

Location, Location, (ChIP-)Location! Mapping Chromatin Landscapes One Immunoprecipitation at a Time

Benjamin P. Berman,^{1*} Baruch Frenkel,^{2,3,4} and Gerhard A. Coetzee^{6,7,5}

¹USC Epigenome Center, University of Southern California, Los Angeles, California

²Department of Orthopaedic Surgery, University of Southern California, Los Angeles, California

³Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, California

⁴Institute of Genetic Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California

⁵Norris Cancer Center, University of Southern California, Los Angeles, California

⁶Department of Urology, University of Southern California, Los Angeles, California

⁷Department Preventive Medicine, University of Southern California, Los Angeles, California

ABSTRACT

A small fraction of the typical animal genome (<5% in humans) codes for the organism's collection of proteins, yet the study of protein coding sequences dominated the early years of genomics research. In the decade since the sequencing of complete eukaryotic genomes, however, genomic techniques have shed a great deal of light on the non-coding DNA making up the remainder. A single molecular technique, Chromatin Immuno-Precipitation (ChIP) location analysis, has had a profound impact and has made possible the study of an incredible range of biology. This issue of The Journal of Cellular Biochemistry aims to put into context advancements made possible by the ChIP-location revolution, while at the same time highlighting some of the most important technical aspects and challenges along with some of the work yet to come. *J. Cell. Biochem.* 107: 1–5, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: CHROMATIN IMMUNOPRECIPITATION; ChIP-on-chip; ChIP-seq; CHROMATIN CONFORMATION CAPTURE; EPIGENETICS, GENE REGULATION

A small fraction of the typical animal genome (<5% in humans) codes for the organism's collection of proteins, yet the study of protein coding sequences dominated the early years of genomics research. In the decade since the sequencing of complete eukaryotic genomes, however, attention has turned to the non-coding "junk" DNA that makes up the remainder. The ability to move beyond the scant few large genomic regions studied prior to that point has shed a great deal of light on the complex interactions between DNA and nuclear proteins that control genomic regulatory processes. These advances would not have been possible, of course, without complete genome sequences and resulting technologies such as the DNA hybridization microarray. But a single molecular technique, Chromatin Immuno-Precipitation (ChIP), has more than any other come to dominate the field and make possible the study of an incredible range of biology. As one measure of the success of this technique, the number of scientific publications utilizing some form of ChIP

has increased nearly 20-fold in the past decade, from 67 articles in 2000 to 1,025 in 2008 (Fig. 1). This issue of The Journal of Cellular Biochemistry aims to put into context advancements made possible by the ChIP-location revolution, while at the same time highlighting some of the most important technical aspects and challenges. Finally, it gives us an idea of the exciting work yet to come—most would agree that we have only seen the very tip of the ChIP-location iceberg.

Wacker and Kim [this issue] begin by giving a detailed historical overview of ChIP techniques beginning with initial experiments by Alexander Varshavsky and colleagues using formaldehyde to chemically cross-link chromosomal DNA to proteins (Fig. 2), and proceeding all the way through modern incarnations that use microarray (ChIP-on-chip) and massively parallel sequencing technologies (ChIP-Seq) to read out the locations of the bound regions. A conceptual framework is developed for understanding the three major processes under investigation: (1) interactions between

Grant sponsor: NIH; Grant numbers: R01 CA109147, R01 DK071122; Grant sponsor: Prostate Cancer Foundation; Grant sponsor: Whittier Foundation.

*Correspondence to: Benjamin P. Berman, PhD, Research Associate, USC Epigenome Center, 1450 Biggy St. Room #G511e, MC 9601, Los Angeles, CA 90033. E-mail: bberman@usc.edu

Received 10 February 2009; Accepted 12 February 2009 • DOI 10.1002/jcb.22133 • 2009 Wiley-Liss, Inc.

Published online 23 March 2009 in Wiley InterScience (www.interscience.wiley.com).

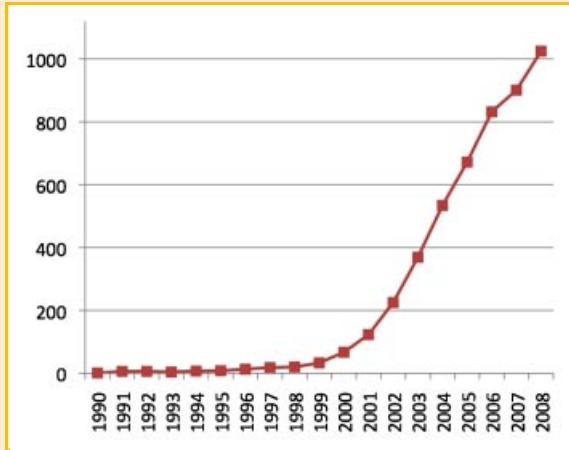


Fig. 1. Publication count using the search term "Chromatin Immunoprecipitation" in January 2009 at the ISI Web of Science (<http://apps.isiknowledge.com>). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

sequence-specific transcription factor complexes and their direct DNA binding targets and core RNA polymerase machinery, (2) epigenetic marks, most importantly histone tail modifications critical for the regulation of chromatin structure and function, and (3) localization of protein complexes that mediate long-range intra- and inter-chromosomal interactions and interact with the structural architecture of the nucleus.

