REVIEW

Recent advances in functional genome analysis [version 1; referees: 2 approved]

Roderic Guigo [1], Michiel de Hoon [2]

[1]Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain
[2]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

v1 **First published:** 21 Dec 2018, **7**(F1000 Faculty Rev):1968 (https://doi.org/10.12688/f1000research.15274.1)
Latest published: 21 Dec 2018, **7**(F1000 Faculty Rev):1968 (https://doi.org/10.12688/f1000research.15274.1)

Abstract
At the beginning of this century, the Human Genome Project produced the first drafts of the human genome sequence. Following this, large-scale functional genomics studies were initiated to understand the molecular basis underlying the translation of the instructions encoded in the genome into the biological traits of organisms. Instrumental in the ensuing revolution in functional genomics were the rapid advances in massively parallel sequencing technologies as well as the development of a wide diversity of protocols that make use of these technologies to understand cellular behavior at the molecular level. Here, we review recent advances in functional genomic methods, discuss some of their current capabilities and limitations, and briefly sketch future directions within the field.

Keywords
functional genomics, massively paralel sequencing

Open Peer Review

Referee Status: ✓ ✓

	Invited Referees 1	2
version 1 published 21 Dec 2018	✓	✓

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

1 **Feng Zhang**, Xi'an Jiaotong University, China

2 **Kerstin Lindblad-Toh**, Uppsala University, Sweden
Broad Institute of MIT and Harvard, USA

Discuss this article

Comments (0)

Corresponding authors: Roderic Guigo (roderic.guigo@crg.cat), Michiel de Hoon (michiel.dehoon@riken.jp)

Author roles: Guigo R: Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **de Hoon M**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Roderic Guigo acknowledges the support of the "Salvador de Madariaga" program from the Spanish Ministry of Education, Culture and Sports during his sabbatical at the Riken; the support of the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa", and the EMBL partnership; and the support of the CERCA Programme/Generalitat de Catalunya. Michiel de Hoon is supported by a research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) to the RIKEN Center for Integrative Medical Sciences.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Guigo R and de Hoon M. **Recent advances in functional genome analysis [version 1; referees: 2 approved]** *F1000Research* 2018, **7**(F1000 Faculty Rev):1968 (https://doi.org/10.12688/f1000research.15274.1)

First published: 21 Dec 2018, **7**(F1000 Faculty Rev):1968 (https://doi.org/10.12688/f1000research.15274.1)

Introduction

Progress in genomics during the last decade has been largely dependent on the development of methods for massively parallel sequencing of large numbers (up to a few hundred million per experiment in some cases) of short DNA sequences, where a single sequence read is about 50 to 150 base pairs. These methods—often referred to as next-generation sequencing (NGS)—are used for genome sequencing, but their high throughput also makes them very attractive for functional genomics assays, where the primary goal is to annotate genome regions with their biological function. This often requires quantitative estimates, which typically are derived from read counts. A myriad of experimental protocols and the accompanying bioinformatic pipelines have been designed that leverage the high-throughput capabilities of NGS to monitor a wide variety of molecular phenotypes reflecting very diverse genomic functions.

In this review, we summarize the protocols that are most frequently applied in functional genomics. Although oligonucleotide array-based methods for genotyping and gene expression are still widely used for large population screens because of the low cost and straightforward bioinformatics, genome tiling arrays, which were instrumental in the emergence of functional genomics[1], have largely been substituted by sequencing-based methods. Our review thus focuses on these. We also discuss recently developed methodologies for large-scale perturbation experiments that can probe the functionality of genomic loci in a massively parallel fashion. We discuss the limitations of short read length, efforts toward developing long read sequencing technologies, and novel approaches to analyze the behavior of single cells and cells in the context of their tissue of origin. We conclude by discussing the role of large-scale international consortia and future directions in functional genomics.

Genomic protocols

Chromatin can be divided into the transcriptionally active open euchromatin and the transcriptionally silent heterochromatin or closed chromatin. The ATAC-seq (assay for transposase-accessible chromatin using sequencing) protocol uses transposases acting on accessible chromatin to create DNA fragments, which then are sequenced to identify open genomic regions in a particular cell type or condition[2]. The number of cells from which the DNA is extracted is important, as too few cells may cause excessive digestion, leading to reads mapping to closed chromatin regions, whereas an excess of cells can cause insufficient digestion, creating long fragments that may be difficult to sequence[3]. The methylation status of DNA can be assayed by using endonuclease digestion by methylation-dependent restriction enzymes, by using bisulfite treatment to convert unmethylated cytosine to uracil (which is recognized as thymine during sequencing), or by using methylation-dependent antibodies to immunoprecipitate methylated chromatin followed by sequencing (ChIP-seq)[4]. False positives may occur if unmethylated cytosines fail to convert to uracil and are interpreted as methylated cytosines. It is essential to minimize DNA degradation during bisulfite treatment, as DNA fragmentation will hamper polymerase chain reaction (PCR) amplification. Tet-assisted bisulfite sequencing[5] can be used to resolve 5-methylcytosine and 5-hydroxymethylcytosine, both of which are found in mammalian genomes[6,7] but are indistinguishable in traditional bisulfite sequencing. ChIP-seq-based approaches do not give single-nucleotide resolution of DNA methylation and are affected by genomic variations in CpG density.

