Published online 20 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D925–D931***
doi: 10.1093/nar/gkv1263

SBR-Blood: systems biology repository for hematopoietic cells

Jens Lichtenberg[1,*], Elisabeth F. Heuston[1], Tejaswini Mishra[2], Cheryl A. Keller[2], Ross C. Hardison[2] and David M. Bodine[1]

[1]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA and [2]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA

Received August 13, 2015; Revised October 30, 2015; Accepted November 04, 2015

ABSTRACT

Extensive research into hematopoiesis (the development of blood cells) over several decades has generated large sets of expression and epigenetic profiles in multiple human and mouse blood cell types. However, there is no single location to analyze how gene regulatory processes lead to different mature blood cells. We have developed a new database framework called hematopoietic Systems Biology Repository (SBR-Blood), available online at http://sbrblood.nhgri.nih.gov, which allows user-initiated analyses for cell type correlations or gene-specific behavior during differentiation using publicly available datasets for array- and sequencing-based platforms from mouse hematopoietic cells. SBR-Blood organizes information by both cell identity and by hematopoietic lineage. The validity and usability of SBR-Blood has been established through the reproduction of workflows relevant to expression data, DNA methylation, histone modifications and transcription factor occupancy profiles.

INTRODUCTION

Hematopoiesis is the process by which pluripotent stem cells divide and differentiate to generate the many types of circulating blood cells. A model of the stages leading to the formation of red blood cells (erythropoiesis) and platelets (megakaryopoiesis) in the hematopoietic hierarchy is provided in Figure 1. Cells at each stage of hematopoiesis are defined by a cohort of cell surface markers. These cell surface markers can be used to isolate different populations using fluorescence activated cell sorting (FACS). The availability of *in vitro* culture systems and animal models allows for a comprehensive analysis of these populations with regard to their regulation, expression, and function (1). The extensive knowledge of regulatory mechanisms in many hematopoietic cell types make hematopoiesis an excellent system in which to study regulatory correlations and their effects on systemic disorders like anemia, bone marrow failures and leukemia (2,3).

Hematopoietic cells vary significantly in their mRNA expression, DNA methylation, and histone modification profiles (4–6). These profiles not only provide insight into the properties of the cell types, but they can also be compared against each other to infer relationships between hematopoietic lineages (subsystems). For example, correlation of RNA-Seq and ChIP-Seq profiles during erythroid differentiation on a small scale showed that the protein KLF1 activates transcription in erythroblasts when bound to a target gene's promoter region (7). Performing such analyses on a larger scale, however, requires specialized database considerations.

Existing hematopoietic databases (Table 1) are each valuable, but they differ in their focus on specific cell types, omics applications or platform sources. Repositories such as BloodExpress (8) and ImmGen (9) allow analysis of multiple hematopoietic lineages but are restricted to expression data. The NCBI Gene Expression Omnibus (GEO) (10) and the ENCODE project (11) host epigenetic next-generation sequencing data but lack the functionality to directly compare data sets across different platforms and cell types. Efforts such as HAEMCODE (12) and CODEX (13) are compiling hematopoietic next-generation sequencing data comprehensively, along with a set of analysis tools. Their interface is experiment-driven rather than cell type focused, meaning that for comprehensive comparisons between cell types users must collect the relevant subset of experiments for further analysis. Tools currently deployed at these resources are not designed to study of correlations between different epigenetic marks or expression profiles. The BloodChIP database (14) allows for correlation between epigenetic data and expression data, but it is limited to only microarray-based expression profiles and does not allow cell type-driven analyses.

