Hindawi Publishing Corporation
Scientifica
Volume 2012, Article ID 437956, 11 pages
http://dx.doi.org/10.6064/2012/437956

Review Article

Genome-Scale Technology Driven Advances to Research into Normal and Malignant Haematopoiesis

Berthold Göttgens

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

Correspondence should be addressed to Berthold Göttgens; bg200@cam.ac.uk

Received 30 September 2012; Accepted 16 December 2012

Academic Editors: J. L. Badano, F. Chiarini, J. Kennell, T. Te Kronnie, H.-Y. Wang, and G. Zafarana

Haematopoiesis or blood development has long served as a model system for adult stem cell biology. Moreover, when combined, the various cancers of the blood represent one of the commonest human malignancies. Large numbers of researchers have therefore dedicated their scientific careers to studying haematopoiesis for more than a century. Throughout this period, many new technologies have first been applied towards the study of blood cells, and the research fields of normal and malignant haematopoiesis have also been some of the earliest adopters of genome-scale technologies. This has resulted in significant new insights with implications ranging from basic biological mechanisms to patient diagnosis and prognosis and also produced lessons likely to be relevant for many other areas of biomedical research. This paper discusses the current state of play for a range of genome-scale applications within haemopoiesis research, including gene expression profiling, ChIP-sequencing, genomewide association analysis, and cancer genome sequencing. A concluding outlook section explores likely future areas of progress as well as potential technological and educational bottlenecks.

1. Introduction

Haematopoiesis represents the process whereby multipotential blood stem and progenitor cells differentiate into more than 10 distinct mature blood cell types. Research over the last 30 years has led to the development of purification protocols that permit the isolation of many of these progenitor and all mature cell types at close to 100% purity. Moreover, biological assays have been developed to validate the functional properties for most of these different cell types including the many progenitors at various stages of maturity. Consequently, differentiation of the blood system is better defined than any other mammalian organ system and has hence become a model system for the wider field of stem cell biology.

Since many of the mature blood cell types are short lived, they need to be constantly replenished throughout adult life, with the consequence that the blood system has one of the fastest turnovers of all human organ systems. Production of the various types of mature blood cells is tightly controlled, with transcription factor and signalling proteins playing particularly prominent roles [1–5]. Long-term formation of mature blood cells from blood stem cells also forms the basis of successful bone marrow transplantation, which therefore represents one of the most widely used stem cell treatments currently in use. Transplantation of blood stem cells has also been used as a powerful assay when applied to experimental animals, in particular rodents. Here it allows for the detection of the presence of blood stem cells in complex mixtures of cells, with the most advanced protocols allowing for the transplantation of a single blood stem cell to give rise to long-term donor-derived haematopoiesis in the transplant recipient [6].

The various types of human leukaemias all share the property of perturbed blood cell production, often with an accumulation of the so-called blast cells that resemble immature blood progenitor cells [7]. With transcription factor and signalling genes being key to normal blood development, it is perhaps no surprise that acquired mutations in these categories of genes are now recognised as one of the commonest causes of leukaemia development [8–11]. Below I will outline how a range of genome-scale approaches has been employed

to provide significant advances to our understanding of both normal and malignant haematopoiesis. This will be followed by a brief outlook on likely future developments and relevance beyond the field of haematopoiesis research.

2. Gene Expression Profiling for Network Inference and Disease Classification

The relative ease of accessing blood cells compared with most other human tissues likely is a major reason as to why several advanced approaches for the extraction of new biological knowledge from large-scale gene expression profiling datasets have been pioneered in studies using blood cells. Below I focus in particular on gene expression profiling studies aimed at regulatory network reconstruction and disease classification.

With the ever-increasing momentum of genome-scale science driven by, for example, human genome project [12, 13], gene expression profiling has rapidly been recognised as a powerful means to define the phenotype of a given cell population. With differentiation not only entailing but most likely being driven by changes in gene expression profiles, generating gene expression profiles for a range of different but related cell types has the potential to identify those aspects of a given expression profile that are characteristic for a given cell type. Moreover, large-scale analysis across multiple different cell types and lineages can be used to define coexpressed gene clusters, which through the use of reverse-engineering approaches can be utilised further for the reconstruction of likely regulatory hierarchies and networks. An early example of this approach was the development of the ARACNE (algorithm for the reconstruction of accurate cellular networks algorithm [14]). In this study, the authors reported the reconstruction of regulatory networks from expression profiles of human B cells, which suggested the existence of hierarchical, scale-free networks, where a few highly interconnected hub genes account for most of the interactions. The authors also identified the MYC protein as a major hub controlling a network of known and previously unknown MYC target genes, some of which represented major hubs themselves. This and related approaches have subsequently been used for the analysis of normal and pathologic networks in a range of mammalian cells [15–24].

Coexpression across more extended differentiation hierarchies has more recently been utilised for the definition of haematopoietic gene sets as well as the inference of regulatory networks [25]. Here the authors generated gene expression profiles for 38 distinct purified human hematopoietic cell populations. Subsequent use of probabilistic models and analysis of *cis*-elements were employed to further define the regulatory circuitry, which lead to the definition of densely interconnected *cis*-regulatory circuits and a number of transcription factors where their differential expression across the different hematopoietic lineages was inferred to be involved in the generation of distinct cellular states. Together with earlier human studies [26] as well as comprehensive expression profiling in the mouse [27–30], this study provides a rich resource for hypothesis generation for future research into the molecular controls of blood cell differentiation.

Gene expression profiling has also been employed widely across a range of different haematological malignancies, in order to identify new classification schemes with potential diagnostic, prognostic, and/or therapeutic value. One of the early protagonists in the field has been the group of Louis Staudt, who in 2000 reported distinct types of diffuse large B-cell lymphoma through the use of gene expression profiling [31]. Diffuse large B-cell lymphoma represents the most common subtype of non-Hodgkin's lymphoma and was known to be clinically heterogeneous. Using extensive expression profiling, the authors identified two molecularly distinct disease forms with gene expression patterns indicative of different stages of B-cell maturation, and differential overall survival. Similar studies have subsequently been performed on a wide range of leukaemias [32–49] and more recently also at very large scale in other malignancies [50]. Molecular classification of tumours on the basis of gene expression therefore has the capacity to identify previously undetected and clinically significant subtypes of cancer.

3. Gene Expression Profiling Compendia as Resources for Hypothesis Generation

With the ever-increasing availability of expression profiling datasets, efforts have been initiated to collate expression data from diverse studies into unified data collections with a view to providing powerful data mining platforms. One example is represented by BloodExpress, an expression profiling resource for mouse haematopoiesis [28]. Through a user-friendly web interface, BloodExpress allows for searching of uniformly processed microarray datasets. BloodExpress covers both mature and progenitor populations and indeed includes expression data for the majority of all mouse blood cell types. Identification of dynamic gene expression changes is therefore facilitated, as expression data can be retrieved for a variety of differentiation pathways within the well-defined haematopoietic differentiation tree. Both gene-centric and cell-type centric interfaces are implemented, with the latter also allowing for filtering by specific gene functional categories, and thus further facilitating the use of BloodExpress for building novel hypotheses [51–53].