Based on the availability of a suitable antibody, ChIP-seq can also be used to assay histone modifications[8] or the binding affinities of transcription factors to genomic regions[9,10]. Improvements to the ChIP-seq protocol have allowed greater resolution while reducing the number of cells required[11,12]. The use of a sensitive but specific antibody is essential for obtaining high-quality ChIP-seq data. Genomic assays typically produce an "epigenomic profile" along the genome (by mapping reads obtained by the assay to the genome). To combine and integrate epigenomic profiles produced by different assays (for instance, ChIP-seq from multiple histone modifications), segmentation algorithms are employed (for instance, 13,14). These algorithms partition the genome into regions (segments) of similar profiles over all assays.

Transcriptome protocols

Quantitative profiling of the transcriptional output of the genome relies on sequencing libraries of cDNA fragments derived from RNA (RNA-seq)[15]. Sequence reads are subsequently assembled bioinformatically to reconstruct the full-length transcript sequence[16,17] and to quantify the expression levels of annotated genes[16,17]. Cap analysis gene expression (CAGE), in contrast, specifically sequences the 5′ end of transcripts to pinpoint the transcription start site and therefore the promoter region of each transcript[18]. CAGE uses a random oligonucleotide as the primer for the reverse transcription reaction, instead of an oligo-dT primer against the poly(A)-tail, allowing both poly(A)+ and poly(A)− transcripts—such as certain long non-coding RNA (lncRNAs)—to be profiled. Specific RNA fractions can be selected for inclusion in the sequencing libraries depending on their size, the chemical group at their 5′ end, the presence or absence of a poly(A) tail, or subcellular localization in the cytoplasm, nucleoplasm, or chromatin fraction[19–21]. These fractions may correspond to different classes of RNA or different stages in the pathway of RNA synthesis and post-processing. Protocols to assess the RNA-interacting partners include ribosome profiling[22] to identify mRNAs undergoing translation as well as various protocols based on high-throughput sequencing of RNA after cross-linking and immunoprecipitation (CLIP-seq and its variants)[23] to find transcripts bound to a specific protein. RNA modifications[24] such as methylation of adenosine nucleotides[25] can be identified by methylation-dependent immunoprecipitation of RNA followed by sequencing[26,27].

Profiling of short non-coding RNAs—in particular, microRNAs (miRNAs)—typically relies on the ligation of oligonucleotide adapters to the 3′ and 5′ end of the short RNA followed by reverse transcription and PCR amplification making use of primer sequences in the adapters[28]. Performing ligation of the 5′ adapter after reverse transcription allows ligation

to the 5′ end of the cDNA, avoiding biases due to the presence of any modifications at the 5′ end of the short RNA molecule[29]. To avoid biases introduced by oligonucleotide adapter ligation to the 3′ end, polyadenylation of the short RNA may be used instead. However, a disadvantage of this approach is that it does not allow the identification of the exact 3′ end of the short RNA, which is unfortunate, as addition and removal of nucleotides—in particular, adenosines—to the 3′ end of miRNAs are prevalent and may play a regulatory role in miRNA stabilization and degradation[30–32].

Regulation of expression in the 3D context of the genome

The 3D structure of the genome in the nucleus provides the physical architecture for expression regulation. Hi-C libraries[33], which are produced by chemical cross-linking, DNA fragmentation, proximity ligation, and high-throughput sequencing of ligated fragments, have shown that chromatin is organized in topologically associating domains, each consisting of chromosome regions that frequently interact with each other, and chromatin looping in which distal regulatory regions such as enhancers and promoters come in close physical proximity to each other. ChIA-PET (chromatin interaction analysis by paired-end tag sequencing)[34] combines proximity ligation of DNA fragments with chromatin immunoprecipitation of specific proteins to globally detect the chromatin interactions in which they participate.

As transcripts produced at regulatory sites[35,36] may mediate distal interactions[37], identifying the chromatin binding sites of regulatory RNAs is of central importance. Chromatin isolation by RNA purification (ChIRP-seq)[38] uses tiling oligonucleotides to pull down a specific lncRNA together with the protein and DNA bound to it, followed by sequencing of the DNA component to determine the exact binding positions of the lncRNA to chromatin. For unbiased mapping of RNA–chromatin interactions, on a genome-wide scale, several protocols related to Hi-C have been developed that rely on proximity ligation of DNA to RNA using a bivalent oligonucleotide linker molecule: MARGI (mapping RNA–genome interactions)[39], global RNA interactions with DNA by deep sequencing (GRID-seq)[40], and chromatin-associated RNA sequencing (ChAR-seq)[41].

High-throughput genomic perturbations

Although chromatin features and transcriptional activity at a genomic region can be indicative of a biological role, ultimately the function of a genomic region is defined by its contribution to the phenotype. Perturbation studies, in which activation or inhibition of specific genomic regions or their products is followed by phenotypic or molecular assays of the cellular response, can be used to elucidate the functional role of genomic loci.

Recent developments in perturbation technologies now allow genomic perturbation experiments to be highly multiplexed, with thousands of knockdowns or activations occurring in parallel in a population of cells. High-throughput perturbation experiments have become feasible in particular because of the development of the CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/Cas9) technology[42] to edit or otherwise modify the genome at specific sites. Whereas programming zinc finger nucleases[43] and transcription activator-like effector nucleases (TALENs)[44] to target specific genomic sites required extensive protein engineering that limited the scalability of these approaches, CRISPR/Cas9 recognizes and cleaves a particular DNA site using only a single guide RNA, which can easily be mass-produced. Nucleotide insertions and deletions introduced near the double-strand break by non-homologous end-joining can abrogate the expression or function of protein-coding genes but may not be sufficient to fully abolish the function of lncRNAs. Therefore, dual-CRISPR systems have been developed that use a pair of guide RNAs to generate double-strand breaks on both sides of a gene of interest, followed by non-homologous end-joining to repair the lesion, resulting in a full deletion of the gene[45,46].

