens, B. (2013) 'Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis', *Nat Cell Biol,* 15(4), pp. 363-72.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., Buettner, F., Macaulay, I. C., Jawaid, W., Diamanti, E., Nishikawa, S., Piterman, N., Kouskoff, V., Theis, F. J., Fisher, J. and Göttgens, B. (2015) 'Decoding the regulatory network of early blood development from single-cell gene expression measurements', *Nat Biotechnol,* 33(3), pp. 269-76.

Morris, S. A., Cahan, P., Li, H., Zhao, A. M., San Roman, A. K., Shivdasani, R. A., Collins, J. J. and Daley, G. Q. (2014) 'Dissecting engineered cell types and enhancing cell fate conversion via CellNet', *Cell,* 158(4), pp. 889-902.

Mullen, A. C., Orlando, D. A., Newman, J. J., Lovén, J., Kumar, R. M., Bilodeau, S., Reddy, J., Guenther, M. G., DeKoter, R. P. and Young, R. A. (2011) 'Master transcription factors determine cell-type-specific responses to TGF-β signaling', *Cell,* 147(3), pp. 565-76.

Murry, C. E. and Keller, G. (2008) 'Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development', *Cell,* 132(4), pp. 661-80.

Müller, R. (1995) 'Transcriptional regulation during the mammalian cell cycle', *Trends Genet,* 11(5), pp. 173-8.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A. and Fraser, P. (2013) 'Single-cell Hi-C reveals cell-to-cell variability in chromosome structure', *Nature,* 502(7469), pp. 59-64.

Ng, H. H. and Surani, M. A. (2011) 'The transcriptional and signalling networks of pluripotency', *Nat Cell Biol,* 13(5), pp. 490-6.

Ng, S. W., Mitchell, A., Kennedy, J. A., Chen, W. C., McLeod, J., Ibrahimova, N., Arruda, A., Popescu, A., Gupta, V., Schimmer, A. D., Schuh, A. C., Yee, K. W., Bullinger, L., Herold, T., Görlich, D., Büchner, T., Hiddemann, W., Berdel, W. E., Wörmann, B., Cheok, M., Preudhomme, C., Dombret, H., Metzeler, K., Buske, C., Löwenberg, B., Valk, P. J., Zandstra, P. W., Minden, M. D., Dick, J. E. and Wang, J. C. (2016) 'A 17-gene stemness score for rapid determination of risk in acute leukaemia', *Nature,* 540(7633), pp. 433-437.

Noonan, J. P. and McCallion, A. S. (2010) 'Genomics of long-range regulatory elements', *Annu Rev Genomics Hum Genet,* 11, pp. 1-23.

Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H. and Grimes, H. L. (2016) 'Single-cell analysis of mixed-lineage states leading to a binary cell fate choice', *Nature,* 537(7622), pp. 698-702.

Orkin, S. H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T. W., Shen, X. and Levasseur, D. N. (2008) 'The transcriptional network controlling pluripotency in ES cells', *Cold Spring Harb Symp Quant Biol,* 73, pp. 195-202.

Osawa, M., Hanada, K., Hamada, H. and Nakauchi, H. (1996) 'Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell', *Science,* 273(5272), pp. 242-5.

Parnas, O., Jovanovic, M., Eisenhaure, T. M., Herbst, R. H., Dixit, A., Ye, C. J., Przybylski, D., Platt, R. J., Tirosh, I., Sanjana, N. E., Shalem, O., Satija, R., Raychowdhury, R., Mertins, P., Carr, S. A., Zhang, F., Hacohen, N. and Regev, A. (2015) 'A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks', *Cell,* 162(3), pp. 675-86.

Pauklin, S. and Vallier, L. (2013) 'The cell-cycle state of stem cells determines cell fate propensity', *Cell,* 155(1), pp. 135-47.

Picelli, S., Björklund, Å., Faridani, O. R., Sagasser, S., Winberg, G. and Sandberg, R. (2013) 'Smart-seq2 for sensitive full-length transcriptome profiling in single cells', *Nat Methods,* 10(11), pp. 1096-8.