Building on established approaches and compiled data while providing missing functionality will enable the use of hematopoietic array and sequencing data for sophisti-

*To whom correspondence should be addressed. Tel: +1 301 435 2250; Fax: +1 301 402 0902; Email: lichtenbergj@mail.nih.gov

Table 1. Data repositories containing omics data relevant to hematopoietic differentiation that comprise the foundation of SBR-Blood

Repository, URL	Lineage Focus Data Focus Analysis Focus
BloodChIP (14), http://149.171.101.136/python/BloodChIP/index.html	Hematopoiesis (Human) Expression/Epigenetic/Annotations Information Lookup
BloodExpress (8), http://hscl.cimr.cam.ac.uk/bloodexpress/	Hematopoiesis (Mouse) Expression/Annotations Information Lookup, Population Correlations
CODEX (13), http://codex.stemcells.cam.ac.uk	Hematopoiesis (Human/Mouse) Expression/Epigenetic/Annotations Information Lookup, Experiment Correlations
ENCODE (11), http://encodeproject.org/ENCODE/	Hematopoiesis (Human) Expression/Epigenetic/Annotations Information Lookup, Data Storage
EpoDB (15–17), http://www.cbil.upenn.edu/EpoDB/	Erythropoiesis (Vertebrates) Sequence/Annotations Information Lookup, Sequence Analysis
ErythronDB (18,19), http://www.cbil.upenn.edu/ErythronDB/	Erythropoiesis (Mouse) Expression/Regulation/Annotations Information Lookup
HAEMCODE (12), http://haemcode.stemcells.cam.ac.uk	Hematopoiesis (Mouse) Epigenetic/Annotations Information Lookup, Experiment Correlations
Hembase (20), http://hembase.niddk.nih.gov/	Erythropoiesis (Human) Expression/Annotations Information Lookup
HemoPDB (21), http://bioinformatics.wistar.upenn.edu/HemoPDB	Hematopoiesis (Vertebrates) Regulation/Annotations Information Lookup
ImmGen (9), http://www.immgen.org/	Hematopoiesis (Human/Mouse) Expression/Annotations Information Lookup, Data Storage, Population Correlations
LymphTF-DB (22), http://www.iupui.edu/~tfinterx/	Lymphopoiesis (Mouse) Regulation/Annotations Information Lookup
NCBI GEO (10), http://www.ncbi.nlm.nih.gov/geo/	Hematopoiesis (Vertebrates) Expression/Epigenetic/Annotations Information Lookup, Data Storage

cated analyses and more detailed hypothesis generation. To do this, we developed the hematopoietic Systems Biology Repository (SBR-Blood), a database that integrates both array-based expression and epigenetic data from existing hematopoietic repositories with published and user-generated next-generation sequencing data. SBR-Blood provides a web-interface to correlate and evaluate data associated with the various cell types and hematopoietic subsystems. By curating and normalizing the included data, SBR-Blood enables cross-correlation analyses between microarray and RNA-Seq data. It also provides tools to concurrently assess changes in the methylation, histone modification, and transcription factor binding profiles. An overview of the functionality, design and data provided through the SBR-Blood database is provided in Figure 2.

A common workflow in epigenetic data analysis consists of the compilation and integration of data from multiple experiments into a single model through many format conversions, normalizations, and set operations. Our goal in designing SBR-Blood is to provide users with the ability to make informed and appropriate analytical decisions and to focus on the important biological questions relevant to hematopoiesis without having to install and master different sets of bioinformatics tools. SBR-Blood makes the underlying analytical process apparent and allows transparency of data generation and parameter settings.

Currently, SBR-Blood content is focused on mouse hematopoietic cells, but the application is highly adaptable and can be expanded to support different organisms, biological systems, and disease states.

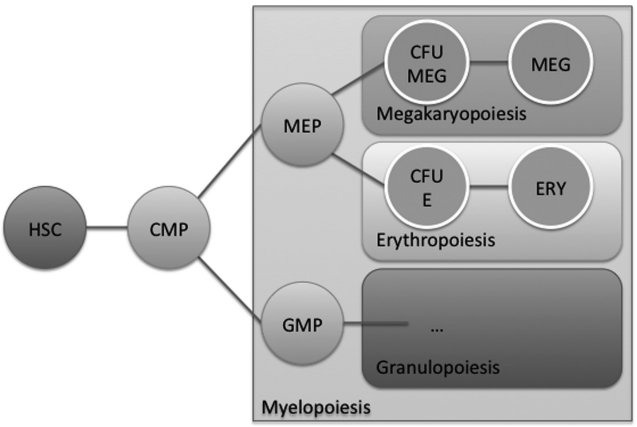

Figure 1. Cell types involved in hematopoietic stem cell differentiation. Cells are grouped by color into lineages. Hematopoietic stem cells (HSC) give rise to common myeloid progenitors (CMP), which differentiate into megakaryocyte erythroid progenitors (MEP) and granulocyte macrophage progenitors (GMP). GMPs give rise to the granulopoiesis lineage and MEPs to both megakaryopoiesis and erythropoiesis, which produces megakaryocytes (MEG), erythrocytes (ERY), respectively and their corresponding colony forming units (CFU).