The applicability of this technology can be extended beyond genome editing by replacing the endonuclease Cas9 with catalytically inactive Cas9 (dead Cas9, or dCas9), which binds to specific genomic sites without inducing cleavage. In CRISPR interference (CRISPRi)[47], tight binding of CRISPR-dCas9 to DNA prevents binding of the transcriptional machinery, effectively shutting down transcript production at the targeted site. Improvements to the knockdown efficiency were achieved by fusing a Krüppel-associated box (KRAB) repressor domain to Cas9[48], inducing histone modification changes associated with repression of expression. Similarly, dCas9 fused to a transactivating domain can be used to induce gene expression at specific sites[48]. Epigenome editing can be performed by fusing dCas9 with epigenome writers or erasers (for example, a histone demethylase)[49].

Second-generation CRISPR editing tools allow the conversion of specific nucleotides at the genomic target site without inducing a double-strand break (for example, to create a novel stop codon)[50].

Large-scale CRISPR-based screening relies on designing guide RNAs in large quantities and introducing them, usually using a lentiviral vector, into a population of cells at a sufficiently low multiplicity of infection, such that most cells will receive either zero or one guide RNA; puromycin selection is applied to exclude cells that were not infected. Typically, up to 10 guide RNAs are used for a single locus, allowing checks for reproducibility and off-target effects due to inadvertent binding of guide RNAs elsewhere in the genome. During cell culture, the proportion of cells containing guide RNAs that inhibit the expression of genes essential for cell growth will decrease. Therefore, genes involved in cell growth can be identified by quantifying, using NGS, the relative enrichment of each guide RNA after culturing[51]. CRISPRi-mediated knockdowns followed by single-cell genome-wide expression profiling (Perturb-seq)[52,53] can be employed to obtain detailed information on the global transcriptome response to the perturbation and to observe cell state changes that are not easily assessed in a straightforward cellular assay.

For functional genomics, perturbation technologies that target DNA have one inherent drawback: it is unclear whether any observed phenotype is due to the inhibition of regulatory DNA elements at the targeted locus or to any RNA product originating there. This ambiguity can be avoided by targeting specific transcripts directly using locked nucleotide antisense oligonucleotides (ASO gapmers) or by using small interfering RNAs (siRNAs) or small hairpin RNAs (shRNAs). Target RNAs bound by an ASO gapmer are recognized and cleaved by endonuclease RNase H, followed by degradation of the RNA. In contrast, the RNA interference (RNAi) pathway relies on processing of siRNAs and shRNAs by Dicer and uptake into the RNA-induced silencing (RISC) complex, which binds to the target RNA, causing its cleavage by Argonaute and degradation by cellular exonucleases. The relative performance of RNAi or RNase H-mediated knockdown depends in part on the cellular localization of the targeted transcripts in either the nucleus or the cytoplasm[54].

Long read sequencing

Reconstruction of the full-length sequence of long DNA or RNA molecules from NGS sequence reads is challenging because of their short read length. Two main technologies are being developed for long read sequencing (that is, sequencing of single molecules several thousand base pairs long). The first relies on the ability to detect the incorporation of a single nucleotide in an elongating DNA chain, which is synthesized by a DNA polymerase attached to the template DNA molecule to be sequenced[55]. The second makes use of the ability to detect changes in an electrical field that are induced when a single nucleotide in the DNA molecule passes through a pore structure[56]. Long read technologies have been used for "de novo" genome sequencing[57], resequencing of the human genome[58], and phased diploid genome assembly[59]. They have also been used for transcript characterization[60], since short reads are unable to fully resolve the exonic structure of alternative transcript isoforms, which usually share a substantial fraction of the sequence. One advantage of long read technologies is that they are able to directly sequence RNA molecules[26,61], eliminating the biases induced during the generation of the cDNA libraries inherent to short read NGS.

Capture methods

Because of the large dynamic range of genome activity (for instance, gene expression), unbiased sequencing of samples to interrogate function genome-wide may be ineffective in uncovering sites within the genome with weak activity. Therefore, depending on the purpose, NGS analysis may be restricted to targeted genomic loci captured using oligonucleotide probes. Capture has been performed, for instance, to identify and characterize the exonic structure of lowly abundant transcripts (lncRNAs, in particular) using both short[62] and long[63] read sequencing.

Spatial transcriptomics

Protocols to prepare libraries for NGS typically require a fair amount of input material (RNA or chromatin), necessitating large numbers of cells to produce a single library. If the samples are obtained from heterogeneous biological sources, such as tissues and organs, they are likely to contain multiple distinct cell types, which interact in a highly structured anatomical environment. Bulk sequencing of these samples reflects an average across the populations of cells used and is unable to resolve the underlying cell types and their position within the tissue sample. Recently, spatially resolved omics methods have been developed with the aim of performing functional genomics assays while retaining positional information[64,65]. Traced back to single-molecule fluorescence *in situ* hybridization (FISH)[66], these methods have dramatically increased throughput by using massively parallel sequencing. For instance, histological sections have been positioned on arrayed reverse transcription primers with unique positional barcodes to maintain the two-dimensional positional information in subsequent RNA-seq analysis[67].

