View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by RERO DOC Digital Library

Biol. Chem., Vol. 390, pp. 1139–1144, November 2009 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2009.116

Minireview

Dissection of gene regulatory networks in embryonic stem cells by means of high-throughput sequencing

Christian Beisel* and Renato Paro

Department Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, CH-4058 Basel, Switzerland

* Corresponding author
e-mail: christian.beisel@bsse.ethz.ch

Abstract

Transcription factor regulation of gene expression and chromatin-controlled epigenetic memory systems are closely cooperating in establishing the pluripotent state of embryonic stem (ES) cells and maintaining cell fate decisions throughout development of an organism. A thorough understanding of the regulatory transcriptional circuitry that rules the underlying plastic yet heritable gene expression programs in ES cells is of great importance. With the advent of next-generation sequencing technologies facilitating the quantitative assessment of functional genomics assays it is now feasible to interrogate transcription networks at a genome-wide scale. Here, we discuss the application of next-generation sequencing in elucidating the molecular mechanisms underlying ES cell function.

Keywords: cellular memory; ChIP-seq; chromatin; deep sequencing; epigenetics; genomics.

Introduction: memory of gene expression programs in the development of an organism

We all originate from a single cell, the fertilized egg. During the process of development – cell number increases – cells start to become different and specialize. This is a hierarchical process with progressive restriction of developmental potentials of cells in a lineage from intermediate stages down to the fully committed specific cell type (Reik, 2007). Although all different cell types (with few exceptions) in a mammalian organism contain the same genetic content, the fate of a cell is defined by its specific settings of gene expression. Once a cell is directed towards a particular developmental route the gene expression program characterizing its identity has to be memorized over the following cell divisions until the terminal differentiation state is reached. The impact of such a memory system becomes obvious when cells escape their fate to enter the detrimental track of tumor formation, a process in which the cancer cells have forgotten their original destiny.

As there are no genetic changes involved in cell fate determination events, the heritability of transcription states has to occur at an epigenetic level. The term epigenetics describes the inheritance of gene expression patterns independent of the underlying DNA sequence. In vertebrates, epigenetic gene regulation is based on the methylation of DNA at cytosines as well as on the modulation of chromatin structure (Bernstein et al., 2007).

DNA methylation is the best-known epigenetic modification. It plays a major role in basic processes, such as X chromosome inactivation, chromosomal stability and parental imprinting. CpG sites, as targets for the methylation, are not randomly distributed in the genome; instead, there are CpG-rich regions known as CpG islands located in the promoter regions of a large fraction of genes (Gardiner-Garden and Frommer, 1987). CpG islands are usually not modified in the early stage of development but can become methylated during cellular differentiation leading to a repression of the targeted genes (Meissner et al., 2008; Mohn et al., 2008). The propagation of DNA methylation has been well-characterized as a replication-coupled process involving a maintenance DNA methyltransferase that copies the methylation mark present at the parent strand to the daughter strand, thereby assuring the epigenetic inheritance of the signal from one cell generation to the following one (Goll and Bestor, 2005).

The central building block of chromatin is the nucleosome consisting of a defined stretch of DNA wrapped around an octamer of four different histone proteins. The packaging of DNA into nucleosomes constitutes a constraint to processes where protein factors need to have access to DNA (Struhl, 1999). However, chromatin is not a simple barrier for DNA dependent processes but is a regulatory platform that ensures the tight control of gene expression states. Biochemical mechanisms have evolved that employ chromatin functionality to regulate nuclear processes, such as gene expression. Protein composites known as chromatin remodeling complexes can physically rearrange nucleosomes on DNA to generate closed (repressed) or open (active) chromatin structures influencing the accessibility for DNA binding proteins (Cairns, 2005). Additionally, there is a multifaceted interplay between enzymes that can modify particular amino acid residues in the tail regions of nucleosomal histone proteins and those that remove the modifications (Kouzarides, 2007). These histone modifications are implicated in controlling gene expression and genome function by establishing dedicated chromatin environments and orchestrating DNA dependent processes. Especially methylation of lysine residues has been considered, similar to cytosine methylation, a potential mark for carrying epigenetic information that is stably maintained through cell divisions. A current model for the propagation of histone marks suggests a self-

perpetuation after the deposition of new nucleosomes behind the replication fork by means of positive-feedback loops (Hansen et al., 2008).

