

DNA sequencing methods in human genetics and disease research

Hans Lehrach

Addresses: Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195, Berlin, Germany; Dahlem Centre for Genome Research and Medical Systems Biology, Fabeckstrasse 60-62, 14195 Berlin, Germany; Alacris Theranostics GmbH, Fabeckstrasse. 60-62, 14195 Berlin, Germany

Email: Lehrach@molgen.mpg.de

F1000Prime Reports 2013, **5**:34 (doi:10.12703/P5-34)

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/legalcode>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You may not use this work for commercial purposes.

The electronic version of this article is the complete one and can be found at: <http://f1000.com/prime/reports/b/5/34>

Abstract

DNA sequencing has revolutionized biological and medical research, and is poised to have a similar impact in medicine. This tool is just one of a number of developments in our capability to identify, quantitate and functionally characterize the components of the biological networks keeping us healthy or making us sick, but in many respects it has played the leading role in this process. The new technologies do, however, also provide a bridge between genotype and phenotype, both in man and model (as well as all other) organisms, revolutionize the identification of elements involved in a multitude of human diseases or other phenotypes, and generate a wealth of medically relevant information on every single person, as the basis of a truly personalized medicine of the future.

The human genome

Starting from having little knowledge of any of the information in the human genome a few decades ago, the combination of cloning [1] and sequencing [2,3] gave us our first access to (initially very small) parts of the human/mouse genome [4]. Through the development of automated sequencing machines [5], this first phase of technology development culminated in the first sequence(s) of the human genome as the result of the human genome project [6,7] followed up by a number of single genomes, all but the first [8] sequenced on different next generation sequencing platforms [9-15].

The variation between the different individuals and their haplotypes was first addressed systematically in the HapMap project [16-19], resulting in the identification of 3.1 million human single nucleotide polymorphisms (SNPs) typed in 270 individuals from 4 major populations, still based on a combination of Sanger sequencing with chip-based genotyping approaches.

With the availability of next generation sequencing platforms [10,14,15,20-24] (summarized in [25]), much

larger scale analyses of genomes and genome variations became possible. The 1000genomes project [26,27] aims to provide information on rarer single nucleotide and structural variations in the human genome, by combining medium deep (typically defined as 4× coverage) genome and complete exome coverage of about 2500 individuals from 27 populations, combined with deep sequencing (>30× coverage) of a limited number of individuals/trios. In parallel, Grand Opportunity Exome sequencing project (GO-ESP), a project to sequence the exomes of 6700 individuals funded by the National Heart, Lung and Blood Institute (NHLBI) has focused specifically on the variations within the coding regions in specific patient groups with over 80 heart, lung and blood-related traits and other diseases of major importance [28].

In particular, the combination of genome, exome and transcriptome analysis is playing a key role in our understanding of mechanisms underlying cancer development, addressed particularly by the International Cancer Genome Consortium (ICGC, www.icgc.org) [29,30] and The Cancer Genome Atlas (TCGA, cancergenome.nih.gov), an analogous US-only project [31], generating comprehensive

catalogues of the somatic changes in different tumor entities by genome/exome sequencing (plus often additional information) of both tumor and germ line cells.

In contrast to these projects, which typically only collect very limited phenotypic information on the individuals analysed, the Personal Genome Project (PGP, www.personalgenomes.org) aims to sequence the genomes of up to 100,000 volunteers and to combine this information with a wide range of their phenotypic/disease history information [32].

The functional consequences of base pair changes or small deletions or insertions in the coding regions, which change the amino acid sequence of the protein, are typically easier to predict than for non-coding (e.g. regulatory) sequences. Different procedures [33-37] have therefore been used to enrich and sequence the exome (or other relevant regions of the genome), either alone, or in combination with more limited coverage of the entire genome. As an interesting variant [38], sequencing can also be targeted directly to specific sequences, by modifying the oligonucleotides on the sequencing chip.