Single-cell technologies

Population averages, such as obtained from bulk tissue sequencing, may further obscure the biological reality of cellular heterogeneity even among cells of the same type. For example, genes that appear to be lowly expressed on the basis of their population average may be highly expressed in a few cells in the population and exert their biological role only there. To capture the heterogeneity among cells in a population, newer protocols have been developed that dramatically lower the amount of input material required, allowing profiling of individual cells[68]. Genome sequencing of single cells has been used to identify genomic changes in tumor samples and to track their evolution[69–74].

Single-cell RNA profiling can assay 50 to 10^4 cells in parallel per experiment, depending on the protocol used; typical coverages are 1,000 to 6,000 genes per cell in case of primary cells and 5,000 to 10,000 genes for cell lines[75]. In plate-based methods such as STRT (single-cell tagged reverse transcription)[76], Smart-seq[77], and Smart-seq2[78], 50 to 500 single cells are placed in individual wells in 96-well plates, and cell lysis and reverse transcription occur in each well. Alternatively, a microfluidic chip can be used for cell placement and reactions[79].

Cell selection is biased by the cellular size and condition, which affects the probability of a cell to be placed successfully in a well. In pooled approaches such as CEL-seq[80,81], a unique barcode is applied to individual cells in the initial stages in the protocol. As the barcode is incorporated in the cDNA during reverse transcription, the cDNAs originating from different cells are distinguishable and can be pooled for PCR amplification. To further increase throughput and reproducibility, in massively parallel single-cell RNA-seq, cells are sorted by fluorescence-activated cell sorting (FACS) into more homogeneous subpopulations in a 384-well plate before single-cell sequencing[82]. As a poly-T primer is typically used for reverse transcription to selectively include poly(A)-tailed transcripts into the library while avoiding ribosomal RNA, other untailed transcripts such as some lncRNAs may be missed. Alternative strategies to reduce the ribosomal RNA component include duplex-specific nuclease treatment[83] and using a template-switching reaction to enrich for 5′-capped

transcripts[84]. Droplet-based isolation of single cells[85,86] enabled massive parallelization with numbers of cells in the range of 10,000 or more, albeit at lower sensitivity.

As an alternative to single-cell analysis, deconvolution methods can be employed to estimate the proportions of constituent cell types in bulk tissue transcriptomic samples on the basis of the expression of known cell type marker genes[82,87].

Functional genomics across species

Comparative genomics methods have also been widely used to identify and catalogue functional regions in genomes. The underlying assumption is that functional regions are more conserved through evolution than genomic regions devoid of function. Genome sequence comparisons are widely used, in particular, to annotate protein-coding genes in newly sequenced genomes. However, while protein coding genes are under strong selective pressure, other functional regions, such as lncRNA genes and regulatory regions, have levels of conservation that are often barely distinguishable from background conservation[88] and therefore genome comparisons are, in general, less helpful to annotate them. For instance, whereas 80% of human protein-coding genes can be detected in mouse, fewer than 10% of lncRNAs are conserved between the two species[89]. Comparative transcriptomic studies help to further understand to what extent the patterns of gene expression and the regulatory networks are conserved across species[90–92]. Particular attention has been paid to comparisons between human and mouse given the unique role of the mouse as a model of human biology[93,94]. While gene expression levels averaged over RNA samples are correlated between human and mouse[89], substantial differences are found when comparing the transcriptome of homologous cell types[95–97]. Overall, conservation of the patterns of expression between human and mouse seems to largely depend on the gene. The expression of some genes varies considerably across species and little across organs and that of other genes varies more across organs than across species. The mouse is a particularly good model for the biology of these genes in humans[98].

Outlook

Massively parallel sequencing has been at the core of most large-scale functional genomic projects to date[36,99–103]. These projects have not only generated data resources that are widely used by the scientific community but also defined "de facto" community standards for experimental protocols and bioinformatic pipelines, placing genomics in general and functional genomics in particular within the reach of individual laboratories. As a result of advances in experimental and computational methods, functional genomics assays are rapidly expanding beyond cell lines and tissue samples, which were mostly profiled in these projects, to complex organs, conditions, primary cells, and even single cells. Indeed, the Human Cell Atlas consortium has formed to characterize each individual cell type in the human body[104]. Functional genomics assays are also being increasingly performed across multiple individuals within a species, allowing investigators to specifically link genetic variants to molecular phenotypes, as well as across different species. Most of these analyses to date have focused on the identification of single-nucleotide polymorphisms (SNPs) that affect gene expression (expression quantitative trait loci)[103,105] but other molecular phenotypes, such as methylation and histone modifications, have also been monitored[106]. Moving toward personalized genomics, the EN-TEx pilot project is currently carrying out multiple functional genomics assays in multiple tissues extracted from a few donors[107].

How these molecular phenotypes propagate in turn through intermediate levels of organizational complexity (cells, tissue, organs, and so on) determines how genetic variation ultimately impacts the phenotypic traits of organisms. Thanks to advances in digital imaging, data on intermediate phenotypes, such as tissues and organs, have recently been made available with associated genomics and molecular data in areas as diverse as cancer[108] and neuropathology[109]. The GTEx project[103] has produced high-resolution histological images for 20,000 human tissue samples, together with matching genome and transcriptome sequencing data. Methods to identify SNPs associated with image-derived organ phenotypes such as neuroanatomical traits (that is, cortical thickness[110]) have been developed, and an upcoming challenge is to relate molecular phenotypes, as monitored by functional genomics, to organ phenotypes, as monitored by imaging technologies.