Pombo, A. and Dillon, N. (2015) 'Three-dimensional genome architecture: players and mechanisms', *Nat Rev Mol Cell Biol,* 16(4), pp. 245-57.

Rackham, O. J., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., Petretto, E., Forrest, A. R., Hayashizaki, Y., Polo, J. M., Gough, J. and Consortium, F. (2016) 'A predictive computational framework for direct reprogramming between human cell types', *Nat Genet,* 48(3), pp. 331-5.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J. (2011) 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature,* 470(7333), pp. 279-83.

Reynolds, N., O'Shaughnessy, A. and Hendrich, B. (2013) 'Transcriptional repressors: multifaceted regulators of gene expression', *Development,* 140(3), pp. 505-12.

Riddell, J., Gazit, R., Garrison, B. S., Guo, G., Saadatpour, A., Mandal, P. K., Ebina, W., Volchkov, P., Yuan, G. C., Orkin, S. H. and Rossi, D. J. (2014) 'Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors', *Cell,* 157(3), pp. 549-64.

Rotem, A., Ram, O., Shoresh, N., Sperling, R. A., Goren, A., Weitz, D. A. and Bernstein, B. E. (2015) 'Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state', *Nat Biotechnol,* 33(11), pp. 1165-72.

Sanchez-Freire, V., Ebert, A. D., Kalisky, T., Quake, S. R. and Wu, J. C. (2012) 'Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns', *Nat Protoc,* 7(5), pp. 829-38.

Sandler, V. M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J. M., Scandura, J. M. and Rafii, S. (2014) 'Reprogramming human endothelial cells to haematopoietic cells requires vascular induction', *Nature,* 511(7509), pp. 312-8.

Sanjana, N. E., Shalem, O. and Zhang, F. (2014) 'Improved vectors and genome-wide libraries for CRISPR screening', *Nat Methods,* 11(8), pp. 783-4.

Satija, R. and Shalek, A. K. (2014) 'Heterogeneity in immune responses: from populations to single cells', *Trends Immunol,* 35(5), pp. 219-29.

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L. and Ren, B. (2016) 'A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome', *Cell Rep,* 17(8), pp. 2042-2059.

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N. K., Riepsaame, J., Calero-Nieto, F. J., Moignard, V., Basilico, S., Kinston, S. J., Hannah, R. L., Chan, M. C., Nürnberg, S. T., Ouwehand, W. H., Bonzanni, N., de Bruijn, M. F. and Göttgens, B. (2016) 'An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability', *Elife,* 5, pp. e11469.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N. K., Macaulay, I. C., Marioni, J. C. and Göttgens, B. (2016) 'Resolving early mesoderm diversification through single-cell expression profiling', *Nature,* 535(7611), pp. 289-93.

Sigvardsson, M. (2012) 'Transcription factor dose links development to disease', *Blood,* 120(18), pp. 3630-1.

Sive, J. and Gottgens, B. (2014) 'Transcriptional network control of normal and leukaemic haematopoiesis', *Exp Cell Res*.

Stergachis, A. B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A. P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., Thurman, R. E., Vong, S., Bates, D., Neri, F., Diegel, M., Giste, E., Dunn, D., Vierstra, J., Hansen, R. S., Johnson, A. K., Sabo, P. J., Wilken, M. S., Reh, T. A., Treuting, P. M., Kaul, R., Groudine, M., Bender, M. A., Borenstein, E. and Stamatoyannopoulos, J. A. (2014) 'Conservation of trans-acting circuitry during mammalian regulatory evolution', *Nature,* 515(7527), pp. 365-70.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', *Cell,* 126(4), pp. 663-76.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., Lao, K. and Surani, M. A. (2010) 'RNA-Seq analysis to capture the transcriptome landscape of a single cell', *Nat Protoc,* 5(3), pp. 516-35.

Tapscott, S. J. (2005) 'The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription', *Development,* 132(12), pp. 2685-95.

Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P. F., Weintraub, H. and Lassar, A. B. (1988) 'MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts', *Science,* 242(4877), pp. 405-11.

Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., DiBiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L., Li, P., Durand, E. M., Mosimann, C., Heffner, G. C., Daley, G. Q., Paulson, R. F., Young, R. A. and Zon, L. I. (2011) 'Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration', *Cell,* 147(3), pp. 577-89.

Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., Gnirke, A. and Meissner, A. (2015) 'Transcription factor binding dynamics during human ES cell differentiation', *Nature,* 518(7539), pp. 344-9.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., Gozdecka, M., Ohnishi, S., Cooper, J., Patel, M., McKerrell, T., Chen, B., Domingues, A. F., Gallipoli, P., Teichmann, S., Ponstingl, H., McDermott, U., Saez-Rodriguez, J., Huntly, B. J., Iorio, F., Pina, C., Vassiliou, G. S. and Yusa, K. (2016) 'A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia', *Cell Rep,* 17(4), pp. 1193-1205.

Van Handel, B., Montel-Hagen, A., Sasidharan, R., Nakano, H., Ferrari, R., Boogerd, C. J., Schredelseker, J., Wang, Y. L., Hunter, S., Org, T., Zhou, J., Li, X. M., Pellegrini, M., Chen, J. N., Orkin, S. H., Kurdistani, S. K., Evans, S. M., Nakano, A. and Mikkola, H. K. A. (2012) 'Scl Represses Cardiomyogenesis in Prospective Hemogenic Endothelium and Endocardium', *Cell,* 150(3), pp. 590-605.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. and Luscombe, N. M. (2009) 'A census of human transcription factors: function, expression and evolution', *Nat Rev Genet,* 10(4), pp. 252-63.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., Park, T. J., Deaville, R., Erichsen, J. T., Jasinska, A. J., Turner, J. M., Bertelsen, M. F., Murchison, E. P., Flicek, P. and Odom, D. T. (2015) 'Enhancer evolution across 20 mammalian species', *Cell,* 160(3), pp. 554-66.

Wek, R. C., Jiang, H. Y. and Anthony, T. G. (2006) 'Coping with stress: eIF2 kinases and translational control', *Biochem Soc Trans,* 34(Pt 1), pp. 7-11.

White, A. K., VanInsberghe, M., Petriv, O. I., Hamidi, M., Sikorski, D., Marra, M. A., Piret, J., Aparicio, S. and Hansen, C. L. (2011) 'High-throughput microfluidic single-cell RT-qPCR', *Proc Natl Acad Sci U S A,* 108(34), pp. 13999-4004.

Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., Cowley, S. M. and Young, R. A. (2012) 'Enhancer decommissioning by LSD1 during embryonic stem cell differentiation', *Nature,* 482(7384), pp. 221-5.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell,* 153(2), pp. 307-19.

Wilkinson, A. C. and Gottgens, B. (2013) 'Transcriptional regulation of haematopoietic stem cells', *Adv Exp Med Biol,* 786, pp. 187-212.

Wilkinson, A. C., Kawata, V. K., Schütte, J., Gao, X., Antoniou, S., Baumann, C., Woodhouse, S., Hannah, R., Tanaka, Y., Swiers, G., Moignard, V., Fisher, J., Hidetoshi, S., Tijssen, M. R., de Bruijn, M. F., Liu, P. and Göttgens, B. (2014) 'Single-cell analyses of regulatory network perturbations using enhancer-targeting TALEs suggest novel roles for PU.1 during haematopoietic specification', *Development,* 141(20), pp. 4018-30.

Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Sánchez Castillo, M., Oedekoven, C. A., Diamanti, E., Schulte, R., Ponting, C. P., Voet, T., Caldas, C., Stingl, J., Green, A. R., Theis, F. J. and Göttgens, B. (2015) 'Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations', *Cell Stem Cell,* 16(6), pp. 712-24.

Wilson, N. K., Schoenfelder, S., Hannah, R., Sánchez Castillo, M., Schütte, J., Ladopoulos, V., Mitchelmore, J., Goode, D. K., Calero-Nieto, F. J., Moignard, V., Wilkinson, A. C., Jimenez-Madrid, I., Kinston, S., Spivakov, M., Fraser, P. and Göttgens, B. (2016) 'Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model', *Blood,* 127(13), pp. e12-23.