While BloodExpress demonstrated the feasibility of integrating gene expression profiles from diverse laboratories, cross-dataset normalisation did represent a formidable challenge, so that much of the expression information was eventually discretised in a binary "on/off" fashion, thus giving some confidence to expression states but also losing potentially vital information in quantitative expression changes. Parallel efforts have therefore been initiated to generate unified expression profiling resources generated by a single laboratory or facility. One particularly prominent example here is the hematopoietic fingerprints database from the Goodell lab [27]. The hematopoietic fingerprints database contains expression profiles for hematopoietic stem cells as well as their differentiated progeny, such as granulocytes, erythrocytes, natural killer cells, monocytes, activated, and naive T cells and also B cells. The database can be downloaded or accessed through the web, or even through a smartphone application. For their publication, the authors also used the

database to generate novel hypotheses on the transcriptional control of the NK cell lineage and the monocyte lineage, where they were able through functional experiments to implicate the transcription factors Zfp105 and Ets1, respectively, in the differentiation of these two lineages. These results together with many subsequent citations [54–59] of their paper proved the extraordinary utility of this resource.

Comparable expression profiling compendia have more recently also been generated for the human haematopoietic system. The first such study generated a web-accessible resource termed the HaemAtlas [26]. This resource contains gene expression profiles generated from a single centre that cover human erythroblasts, megakaryocytes, B cells, cytotoxic and helper T cells, natural killer cells, granulocytes, and monocytes. Bioinformatic analysis focussing on specific functional categories reported cell type specific characteristics related to transcription factor genes as well as immunoglobulin superfamily members. As the above resources, the HaemAtlas is freely accessible and has therefore played an important role in accelerating hypotheses generation, particularly within the context of genomewide association studies [60–65]. A more extended human expression profiling resource was published subsequently by the lab of Ben Ebert [25]. A total of 38 different human progenitor and mature blood populations are contained within this resource, which also has a web portal known as DMap or differentiation map portal. The free accessibility of this resource again has meant that it has been rapidly used by the community [66–69]. Moreover, a compendium of gene expression profiles for malignant blood cells has also recently been reported [70], which again through a user-intuitive website facilitate data analysis and hypothesis generation. Extensive analysis and visualization tools allow for the integrated analysis of more than 5,800 leukaemia and normal haematopoiesis samples, with easy data retrieval obviating the need for potentially redundant investigations.

A particularly exciting recent development has been an attempt to generate an expression profiling compendia that will allow for an absolute expression quantification [29]. Here more than 10,000 different gene expression profiles were integrated to explore the previously unknown and variable sensitivities of each probe set. The resulting Gene Expression Commons database utilises statistical attributes of each microarray probe (e.g., dynamic range and also threshold) to define absolute expression levels for each gene. The web-based platform represents an implementation with 39 highly purified mouse blood stem/progenitor/differentiated populations and covers almost all of the mouse haematopoietic system. The software is implemented as an open platform, so that individual investigators cannot only explore expression levels of genes or gene families, but also upload their own datasets into the database for cross-comparisons. It is likely that this approach will not only contribute to a very significant progress in haematopoiesis research, but also be applied to many other biomedical research fields, since it appears to overcome some of the previous limitations inherent to cross-comparisons of microarray datasets.

4. Epigenomic Analysis

While a universally accepted definition of epigenetics as well as epigenomics has this far eluded the field, the current consensus view is that modification to the DNA sequence (e.g., methylation) as well as posttranslational modification to chromatin proteins (e.g., histone modifications) represents key aspects of the epigenetic control of gene regulation. Following the advent of high-throughput sequencing technologies, it was quickly realised that genomewide analysis of histone modification status is particularly amenable to this new technology. To this end, the sheared chromatin fragments are subjected to sequencing following immunoprecipitation with suitable antibodies (chromatin immunoprecipitation or ChIP), with the whole technique commonly referred to as ChIP-Seq. Posttranslational modifications indicating both active and repressed transcriptional status are well recognized and have been mapped at genome scale in a variety of both mouse and human blood cell types [71–80].

It is hoped that genome-scale chromatin maps will eventually provide information complementary to gene expression profiling, with implication for human disease prediction, diagnosis, prognosis, and treatment. Indeed, the European Union invested close to 30 million euro in 2011 into its new BLUEPRINT initiative [81]. BLUEPRINT brings together 41 leading European universities, research institutes, and industry entrepreneurs, with the primary goal of mapping chromatin status in human blood cells from healthy and diseased individuals and to provide at least 100 reference epigenomes to the scientific community. Resource-generating activity will be complemented by research into blood-based diseases, including common leukaemias and autoimmune disease (Type 1 Diabetes). The recurring theme therefore is that once again a new ground-breaking initiative uses blood cells as their experimental model of choice.

Genome-scale information on DNA methylation can be obtained using a variety of approaches, with direct sequencing of the entire genomic DNA following bisulfite treatment perhaps representing the most comprehensive approach. Indeed, dramatic improvements and falling costs of high throughput sequencing have made bisulfite sequencing (BS-Seq) a viable option for the global analysis of DNA methylation [82–86]. The well-understood nature of the haematopoietic differentiation tree has again made application of these technologies to blood cells the model of choice for studying the nature of dynamic changes in DNA methylation [83, 87, 88]. Lastly, recent studies demonstrated that ten-eleven translocation (Tet) proteins can catalyze 5methyl-cytosine (5mC) oxidation and generate 5mC derivatives, including 5-hydroxymethylcytosine (5hmC). Importantly, Tet family proteins and 5hmC appear to play key roles in normal development as well as in many diseases, with mutations in human leukaemia patients representing a particularly prominent example [89–103]. Concerted research efforts are currently directed therefore at generating genomewide maps of both the locations of 5hmC as well as the binding sites of Tet proteins [96, 97, 104–110].

5. Genome-Scale Transcription Factor Maps

When using antibodies that specifically recognise transcription factor proteins, the ChIP-Seq technology is also readily amenable to determining genomewide transcription factor binding maps. Indeed, one of the earliest reports using this technology mapped the locations of the NRSF transcriptional regulator in a T-lymphoid cell line [111]. Transcription factor ChIP-Seq studies have now been performed for more than 30 different factors across all major branches of the haematopoietic differentiation tree, as well as a range of mouse and human leukaemia cells and cell lines [51, 68, 76, 78, 79, 112–123]. Important lessons that have been learned from these studies include that (1) high confidence transcription factor binding events highlight functional gene regulatory sequences [117], (2) multifactor studies can reveal previously unrecognised combinatorial interactions between pairs or groups of TFs [118], (3) combinatorial binding events can be used to localise genes that are candidate regulators of developmental processes such as blood cell differentiation [78], and (4) studies from different laboratories can be readily integrated and exploited to perform bioinformatic searches across large numbers of different studies [113].