To understand the gene regulatory networks that govern cell fate determination and maintenance, it is crucial to reveal the mutual functional dependency of sequence specific transcription factors, the RNA polymerase machinery and the epigenetic regulatory system. Experimentally this requires addressing the questions of where the transcription factors and chromatin regulators are located within the genome and how changes in the protein binding or histone modification patterns correlate with changes in the transcriptional program and cell state.

Principles of parallel sequencing and its application to quantify functional states

The key methodology to detect direct physical interactions between proteins and DNA *in vivo* is chromatin immunoprecipitation (ChIP). In ChIP experiments, an antibody specific for a DNA binding factor or histone modification is used to enrich for target DNA sites to which the protein was bound in a living cell (Figure 1). The enriched DNA sites are then identified and quantified. For the large genomes of mammals, it has been difficult to achieve ChIP measurements that combine high accuracy, the comprehensiveness of an entire genome, and high binding-site resolution.

Recently, the development of new high-throughput sequencing technologies – generally referred to as 'deep', 'high-throughput', '(massive) parallel' or 'next-generation' sequencing technologies – enable a new way of quantifying genome-scale functional assays, such as ChIP or transcription profiling (Holt and Jones, 2008). Three different approaches represented by the technologies from Roche, Illumina, and ABI have recently been extensively applied for these purposes. In general, a DNA library with immunoprecipitated fragments, combined with specific adaptor sequences at both sides, becomes sequenced from one end or from both ends in a highly paralleled shotgun fashion without a detour via bacterial cloning steps (Figure 1). In particular, the Illumina Genome Analyzer and the ABI Solid system have proven to be well suited for the analysis requirements of functional genomics applications. They generate millions of short sequence reads of ~30 nucleotides per sequencing run which can be mapped to the reference genome. Owing to the huge amount of reads, it is possible to reliably calculate the frequency of fragments in a complex mixture of DNA molecules (Wold and Myers, 2008). Desired levels of sensitivity and statistical certainty required to detect rare molecules can be achieved by adjusting the total number of reads. In this way, ChIP-enriched DNA can be quantitatively analyzed to identify overrepresented fragments in the mixture that might correspond to genomic locations where the protein of interest was bound. The analysis of ChIP assays by next-generation sequencing, called ChIP-seq, offers a higher resolution than microarray based analyses (ChIP-chip). Additionally, ChIP-seq offers an affordable genome-wide output for mammalian genomes with minimal hands-on processing using less input material and a rapid analysis pipeline that relies on 'simple' counting sequenced reads (Mardis, 2007). Next-generation sequencing analysis can be applied to the quantification of any DNA or any other nucleic acid sample that can be converted to DNA. Besides ChIP-seq, currently the most prominent approaches comprise profiling of mRNA populations (RNA-seq) and the identification of small RNA samples, such as miRNAs, which play a crucial role in post-transcriptional regulation (Morin et al., 2008; Mortazavi et al., 2008). In both cases, the sequencing adaptors are incorporated during the process of cDNA library generation. In the case of transcription profiling, RNA-seq shows a broader dynamic signal range of expression levels with up to five orders of magnitude and seems to be much more sensitive than traditional microarray applications (Mortazavi et al., 2008). Recent studies on