We believe that we are still in the infancy of genomic technologies and there is ample opportunity for progress. The ultimate goal is the real-time *in vivo* multi-omics characterization of all of the cells within a multicellular organism—without interfering with the system. In the path toward this ideal, less-invasive protocols and protocols to probe the multiple facets of genome function at single-cell resolution (that is, protocols to perform multiple functional genomics assays in the same cell) will certainly be developed, as will methods able to simultaneously integrate data monitoring the behavior of millions of cells. The computational challenges associated with the exponential growth of data, driven by genomic technologies, will be enormous. New algorithms as well as new developments in hardware that overcome the limitations of current technological solutions may need to be developed.

Grant information

Roderic Guigo acknowledges the support of the "Salvador de Madariaga" program from the Spanish Ministry of Education, Culture and Sports during his sabbatical at the Riken; the support of the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa", and the EMBL partnership; and the support of the CERCA Programme/Generalitat de Catalunya. Michiel de Hoon is supported by a research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) to the RIKEN Center for Integrative Medical Sciences.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

F1000 recommended

1. F ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, *et al.*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature.* 2007; **447**(7146): 799–816.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

2. F Buenrostro JD, Giresi PG, Zaba LC, *et al.*: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** *Nat Methods.* 2013; **10**(12): 1213–8.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

3. Buenrostro JD, Wu B, Chang HY, *et al.*: **ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.** *Curr Protoc Mol Biol.* 2015; **109**(1): 21.29.1–9.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Laird PW: **Principles and challenges of genomewide DNA methylation analysis.** *Nat Rev Genet.* 2010; **11**(3): 191–203.
PubMed Abstract | Publisher Full Text

5. Yu M, Hon GC, Szulwach KE, *et al.*: **Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine.** *Nat Protoc.* 2012; **7**(12): 2159–70.
PubMed Abstract | Publisher Full Text | Free Full Text

6. F Tahiliani M, Koh KP, Shen Y, *et al.*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science.* 2009; **324**(5929): 930–5.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

7. F Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science.* 2009; **324**(5929): 929–30.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

8. F Mikkelsen TS, Ku M, Jaffe DB, *et al.*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature.* 2007; **448**(7153): 553–60.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

9. F Johnson DS, Mortazavi A, Myers RM, *et al.*: **Genome-wide mapping of *in vivo* protein-DNA interactions.** *Science.* 2007; **316**(5830): 1497–502.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

10. Valouev A, Johnson DS, Sundquist A, *et al.*: **Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data.** *Nat Methods.* 2008; **5**(9): 829–34.
PubMed Abstract | Publisher Full Text | Free Full Text

11. Furey TS: **ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions.** *Nat Rev Genet.* 2012; **13**(12): 840–52.
PubMed Abstract | Publisher Full Text | Free Full Text

12. F Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, *et al.*: **Immunogenetics. Chromatin state dynamics during blood formation.** *Science.* 2014; **345**(6199): 943–9.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

13. Ernst J, Kellis M: **ChromHMM: automating chromatin-state discovery and characterization.** *Nat Methods.* 2012; **9**(3): 215–6.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Hoffman MM, Buske OJ, Wang J, *et al.*: **Unsupervised pattern discovery in human chromatin structure through genomic segmentation.** *Nat Methods.* 2012; **9**(5): 473–6.
PubMed Abstract | Publisher Full Text | Free Full Text

15. F Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet.* 2009; **10**(1): 57–63.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

16. Martin JA, Wang Z: **Next-generation transcriptome assembly.** *Nat Rev Genet.* 2011; **12**(10): 671–82.
PubMed Abstract | Publisher Full Text

17. F Mortazavi A, Williams BA, McCue K, *et al.*: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods.* 2008; **5**(7): 621–8.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

18. Takahashi H, Lassmann T, Murata M, *et al.*: **5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing.** *Nat Protoc.* 2012; **7**(3): 542–61.
PubMed Abstract | Publisher Full Text | Free Full Text

19. F Affymetrix ENCODE Transcriptome Project; Cold Spring Harbor Laboratory ENCODE Transcriptome Project: **Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs.** *Nature.* 2009; **457**(7232): 1028–32.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

20. F Tilgner H, Knowles DG, Johnson R, *et al.*: **Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs.** *Genome Res.* 2012; **22**(9): 1616–25.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

21. F Djebali S, Davis CA, Merkel A, *et al.*: **Landscape of transcription in human cells.** *Nature.* 2012; **489**(7414): 101–8.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

22. F Ingolia NT, Ghaemmaghami S, Newman JR, *et al.*: **Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling.** *Science.* 2009; **324**(5924): 218–23.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

23. F Wheeler EC, van Nostrand EL, Yeo GW: **Advances and challenges in the detection of transcriptome-wide protein-RNA interactions.** *Wiley Interdiscip Rev RNA.* 2018; **9**(1): e1436.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

24. Nachtergaele S, He C: **The emerging biology of RNA post-transcriptional modifications.** *RNA Biol.* 2017; **14**(2): 156–63.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Fu Y, Dominissini D, Rechavi G, *et al.*: **Gene expression regulation mediated through reversible m^{6}A RNA methylation.** *Nat Rev Genet.* 2014; **15**(5): 293–306.
PubMed Abstract | Publisher Full Text

26. F Garalde DR, Snell EA, Jachimowicz D, *et al.*: **Highly parallel direct RNA sequencing on an array of nanopores.** *Nat Methods.* 2018; **15**(3): 201–6.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