Mouse and human embryonic stem cells have been one of the systems most extensively studied by ChIP-location, and Wacker and Kim [this issue] and Zechini and Mills [this issue] outline how it has substantially added to our understanding of the regulatory processes involved in pluripotency and lineage specification. Importantly,

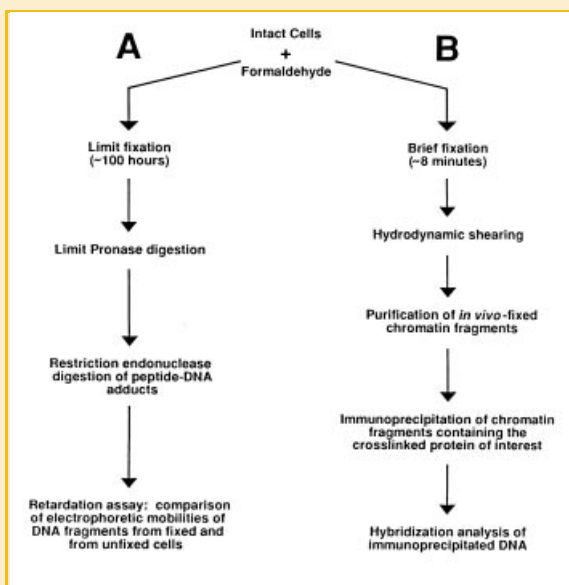


Fig. 2. In vivo chromatin mapping with formaldehyde [reproduced with permission from Solomon et al., 1988].

studying transcriptional control and epigenetic changes using ChIP-chip and ChIP-seq has aided in the development of artificially induced pluripotent stem cells (iPS) in mammals, a development which holds great promise for regenerative medicine. Some of the most interesting stem cell findings challenge older concepts about how global epigenetic marks and the recruitment of RNA Polymerase II to promoters influence transcription. The past several years of intense study have established that histone marks once considered "active" and "repressive" actually co-occur at many promoters during stem cell development and interact in ways that we still do not fully understand. We have also learned that RNA polymerase II is present at the promoters of many inactive genes, possibly maintaining a state of competency to allow for future activation and elongation [Muse et al., 2007; Zeitlinger et al., 2007; Wu and Snyder, 2008]. ChIP-location has recently been used to identify a large number of long non-coding RNAs expressed in stem cells [Guttman et al., 2009], some of which may play critical roles in the establishment of these chromatin domains [Rinn et al., 2007; Zhao et al., 2008].

ChIP-location studies are also facilitating new insights into the biology of cancer. Zechini and Mills [this issue] describe how the study of genomic regulatory processes involving nuclear hormone receptors has led to a number of important insights, including the computational identification of key DNA-binding motifs and novel transcriptional co-factors such as FoxA1 [Carroll et al., 2005], as well as the connection between transcription factor DNA binding and the recruitment of both histone modifying enzymes and their associated histone tail modifications [Strahl and Allis, 2000; Bernstein et al., 2007; Heintzman et al., 2007]. Identification and interrogation of regulatory control elements lying downstream of key signaling pathways such as these will prove invaluable for interpreting the results of genome wide polymorphism association studies (GWAS), which are increasingly implicating the non-coding sequences of the genome in disease risk. Examples are the role of a common, non-protein coding regulatory variant in the *RET* gene that explains much of Hirschprung's disease [Sancandi et al., 2003; Fitze et al., 2003a,b], and more recently, the association of non protein-coding variation in chromosome 8q24 with prostate and other cancers [Amundadottir et al., 2006; Freedman et al., 2006; Gudmundsson et al., 2007; Haiman et al., 2007; Yeager et al., 2007], and in the *FGFR2* intron with breast cancer [Meyer et al., 2008]. Other unexpected observations are made possible by global ChIP-location, for instance that most known translocation breakpoints in T-cell cancer are located in regions of active and open chromatin based on histone profiling [Barski et al., 2007]. Cancer research also underscores the need to reduce the amount of starting material necessary to perform ChIP analyses—current protocols require millions of cells, but new techniques, requiring 10,000 cells or less [Acevedo et al., 2007], will be necessary to move from cell lines to primary tumors and cell-sorted subpopulations of cancer stem cells.