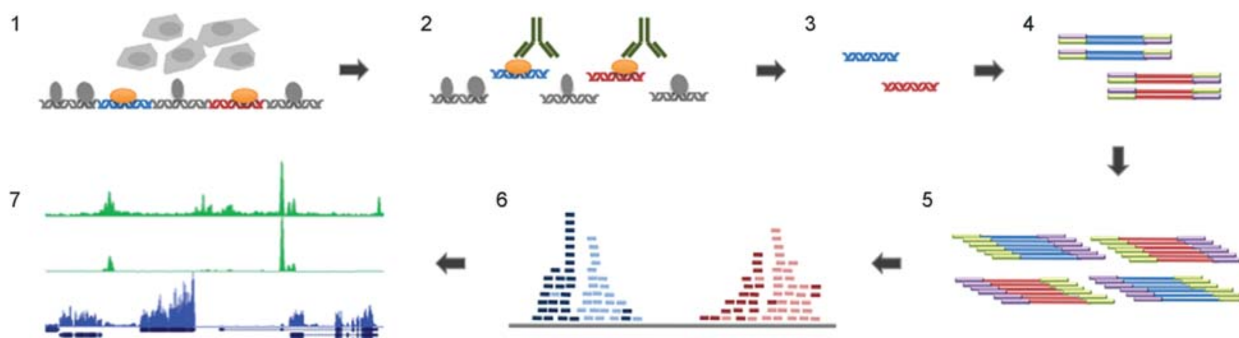

Figure 1 Outline of ChIP-seq.
Cells (most ideally a homogeneous cell type, cell cycle stage, etc.) are fixed by formaldehyde resulting in covalent DNA-protein crosslinks (1). The fixed cells are sonicated to isolate and shear chromatin to the size of 150–1000 bp fragments. To enrich and purify the chromatin fragments bound by the protein of interest, a specific antibody is added for immunoprecipitation (2). The crosslinks are reversed and the DNA purified (3). A DNA library is generated by end repair, adaptor ligation, and limited PCR amplification (4). Subsequently, the DNA fragments undergo a clonal amplification step specific to the sequencing platform used (5). After parallel sequencing millions of fragments enriched in the ChIP sample the sequence reads are mapped to the reference genome (6). Two enriched genomic regions are shown and the directionality of sequenced reads is indicated by dark and light blue and red. In (7) calculated ChIP-seq profiles of two proteins are depicted (green) together with the RNA profile generated by RNA-seq of the poly(A)-mRNA fraction (blue).

different mammalian cell systems reported 25–75% more identified transcripts (Cloonan et al., 2008; Sultan et al., 2008; Tang et al., 2009). By aligning the reads to reference transcripts or their *de novo* assembly, new splice variants can be annotated and thus one gains exact information on its structure in addition to the expression level of a transcript. Small RNA species can now be identified by simple gel purification of RNA samples migrating at the nucleotide length of interest, usually 20–30 nt, and sequencing the sample. All functional assays analyzed by next-generation sequencing have in common that an *a priori* knowledge of the genomic regions of interest is not required as the reads can be mapped to the entire reference genome. This will ultimately lead to a complete annotation of coding and non-coding transcription units of the human genome as well as of the genomes of the many organisms of interest.

Technical details of the different next-generation sequencing platforms and the details on the different methods, including ChIP-seq, RNA-seq, and others, have been reviewed elsewhere (Mardis, 2007; Holt and Jones, 2008; Wang et al., 2009). In this review, we will focus on the application of next-generation sequencing in the dissection of regulatory circuits controlling embryonic stem (ES) cell function which has been the focus of much attention in recent publications.

Next-generation ES cell research

Many studies over the recent years have acclaimed the exceptional role of ES cells for basic and applied research. ES cells are derived from the inner cell mass of the mammalian blastocyst. They are pluripotent and can give rise to all cell types of the adult organism. At the same time they have the potential to self-renew and can be kept for extended periods of time in culture, making them an important model for developmental biology and a promising resource for regenerative medicine (Jaenisch and Young, 2008). The molecular mechanisms underlying the establishment and maintenance of pluripotency in ES cells are of great interest. The knowledge of the transcription factors and epigenetic modifications operating in a regulatory network will facilitate both directed programming of ES cells to specific lineages and the reprogramming of somatic cells to an ES-like state. The pluripotent state of ES cells is established by the action of key transcription factors, such as Oct4, Sox2, and Nanog. Cell fate decisions are accompanied by epigenetic modifications, which ensure the stable inheritance of the underlying gene expression patterns. In this regard, it is important to determine how transcription factor networks interact with different epigenetic regulatory systems to achieve this stability.

Until recently, attention has been focused almost exclusively on the role of the transcription factors Oct4, Sox2, and Nanog (Orkin, 2005). In a study published in 2008 Chen and colleagues demonstrated the power of next-generation sequencing technology by lifting the analysis of the transcription factor network of mouse ES cells to the next level of complexity (Chen et al., 2008). They produced ChIP-seq maps of 13 transcription factors well known for their functional roles in mouse ES cells and found 1100–40 000 binding sites with specificity greater than 95% determined by parallel ChIP-qPCR analysis (Table 1). Most interestingly, the study showed two distinct regulatory networks with one defined by the clustered binding of Oct4, Nanog, Sox2, Smad1, and STAT3. Smad1 and STAT3 are the effector molecules of two signaling pathways crucial for self-renewal. This clearly illustrates the convergence of these signaling pathways with the core ES cell transcriptional circuitry defined by Oct4, Nanog, and Sox2 at common target genes. The striking feature of ChIP-seq to interrogate the entire genome without being limited to promoter regions also showed benefits as the majority of Oct4-Nanog-Sox2 binding sites were found outside promoter regions. Intriguingly, the authors could demonstrate with 25 out of 25 regions tested, that these sequences show ES cell specific enhancer activity. Other transcription factors analyzed in this study comprised c-Myc, Klf4, Zfx, and Esrrb which have been previously shown to be important for self-renewal of ES cells and for reprogramming of somatic cells to induced pluripotent stem cells. In this way, Chen and colleagues provided an unprecedented dataset enabled by the use of massive parallel sequencing technology, which paves the way for the understanding of the transcription regulatory network in ES cells and for the identification of additional factors required for self-renewal, pluripotency, and reprogramming functions.