Transcription factor ChIP-Seq maps have also been utilised to map the genomewide locations of leukaemogenic factors. One recent example mapped the locations of the RUNX1/ETO fusion protein, which is a leukaemia-initiating transcription factor that interferes with RUNX1 function [124]. Global analysis of chromatin status, transcription factor binding, and gene expression showed that RUNX1/ETO controls important regulators of hematopoietic differentiation and self-renewal. Moreover, the removal of RUNX1/ETO re-established the RUNX1 binding profile seen in normal blood cells and also caused the inhibition of leukemic proliferation and self-renewal, thus highlighting the potential therapeutic value of targeting abnormal transcriptional processes in cancer. Other studies have mapped the genomewide locations of the PML-RAR leukaemia oncogene [77, 125], as well as mutant Notch1 [126], which is recognised as one the most commonly mutated genes in T-cell leukaemia [127]. Collectively, these studies have already demonstrated that application of ChIP-Seq technology to the study of transcription factor oncogenes has the potential to provide new mechanistic insights with potential therapeutic value.

6. Genomewide Association Studies

Genomewide association studies (GWASs) examine many common genetic variants in different individuals to see if any variant is associated with a particular trait. GWAS surveys typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits for major diseases, but can also identify variants that are linked to phenotypic variation seen in normal populations. Blood parameters such as the numbers of a particular blood cell type per millilitre of blood vary between different individuals, and much of this variation is thought to be genetic. Since elevated or reduced levels of certain blood cell types can predispose to diseases, GWAS studies have been performed to identify some of the underlying variants. In particular, blood platelets play a pivotal role in cardiovascular disease through their involvement in atherothrombosis following coronary artery plaque rupture. Not only does the platelet response to such an event vary between individuals, but also this variation is largely genetically controlled. A recent integrated systems biology approach performed high density genotyping of 110 genes in a cohort of more than 500 individuals with known levels of platelet response, which was followed by gene expression profiling and proteomics studies on platelets from individuals with the so-called "extreme end" response phenotypes [128]. With the advent of high density (i.e., 500,000 SNPs) genotyping arrays, a large number of case and control samples can now be tested at an affordable cost. By making use of the recently completed Wellcome Trust Case Control Consortium (WTCCC) study, common sequence variants that are associated with abnormal blood parameters and/or confer risk to myocardial infarction have been identified. Total white blood cell count and its constituent subtypes were recently used to study several 10,000 subjects [129, 130], to determine genetic factors influencing variability within the blood parameters. Ten variants associated with total white blood cell count were identified with additional variants associated with variation in specific blood cell types. Possible functional relationships between the relevant candidate genes were predicted by integrating gene expression and pathways-based analyses, which revealed functional connectivity among implicated loci.

GWAS studies are commonly followed up by meta-analysis of several independent studies. Analysis of 66,867 individuals of European ancestry followed by extensive biological and functional assessment identified 68 loci reliably associated with platelet count and volume [130]. Expression profiling demonstrated a tendency for lineage specific expression within the haematopoietic differentiation tree for the genes identified by GWAS. Using a functional followup in zebrafish and drosophila, 11 of these genes were validated as novel regulators of blood cell formation, thus providing an example of successful translation of GWAS studies to generating new functional insights.

7. Cancer Genome Sequencing

Identification of the whole complement of genes that are mutated in human cancers and therefore drive malignant transformation has been a central aim of cancer research ever since the advent of recombinant DNA technology. Following the development of ultra-high-throughput sequencing technologies and the subsequent reductions in the costs of sequencing whole genomes, projects have been initiated in Europe, America, and Asia to decode entire genomes from multiple individuals for a wide range of human cancers. Identification of somatically acquired mutations is thought to provide the most direct route towards delineating genes critical to the development of human cancers, yet it is not entirely clear at this stage what will be the most reliable techniques to distinguish these so-called "driver" mutations from the so-called "passenger" mutations; that are also found

in the tumour but occurred by chance without providing a selective advantage to tumour growth and/or survival.

Within haematopoiesis research, cancer genome sequencing has now been applied to several different haematological malignancies [69, 131–145]. For example, multiple myeloma which is an incurable malignancy of plasma cells, was investigated by the sequencing of 38 tumour genomes and their comparison to matched normal DNAs, which revealed several new and unexpected putative oncogenic mechanisms [146]. These included mutations of genes involved in protein translation, histone methylation, and blood coagulation, thus demonstrating that cancer genome sequencing of large collections of samples can yield new and previously unsuspected insights into cancer.

A more recent study investigating cancer genomes in patients with acute myeloid leukaemia (AML) specifically addressed the question of passenger and driver mutations [147]. Of note, normal karyotypes are common and genomic instability is unusual in AML. By comparing AML samples with a known initiating event (PML-RARA) with normal karyotype AML samples and nonleukaemic blood stem/progenitor cells, this study suggests that most mutations in AML genomes are random events that occurred prior to the initiating leukaemogenic mutation, and that in many cases, only one or two additional, cooperating mutations are needed to generate the malignant founding clone. Cells from the founding clone can acquire additional cooperating mutations, yielding subclones that can contribute to disease progression and/or relapse, thus making clonal analysis of cancers an important goal of current research efforts. Indeed, it was again an analysis using blood cells as a model which provided an important recent breakthrough. The particular study in question reported a method for sequencing complete exomes from single cells [148]. The authors went on to use this method to perform whole-exome single-cell sequencing for 58 single cells from a patient with myeloproliferative neoplasm. This analysis suggested that in this particular patient, the neoplasm followed the pathway of monoclonal evolution. Importantly, this technological breakthrough now sets the scene for similar analyses in other human leukaemias as well as solid cancers.

8. Outlook

The application of new genome-scale technologies has resulted in an unprecedented increase in the amounts of data available for hypothesis generation in biomedical research. However, the extent to which this explosion in data has contributed to an actual increase in mechanistic understanding has so far been less impressive. To some extent this is to be expected, since the new datasets have been available at most for a few years, and mechanistic studies often have longer timescales. However, there are also cultural, training/educational, and technical issues that need to be overcome to accelerate the exploitation of new datasets.

In terms of research culture, it has become apparent that many investigators have found a comfortable niche for themselves, where essentially they end up with repeating the same data collection exercise every couple of years, each time with the next version of whatever genome-scale technology they have become an expert in (an example here would be remapping SNPs when the next version SNP array becomes available, or performing ChIP-on-chip analysis one year and follow this with ChIP-Seq of more or less the same samples the next years). This however avoids asking the much harder question whether it is possible to decipher the meaning of any underlying biology from a genome-scale observation and how this might be exploited to deliver actual biomedical advances. Journals and funding agencies should pay more attention to this issue and provide better rewards to those investigators willing to tackle the harder mechanistic questions. Just because an experiment has cost millions of dollars and used the latest technology, this on its own, does not mean that it will have a lasting value.

The educational issue is that the exploitation of genome-scale datasets requires biologists with a robust understanding of bioinformatics and statistics and ideally also a significant knowledge of computer programming languages. Universities are increasingly offering relevant courses, yet the demand for bioinformatically trained biologists is still an outstripping supply. Moreover, the longer-term career path for such scientists is not clear either, at least in academic settings, because of the tensions between the need to performing primary, intellectual novel research in order to achieve academic promotions, which does not blend well with carrying out a bioinformatic support function.