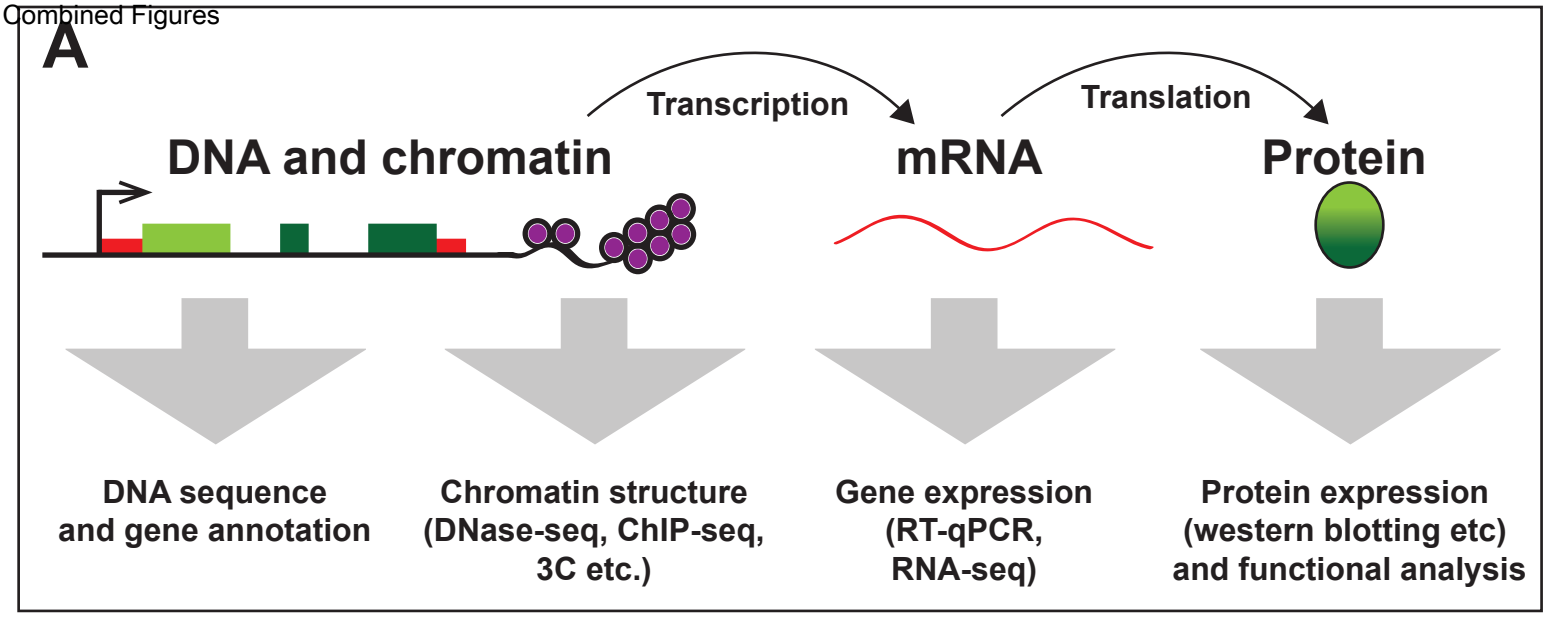
Woodhouse, S., Moignard, V., Göttgens, B. and Fisher, J. (2016) 'Processing, visualising and reconstructing network models from single-cell data', *Immunol Cell Biol,* 94(3), pp. 256-65.

Xie, H., Ye, M., Feng, R. and Graf, T. (2004) 'Stepwise reprogramming of B cells into macrophages', *Cell,* 117(5), pp. 663-76.

Xu, H., Ang, Y. S., Sevilla, A., Lemischka, I. R. and Ma'ayan, A. (2014) 'Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells', *PLoS Comput Biol,* 10(8), pp. e1003777.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H. and Nakauchi, H. (2013) 'Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells', *Cell,* 154(5), pp. 1112-26.