Table 1 Sequencing depth and mapping summary of ChIPped proteins from mouse ES cells.

Protein mapped	# total reads*	# mapped reads*	Sites bound
Marson et al., 2008			
Oct4	8.76	3.37	17 255
Sox2	25.44	5.67	15 085
Nanog	17.80	6.03	16 688
Suz12	7.69	2.51	3434
Tcf3	11.54	6.11	6291
H3K4me3	10.92	6.94	19 632
H3K79me2	7.6	4.18	28 694
H3K36me3	9.4	4.53	15 463
Chen et al., 2008			
Oct4	9.31	5.69	3761
Sox2	9.27	5.78	4526
Nanog	12.76	8.82	10 343
Suz12	8.77	6.12	4215
STAT3	9.89	6.23	2546
Smad1	5.32	3.39	1126
Tcfcp2l1	12.62	9.47	26 910
CTCF	5.82	3.80	39 609
Zfx	6.67	4.62	10 338
Klf4	7.40	4.34	10 875
Esrrb	10.70	7.94	21 647
c-Myc	10.69	7.20	3422
n-Myc	10.19	5.52	7182
E2f1	12.46	8.97	20 699
p300	8.83	5.69	524

Proteins indicated in bold font have been analyzed in both studies. The differences in the numbers of binding sites result most probably because of the use of different antibodies and different criteria to define thresholds. All ChIPs have been analyzed with an Illumina GA1 with a read length of 26 bases. *Numbers are reported as millions of reads.

ChIP-seq'ing chromatin signals of ES cells

To understand the role of histone modifications in gene expression and cell fate decisions or – conversely – to enable the prediction of gene expression states and to determine the fate of a cell by looking at histone modification patterns, Mikkelsen et al. started to generate comprehensive maps of certain histone modifications across the whole genome of three mouse pluripotent and lineage-committed cell lines by means of ChIP-seq (Mikkelsen et al., 2007). They concentrated on histone modifications for which the function had been well defined previously. Trimethylation of histone H3 at lysine 27 (H3K27me3) is a mark for repressive chromatin set by Polycomb group proteins (PcG). PcG proteins play a fundamental and evolutionary highly conserved role in the epigenetic regulation of gene expression during development by heritably maintaining the repressed state of major developmental genes (Ringrose and Paro, 2004). Conversely H3K4me3 is generally associated with active chromatin. Previous ChIP-chip studies showed a colocalization of these histone modifications in ES cells, which was confirmed by Mikkelsen et al. using ChIP-seq (Mikkelsen et al., 2007; Sharov and Ko, 2007). These regions have been referred to as bivalent domains. However, in the meantime several studies demonstrated that bivalent domains are not specific for ES cells but rather a function of the CpG content of the underlying promoter (Mohn and Schübeler, 2009). All CpG-rich promoters are targeted by H3K4 methylation in ES cells whereby repressed developmental genes are concomitantly methylated at H3K27. Mikkelsen et al. and others have demonstrated that differentiation to a multipotent progenitor state can lead to dynamic changes at bivalent domains with the loss of H3K4me3 or H3K27me3 at then repressed or active promoters. As such, the tracking of these histone modifications at key regulatory genes might help to distinguish developmental potential and state of a certain cell. In this regard, the same group of researchers followed H3K4me3 and H3K27me3 by ChIP-seq analysis in the process of induced reprogramming of somatic cells (Mikkelsen et al., 2008). Such a genome-wide analysis for different cell states would have been virtually impossible before the advent of next-generation sequencing. Interestingly, ChIP-seq analysis could also be used to identify allele specific transcription (Mikkelsen et al., 2007). The co-occurrence of H3K9me3 with H3K4me3 delineates imprinting control regions in ES cells. This knowledge they applied to screen the genome of ES cells derived from a cross of two distantly related mouse strains providing 3.5 million single nucleotide polymorphisms (SNPs) for imprinting control regions by means of H3K9me3 and H3K4me3 ChIP-seq analysis. Then, they utilized ChIP-seq of H3K36me3 to look at active transcription units and searched for allelic imbalance of SNP distribution. Using this approach, the authors could identify novel regions showing allele-specific gene expression, suggesting that ChIP-seq is a powerful tool for this purpose – under the prerequisite of having sufficient sequencing coverage and dense maps of SNPs.