Finally, there are also scientific conceptual issues that impede interpretation and thus exploitation of genome-scale datasets. For example, it is clear that cell fate decisions are made by individual cells, and that indeed there is substantial heterogeneity within a given biological cell population [149–152]. Genome-scale technologies however often require the need to generate material from many thousands to millions of cells and thus can only report population averages. Some recent progress has been achieved at the level of single-cell RNA sequencing and exome sequencing [148, 153–155]. However, data will need to be generated for 100s if not 1,000s of single cells to make sure that the full heterogeneity of a given population has been sampled. Another important issue is that the genome is not a linear sequence, but instead organised in a complex 3-dimensional fashion [156–158]. Chromatin maps as well as transcription factor maps and gene expression data therefore need to be integrated with information on the 3-dimensional structure of the genome. Importantly, there appears to be a rapid progress in our ability to comprehensively map chromosome conformation at genomic scales [159–163]. However, it is likely that many long-range chromosome interactions are not particularly rigid and may indeed be rather transient. Measurements generated from cell populations therefore likely reflect the ensemble of common interactions and will therefore need to be deconvoluted so as to obtain not just one possible "solution" but instead a number of possible 3-D interaction maps. And as with many of the other techniques covered in this paper, the data generation phase is of a descriptive rather than functional nature, so that the mere description of chromosome loops will not provide direct proof of functional relevance.

Given the rapid pace of recent technological innovations particularly in terms of generating descriptive data (the various genome-scale maps), the major roadblocks will be to enhance the throughput of downstream functional studies. Importantly, many of the technologies developed to generate the largely descriptive genome-scale maps can also be adapted to multiplex and thus accelerate the analysis of downstream functional assays. For example, high-throughput next generation sequencing can be adapted to many biological assays that require counting. Reporter gene assays can be replaced by measuring transcript abundance rather than luciferase or lacZ enzyme activity, and through the inclusion of sequence tags, many different promoters can be assayed simultaneously [164]. Another exciting area of recent progress concerns the application of transcription activator-like effector nucleases (TALENs) to perform highly efficient modifications of the genomes, and thus generates all sorts of mutant alleles both in cell lines and also in vivo [165, 166]. The likelihood therefore is that the next decade will not only bring an ever increased rate of generating descriptive whole-genome datasets, but also show substantial acceleration in the generation of new biological insights.

References

[1] D. Bryder, D. J. Rossi, and I. L. Weissman, "Hematopoietic stem cells: the paradigmatic tissue-specific stem cell," *American Journal of Pathology*, vol. 169, no. 2, pp. 338–346, 2006.

[2] D. Miranda-Saavedra and B. Göttgens, "Transcriptional regulatory networks in haematopoiesis," *Current Opinion in Genetics and Development*, vol. 18, no. 6, pp. 530–535, 2008.

[3] S. H. Orkin and L. I. Zon, "Hematopoiesis: an evolving paradigm for stem cell biology," *Cell*, vol. 132, no. 4, pp. 631–644, 2008.

[4] J. E. Pimanda and G. Berthold, "Gene regulatory networks governing haematopoietic stem cell development and identity," *International Journal of Developmental Biology*, vol. 54, no. 6-7, pp. 1201–1211, 2010.

[5] N. K. Wilson, F. J. Calero-Nieto, R. Ferreira, and B. Göttgens, "Transcriptional regulation of haematopoietic transcription factors," *Stem Cell Research & Therapy*, vol. 2, no. 1, article 6, 2011.

[6] M. Osawa, K. I. Hanada, H. Hamada, and H. Nakauchi, "Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell," *Science*, vol. 273, no. 5272, pp. 242–245, 1996.

[7] D. G. Tenen, "Disruption of differentiation in human cancer: AML shows the way," *Nature Reviews Cancer*, vol. 3, no. 2, pp. 89–101, 2003.

[8] F. Rosenbauer and D. G. Tenen, "Transcription factors in myeloid development: balancing differentiation with transformation," *Nature Reviews Immunology*, vol. 7, no. 2, pp. 105–117, 2007.

[9] D. G. Tenen, R. Hromas, J. D. Licht, and D. E. Zhang, "Transcription factors, normal myeloid development, and leukemia," *Blood*, vol. 90, no. 2, pp. 489–519, 1997.

[10] S. Koschmieder, F. Rosenbauer, U. Steidl, B. M. Owens, and D. G. Tenen, "Role of transcription factors C/EBPα and PU.1 in normal hematopoiesis and leukemia," *International Journal of Hematology*, vol. 81, no. 5, pp. 368–377, 2005.

[11] Y. Shima and I. Kitabayashi, "Deregulated transcription factors in leukemia," *International Journal of Hematology*, vol. 94, no. 2, pp. 134–141, 2011.

[12] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[13] J. C. Venter, M. D. Adams, E. W. Myers et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[14] K. Basso, A. A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, and A. Califano, "Reverse engineering of regulatory networks in human B cells," *Nature Genetics*, vol. 37, no. 4, pp. 382–390, 2005.

[15] L. Agnelli, M. Forcato, F. Ferrari et al., "The reconstruction of transcriptional networks reveals critical genes with implications for clinical outcome of multiple myeloma," *Clinical Cancer Research*, vol. 17, no. 23, pp. 7402–7412, 2011.

[16] J. D. Allen, Y. Xie, M. Chen, L. Girard, and G. Xiao, "Comparing statistical methods for constructing large scale gene networks," *PLoS One*, vol. 7, no. 1, article e29348, 2012.

[17] G. Altay and F. Emmert-Streib, "Inferring the conservative causal core of gene regulatory networks," *BMC Systems Biology*, vol. 4, article 132, 2010.

[18] K. Basso, M. Saito, P. Sumazin et al., "Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells," *Blood*, vol. 115, no. 5, pp. 975–984, 2010.

[19] M. Cadeiras, M. von Bayern, A. Sinha et al., "Drawing networks of rejection—a systems biological approach to the identification of candidate genes in heart transplantation," *Journal of Cellular and Molecular Medicine*, vol. 15, no. 4, pp. 949–956, 2011.

[20] M. Grimaldi, R. Visintainer, and G. Jurman, "RegnANN: reverse engineering gene networks using artificial neural networks," *PLoS One*, vol. 6, no. 12, article e28646, 2011.

[21] R. Mahdi, A. S. Madduri, G. Wang et al., "Empirical Bayes conditional independence graphs for regulatory network recovery," *Bioinformatics*, vol. 28, no. 15, pp. 2029–2036, 2012.

[22] A. A. Margolin, I. Nemenman, K. Basso et al., "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics*, vol. 7, supplement 1, article S7, 2006.

[23] A. A. Margolin, K. Wang, W. K. Lim, M. Kustagi, I. Nemenman, and A. Califano, "Reverse engineering cellular networks," *Nature Protocols*, vol. 1, no. 2, pp. 662–671, 2006.

[24] P. E. Meyer, K. Kontos, F. Lafitte, and G. Bontempi, "Information-theoretic inference of large transcriptional regulatory networks," *Eurasip Journal on Bioinformatics and Systems Biology*, vol. 2007, Article ID 79879, 2007.

[25] N. Novershtern, A. Subramanian, L. N. Lawton et al., "Densely interconnected transcriptional circuits control cell states in human hematopoiesis," *Cell*, vol. 144, no. 2, pp. 296–309, 2011.

[26] N. A. Watkins, A. Gusnanto, B. De Bono et al., "A HaemAtlas: characterizing gene expression in differentiated human blood cells," *Blood*, vol. 113, no. 19, pp. e1–e9, 2009.