Two articles focus on technical aspects of emerging ChIP techniques. Barski and Zhao [this issue] discuss ChIP-seq, where ChIP isolated fragments are directly sequenced using high-density ("2nd generation") sequencers capable of generating hundreds of millions of sequence tags in a single experiment. ChIP-seq has a number of advantages over its predecessor, ChIP-on-chip: First, it does not suffer from hybridization artifacts that confound all

microarray-based assays. Second, while microarray approaches can typically cover roughly the half of they typical mammalian genome that does not cross-hybridize to oligonucleotide probes, ChIP-seq can currently cover over 70% [Mikkelsen et al., 2007] and will very quickly improve to cover the vast majority including many transposable elements. While some consider repetitive sequences to be “selfish” DNA that does not contribute to genomic regulation, there is little direct evidence for this. To the contrary, significant evidence exists that transposable elements can have a direct effect on regulatory processes [Bejerano et al., 2006] and contribute significantly to the evolutionary processes that shape *cis*-regulatory sequences [Feschotte, 2008; Xie et al., 2006]. Another important benefit of ChIP-seq is that it has the ability to detect allele-specific binding [Mikkelsen et al., 2007].

Distinguishing signal from noise in whole genome ChIP-seq and ChIP-chip data is a difficult statistical problem, and Barski and Zhao [this issue] provide a substantive discussion of the issues involved. While a number of techniques using sliding window averages [Jothi et al., 2008] and hidden markov models [Ji and Wong, 2005; Du et al., 2006] have been developed, all have substantial error rates and most are poorly suited to the extended domains characteristic of many epigenetic marks. Zechini and Mills [this issue] highlight this problem in a discussion of the troubling lack of replicability in whole-genome ChIP-on-chip experiments when performed by different labs. While some of this variability might be truly biological, it is clear that at least some is technical in nature [Johnson et al., 2008] and is highly dependent on selecting an appropriate underlying statistical model and enrichment cutoff. Another emerging factor is that ChIP assays seem to pick up bona fide low-level binding events that do not appear to have any physiological significance. A recent study [Li et al., 2008] made the interesting observation that the most highly bound regions were a great deal better at predicting gene expression patterns than the more weakly bound regions, even within a set of highly statistically significant binding regions that were replicatable using an independent antibodies. This suggests that while a ChIP-location experiment for a single factor in isolation might not be an extremely good predictor of regulatory function, the addition of ChIP data for a second related protein, such as a co-factor or a related epigenetic mark, may multiply the predictive power. Indeed, when the epigenetic activation mark for histone H3 acetylation was added to ChIP data for the sequence specific transcription factor AR, correlation with regulatory function was exponentially better [Jia et al., 2008].

Another confounding factor in the analysis of ChIP-seq data is how to deal with copy number changes and other chromosomal aberrations common in cancer genomes. Most ChIP-chip study designs deal with this by normalizing against a control sample consisting of ChIP input DNA or a ubiquitous mark such as unmodified histone H3. We believe it is always a good idea to follow such strategy with ChIP-seq, because it is now well established that ChIP-seq can have a tendency to enrich non-specifically for many nucleosome-free hypersensitive sites in the genome [Giresi et al., 2007] (particularly at active promoters). Because cancer genomes often contain many regions with increased copy number, they present an even larger background problem and many more

sequencing tags will be necessary to control for these effects and prevent false positives. While this may not be a major concern when ChIPping a factor with very high enrichment over a relatively small fraction of the genome (like NRSF [Johnson et al., 2007], or STAT1 [Robertson et al., 2007]), it could present serious issues when trying to identify domains with low enrichment covering large stretches of the genome, for example histone marks H3K36me3 and H3K27me3.

Barski and Zhao [this issue] highlight the data processing requirements of generating and analyzing high throughput ChIP-location data. Development of a bioinformatics infrastructure for storage, analysis, and exchange of this data will be a crucial challenge for the genomics community. While repositories for proteins, genes, and promoters (Genbank) and gene expression microarray data (GEO) have existed for a number of years, ChIP-seq datasets are several orders of magnitude bulkier than their microarray analogs. Central resources such as NCBI’s new short read archive [Wheeler et al., 2008] will be critical for warehousing such experiments, but we will need analysis tools that make it possible to compare new or private ChIP datasets (at base pair resolution) to previously published ones. The UCSC and ENSEMBL genome browser groups have provided some early tools for these types of analysis, but they do not really fill the needs of modern ChIP-based research. Large collaborative projects such as the ENCODE Consortium will undoubtedly help to spur such developments [Thomas et al., 2007].