H3K36me3 had already previously been linked to transcriptional elongation. Remarkably, Mikkelsen et al. could show that H3K36me3 in combination with H3K4me3 seems to be ideally suited to annotate unknown 3′- or 5′-regions of genes or entire new transcription units (Mikkelsen et al., 2007). A peak of H3K4me3 indicates the transcription start site and H3K36me3 marks the entire transcribed gene body. This finding promoted a study to use 'K4–K36 domains' identified by ChIP-seq analysis to screen the mouse genome for the presence of large intervening non-coding RNAs (lincRNAs) residing outside of protein coding genes (Guttman et al., 2009). This was an interesting approach as many studies over recent years concentrated on the identification and functional dissection of such non-coding RNAs, demonstrating that these transcripts play important roles in imprinting and gene regulation. Guttman and coworkers identified more than 1200 K4–K36 domains with at least 5 kb in size. Further functional and bioinformatics examinations revealed that the majority of these domains produce multiexonic, non-protein coding RNAs expressed in different cell types. Comparison with ChIP-seq data for the pluripotency factors Oct4 and Nanog indicated that these proteins bind to 118 lincRNA promoters, which also show an expression profile characteristic for genes involved in maintaining the pluripotent state of ES cells. They also noted that one of these lincRNAs had been uncovered in an RNAi screening approach (by serendipity as the study aimed for protein-coding transcripts) to identify genes involved in mouse ES cell proliferation and was found as one of the top hits. This promises that we can expect many new regulatory pathways to be revealed in ES cells involving non-coding RNAs which have been discovered by means of functional genomics approaches using the new massive parallel sequencing technologies.

A similar approach to identify another class of non-coding RNAs has been undertaken by Marson et al. (2008). They concentrated on micro-RNAs (miRNAs) as potential major players in the regulatory circuitry underlying ES cell self-renewal and pluripotency. Several hundred of mature miRNAs have been identified in the human genome previously. Information on their transcriptional regulatory regions, however, is lacking in most cases because they are expressed as polycistronic RNAs (primary miRNAs), which become rapidly processed. To assess their role in the core transcriptional circuitry of ES cells, it is essential to identify their promoter regions and thereby binding sites for potential transcription factors regulating miRNA expression. Utilizing genome-wide histone modification landmarks, such as H3K4me3 and H3K36me3, from several other studies Marson and colleagues identified promoters of 185 primary miRNAs in mouse and of 294 in human corresponding to 336 and 441 mature miRNAs, respectively. To find connections between miRNA expression regulation and the core factors specifying ES cell functions, the authors used ChIP-seq to identify the binding sites of Oct4, Sox2, Nanog, and Tcf3 in mouse ES cells (Tcf3 has been shown to repress pluripotency genes in ES cells under certain culture conditions). Here, it will also be interesting to correlate this dataset with the ChIP-seq data generated by

Chen et al. (2008) to gain a more comprehensive overview of the transcription factor network (Table 1). Further functional approaches comprising miRNA expression profiling by quantitative massive parallel sequencing in ES cells and more differentiated cell states indicated that miRNAs bound by the investigated transcription factor quartet are regulated by these factors in pluripotent cells. Furthermore, this study revealed that Oct4-Sox2-Nanog-Tcf3 bound miRNA gene loci silenced in ES cells are also bound by the PcG protein SUZ12 and are marked by H3K27me3. Upon differentiation into certain lineages they lose H3K27me3 and become expressed, whereas in other tissues they stay silenced. This demonstrates that the PcG system is involved in regulating and maintaining the expression state of tissue specific miRNAs. In summary, Marson et al. impressively enabled the integration of different layers of molecular regulatory mechanisms – DNA sequence specific transcription factors, chromatin modifiers, and miRNAs – governing ES cell identity at the transcriptional and post-transcriptional level.