[27] S. M. Chambers, N. C. Boles, K. Y. K. Lin et al., "Hematopoietic fingerprints: an expression database of stem cells and their progeny," *Cell Stem Cell*, vol. 1, no. 5, pp. 578–591, 2007.

[28] D. Miranda-Saavedra, S. De, M. W. Trotter, S. A. Teichmann, and B. Göttgens, "BloodExpress: a database of gene expression in mouse haematopoiesis," *Nucleic Acids Research*, vol. 37, no. 1, pp. D873–D879, 2009.

[29] J. Seita, D. Sahoo, D. J. Rossi et al., "Gene expression commons: an open platform for absolute gene expression profiling," *PLoS One*, vol. 7, no. 7, article e40321, 2012.

[30] F. O. Bagger, N. Rapin, K. Theilgaard-Mönch et al., "HemaExplorer: a Web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis," *Blood*, vol. 119, no. 26, pp. 6394–6395, 2012.

[31] A. A. Alizadeh, M. B. Elsen, R. E. Davis et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[32] A. Andersson, C. Ritz, D. Lindgren et al., "Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status," *Leukemia*, vol. 21, no. 6, pp. 1198–1203, 2007.

[33] B. V. Balgobind, M. M. van den Heuvel-Eibrink, R. X. De Menezes et al., "Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia," *Haematologica*, vol. 96, no. 2, pp. 221–230, 2011.

[34] A. Barski, S. Cuddapah, K. Cui et al., "High-resolution profiling of histone methylations in the human genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.

[35] K. M. Bernt, N. Zhu, A. U. Sinha et al., "MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L," *Cancer Cell*, vol. 20, no. 1, pp. 66–78, 2011.

[36] M. Camós, J. Esteve, P. Jares et al., "Gene expression profiling of acute myeloid leukemia with translocation t(8;16)(p11;p13) and MYST3-CREBBP rearrangement reveals a distinctive signature with a specific pattern of HOX gene expression," *Cancer Research*, vol. 66, no. 14, pp. 6947–6954, 2006.

[37] E. Chen, P. A. Beer, A. L. Godfrey et al., "Distinct clinical phenotypes associated with JAK2V617F reflect differential STAT1 signaling," *Cancer Cell*, vol. 18, no. 5, pp. 524–535, 2010.

[38] E. Coustan-Smith, C. G. Mullighan, M. Onciu et al., "Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia," *The Lancet Oncology*, vol. 10, no. 2, pp. 147–156, 2009.

[39] N. C. Gutiérrez, R. López-Pérez, J. M. Hernández et al., "Gene expression profile reveals deregulation of genes with relevant functions in the different subclasses of acute myeloid leukemia," *Leukemia*, vol. 19, no. 3, pp. 402–409, 2005.

[40] I. Homminga, R. Pieters, A. W. Langerak et al., "Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia," *Cancer Cell*, vol. 19, no. 4, pp. 484–497, 2011.

[41] L. Huang, K. Zhou, Y. Yang et al., "FLT3-ITD-associated gene-expression signatures in NPM1-mutated cytogenetically normal acute myeloid leukemia," *International Journal of Hematology*, vol. 96, no. 2, pp. 234–240, 2012.

[42] A. Kohlmann, C. Schoch, S. Schnittger et al., "Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients," *Leukemia*, vol. 18, no. 1, pp. 63–71, 2004.

[43] Z. Li, W. Zhang, M. Wu et al., "Gene expression-based classification and regulatory networks of pediatric acute lymphoblastic leukemia," *Blood*, vol. 114, no. 20, pp. 4486–4493, 2009.

[44] J. Nordlund, A. Kiialainen, O. Karlberg et al., "Digital gene expression profiling of primary acute lymphoblastic leukemia cells," *Leukemia*, vol. 26, no. 6, pp. 1218–1227, 2012.

[45] I. Oschlies, W. Klapper, M. Zimmermann et al., "Diffuse large B-cell lymphoma in pediatric patients belongs predominantly to the germinal-center type B-cell lymphomas: a clinicopathologic analysis of cases included in the German BFM (Berlin-Frankfurt-Münster) Multicenter Trial," *Blood*, vol. 107, no. 10, pp. 4047–4052, 2006.

[46] A. Rosenwald, A. A. Alizadeh, G. Widhopf et al., "Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia," *Journal of Experimental Medicine*, vol. 194, no. 11, pp. 1639–1647, 2001.

[47] M. E. Ross, X. Zhou, G. Song et al., "Classification of pediatric acute lymphoblastic leukemia by gene expression profiling," *Blood*, vol. 102, no. 8, pp. 2951–2959, 2003.

[48] C. Schoch, W. Kern, A. Kohlmann, W. Hiddemann, S. Schnittger, and T. Haferlach, "Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile," *Genes Chromosomes and Cancer*, vol. 43, no. 3, pp. 227–238, 2005.

[49] F. P. G. Silva, S. M. A. Swagemakers, C. Erpelinck-Verschueren et al., "Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status," *Blood*, vol. 114, no. 14, pp. 3001–3007, 2009.

[50] C. Curtis, S. P. Shah, S. F. Chin et al., "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, no. 7403, pp. 346–352, 2012.

[51] F. J. Calero-Nieto, A. D. Wood, N. K. Wilson, S. Kinston, J. R. Landry, and B. Göttgens, "Transcriptional regulation of Elf-1: locus-wide analysis reveals four distinct promoters, a tissue-specific enhancer, control by PU.1 and the importance of Elf-1 downregulation for erythroid maturation," *Nucleic Acids Research*, vol. 38, no. 19, pp. 6363–6374, 2010.

[52] R. K. Hyde, Y. Kamikubo, S. Anderson et al., "Cbfb/Runx1 repression-independent blockage of differentiation and accumulation of Csf2rb-expressing cells by Cbfb-MYH11," *Blood*, vol. 115, no. 7, pp. 1433–1443, 2010.

[53] M. R. Kantorovitz, M. Kazemian, S. Kinston et al., "Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse," *Developmental Cell*, vol. 17, no. 4, pp. 568–579, 2009.

[54] W. W. Pang, E. A. Price, D. Sahoo et al., "Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age," *Proceedings of the National Academy of Sciences of USA*, vol. 108, no. 50, article 20012-7, 2011.

[55] X. Wang, M. B. Werneck, B. G. Wilson et al., "TCR-dependent transformation of mature memory phenotype T cells in mice," *The Journal of Clinical Investigation*, vol. 121, no. 10, pp. 3834–3845, 2011.

[56] Y. Benita, Z. Cao, C. Giallourakis, C. Li, A. Gardet, and R. J. Xavier, "Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor," *Blood*, vol. 115, no. 26, pp. 5376–5384, 2010.

[57] D. Laifenfeld, A. Gilchrist, D. Drubin et al., "The role of hypoxia in 2-butoxyethanol-Induced hemangiosarcoma," *Toxicological Sciences*, vol. 113, no. 1, pp. 254–266, 2009.