27. F Meyer KD, Saletore Y, Zumbo P, *et al.*: **Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons.** *Cell.* 2012; **149**(7): 1635–46.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

28. Hafner M, Landgraf P, Ludwig J, *et al.*: **Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing.** *Methods.* 2008; **44**(1): 3–12.
PubMed Abstract | Publisher Full Text | Free Full Text

29. F Pak J, Fire A: **Distinct populations of primary and secondary effectors during RNAi in *C. elegans*.** *Science.* 2007; **315**(5809): 241–4.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

30. Burroughs AM, Ando Y, de Hoon MJ, *et al.*: **A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness.** *Genome Res.* 2010; **20**(10): 1398–410.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Katoh T, Sakaguchi Y, Miyauchi K, *et al.*: **Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2.** *Genes Dev.* 2009; **23**(4): 433–8.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Boele J, Persson H, Shin JW, *et al.*: **PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease.** *Proc Natl Acad Sci U S A.* 2014; **111**(31): 11467–72.
PubMed Abstract | Publisher Full Text | Free Full Text

33. F Eagen KP: **Principles of Chromosome Architecture Revealed by Hi-C.** *Trends Biochem Sci.* 2018; **43**(6): 469–78.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

34. F Fullwood MJ, Liu MH, Pan YF, *et al.*: **An oestrogen-receptor-alpha-bound human chromatin interactome.** *Nature.* 2009; **462**(7269): 58–64.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

35. Ørom UA, Shiekhattar R: **Long non-coding RNAs and enhancers.** *Curr Opin Genet Dev.* 2011; **21**(2): 194–8.
PubMed Abstract | Publisher Full Text | Free Full Text

36. F Hon CC, Ramilowski JA, Harshbarger J, *et al.*: **An atlas of human long non-coding RNAs with accurate 5' ends.** *Nature.* 2017; **543**(7644): 199–204.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

37. F Li W, Notani D, Ma Q, *et al.*: **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.** *Nature.* 2013; **498**(7455): 516–20.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

38. F Chu C, Qu K, Zhong FL, *et al.*: **Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.** *Mol Cell.* 2011; **44**(4): 667–78.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

39. F Sridhar B, Rivas-Astroza M, Nguyen TC, *et al.*: **Systematic Mapping of RNA-Chromatin Interactions *In Vivo*.** *Curr Biol.* 2017; **27**(4): 602–9.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

40. F Li X, Zhou B, Chen L, *et al.*: **GRID-seq reveals the global RNA-chromatin interactome.** *Nat Biotechnol.* 2017; **35**(10): 940–50.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

41. F Bell JC, Jukam D, Teran NA, *et al.*: **Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts.** *eLife.* 2018; **7**: pii: e27024.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

42. Hsu PD, Lander ES, Zhang F: **Development and applications of CRISPR-Cas9 for genome engineering.** *Cell.* 2014; **157**(6): 1262–78.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Urnov FD, Rebar EJ, Holmes MC, *et al.*: **Genome editing with engineered zinc finger nucleases.** *Nat Rev Genet.* 2010; **11**(9): 636–46.
PubMed Abstract | Publisher Full Text

44. F Joung JK, Sander JD: **TALENs: a widely applicable technology for targeted genome editing.** *Nat Rev Mol Cell Biol.* 2013; **14**(1): 49–55.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

45. Aparicio-Prat E, Arnan C, Sala I, *et al.*: **DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs.** *BMC Genomics.* 2015; **16**: 846.
PubMed Abstract | Publisher Full Text | Free Full Text

46. F Zhu S, Li W, Liu J, *et al.*: **Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library.** *Nat Biotechnol.* 2016; **34**(12): 1279–86.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

47. F Qi LS, Larson MH, Gilbert LA, *et al.*: **Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.** *Cell.* 2013; **152**(5): 1173–83.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

48. F Gilbert LA, Larson MH, Morsut L, *et al.*: **CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.** *Cell.* 2013; **154**(2): 442–51.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

49. F Kearns NA, Pham H, Tabak B, *et al.*: **Functional annotation of native enhancers with a Cas9-histone demethylase fusion.** *Nat Methods.* 2015; **12**(5): 401–3.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

50. F Billon P, Bryant EE, Joseph SA, *et al.*: **CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons.** *Mol Cell.* 2017; **67**(6): 1068–1079.e4.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

51. F Liu SJ, Horlbeck MA, Cho SW, *et al.*: **CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells.** *Science.* 2017; **355**(6320): pii: aah7111.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

52. F Dixit A, Parnas O, Li B, *et al.*: **Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens.** *Cell.* 2016; **167**(7): 1853–1866.e17.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

53. F Adamson B, Norman TM, Jost M, *et al.*: **A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response.** *Cell.* 2016; **167**(7): 1867–1882.e21.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

54. F Lennox KA, Behlke MA: **Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides.** *Nucleic Acids Res.* 2016; **44**(2): 863–77.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

55. F Eid J, Fehr A, Gray J, *et al.*: **Real-time DNA sequencing from single polymerase molecules.** *Science.* 2009; **323**(5910): 133–8.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

56. F Clarke J, Wu HC, Jayasinghe L, *et al.*: **Continuous base identification for single-molecule nanopore DNA sequencing.** *Nat Nanotechnol.* 2009; **4**(4): 265–70.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