Yamazaki, H., Suzuki, M., Otsuki, A., Shimizu, R., Bresnick, E. H., Engel, J. D. and Yamamoto, M. (2014) 'A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression', *Cancer Cell,* 25(4), pp. 415-27.


A
Transcription
Translation
DNA and chromatin
mRNA
Protein
DNA sequence and gene annotation
Chromatin structure (DNase-seq, ChIP-seq, 3C etc.)
Gene expression (RT-qPCR, RNA-seq)
Protein expression (western blotting etc) and functional analysis

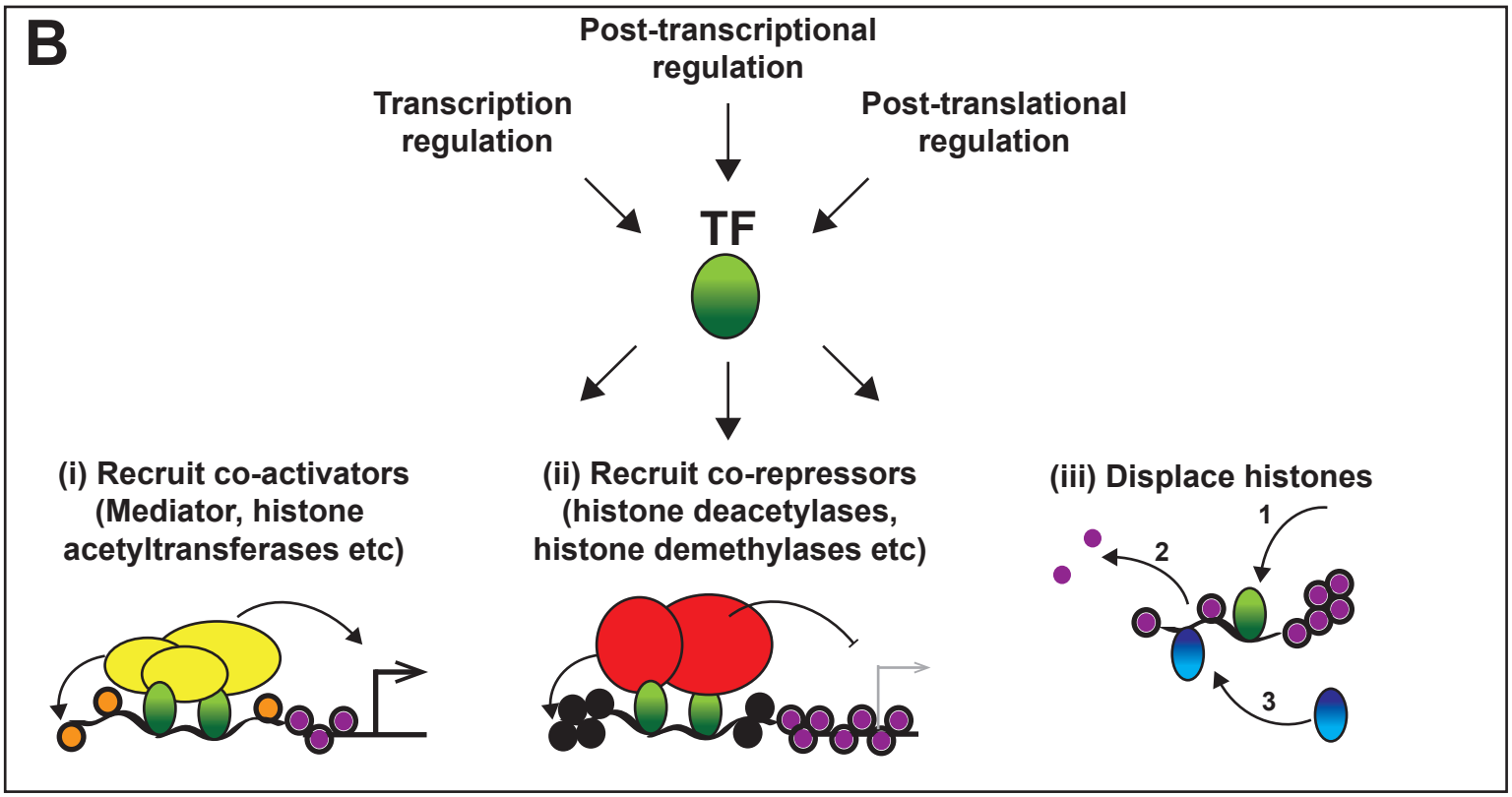
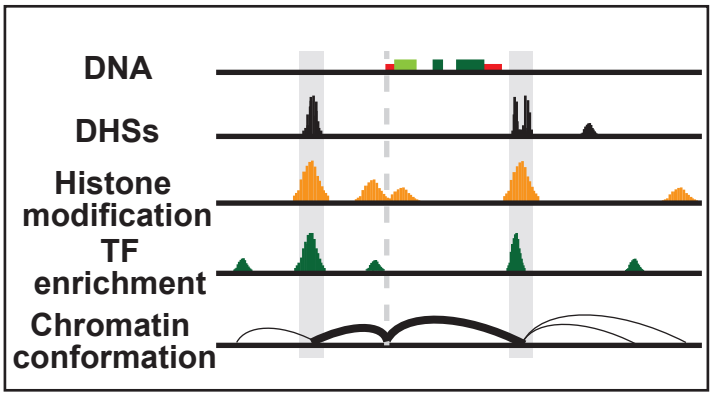