[58] K. S. Poland, D. L. Shardy, M. Azim et al., "Overexpression of ZNF342 by juxtaposition with MPO promoter/enhancer in the novel translocation t(17;19)(q23;q13.32) in pediatric acute myeloid leukemia and analysis of ZNF342 expression in

leukemia," *Genes Chromosomes and Cancer*, vol. 48, no. 6, pp. 480–489, 2009.

[59] A. Gerrits, B. Dykstra, M. Otten, L. Bystrykh, and G. de Haan, "Combining transcriptional profiling and genetic linkage analysis to uncover gene networks operating in hematopoietic stem cells and their progeny," *Immunogenetics*, vol. 60, no. 8, pp. 411–422, 2008.

[60] R. Qayyum, B. M. Snively, E. Ziv et al., "A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans," *PLOS Genetics*, vol. 8, no. 3, Article ID e1002491, 2012.

[61] S. Sivapalaratnam, H. Basart, N. A. Watkins et al., "Monocyte gene expression signature of patients with early onset coronary artery disease," *PLoS One*, vol. 7, no. 2, article e32166, 2012.

[62] M. Rotival, T. Zeller, P. S. Wild et al., "Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans," *PLOS Genetics*, vol. 7, no. 12, Article ID e1002367, 2011.

[63] R. S. Fehrmann, R. C. Jansen, J. H. Veldink et al., "Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA," *PLOS Genetics*, vol. 7, no. 8, Article ID e1002197, 2011.

[64] A. H. Goodall, P. Burns, I. Salles et al., "Transcription profiling in human platelets reveals LRRFIP1 as a novel protein regulating platelet function," *Blood*, vol. 116, no. 22, pp. 4646–4656, 2010.

[65] S. K. Ganesh, N. A. Zakai, F. J. A. Van Rooij et al., "Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium," *Nature Genetics*, vol. 41, no. 11, pp. 1191–1198, 2009.

[66] D. Wang, A. Rendon, W. Ouwehand, and L. Wernisch, "Transcription factor co-localization patterns affect human cell type-specific gene expression," *BMC Genomics*, vol. 13, article 263, 2012.

[67] T. Suzuki, M. Nakano-Ikegaya, H. Yabukami-Okuda et al., "Reconstruction of monocyte transcriptional regulatory network accompanies monocytic functions in human fibroblasts," *PLoS One*, vol. 7, no. 3, article e33474, 2012.

[68] S. N. Wontakal, X. Guo, C. Smith et al., "A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation," *Proceedings of the National Academy of Sciences of USA*, vol. 109, no. 10, pp. 3832–3837, 2012.

[69] J. Zhang, L. Ding, L. Holmfeldt et al., "The genetic basis of early T-cell precursor acute lymphoblastic leukaemia," *Nature*, vol. 481, no. 7380, pp. 157–163, 2012.

[70] K. Hebestreit, S. Gröttrup, D. Emden et al., "Leukemia gene atlas—a public platform for integrative exploration of genome-wide molecular data," *PLoS One*, vol. 7, no. 6, article e39148, 2012.

[71] M. Adli, J. Zhu, and B. E. Bernstein, "Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors," *Nature Methods*, vol. 7, no. 8, pp. 615–618, 2010.

[72] N. Bonadies, S. D. Foster, W. I. Chan et al., "Genome-wide analysis of transcriptional reprogramming in mouse models of acute myeloid leukaemia," *PLoS One*, vol. 6, no. 1, article e16330, 2011.

[73] Y. Cheng, W. Wu, S. A. Kumar et al., "Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression," *Genome Research*, vol. 19, no. 12, pp. 2172–2184, 2009.

[74] M. A. Dawson, A. J. Bannister, B. Göttgens et al., "JAK2 phosphorylates histone H3Y41 and excludes HP1α from chromatin," *Nature*, vol. 461, no. 7265, pp. 819–822, 2009.

[75] Y. He, Y. Yu, Y. Zhang et al., "Genome-wide bovine H3K27me3 modifications and the regulatory effects on genes expressions in peripheral blood lymphocytes," *PLoS One*, vol. 7, no. 6, article e39094, 2012.

[76] P. C. Hollenhorst, K. J. Chandler, R. L. Poulsen, W. E. Johnson, N. A. Speck, and B. J. Graves, "DNA specificity determinants associate with distinct transcription factor functions," *PLoS Genetics*, vol. 5, no. 12, Article ID e1000778, 2009.

[77] S. Saeed, C. Logie, K. J. Francoijs et al., "Chromatin accessibility, p300 and histone acetylation define PML-RARα and AML1-ETO binding sites in acute myeloid leukemia," *Blood*, vol. 120, no. 15, pp. 3058–3068, 2012.

[78] M. R. Tijssen, A. Cvejic, A. Joshi et al., "Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators," *Developmental Cell*, vol. 20, no. 5, pp. 597–609, 2011.

[79] J. A. Zhang, A. Mortazavi, B. A. Williams, B. J. Wold, and E. V. Rothenberg, "Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity," *Cell*, vol. 149, no. 2, pp. 467–482, 2012.

[80] P. Wong, S. M. Hattangadi, A. W. Cheng, G. M. Frampton, R. A. Young, and H. F. Lodish, "Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes," *Blood*, vol. 118, no. 16, pp. e128–e138, 2011.

[81] D. Adams, L. Altucci, S. E. Antonarakis et al., "BLUEPRINT to decode the epigenetic signature written in blood," *Nature Biotechnology*, vol. 30, no. 3, pp. 224–226, 2012.

[82] C. Bock, E. Kiskinis, G. Verstappen et al., "Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines," *Cell*, vol. 144, no. 3, pp. 439–452, 2011.

[83] C. Bock, I. Beerman, W. H. Lien et al., "DNA methylation dynamics during in vivo differentiation of blood and skin stem cells," *Molecular Cell*, vol. 47, no. 4, pp. 633–647, 2012.

[84] A. Meissner, T. S. Mikkelsen, H. Gu et al., "Genome-scale DNA methylation maps of pluripotent and differentiated cells," *Nature*, vol. 454, no. 7205, pp. 766–770, 2008.

[85] R. Lister, M. Pelizzola, R. H. Dowen et al., "Human DNA methylomes at base resolution show widespread epigenomic differences," *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.

[86] R. Lister, M. Pelizzola, Y. S. Kida et al., "Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells," *Nature*, vol. 471, no. 7336, pp. 68–73, 2011.

[87] E. Hodges, A. Molaro, C. O. Dos Santos et al., "Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment," *Molecular Cell*, vol. 44, no. 1, pp. 17–28, 2011.

[88] A. Hogart, J. Lichtenberg, S. S. Ajay et al., "Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites," *Genome Research*, vol. 22, no. 8, pp. 1407–1418, 2012.

[89] P. A. Beer, F. Delhommeau, J. P. LeCouédic et al., "Two routes to leukemic transformation after a JAK2 mutation-positive myeloproliferative neoplasm," *Blood*, vol. 115, no. 14, pp. 2891–2900, 2010.

[90] W. C. Chou, S. C. Chou, C. Y. Liu et al., "TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics," *Blood*, vol. 118, no. 14, pp. 3803–3810, 2011.

[91] F. Delhommeau, S. Dupont, V. Della Valle et al., "Mutation in TET2 in myeloid cancers," *The New England Journal of Medicine*, vol. 360, no. 22, pp. 2289–2301, 2009.