Different genomes (and particularly cancer genomes) do not just differ in their genomic sequence, but are often also characterized by translocations, copy number variation and loss of heterozygosity. Specific translocations, for example, have been identified early as characteristic for specific types of tumors [39]. Next generation sequencing of the genome by so-called mate-pairs (sequencing the ends of large, circularised DNA fragments on a single fragment) has proven an effective technique to identify such rearrangements [40]. Similarly, translocations resulting in the fusion of transcripts observed, for example, in the case of fusion proteins, can be identified by RNAseq analysis [41,42].

The identification of larger scale copy number changes, first identified by comparative genome hybridization on chromosomes (CGH) [43,44], and then at higher resolution by array analysis (array-CGH) [45], has essentially been overtaken by sequence analyses [46-49], providing much better resolution, but also information on the copy number changes of the two haplotypes separately [50,51].

Given the same overall sequence composition, the biological function of the genome depends critically on the haplotypes contained within it, illustrated by the original definition of a gene in bacteria and phages as cistron on the basis of a cis-trans test (two mutations were considered to be in the same gene, if the phenotype

of the cell or phage was different, if the mutations were carried on the same DNA [in *cis*] or on two different fragments [in *trans*]). Although this definition is not possible in eucaryotes, due to alternative splicing, mixed diploid sequencing will usually still not be able to determine, if, for example, two heterozygous loss-of-function mutations are in *cis*, with two mutations in one copy of the gene, but leaving the other copy intact, or in *trans*, inactivating both copies of the gene. However, this information can be gained by statistical analyses [52] or experimentally by an increasing number of experimental strategy approaches, for example based on the sequencing of pools of fosmid clones, sorted chromosomes or longer DNA fragments [53-57].

Difficult materials

Beyond the different platforms used in these analyses (Illumina, Solid, 454, PacBio, Complete Genomics, Ion Torrent) with more still under development (e.g. Oxford Nanopore), there are a number of technical variations focusing on specific aspects of the sequencing process. An important burden in the analysis is, for example, the still relatively high error rate of many of these sequencing platforms, as well as, in some cases, the effect of partially damaged DNA. A major step to address this problem has been taken by Schmitt *et al.* who labeled every DNA fragment with a random tag during library construction. In the case of errors in the sequencing process or of damage in the original DNA fragment, the sequences of the two strands of the original DNA will differ, flagging these variants as being caused either by damage to the original DNA or errors in one of the amplification steps in the sequencing process [58]. Special technologies also had to be developed for the analysis of badly preserved DNA, e.g. due to the age of the material [59,60] or to the formaldehyde action in formalin fixed paraffin embedded (FFPE) material [61].

Another difficult challenge with current technologies lies in the sequencing of very small amounts of DNA, e.g. from free DNA in serum, or from individual cells (e.g. circulating tumor cells) [56,62,63], a problem that could be simplified in the future by techniques able to analyse un-modified, un-amplified DNA samples.

Functional analysis of the genome

Ultimately, the sequence of the DNA has to be understood in terms of its many functions. With the recent publications on the Encyclopedia of DNA Elements (ENCODE, <http://www.genome.gov/10005107>) project, a large amount of information has been created. This work illustrates many of the techniques available to functionally characterize the genome [64-72] etc.