MATERIALS AND METHODS

SBR-Blood is designed as a three-tier web application. The first is a server-based presentation tier that allows the user to select, view and mine information stored in the repository. The second tier, the logic tier, implements the database management and query system, which operate on the third (the data tier) and are responsible for integrating the different hematopoietic datasets into a single repository. This design, common to most bioinformatics web-services, provides users with easy access to the application without concern for system requirements and availability of the database. Ensuring a very basic server infrastructure, using the common gateway interface, the Perl programming environment and the standard Perl database interface (DBI) available to Unix/Linux, Macintosh and Windows platforms, allows portability of the repository in terms of hosting requirements. By ensuring that SBR-Blood is not only open source but also compatible with a large range of data center configurations, it can be adapted to a variety of organisms, biological systems and processes.

Integration of existing data

To provide a unified platform for mouse hematopoietic data, SBR-Blood incorporates the existing body of data, including microarrays and next-generation sequencing datasets (derived from RNA-Seq, Methyl-Seq, and ChIP-Seq). The database contains information from 228 experiments for 23,213 mRNA transcripts and 3227 long non-coding RNAs (lncRNA). The data are broken down into 120 expression, 44 methylation, 32 transcription factor binding and 32 histone-modification assays. Each assay has various layers of relevant information that provide important metadata regarding the experiment. Expression levels and epigenetic modifications must be associated with transcripts of protein-coding mRNA and non-coding RNA. We assigned default ranges in and around genes to allow grouping of epigenetic signals into different genomic partitions relative to the transcription start and end sites of known genes (Figure 3).

Information in the data tier is organized into five explicit layers. Each datum, represented as a measurement associated with an Ensembl transcript ID and its regulatory region (e.g. promoter or gene body), can be linked to an experimental replicate ID. In general, current experimental methodology requires several measurements for expression and intensity of an epigenetic mark in order to test the statistical significance of the observation. Each set of replicates is then associated with a specific cell type. Multiple cell types can be associated with similar experimental types (e.g. DNA methylation profiling) in order to analyze them according to a specific experimental procedure (e.g. lineage commitment or disease state).

In addition to the required data, an experiment is also annotated with metadata describing the animal model, disease state or the sample's tissue of origin, as well as the publication where the data are presented. Cell types may be further assigned to different hematopoietic lineages/subsystems and the sets of cell surface markers that are used to identify it. Finally, the database contains information about the genomic coordinates of transcript IDs, the targets assessed in epigenetic experiments, and the genomic partitions containing these marks.

Microarray expression data were retrieved from the Gene Expression Omnibus (GEO) (10) and the BloodExpress repository (8). RNA-Seq, epigenetic array and sequencing data were retrieved from GEO or the mouseENCODE portal (http://mouseencode.org). In order to provide concurrent expression profiles, mouseENCODE and GEO are mined for new hematopoietic data sets automatically on a monthly basis, processing newly discovered sets automatically and providing them for curation into SBR. Array-based data were processed using the bioconductor library affy in R (23), and sequencing-based expression data were aligned using STAR (24) and profiled with the RSEM toolkit (25). Epigenetic profiles were constructed by mapping the next-generation sequencing reads using Bowtie2 (26) and peak calling using MACS (27), ERANGE (28) and SICER (29) via the SigSeeker ensemble (30).

Quality control annotations

In addition to experimental quantification, SBR-Blood also assesses data quality. Array-based quality is analzyed with the MetaQC package for R (31), which uses an internal quality control index to characterize the coexpression behavior across different studies and remove potentially inconsistent experiments from the analysis. Next-generation sequencing quality is analyzed via the FastQC module (32), which provides information about the quality of individual bases (e.g. GC content) and the entire sequence (e.g. read duplication levels). Read complexity is analyzed through the samtools flagstat method (33).