57. F Nowoshilow S, Schloissnig S, Fei JF, *et al.*: **The axolotl genome and the evolution of key tissue formation regulators.** *Nature.* 2018; **554**(7690): 50–5.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

58. F Jain M, Koren S, Miga KH, *et al.*: **Nanopore sequencing and assembly of a human genome with ultra-long reads.** *Nat Biotechnol.* 2018; **36**(4): 338–45.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

59. F Chin CS, Peluso P, Sedlazeck FJ, *et al.*: **Phased diploid genome assembly with single-molecule real-time sequencing.** *Nat Methods.* 2016; **13**(12): 1050–4.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

60. Tilgner H, Grubert F, Sharon D, *et al.*: **Defining a personal, allele-specific, and single-molecule long-read transcriptome.** *Proc Natl Acad Sci U S A.* 2014; **111**(27): 9869–74.
PubMed Abstract | Publisher Full Text | Free Full Text

61. Vilfan ID, Tsai YC, Clark TA, *et al.*: **Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription.** *J Nanobiotechnology.* 2013; **11**: 8.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Mercer TR, Gerhardt DJ, Dinger ME, *et al.*: **Targeted RNA sequencing reveals the deep complexity of the human transcriptome.** *Nat Biotechnol.* 2011; **30**(1): 99–104.
PubMed Abstract | Publisher Full Text | Free Full Text

63. F Lagarde J, Uszczynska-Ratajczak B, Carbonell S, *et al.*: **High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing.** *Nat Genet.* 2017; **49**(12): 1731–40.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

64. F Moor AE, Itzkovitz S: **Spatial transcriptomics: paving the way for tissue-level systems biology.** *Curr Opin Biotechnol.* 2017; **46**: 126–33.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

65. Crosetto N, Bienko M, van Oudenaarden A: **Spatially resolved transcriptomics and beyond.** *Nat Rev Genet.* 2015; **16**(1): 57–66.
PubMed Abstract | Publisher Full Text

66. Raj A, van den Bogaard P, Rifkin SA, *et al.*: **Imaging individual mRNA molecules using multiple singly labeled probes.** *Nat Methods.* 2008; **5**(10): 877–9.
PubMed Abstract | Publisher Full Text | Free Full Text

67. F Ståhl PL, Salmén F, Vickovic S, *et al.*: **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.** *Science.* 2016; **353**(6294): 78–82.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

68. F Tanay A, Regev A: **Scaling single-cell genomics from phenomenology to mechanism.** *Nature.* 2017; **541**(7637): 331–8.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

69. F Navin N, Kendall J, Troge J, *et al.*: **Tumour evolution inferred by single-cell sequencing.** *Nature.* 2011; **472**(7341): 90–4.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

70. Zong C, Lu S, Chapman AR, *et al.*: **Genome-wide detection of single-nucleotide and copy-number variations of a single human cell.** *Science.* 2012; **338**(6114): 1622–6.
PubMed Abstract | Publisher Full Text | Free Full Text

71. Xu X, Hou Y, Yin X, *et al.*: **Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor.** *Cell.* 2012; **148**(5): 886–95.
PubMed Abstract | Publisher Full Text

72. F Hou Y, Song L, Zhu P, *et al.*: **Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm.** *Cell.* 2012; **148**(5): 873–85.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

73. F Evrony GD, Cai X, Lee E, *et al.*: **Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain.** *Cell.* 2012; **151**(3): 483–96.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

74. F Rotem A, Ram O, Shoresh N, *et al.*: **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state.** *Nat Biotechnol.* 2015; **33**(11): 1165–72.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

75. F Papalexi E, Satija R: **Single-cell RNA sequencing to explore immune cell heterogeneity.** *Nat Rev Immunol.* 2018; **18**(1): 35–45.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

76. Islam S, Zeisel A, Joost S, *et al.*: **Quantitative single-cell RNA-seq with unique molecular identifiers.** *Nat Methods.* 2014; **11**(2): 163–6.
PubMed Abstract | Publisher Full Text

77. Ramsköld D, Luo S, Wang YC, *et al.*: **Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.** *Nat Biotechnol.* 2012; **30**(8): 777–82.
PubMed Abstract | Publisher Full Text | Free Full Text

78. Picelli S, Björklund ÅK, Faridani OR, *et al.*: **Smart-seq2 for sensitive full-length transcriptome profiling in single cells.** *Nat Methods.* 2013; **10**(11): 1096–8.
PubMed Abstract | Publisher Full Text

79. F Shalek AK, Satija R, Shuga J, *et al.*: **Single-cell RNA-seq reveals dynamic paracrine control of cellular variation.** *Nature.* 2014; **510**(7505): 363–9.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

80. Hashimshony T, Wagner F, Sher N, *et al.*: **CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification.** *Cell Rep.* 2012; **2**(3): 666–73.
PubMed Abstract | Publisher Full Text

81. F Hashimshony T, Senderovich N, Avital G, *et al.*: **CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq.** *Genome Biol.* 2016; **17**: 77.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

82. F Jaitin DA, Kenigsberg E, Keren-Shaul H, *et al.*: **Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types.** *Science.* 2014; **343**(6172): 776–9.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

83. F Sheng K, Cao W, Niu Y, *et al.*: **Effective detection of variation in single-cell transcriptomes using MATQ-seq.** *Nat Methods.* 2017; **14**(3): 267–70.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

84. Kouno T, Moody J, Kwon A, *et al.*: **C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution.** *bioRxiv.* 2018.
Publisher Full Text