The epigenome/methylome

Methylation and other modifications of DNA play a major role in affecting its transcription [73]. A number of different approaches have been used to detect such DNA modifications [74-82]. The “gold standard”, based on treating the DNA with bisulfite, a reagent converting cytosine into uracil, while leaving 5-methylcytosine unchanged, followed by whole-genome sequencing, gives by far the largest amount of information, albeit at correspondingly high cost. At lower cost (and generating correspondingly less information), restriction digests can be used to enrich particularly informative segments of the genome for bisulfite sequencing [77], or to focus on specific positions using either next generation sequencing [78] or chip-based sequencing techniques. The Infinium 450K Methylation array allows the determination of the bases remaining after bisulfite treatment of more than 480,000 cytosines in the genome, selected as particularly informative for methylation analyses [83]. Alternative procedures (e.g. MeDip) rely on the selective isolation of DNA fragments carrying the appropriate modified base by antibodies or proteins binding selectively to a particular type of modified DNA, followed by next generation sequencing, or sequencing of short fragments generated by restriction enzymes targeting regions with many GpCs, the dinucleotide sequence carrying most methylation in mammalian DNA. However, methyl C is not the only modified base in mammalian genomes: hydroxymethyl C, an alternatively modified base, cannot be distinguished from methyl C through bisulfite-based analysis methods [84] but can be selectively identified by antibodies recognizing this DNA modification [85]. New sequencing platforms promise to allow direct detection of modified bases, very significantly simplifying this type of analysis [86,87]. An additional ‘epigenetic code’ beyond the direct modification of the DNA is, however, provided by histone modification [88], which can be analysed by chromatin immunoprecipitation (ChIP)-seq techniques [89,90].

Regulatory sequences, genome structure

While it is relatively straightforward to identify protein-coding sequences, the identification of regulatory sequences or other sequence elements still represents a significant challenge. Interesting, new sequence-based approaches to identify such elements include, for example, the identification of sequences containing DNase-sensitive sites [91]; the mapping of open chromatin by crosslinking chromatin with formaldehyde; extracting the DNA cross-linked to protein by phenol extraction, and sequencing the remaining free DNA (FAIRE sequencing [92]); ChIP-seq with antibodies or other specific binding molecules (aptamers, affibodies, etc. [93-96]) against transcription factors [89]; the

identification of binding sites by fusing the protein of interest (in this case a transcription factor) to a Dam methylase gene and then sequencing the DNA protected from digestion by the methylation (DamID, [97]); and the generation of quantitative enhancer maps by cloning the genome in fragments downstream of a minimal promoter, causing the enhancers to be transcribed at a level proportional to their enhance strength (STARR-sequencing, [98]).

A novel variant in the use of next-generation sequencing equipment to identify genomic sequences binding to specific proteins, high throughput sequencing-fluorescent ligand profiling (HiTS-FLIP) [99], takes advantage of the optics and fluidics of an Illumina sequencer to score the binding of a fluorescence-labeled protein or other ligand, as an interesting alternative to previously described protocols [100].

A wide range of new, sequence-based techniques has also been developed to analyze the proximity of different elements of the genome to each other either directly (C3, C4, C5, HiC) or after selection for the presence of specific proteins (ChIP-loop, ChIA-PET), or their proximity to specific structural elements (e.g. nuclear membrane) by DamID or other protocols [101-104]. Similarly, next-generation sequencing has allowed a detailed analysis of the pattern of replication of the genome [105].

Transcriptome analysis

The analysis of transcripts by next-generation sequencing techniques (RNAseq) in its many different forms addresses a key step in the flow of information from the genome (and epigenome) to the phenotype of the organism. It provides information on many different types of transcripts (protein coding, short and long non-coding RNAs, micro RNAs), and has revolutionized the analysis of expression patterns, alternative splicing and allele-specific expression, providing unbiased digital information far beyond the results provided by different hybridization-based platforms [106,107].

In general, RNAseq has been carried out by first converting the RNA into cDNA, which is then sequenced as described above. Direct RNA sequencing has, however, also been described [108]. Another interesting variant is provided by on-flowcell reverse transcription sequencing (FRT-seq), the direct reverse transcription of mRNA on the flowcell, eliminating steps that could lead to specific artifacts [109].

A variety of protocols have been developed to handle more samples in parallel [110]. Earlier protocols did not preserve the information on the strandedness of the transcripts,

neglecting essential information [111,112]. This has been addressed in more recent protocols [113,114].

A wide range of protocols has been developed to focus on particular types of transcripts: coding, long and short non-coding and micro RNA etc, addressed by different protocols (poly-A plus, ribo-minus, etc.) [115]. An interesting variant to the ribo-minus strategy, relying on the removal of ribosomal RNA (and, in newer versions, also mitochondrial RNA) from the library involves “not so random” hexamer primers, selected not to contain sequences able to prime on ribosomal RNA, but to prime cDNA/double strand production [116].