[92] A. Tefferi, K. H. Lim, O. Abdel-Wahab et al., "Detection of mutant *TET2* in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML," *Leukemia*, vol. 23, no. 7, pp. 1343–1345, 2009.

[93] M. E. Figueroa, O. Abdel-Wahab, C. Lu et al., "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation," *Cancer Cell*, vol. 18, no. 6, pp. 553–567, 2010.

[94] L. Holmfeldt and C. G. Mullighan, "The role of TET2 in hematologic neoplasms," *Cancer Cell*, vol. 20, no. 1, pp. 1–2, 2011.

[95] A. M. Jankowska, H. Makishima, R. V. Tiu et al., "Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A," *Blood*, vol. 118, no. 14, pp. 3932–3941, 2011.

[96] M. Ko, Y. Huang, A. M. Jankowska et al., "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2," *Nature*, vol. 468, no. 7325, pp. 839–843, 2010.

[97] M. Ko, H. S. Bandukwala, J. An et al., "Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 35, pp. 14566–14571, 2011.

[98] O. Kosmider, V. Gelsi-Boyer, M. Ciudad et al., "TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia," *Haematologica*, vol. 94, no. 12, pp. 1676–1681, 2009.

[99] S. M. C. Langemeijer, J. H. Jansen, J. Hooijer et al., "TET2 mutations in childhood leukemia," *Leukemia*, vol. 25, no. 1, pp. 189–192, 2011.

[100] C. G. Mullighan, "TET2 mutations in myelodysplasia and myeloid malignancies," *Nature Genetics*, vol. 41, no. 7, pp. 766–767, 2009.

[101] C. Saint-Martin, G. Leroy, F. Delhommeau et al., "Analysis of the ten-eleven translocation 2 (TET2) gene in familial myeloproliferative neoplasms," *Blood*, vol. 114, no. 8, pp. 1628–1632, 2009.

[102] A. H. Shih, O. Abdel-Wahab, J. P. Patel, and R. L. Levine, "The role of mutations in epigenetic regulators in myeloid malignancies," *Nature Reviews Cancer*, vol. 12, no. 9, pp. 599–612, 2012.

[103] S. Weissmann, T. Alpermann, V. Grossmann et al., "Landscape of TET2 mutations in acute myeloid leukemia," *Leukemia*, vol. 26, no. 5, pp. 934–942, 2012.

[104] G. Ficz, M. R. Branco, S. Seisenberger et al., "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation," *Nature*, vol. 473, no. 7347, pp. 398–404, 2011.

[105] S. Ito, A. C. Dalessio, O. V. Taranova, K. Hong, L. C. Sowers, and Y. Zhang, "Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification," *Nature*, vol. 466, no. 7310, pp. 1129–1133, 2010.

[106] K. P. Koh, A. Yabuuchi, S. Rao et al., "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells," *Cell Stem Cell*, vol. 8, no. 2, pp. 200–213, 2011.

[107] A. A. Sérandour, S. Avner, F. Oger et al., "Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers," *Nucleic Acids Research*, vol. 40, no. 17, pp. 8255–8265, 2012.

[108] H. Wu, A. C. D'Alessio, S. Ito et al., "Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells," *Genes and Development*, vol. 25, no. 7, pp. 679–684, 2011.

[109] H. Wu and Y. Zhang, "Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells," *Cell Cycle*, vol. 10, no. 15, pp. 2428–2436, 2011.

[110] Y. Xu, F. Wu, L. Tan et al., "Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells," *Molecular Cell*, vol. 42, no. 4, pp. 451–464, 2011.

[111] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold, "Genome-wide mapping of in vivo protein-DNA interactions," *Science*, vol. 316, no. 5830, pp. 1497–1502, 2007.

[112] T. Fujiwara, H. O'Geen, S. Keles et al., "Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy," *Molecular Cell*, vol. 36, no. 4, pp. 667–681, 2009.

[113] R. Hannah, A. Joshi, N. K. Wilson, S. Kinston, and B. Göttgens, "A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms," *Experimental Hematology*, vol. 39, no. 5, pp. 531–541, 2011.

[114] M. T. Kassouf, J. R. Hughes, S. Taylor et al., "Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells," *Genome Research*, vol. 20, no. 8, pp. 1064–1083, 2010.

[115] L. Li, R. Jothi, K. Cui et al., "Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells," *Nature Immunology*, vol. 12, no. 2, pp. 129–136, 2011.

[116] C. G. Palii, C. Perez-Iratxeta, Z. Yao et al., "Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages," *EMBO Journal*, vol. 30, no. 3, pp. 494–509, 2011.

[117] N. K. Wilson, D. Miranda-Saavedra, S. Kinston et al., "The transcriptional program controlled by the stem cell leukemia gene Scl /Tal1 during early embryonic hematopoietic development," *Blood*, vol. 113, no. 22, pp. 5456–5465, 2009.

[118] N. K. Wilson, S. D. Foster, X. Wang et al., "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators," *Cell Stem Cell*, vol. 7, no. 4, pp. 532–544, 2010.

[119] M. Garber, N. Yosef, A. Goren et al., "A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals," *Molecular Cell*, vol. 47, no. 5, pp. 810–822, 2012.

[120] A. M. Pilon, S. S. Ajay, S. A. Kumar et al., "Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation," *Blood*, vol. 118, no. 17, pp. e139–e148, 2011.

[121] S. N. Wontakal, X. Guo, B. Will et al., "A large gene network in immature Erythroid cells is controlled by the myeloid and B cell transcriptional regulator PU.1," *PLoS Genetics*, vol. 7, no. 6, Article ID e1001392, 2011.

[122] L. Durant, W. T. Watford, H. L. Ramos et al., "Diverse targets of the transcription factor STAT3 contribute to T cell

pathogenicity and homeostasis," *Immunity*, vol. 32, no. 5, pp. 605–615, 2010.

[123] H. Kwon, D. Thierry-Mieg, J. Thierry-Mieg et al., "Analysis of interleukin-21-induced Prdm1 gene regulation reveals functional cooperation of STAT3 and IRF4 transcription factors," *Immunity*, vol. 31, no. 6, pp. 941–952, 2009.

[124] A. Ptasinska, S. A. Assi, D. Mannari et al., "Depletion of RUNX1/ETO in t(8,21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding," *Leukemia*, vol. 26, no. 8, pp. 1829–1841, 2012.

[125] J. H. A. Martens, A. B. Brinkman, F. Simmer et al., "PML-RARα/RXR alters the epigenetic landscape in acute promyelocytic leukemia," *Cancer Cell*, vol. 17, no. 2, pp. 173–185, 2010.

[126] H. Wang, J. Zou, B. Zhao et al., "Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 36, pp. 14908–14913, 2011.

[127] I. Aifantis, E. Raetz, and S. Buonamici, "Molecular pathogenesis of T-cell leukaemia and lymphoma," *Nature Reviews Immunology*, vol. 8, no. 5, pp. 380–390, 2008.

[128] N. Soranzo, T. D. Spector, M. Mangino et al., "A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium," *Nature Genetics*, vol. 41, no. 11, pp. 1182–1190, 2009.