B
Post-transcriptional regulation
Transcription regulation
Post-translational regulation
TF
(i) Recruit co-activators (Mediator, histone acetyltransferases etc)
(ii) Recruit co-repressors (histone deacetylases, histone demethylases etc)
(iii) Displace histones
1
2
3

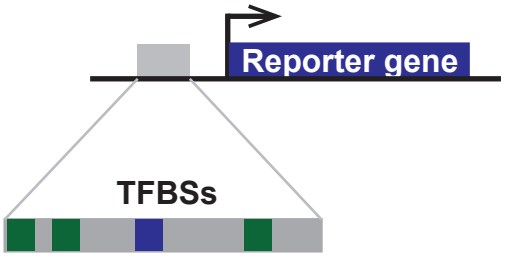
Figure 1

A

1. Identify putative enhancers



2. Determine enhancer activity (and funtional TFBSs)

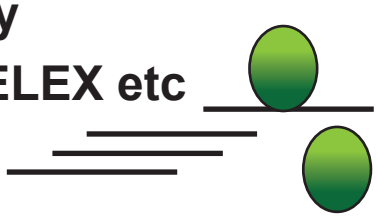


B

1. Identify protein domains

2. Determine DNA binding specificity

EMSA, SELEX etc



C

1. Functional assays

2. Gene expression analysis

3. Perturbation analysis

4. TF network modelling

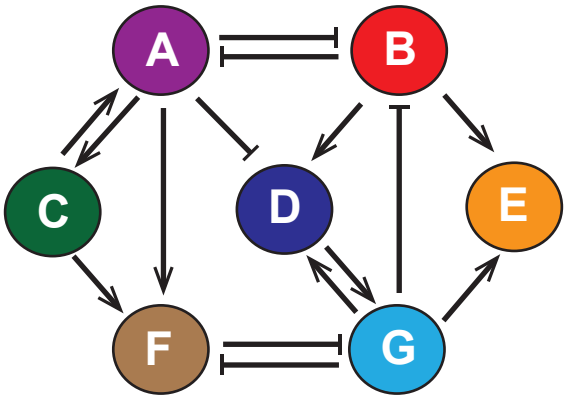


Figure 2

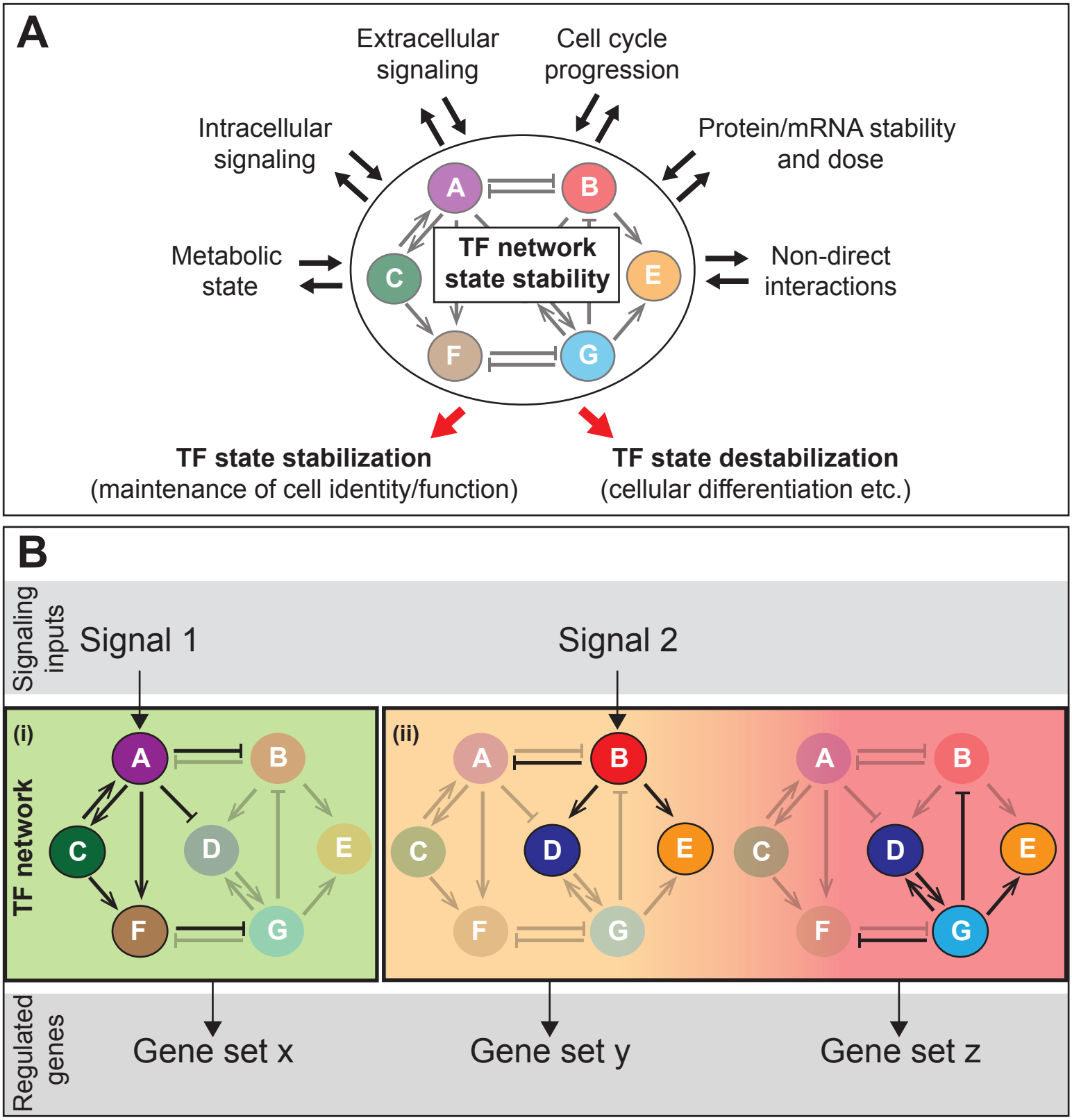

A
Extracellular signaling
Cell cycle progression
Intracellular signaling
Protein/mRNA stability and dose
Metabolic state
Non-direct interactions
TF network state stability
A
B
C
D
E
F
G
TF state stabilization
(maintenance of cell identity/function)
TF state destabilization
(cellular differentiation etc.)
B
Signaling inputs
Signal 1
Signal 2
(i)
(ii)
TF network
Regulated genes
Gene set x
Gene set y
Gene set z