Conclusions and future directions

Without doubt, genome-scale chromatin profiling enhanced by the advent of next-generation sequencers has revolutionized current research in the field of functional genomics. The mere throughput of ChIP samples that can be analyzed facilitates a new way of approaching a comprehensive dissection of entire gene regulatory networks underlying cell fate decisions. Technical improvements of the current sequencing technologies and new developments in this field will increase read number output per sequencing run, which will (1) improve data quality by gaining a higher sequence coverage and (2) boost sample throughput as multiplexing (sequencing of differently 'barcoded' samples in one sequencing run) will become an option.

To fully understand transcriptional networks it is essential to correlate the protein binding and histone modification profiles with gene expression data. Based on cost and throughput, current gene expression profiling relies, in majority, on traditional microarray techniques but several studies using quantitative massive parallel sequencing for the analysis of isolated mRNA have started to challenge the predominance of microarray approaches. The advantages of RNA-seq are becoming obvious: Cloonan et al. (2008) described 59% more transcripts in mouse ES cells and embryoid bodies (a differentiation step after the ES cell state) compared with microarrays. Even more remarkable, Tang et al. demonstrated the feasibility of single cell RNA-seq with an isolated mouse blastomere and reported 75% more genes than microarrays (Tang et al., 2009). In addition, the identification of unknown splicing events is possible – Tang and colleagues revealed more than 1700 new splice junctions – as well as the discovery of transcription units outside of annotated exons – Cloonan et al. found more than 30% of identifiable reads in non-annotated regions. These approaches illustrate that a comprehensive understanding of gene regulation events has to lead via quantitative sequencing measurements of the transcriptome, which provide many advantages over traditional microarray analysis (reviewed by Wang et al., 2009).

All studies described depict the linear distribution of chromatin bound factors along the genome, which is an intrinsic outcome of classical ChIP assays. Genomic DNA, however, is organized in a three-dimensional manner that – most probably – also influences the gene expression program and cellular fate. Methods have been developed to reveal intra- and interchromosomal interactions, such as the 3C or the 6C technologies (Dekker et al., 2002; Tiwari et al., 2008). 3C generates a library of DNA fragments that reflects literally all interactions and could be utilized in combination with massive parallel sequencing for the analysis of the three-dimensional chromatin structure genome-wide. Furthermore, the 6C method combines ChIP with 3C. Thereby, a DNA library becomes generated (susceptible for parallel sequencing analysis) which is enriched for fragments reflecting chromatin loops between binding sites of the ChIP purified protein.

Probably the biggest challenge for the future will be to integrate all functional genomics datasets to develop a comprehensive picture of the transcriptional networks comprising DNA sequence specific transcription factors, epigenetic regulators, and histone modifications, as well as regulatory long non-coding RNAs and post-transcriptionally acting miRNAs. The beginning of this endeavor seems very promising proven by the studies described in this review but many more have to follow. It is unquestionably attractive to pursue the goal towards the full understanding of the gene regulatory network underlying cell fate decisions and their maintenance. It will facilitate the quantitative and predictive description of these processes, which will ultimately lead to a comprehensive understanding of human development, as well as to an efficient application of ES cells in regenerative medicine.

Acknowledgments

We thank members of the Paro laboratory for critical reading of the manuscript. We apologize to all authors whose work was not included here owing to space limitations.

References

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Cairns, B.R. (2005). Chromatin remodeling complexes: strength in diversity, precision through specialization. Curr. Opin. Genet. Dev. *15*, 185–190.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods *5*, 613–619.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Goll, M.G. and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Holt, R.A. and Jones, S.J. (2008). The new paradigm of flow cell sequencing. Genome Res. *18*, 839–846.

Jaenisch, R. and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Mardis, E.R. (2007). ChIP-seq: welcome to the new frontier. Nat. Methods *4*, 613–614.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Mohn, F. and Schübeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet. *25*, 129–136.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell *30*, 755–766.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. *18*, 610–621.

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.

Orkin, S.H. (2005). Chipping away at the embryonic stem cell network. Cell *122*, 828–830.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Ringrose, L. and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

Sharov, A.A. and Ko, M.S. (2007). Human ES cell profiling broadens the reach of bivalent domains. Cell Stem Cell *1*, 237–238.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1–4.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science *321*, 956–960.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tiwari, V., Cope, L., McGarvey, K., Ohm, J., and Baylin, S. (2008). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. Genome Res. *18*, 1171–1179.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Wold, B. and Myers, R. (2008). Sequence census methods for functional genomics. Nat. Methods *5*, 19–21.

Received May 13, 2009; accepted June 23, 2009