85. F Macosko EZ, Basu A, Satija R, *et al.*: **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.** *Cell.* 2015; **161**(15): 1202–14.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

86. F Klein AM, Mazutis L, Akartuna I, *et al.*: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.** *Cell.* 2015; **161**(5): 1187–201.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

87. F Aran D, Hu Z, Butte AJ: **xCell: Digitally portraying the tissue cellular heterogeneity landscape.** *Genome Biol.* 2017; **18**(1): 220.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

88. Marques AC, Ponting CP: **Catalogues of mammalian long noncoding RNAs: Modest conservation and incompleteness.** *Genome Biol.* 2009; **10**(11): R124.
PubMed Abstract | Publisher Full Text | Free Full Text

89. Pervouchine DD, Djebali S, Breschi A, *et al.*: **Enhanced transcriptome maps from**

multiple mouse tissues reveal evolutionary constraint in gene expression. *Nat Commun.* 2015; **6**: 5903.
PubMed Abstract | Publisher Full Text | Free Full Text

90. Gerstein MB, Rozowsky J, Yan KK, *et al.*: **Comparative analysis of the transcriptome across distant species.** *Nature.* 2014; **512**(7515): 445–8.
PubMed Abstract | Publisher Full Text | Free Full Text

91. F Barbosa-Morais NL, Irimia M, Pan Q, *et al.*: **The evolutionary landscape of alternative splicing in vertebrate species.** *Science.* 2012; **338**(6114): 1587–93.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

92. Merkin J, Russell C, Chen P, *et al.*: **Evolutionary dynamics of gene and isoform regulation in mammalian tissues.** *Science.* 2012; **338**(6114): 1593–9.
PubMed Abstract | Publisher Full Text | Free Full Text

93. F Yue F, Cheng Y, Breschi A, *et al.*: **A comparative encyclopedia of DNA elements in the mouse genome.** *Nature.* 2014; **515**(7527): 355–64.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

94. Breschi A, Gingeras TR, Guigó R: **Comparative transcriptomics in human and mouse.** *Nat Rev Genet.* 2017; **18**(7): 425–40.
PubMed Abstract | Publisher Full Text

95. Pishesha N, Thiru P, Shi J, *et al.*: **Transcriptional divergence and conservation of human and mouse erythropoiesis.** *Proc Natl Acad Sci U S A.* 2014; **111**(11): 4103–8.
PubMed Abstract | Publisher Full Text | Free Full Text

96. Schroder K, Irvine KM, Taylor MS, *et al.*: **Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages.** *Proc Natl Acad Sci U S A.* 2012; **109**(16): E944–53.
PubMed Abstract | Publisher Full Text | Free Full Text

97. F Jubb AW, Young RS, Hume DA, *et al.*: **Enhancer Turnover Is Associated with a Divergent Transcriptional Response to Glucocorticoid in Mouse and Human Macrophages.** *J Immunol.* 2016; **196**(2): 813–22.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

98. Breschi A, Djebali S, Gillis J, *et al.*: **Gene-specific patterns of expression variation across organs and species.** *Genome Biol.* 2016; **17**(1): 151.
PubMed Abstract | Publisher Full Text | Free Full Text

99. F ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature.* 2012; **489**(7414): 57–74.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

100. F Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, *et al.*: **Integrative analysis of 111 reference human epigenomes.** *Nature.* 2015; **518**(7539): 317–30.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

101. Martens JH, Stunnenberg HG: **BLUEPRINT: mapping human blood cell epigenomes.** *Haematologica.* 2013; **98**(10): 1487–9.
PubMed Abstract | Publisher Full Text | Free Full Text

102. F FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, *et al.*: **A promoter-level mammalian expression atlas.** *Nature.* 2014; **507**(7493): 462–70.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

103. F GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group, *et al.*: **Genetic effects on gene expression across human tissues.** *Nature.* 2017; **550**(7675): 204–13.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

104. F Regev A, Teichmann SA, Lander ES, *et al.*: **The Human Cell Atlas.** *eLife.* 2017; **6**: pii: e27041.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

105. Lappalainen T, Sammeth M, Friedländer MR, *et al.*: **Transcriptome and genome sequencing uncovers functional variation in humans.** *Nature.* 2013; **501**(7468): 506–11.
PubMed Abstract | Publisher Full Text | Free Full Text

106. F Chen L, Ge B, Casale FP, *et al.*: **Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells.** *Cell.* 2016; **167**(5): 1398–1414.e24.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

107. eGTEx Project: **Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease.** *Nat Genet.* 2017; **49**(12): 1664–70.
PubMed Abstract | Publisher Full Text

108. F Gutman DA, Khalilia M, Lee S, *et al.*: **The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research.** *Cancer Res.* 2017; **77**(21): e75–e78.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

109. Medland SE, Jahanshad N, Neale BM, *et al.*: **Whole-genome analyses of whole-brain data: working within an expanded search space.** *Nat Neurosci.* 2014; **17**(6): 791–800.
PubMed Abstract | Publisher Full Text | Free Full Text

110. Hashimoto R, Ohi K, Yamamori H, *et al.*: **Imaging genetics and psychiatric disorders.** *Curr Mol Med.* 2015; **15**(2): 168–75.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Referee Status: ✔ ✔

Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

Version 1

1 **Kerstin Lindblad-Toh** Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
Competing Interests: No competing interests were disclosed.

2 **Feng Zhang** School of Public Health, Health Science Center, Xi'an Jiaotong University, Shaanxi, China
Competing Interests: No competing interests were disclosed.