In the final article, Fullwood and Ruan [this issue] describe the newest frontier in ChIP-location, and how it is being used to study the regulatory mechanisms involved in long-range chromosomal interactions. Chromatin Conformation Capture (3C) is a method for identifying distinct genomic regions bound by a common protein or protein complex, which brings them in close spatial proximity within the 3D space of the nucleus [Simonis et al., 2007]. The combination of 3C with ChIP has led to the elucidation of chromosomal looping events that bring together distant sites into central hubs or interchromatin granules that are critical for gene activation [Hu et al., 2008]. While these techniques seem poised to emerge as the next major step in the evolution of ChIP, they are fraught with numerous sources of noise and potential bias. Fullwood and Ruan [this issue] describe an emerging methodology including the optimum use of sonication-based chromatin fragmentation, ChIP-based enrichment, chromatin proximity ligation and Paired-End Tag ultra-high-throughput sequencing as the winning strategy to establish DNA/protein interactomes in three-dimensional space.

ChIP-location is also helping define the architecture of the nucleus by mapping the nuclear organizing proteins themselves. Binding sites for the insulator protein CTCF [Cuddapah et al., 2009] and Nuclear Lamina attachment proteins [Guelen et al., 2008] have led to unexpected connections between gene regulation and sub-nuclear chromatin organization. Broad disruption of chromatin looping has been recently implicated in the progression of breast cancer [Han et al., 2008], and it will be exciting to determine what other roles nuclear organization plays in cancer and other diseases. ChIP-location will undoubtedly play a role in this discovery process.

ChIP-location techniques have the potential to help unravel some of the most fundamental and vexing questions in molecular biology, for instance whether the contribution of simple *cis*-acting

regulatory motifs or more indirect, trans-acting factors in the nuclear environment are more central in the regulation of complex transcriptional networks. Zechinni and Mills [this issue] dive into this debate head on, highlighting a recent ground breaking study [Wilson et al., 2008] involving a trisomic mouse model that carries most of human chromosome 21. Using ChIP-on-chip to determine the location of key liver-specific transcription factors and epigenetic histone marks in liver cells, they showed that despite the very different cellular environment, regulatory regions on the heterologous human chromosome have behavior almost identical to that of their native human cell counterparts, but very different from the homologous mouse regions. Gene expression from the transplanted chromosome follows the same pattern, seemingly unaffected by the differences of the mouse cellular environment and arguing for local *cis*-acting sequences playing a dominant role. While this stunning result addresses the question of *cis*-acting sequences and the trans-acting nuclear milieu, a related question is the primacy of transcription factors binding their short recognition motifs versus larger scale epigenomic structural events. Does the former determine the latter via the histone modifying activities of transcription factors, or does epigenomic organization precede and make possible site-specific binding? While the number of well-understood cases is still too small to say for sure, the ability to profile transcription factor occupancy and markers of chromatin structure in parallel as a function of time and development will facilitate better understanding of this relationship. Recent studies in yeast and *Drosophila* [Segal et al., 2008; Gertz et al., 2009] lend support to the notion that a complex arrangement of short protein binding sites can explain much of the complexity observed in transcriptional regulation. How such sequences encode complex animal regulatory programs is an ongoing pursuit, but it is clear that we are only in the early stages of being able to identify and interpret the code within these *cis*-regulatory sequences [Hare et al., 2008]; ChIP-location will play the key role in the pursuit of this *cis*-regulatory “grammar,” by allowing the identification of many thousands of new functional *cis*-regulatory regions.

In the span of a short decade, ChIP-location techniques have allowed genomic maps to progress from one-dimension to richly populated 2D and even 3D landscapes. Although we are still learning the basics of how to navigate in this dynamic chromatin world, it is certainly an exciting time to be a genome explorer.

ACKNOWLEDGMENTS

Work reported from the authors was supported by the NIH (R01 CA109147 to GAC and R01 DK071122 to BF), the Prostate Cancer Foundation (to GAC), the Whittier Foundation (to GAC) and the J. Harold and Edna L. LaBriola Chair in Genetic Orthopaedic Research (held by BF).

REFERENCES

Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, Farnham PJ. 2007. Genome-scale ChIP-chip analysis using 10,000 human cells. *Biotechniques* 43:791–797.

Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediksdottir KR, Cazier JB, Sainz J, Jakobsdottir M,

Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. 2006. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38:652–658.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90.