Correlation of expression patterns with epigenetic data

Expression and epigenetic profiles are normalized between different array-based and sequencing-based profiling approaches using the R quantile normalization technique

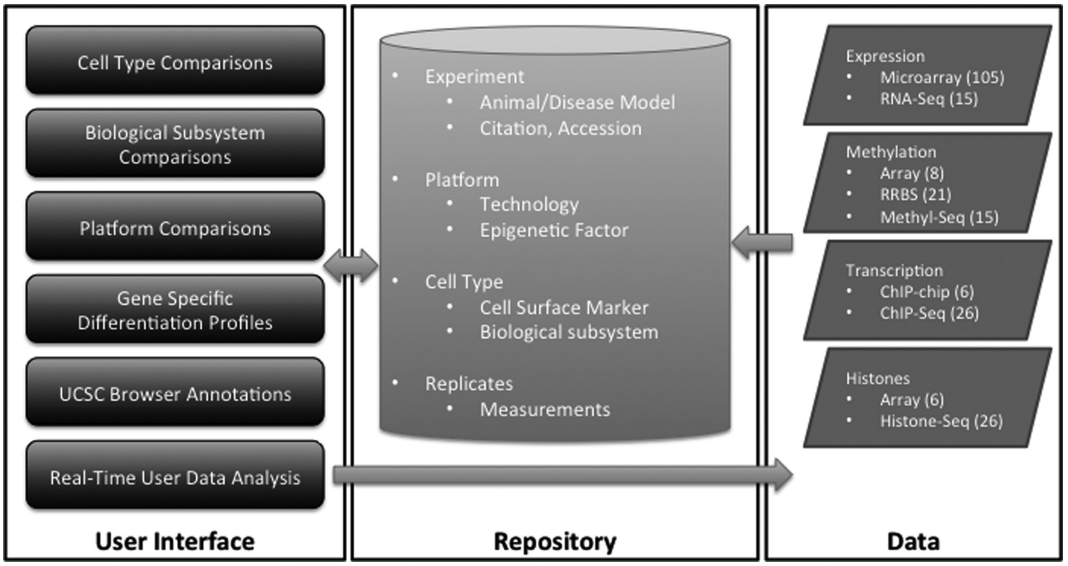

Figure 2. Overview of the functionalities and data integrated into SBR-Blood.

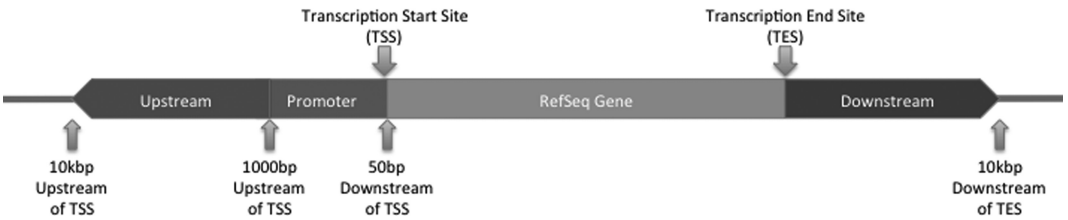

Figure 3. Overview of the relative locations used to characterize the genomic partitions applied in SBR. All partitions are non-overlapping.