Figure 3

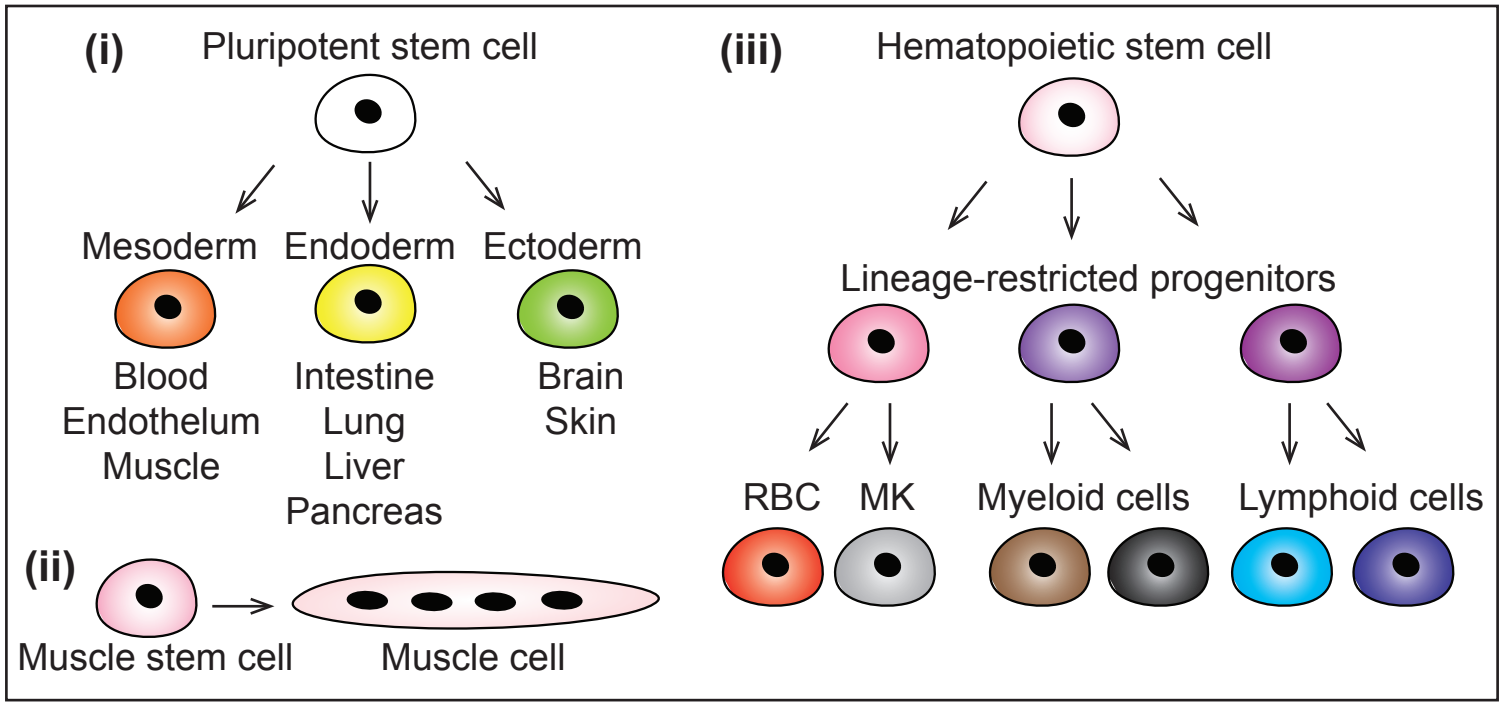

(i)
Pluripotent stem cell
Mesoderm
Endoderm
Ectoderm
Blood
Endothelum
Muscle
Intestine
Lung
Liver
Pancreas
Brain
Skin
(ii)
Muscle stem cell
Muscle cell
(iii)
Hematopoietic stem cell
Lineage-restricted progenitors
RBC
MK
Myeloid cells
Lymphoid cells

Figure 4