To selectively analyse the 5′ end of RNAs from very small amounts of RNA, two new protocols, nanoSCAN and CAGEscan, have been described [117] as modifications of the original CAGE protocol [118,119]. Paired-end (PET) sequencing [120] can be used to advantage to compensate, to some extent, for the typically short read length of most current next-generation sequencing platforms. The analysis of alternative splicing patterns, however, still remains a difficult problem. Longer reads (PacBio, 454) could contribute significantly to the identification of the exact structure of transcripts for every gene expressed in a sample [121-126]. The mapping of branch points also adds relevant information [127]. Analysis of allele-specific expression patterns [128] and changes by RNA editing also provides essential information as to the function of a gene [129], as ultimately the genome acts through the RNA.

Many techniques have been developed to identify and quantitate microRNAs, to identify their targets, and to analyze their functions [130-134]. In addition, more and more long non-coding RNAs have become increasingly associated with many different regulatory processes [135].

There are obvious technical difficulties in cases where only small amounts of RNA, e.g. from single cells, have to be sequenced. Significant progress has been made, but there are obvious limitations due to the inherent technical and biological noise in this type of data [136-140].

Transcript abundance can change due to differences in synthesis or degradation rate. To be able to distinguish between these parameters, different procedures allowing the selection and subsequent sequencing of newly synthesized RNAs have been developed [141-145].

In complex tissues, alternative techniques are needed to selectively analyze transcripts from specific cell types [130,146,147]. As an alternative, *in-situ* transcriptome sequencing could combine spatial resolution with the

information content of transcriptome analysis. Direct *in situ* sequencing protocols, based, for example, on polony sequencing [148,149] have inherent limitations in the combined resolution versus sequencing depth, which can be overcome by the use of spatially encoded oligonucleotide primers (unpublished).

Proteins and protein interactions

While nucleic acids can be essentially detected and characterized down to the level of single molecules, this is typically not the case with proteins. In spite of major progress in mass spectroscopy (e.g. [150-153]) and other analysis techniques (e.g. [154-159]), we are still far from the power provided, for example, by next-generation sequencing in transcriptome analysis. A significant effort has, therefore, been directed at converting the analysis of proteins into a nucleic acid analysis problem. Sequencing the RNA protected by the ribosome does, for example, give detailed information on the protein synthesis, determines translation rates and identifies previously unknown proteins [160].

To be able to apply the sensitivity and throughput of nucleic acid analysis, a number of techniques have been developed to tag proteins, antibodies or other binding agents (or even chemicals) with nucleic acids, which can then be analyzed by deep sequencing. In proximity ligation for example, two or more binders are tagged with different oligonucleotides, which can form amplifiable sequences, if they are held in close proximity, allowing the highly sensitive detection of proteins (two binders to the same protein), protein modifications (one binder to the protein, one to the modification), protein-protein complexes (one binder each for both proteins), or larger structures (multiple binders carrying sequences), which will only give an amplifiable result if all components are present in close proximity [161,162]. Similarly, the results of different types of protein interaction assays (e.g. from a two-hybrid analysis) can be read out by selective amplification and sequence analysis [163-166].

Metagenome analysis, cell phenotypes, and much more

Microbial populations can play an important role in human diseases and other phenotypes. Next generation sequencing has allowed much more detailed analyses of these complex populations [167-169].

Next-generation sequencing techniques can also be used to great advantage to analyze the effect of specific conditions on a cell population marked by specific sequences with little effort. Next-generation sequencing can also help to identify causal variants for interesting phenotypes at the cell level, by either submitting populations of different

cells recognizable by a (introduced or naturally present) specific sequence, followed by analyzing the differences in sequence representation before and after the selection step, or by selection for specific phenotypes, followed by sequencing the genome to identify the causal variants [170-173].