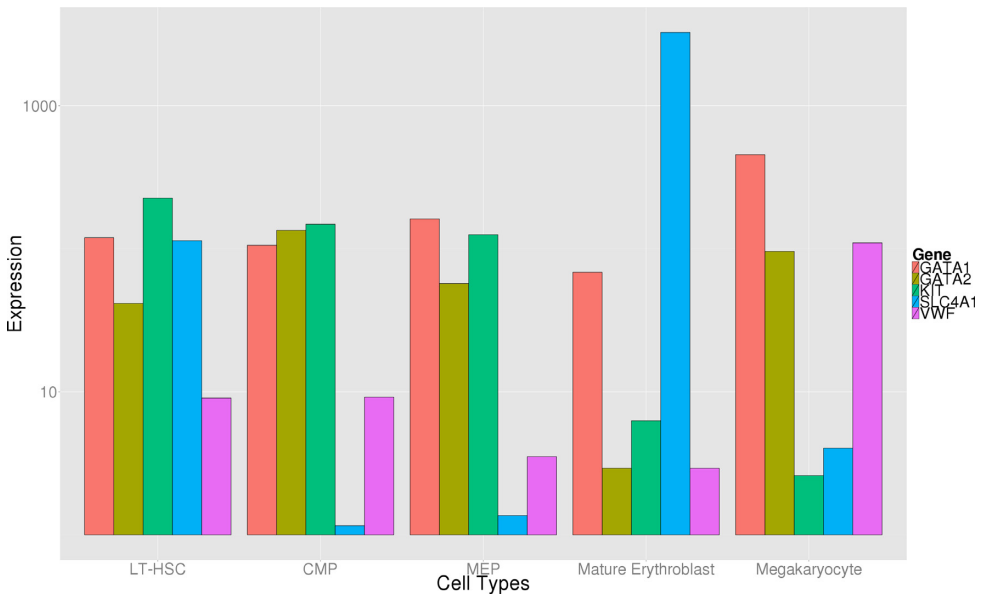

Figure 4. RNASeq mRNA expression profiles for a set of user-specified genes, correlated via the 'Gene Mining' module of SBR. Each cell shows the average relative expression value for a gene in a specific cell type, normalized across the different experiments.

Table 2. Expression and methylation during erythropoiesis

	HSC	CMP	CFU-E	ERY	Common
mRNA					
Expressed	3508	5366	3836	7458	297
Methylated	18437	16094	10337	12923	9549
Methylated and Expressed	2937	4045	1658	4417	133
Methylated and Not Expressed	15500	12049	8679	8506	9416
Not Methylated but Expressed	571	1321	2178	3041	164
Not Methylated and Not Expressed	1268	1753	9040	2832	13367
ncRNA					
Expressed	40	81	39	203	4
Methylated	1025	842	473	594	429
Methylated and Expressed	25	51	18	91	0
Methylated and Not Expressed	1000	791	455	503	429
Not Methylated but Expressed	15	30	21	112	4
Not Methylated and Not Expressed	2162	2304	2715	2430	2794

Expression is determined by microarray andRNA-Seq expression profiles. Methylation is defined as the number of transcripts with a DNA methylation signal in the promoter. We have annotated a complete set of 23 213 mRNA transcripts and 3227 lncRNAs based on (38).

Table 3. Dynamic correlation of methylation data

	Total	HSC	CMP	CFU-E	ERY
Upstream	54	52	14	18	52
Promoter	5	5	0	1	5
RefSeq	804	798	673	609	793
Downstream	137	135	61	53	131

Methylation data that could not be associated with CMP promoters (6) were chosen to generate a custom peak profile for comparison.

available through the preprocessCore package (available through the CRAN repository). SBR-Blood allows users to correlate cell type expression and epigenetic profiles with each other as well as with sets of entire hematopoietic subsystems on the basis of experiment methodology or genomic locus. Correlations can be conducted using genome-wide approaches or genomic partitions, and can be based on raw measurements or statistical significance. The statistical significance of internal comparison variances across measurements for cells is established using the ANOVA statistical test and a user-defined significance threshold (34). Additionally, a variety of multiple test corrections (e.g. Bonferroni or Benjamini-Hochberg) can be applied to modify the stringency of the results. Intersections and unions, computed for user-specified combinations of experimental and repository data, and controlled via various statistical thresholds, can be exported as gene lists, peak lists, or as genomic coordinates in BED format. In addition to providing sequence information about gene sets, each gene in SBR-Blood is annotated with a link to its location in the UCSC Genome Browser (35) and its specific Genbank record (36).

Real-time queries of user-specific data

SBR-Blood is populated with curated published data. Users can apply the database to produce real-time correlations between their own data and these curated sets. The user supplied data are contained in a parallel data tier separated from the main repository in order to maintain database curation. The advantage of this method is that it enables correlation of user-provided data against the database and allows the use of all of SBR-Blood's features to analyze early-stage experimental data.