[129] M. A. Nalls, D. J. Couper, T. Tanaka et al., "Multiple loci are associated with white blood cell phenotypes," *PLOS Genetics*, vol. 7, no. 6, Article ID e1002113, 2011.

[130] C. Gieger, A. Radhakrishnan, A. Cvejic et al., "New gene functions in megakaryopoiesis and platelet formation," *Nature*, vol. 480, no. 7376, pp. 201–208, 2011.

[131] P. A. Greif, A. Dufour, N. P. Konstandin et al., "GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia," *Blood*, vol. 120, no. 2, pp. 395–403, 2012.

[132] H. L. Koskela, S. Eldfors, P. Ellonen et al., "Somatic STAT3 mutations in large granular lymphocytic leukemia," *The New England Journal of Medicine*, vol. 366, no. 20, pp. 1905–1913, 2012.

[133] V. Quesada, L. Conde, N. Villamor et al., "Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia," *Nature Genetics*, vol. 44, no. 1, pp. 47–52, 2012.

[134] L. Wang, M. S. Lawrence, Y. Wan et al., "SF3B1 and other novel cancer genes in chronic lymphocytic leukemia," *The New England Journal of Medicine*, vol. 365, no. 26, pp. 2497–2506, 2011.

[135] V. Grossmann, E. Tiacci, A. B. Holmes et al., "Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype," *Blood*, vol. 118, no. 23, pp. 6153–6163, 2011.

[136] K. Yoshida, M. Sanada, Y. Shiraishi et al., "Frequent pathway mutations of splicing machinery in myelodysplasia," *Nature*, vol. 478, no. 7367, pp. 64–69, 2011.

[137] X. J. Yan, J. Xu, Z. H. Gu et al., "Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia," *Nature Genetics*, vol. 43, no. 4, pp. 309–317, 2011.

[138] R. E. Dickinson, H. Griffin, V. Bigley et al., "Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency," *Blood*, vol. 118, no. 10, pp. 2656–2658, 2011.

[139] E. Tiacci, V. Trifonov, G. Schiavoni et al., "BRAF mutations in hairy-cell leukemia," *The New England Journal of Medicine*, vol. 364, no. 24, pp. 2305–2315, 2011.

[140] B. A. Walker, C. P. Wardell, L. Melchor et al., "Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4,14) and t(11,14) myeloma," *Blood*, vol. 120, no. 5, pp. 1077–1086, 2012.

[141] M. J. Walter, D. Shen, L. Ding et al., "Clonal architecture of secondary acute myeloid leukemia," *The New England Journal of Medicine*, vol. 366, no. 12, pp. 1090–1098, 2012.

[142] D. C. Link, L. G. Schuettpelz, D. Shen et al., "Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML," *Journal of the American Medical Association*, vol. 305, no. 15, pp. 1568–1576, 2011.

[143] L. D. Wartman, D. E. Larson, Z. Xiang et al., "Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression," *The Journal of Clinical Investigation*, vol. 121, no. 4, pp. 1445–1455, 2011.

[144] J. B. Egan, C. X. Shi, W. Tembe et al., "Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides," *Blood*, vol. 120, no. 5, pp. 1060–1066, 2012.

[145] X. S. Puente, M. Pinyol, V. Quesada et al., "Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia," *Nature*, vol. 475, no. 7354, pp. 101–105, 2011.

[146] M. A. Chapman, M. S. Lawrence, J. J. Keats et al., "Initial genome sequencing and analysis of multiple myeloma," *Nature*, vol. 471, no. 7339, pp. 467–472, 2011.

[147] J. S. Welch, T. J. Ley, D. C. Link et al., "The origin and evolution of mutations in acute myeloid leukemia," *Cell*, vol. 150, no. 2, pp. 264–278, 2012.

[148] Y. Hou, L. Song, P. Zhu et al., "Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm," *Cell*, vol. 148, no. 5, pp. 873–885, 2012.

[149] J. P. Glotzbach, M. Januszyk, I. N. Vial et al., "An information theoretic, microfluidic-based single cell analysis permits identification of subpopulations among putatively homogeneous stem cells," *PLoS ONE*, vol. 6, no. 6, article e21211, 2011.

[150] A. Raj and A. van Oudenaarden, "Nature, nurture, or chance: stochastic gene expression and its consequences," *Cell*, vol. 135, no. 2, pp. 216–226, 2008.

[151] S. Itzkovitz, A. Lyubimova, I. C. Blat et al., "Single-molecule transcript counting of stem-cell markers in the mouse intestine," *Nature Cell Biology*, vol. 14, no. 1, pp. 106–114, 2012.

[152] L. Warren, D. Bryder, I. L. Weissman, and S. R. Quake, "Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17807–17812, 2006.

[153] F. Tang, C. Barbacioru, E. Nordman et al., "Deterministic and stochastic allele specific gene expression in single mouse blastomeres," *PLoS ONE*, vol. 6, no. 6, article e21208, 2011.

[154] F. Tang, K. Lao, and M. A. Surani, "Development and applications of single-cell transcriptome analysis," *Nature Methods*, vol. 8, supplement 4, pp. S6–S11, 2011.

[155] X. Xu, Y. Hou, X. Yin et al., "Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor," *Cell*, vol. 148, no. 5, pp. 886–895, 2012.

[156] C. Hou and V. G. Corces, "Throwing transcription for a loop: expression of the genome in the 3D nucleus," *Chromosoma*, vol. 121, no. 2, pp. 107–116, 2012.

[157] A. Sanyal, D. Baù, M. A. Martí-Renom, and J. Dekker, "Chromatin globules: a common motif of higher order chromosome structure?" *Current Opinion in Cell Biology*, vol. 23, no. 3, pp. 325–331, 2011.

[158] L. B. Edelman and P. Fraser, "Transcription factories: genetic programming in three dimensions," *Current Opinion in Genetics & Development*, vol. 22, no. 2, pp. 110–114, 2012.

[159] J. M. Belton, R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan, and J. Dekker, "Hi-C: a comprehensive technique to capture the conformation of genomes," *Methods*, vol. 58, no. 3, pp. 268–276, 2012.

[160] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams et al., "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.

[161] J. Zhang, H. M. Poh, S. Q. Peh et al., "ChIA-PET analysis of transcriptional chromatin interactions," *Methods*, vol. 58, no. 3, pp. 289–299, 2012.

[162] G. Li, X. Ruan, R. K. Auerbach et al., "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation," *Cell*, vol. 148, pp. 184–298, 2012.

[163] G. Li, M. J. Fullwood, H. Xu et al., "ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing," *Genome Biology*, vol. 11, no. 2, article R22, 2010.

[164] A. Melnikov, A. Murugan, X. Zhang et al., "Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay," *Nature Biotechnology*, vol. 30, no. 3, pp. 271–277, 2012.

[165] V. M. Bedell, Y. Wang, J. M. Campbell et al., "In vivo genome editing using a high-efficiency TALEN system," *Nature*, vol. 491, no. 7422, pp. 114–118, 2012.

[166] K. Yusa, S. T. Rashid, H. Strick-Marchand et al., "Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells," *Nature*, vol. 478, no. 7369, pp. 391–394, 2011.