Bernstein BE, Meissner A, Lander ES. 2007. The Mammalian epigenome. *Cell* 128:669–681.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122:33–43.

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19:24–32.

Du J, Rozowsky JS, Korbel JO, Zhang ZD, Royce TE, Schultz MH, Snyder M, Gerstein M. 2006. A supervised hidden markov model framework for efficiently segmenting tiling array data in transcriptional and ChIP-chip experiments: Systematically incorporating validated biological knowledge. *Bioinformatics* 22:3016–3024.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397–405.

Fitze G, Appelt H, Konig IR, Gorgens H, Stein U, Walther W, Gossen M, Schreiber M, Ziegler A, Roesner D, Schackert HK. 2003a. Functional haplotypes of the RET proto-oncogene promoter are associated with Hirschsprung disease (HSCR). *Hum Mol Genet* 12:3207–3214.

Fitze G, Schierz M, Kuhlisch E, Schreiber M, Ziegler A, Roesner D, Schackert HK. 2003b. Novel intronic polymorphisms in the RET proto-oncogene and their association with Hirschsprung disease. *Hum Mutat* 22:177.

Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D. 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 103:14068–14173.

Gertz J, Siggia ED, Cohen BA. 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* 457:215–218.

Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. 2007. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Res* 17:877–885.

Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediksdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JJ, Kiemeny LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. 2007. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39:631–637.

- Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453:948–951.
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*.
- Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D. 2007. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39:638–644.
- Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T. 2008. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* 452:187–193.
- Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. 2008. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet* 4:e1000106.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39:311–318.
- Hu Q, Kwon YS, Nunez E, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG, Fu XD. 2008. Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci USA* 105:19199–19204.
- Ji H, Wong WH. 2005. TileMap: Create chromosomal map of tiling array hybridizations. *Bioinformatics* 21:3629–3636.
- Jia L, Berman BP, Jariwala U, Yan X, Cogan JP, Walters A, Chen T, Buchanan G, Frenkel B, Coetzee GA. 2008. Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity. *PLoS One* 3:e3645.
- Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316:1497–1502.
- Johnson DS, Li W, Gordon DB, Bhattacharjee A, Curry B, Ghosh J, Brizuela L, Carroll JS, Brown M, Flicek P, Koch CM, Dunham I, Bieda M, Xu X, Farnham PJ, Kapranov P, Nix DA, Gingeras TR, Zhang X, Holster H, Jiang N, Green RD, Song JS, McCuine SA, Anton E, Nguyen L, Trinklein ND, Ye Z, Ching K, Hawkins D, Ren B, Scacheri PC, Rozowsky J, Karpikov A, Euskirchen G, Weissman S, Gerstein M, Snyder M, Yang A, Moqtaderi Z, Hirsch H, Shulha HP, Fu Y, Weng Z, Struhl K, Myers RM, Lieb JD, Liu XS. 2008. Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res* 18:393–403.
- Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* 36:5221–5231.
- Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmam R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* 6:e27.
- Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C, Ponder BA. 2008. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* 6:e108.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560.
- Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* 39:1507–1511.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129:1311–1323.
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4:651–657.
- Sancandi M, Griseri P, Pesce B, Patrone G, Puppo F, Lerone M, Martucciello G, Romeo G, Ravazzolo R, Devoto M, Ceccherini I. 2003. Single nucleotide polymorphic alleles in the 5' region of the RET proto-oncogene define a risk haplotype in Hirschsprung's disease. *J Med Genet* 40:714–718.
- Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451:535–540.
- Simonis M, Kooren J, de Laat W. 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4:895–901.
- Solomon MJ, Larsen PL, Varshavsky A. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53:937–947.
- Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* 403:41–45.
- Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, Kuhn RM, Rhead BL, Smith KE, Thakkapallayil A, Zweig AS, Haussler D, Kent WJ. 2007. The ENCODE Project at UC Santa Cruz. *Nucleic Acids Res* 35:D663–D667.
- Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmsberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. 2008. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36:D13–D21.
- Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT. 2008. Species-specific transcription in mice carrying human chromosome 21. *Science* 322:434–438.
- Wu JQ, Snyder M. 2008. RNA polymerase II stalling: Loading at the start prepares genes for a sprint. *Genome Biol* 9:220.
- Xie X, Kamal M, Lander ES. 2006. A family of conserved noncoding elements derived from an ancient transposable element. *Proc Natl Acad Sci USA* 103:11659–11664.
- Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancell-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF, Jr., Hoover R, Hunter DJ, Chanock SJ, Thomas G. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39:645–649.
- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39:1512–1516.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322:750–756.