Database content

Currently, SBR-Blood is populated with mouse hematopoietic data, primarily from healthy C57BL/6J mice and immortalized mouse-derived cell lines. In addition to microarray expression data extracted from the BloodExpress repository, SBR-Blood also contains publicly available Methyl-Seq data (6), RNA-Seq expression data (37), ChIP-Seq for several transcription factors including EKLF, GATA1 and GATA2 (5,7), and Histone-Seq data representing several histone modifications (5). All hematopoietic cell types in SBR-Blood (Supplementary Table S1) are defined by cell surface marker expression and fluorescence activated cytometric sorting.

A web interface depicting the different cell type relationships in the hematopoietic system and subsystems provides the user with the opportunity to query SBR-Blood as a straight-forward information repository. Each of the cell types in the interface can be selected and relevant expression profiles and epigenetic datasets are made available. SBR-Blood correlations enable comparisons between expression and epigenetic profiles associated with specific cell types. In addition to providing population correlations, SBR-Blood provides an interface to query the database for changes in expression and epigenetic profiles of user-specified gene sets. In this process the user enters a set of gene symbols or transcript IDs and selects cell types or hematopoietic subsystems to query. Transcript information in all cases is annotated through links to respective locations and records in the UCSC Genome Browser and in Genbank.

Example applications

Localized methylation mapping constructed via representation bisulfite sequencing and genome-wide mapping constructed via MBD2 pulldown, together with RNA-Seq and microarray expression, are applied in a case study describing the erythroid lineage. The cross correlation feature of SBR-Blood reveals that a subset of the cells involved in erythropoiesis (HSC, CMP, CFU-E and ERY, Figure 1) undergo an increase in overall gene expression and a decrease in promoter DNA methylation as they mature (Table 2). Furthermore, there is a strong core for both methylation and expression, meaning that very few genes pick up methylation or lose expression during the differentiation process.

Using the dynamic correlation feature of SBR-Blood, we used a set of methylation sites identified as specifically absent in promoters of CMP by Hogart *et al.* (6) as an internal validation for SBR-Blood (Table 3). By showing that the methylation signals are apparent in other cell types aside from CMPs we confirmed that they were specifically demethylated in CMPs and remethylated later in the lineage. Furthermore, genes associated with promoter-specific methylation were incrementally re-methylated with increasing maturation. A similar pattern could be observed for upstream regulatory regions, while the methylation sites associated with coding and downstream regions were subjected to further demethylation in CFU-E before recovering in ERY. This indicates that cell-type specific genomic regions experience temporary changes in their epigenetic profiles.

To evaluate individual changes in the mRNA levels of several genes during erythropoiesis and megakaryopoiesis, a progenitor associated gene (*Kit*), an erythroid gene (*Slc4a1*), a megakaryocyte gene (*Vwf*) and two genes involved in the regulation of both erythropoiesis and megakaryopoiesis (*Gata1* and *Gata2*) were chosen to construct a user-specified SBR-Blood query (Figure 4). The levels of *Gata1* mRNA and *Gata2* mRNA have been shown to be inversely correlated, a process known as Gata switching (39). In agreement with these results, the level of *Gata2* mRNA is highest before erythroid commitment, after which it drops off significantly, while *Gata1* is lower than *Gata2* during initial differentiation and experiences a subsequent increase during erythroid commitment (Figure 4). *Kit* is an example of a gene that is highly expressed during the early stages of differentiation but down regulated during differentiation (40). The expression of *Vwf* and *Slc4a1* mRNA is consistent with their known behavior. In particular, the level of *Slc4a1* mRNA is high in stem cells as well as erythroblasts and low in megakaryocytes (41), while *Vwf* mRNA levels are high in megakaryocytes and low in erythroblasts (42).

CONCLUSION

SBR-Blood enables the user to conduct cell type correlations with previously published data sets and data sets currently under investigation. By reproducing concepts manually validated in the community, it has been shown to be correct in terms of functionality (using the integrated data to confirm established knowledge) and provides a useful resource for studying epigenetic profile changes during stem cell differentiation. The structure of SBR-Blood makes it highly adaptable and can support different organisms and biological systems. In particular, SBR-Blood is an excellent tool to compare transcription and epigenetic regulation profiles. The available instance of SBR-Blood is currently being expanded to support variant or disease-related hematopoiesis comparisons in mouse as well as human. This addition will provide further insights into the epigenetic regulatory changes associated with blood-related diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Human Genome Research Institute (NHGRI) intramural funds. Funding for open access charge: NHGRI intramural funds.
Conflict of interest statement. None declared.

REFERENCES

1. Chao,M.P., Seita,J. and Weissman,I.L. (2008) Establishment of a normal hematopoietic and leukemia stem cell hierarchy. *Cold Spring Harb. Symp. Quant. Biol.*, **73**, 439–449.
2. Bonnet,D. and Dick,J.E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.*, **3**, 730–737.
3. Cantor,A.B. and Orkin,S.H. (2002) Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene*, **21**, 3368–3376.
4. Chambers,S.M., Boles,N.C., Lin,K.Y.K., Tierney,M.P., Bowman,T.V., Bradfute,S.B., Chen,A.J., Merchant,A.A., Sirin,O., Weksberg,D.C. *et al.* (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*, **1**, 578–591.
5. Wilson,N.K., Foster,S.D., Wang,X., Knezevic,K., Schuette,J., Kaimakis,P., Chilarska,P.M., Kinston,S., Ouwehand,W.H., Dzierzak,E. *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, **7**, 532–544.
6. Hogart,A., Lichtenberg,J., Ajay,S.S., Anderson,S.M., Margulies,E.H. and Bodine,D.M. (2012) Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal over-representation of ETS transcription factor binding sites. *Genome Res.*, **22**, 1407–1418.
7. Pilon,A.M., Subramanian,S.A., Kumar,S.A., Steiner,L.A., Cherukuri,P., Wincovitch,S., Anderson,S.M., Mullikin,J., Gallagher,P.G., Hardison,R. *et al.* (2011) Genome-wide ChIP-seq reveals a dramatic shift in the binding of the transcription factor erythroid krupple-like factor (EKLF) during erythrocyte differentiation. *Blood*, **118**, e139–e148.
8. Miranda-Saavedra,D., De,S., Trotter,M.W., Teichmann,S.A. and Gottgens,B. (2009) BloodExpress: a database of gene expression in mouse haematopoiesis. *Nucleic Acids Res.*, **37**, D873–D879.
9. Heng,T.S.P., Painter,M.W., Elpek,K., Lukacs-Kornek,V., Mauermann,N., Turley,S.J., Koller,D., Kim,F.S., Wagers,A.J., Asinovski,N. *et al.* (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.*, **9**, 1091–1094.
10. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
11. Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigó,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
12. Ruau,D., Ng,F.S., Wilson,N.K., Hannah,R., Diamanti,E., Lombard,P., Woodhouse,S. and Gottgens,B. (2013) Building an

ENCODE-style data compendium on a shoestring. *Nat. Methods*, **10**, 926.
13. Sanchez-Castillo,M., Ruau,D., Wilkinson,A.C., Ng,F.S., Hannah,R., Diamanti,E., Lombard,P., Wilson,N.K. and Gottgens,B. (2015) CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.*, **43**, D1117–D1123.
14. Chacon,D., Beck,D., Perera,D., Wong,J.W. and Pimanda,J.E. (2014) BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Res.*, **42**, D172–D177.
15. Salas,F., Haas,J., Stoeckert,C.J. and Overton,G.C. (1997) EpoDB: An erythropoiesis gene expression database in progress. *Bioinformatics*, **1278**, 52–61.
16. Salas,F., Haas,J., Brunk,B., Stoeckert,C.J. Jr and Overton,G.C. (1998) EpoDB: a database of genes expressed during vertebrate erythropoiesis. *Nucleic Acids Res.*, **26**, 288–289.
17. Stoeckert,C.J., Salas,F., Brunk,B. and Overton,G.C. (1999) EpoDB: a prototype database for the analysis of genes expressed during vertebrate erythropoiesis. *Nucleic Acids Res.*, **27**, 200–203.
18. Kingsley,P.D., Greenfest-Allen,E., Frame,J.M., Bushnell,T.P., Malik,J., McGrath,K.E., Stoeckert,C.J. and Palis,J. (2013) Ontogeny of erythroid gene expression. *Blood*, **121**, e5–e13.
19. Greenfest-Allen,E., Malik,J., Palis,J. and Stoeckert,C.J. Jr (2013) Stat and interferon genes identified by network analysis differentially regulate primitive and definitive erythropoiesis. *BMC Syst. Biol.*, **7**, 38.
20. Goh,S.H., Lee,Y.T., Bouffard,G.G. and Miller,J.L. (2004) Hembase: browser and genome portal for hematology and erythroid biology. *Nucleic Acids Res.*, **32**, D572–D574.
21. Pohar,T.T., Sun,H. and Davuluri,R.V. (2004) HemoPDB: Hematopoiesis Promoter Database, an information resource of transcriptional regulation in blood cell development. *Nucleic Acids Res.*, **32**, D86–D90.
22. Childress,P., Fletcher,R. and Perumal,N. (2007) LymphTF-DB: a database of transcription factors involved in lymphocyte development. *Genes Immun.*, **8**, 360–365.
23. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
24. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
25. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
26. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
27. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nussbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
28. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
29. Zang,C., Schones,D.E., Zeng,C., Cui,K., Zhao,K. and Peng,W. (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*, **25**, 1952–1958.
30. Lichtenberg,J., Heuston,E.F. and Bodine,D.M. (2014) *Bioinformatics Open Source Conference*, Boston.
31. Kang,D.D., Sibille,E., Kaminski,N. and Tseng,G.C. (2012) MetaQC: objective quality control and inclusion/exclusion criteria for genomic meta-analysis. *Nucleic Acids Res.*, **40**, e15.
32. Andrews,S. (2010) FASTQC. A quality control tool for high throughput sequence data.
33. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
34. Sirbu,A., Ruskin,H.J. and Crane,M. (2010) Cross-platform microarray data normalisation for regulatory network inference. *PLoS One*, **5**, e13822.
35. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M., Haussler,D. and others. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
36. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2011) GenBank. *Nucleic Acids Res.*, **39**, D32–D37.
37. Flygare,J., Estrada,V.R., Shin,C., Gupta,S. and Lodish,H.F. (2011) HIF1$$ synergizes with glucocorticoids to promote BFU-E progenitor self-renewal. *Blood*, **117**, 3435–3444.
38. Paralkar,V.R., Mishra,T., Luan,J., Yao,Y., Kossenkov,A.V., Anderson,S.M., Dunagin,M., Pimkin,M., Gore,M., Sun,D. *et al.* (2014) Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*, **123**, 1927–1937.
39. Snow,J.W., Trowbridge,J.J., Johnson,K.D., Fujiwara,T., Emambokus,N.E., Grass,J.A., Orkin,S.H. and Bresnick,E.H. (2011) Context-dependent function of 'GATA switch' sites in vivo. *Blood*, **117**, 4769–4772.
40. Uoshima,N., Ozawa,M., Kimura,S., Tanaka,K., Wada,K., Kobayashi,Y. and Kondo,M. (1995) Changes in c-Kit expression and effects of SCF during differentiation of human erythroid progenitor cells. *Br. J. Haematol.*, **91**, 30–36.
41. Paw,B.H., Davidson,A.J., Zhou,Y., Li,R., Pratt,S.J., Lee,C., Trede,N.S., Brownlie,A., Donovan,A., Liao,E.C. *et al.*, (2003) Cell-specific mitotic defect and dyserythropoiesis associated with erythroid band 3 deficiency. *Nat. Genet.*, **34**, 59–64.
42. Sanjuan-Pla,A., Macaulay,I.C., Jensen,C.T., Woll,P.S., Luis,T.C., Mead,A., Moore,S., Carella,C., Matsuoka,S., Bouriez Jones,T. *et al.* (2013) Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, **502**, 232–236.