As sequencing costs have dropped (and, in spite of the current reversal, they are likely to continue to drop over the longer term, driven by new technology platforms) and more applications are transformed into DNA sequencing (see [174] for a proposed approach to analyse neuron connectivity by DNA sequencing), we can expect to have increasingly different types of information available, not only for basic research but also for the individualized application of this knowledge for the benefit of individuals. This is likely to be the case, in spite of the currently high analysis costs often dwarfing the costs of data generation [175], as the existing, only partly automated, analysis pipelines mature, with data analysis ultimately limited by the costs of the electricity required for the computation.

Human phenotypes and diseases

At the beginning of positional cloning of Mendelian traits in man and other mammals, it took many years to identify some of the first human and mouse genes defined only by their phenotype [74,176]. Today, this could be carried out within weeks, after the family members have been collected or appropriate crosses have been carried out. Exome sequencing has already allowed the identification of causative mutations in a large number of analyses (e.g. [177-182]). Next-generation sequencing is, in particular, able to also identify causative mutations for diseases or phenotypes, for which no or too little family material is available (e.g. new mutations) [183-185].

The analysis of multifactorial traits by genotyping in general is limited to common alleles (the 'common disease-common allele' hypothesis) [186]. It has, however, become increasingly obvious that many phenotypes are caused by many rare alleles, or copy number variants, leaving exome or genome sequencing as the most obvious analysis route [187,188]. Interestingly, it has been shown that the combination of low coverage sequencing and imputation can be a cost-effective alternative to standard chip-based genotyping techniques [189]. Genome/exome sequencing, therefore, increasingly complements or replaces genotyping-based analyses, and is particularly important for providing clinically relevant information on the individual [190,191].

Next-generation sequencing has also proven itself as a relevant and powerful tool to detect disease-causing

mutations in the genome of the embryo in preimplantation diagnosis [192], or to analyse fetal nucleic acids in maternal plasma (e.g. for diagnosis of trisomy 21) [193,194].

Sequencing and other -omics techniques have proven particularly important for the analysis of tumors, since cancer, in a sense, is a "genomic" disease [29]. Analysis of tumors or tumor-derived cell lines [195] by deep genome/exome and transcriptome sequencing, combined with sequencing of the genome/exome of the patient, plays an increasingly important role in guiding the therapy choice [196-198], either through the identification of "actionable" variants, or, in future, increasingly through computer models [199,200] able to incorporate many different types of information to generate "virtual patient" models, on which the treatment of the individual patient can be optimized [201].

The future

We come from a situation a few decades ago when we hardly knew anything about our genome and its components. Development of cloning, sequencing and polymerase chain reaction (PCR) techniques has allowed us to identify specific genes and analyze their function, or, increasingly, identify the gene responsible for a specific (organismal) phenotype or disease. Completing the sequence of the human genome of 3 billion bases has revolutionized our understanding of biology and medicine, and has made many tasks, which were either impossible or very, very difficult, (relatively) easy. We now know the sequence of thousands of genomes, and are likely to know, sooner or later, the sequence of everybody's genome, complemented by a wide range of different analysis techniques (limited much more by the availability of samples, than by the complexity or cost of the analyses). We have moved from hybridisation-based array/chip analyses [202,203], generating the first "big data" types, more and more to sequence-based analyses, from gene sets, to exomes and whole genomes. Many of these analyses have become fairly straightforward; others, and in particular the analyses of indels/copy-number variations (CNVs), are still difficult, even combining a wide range of different analysis techniques [51,204].

In going from essentially no knowledge to knowing billions of bases of one human genome (as well as a lot of additional information on transcripts, proteins and metabolites, etc.) to many billions of bases of the genome and other -omics data for billions of humans puts us roughly half way (in a log scale) on a long road aiming to use abundant information (and computing power) to optimize treatment, prevention and well-being for everybody.

But what will we do with this abundance of information? A total of 32 (binary) biomarkers could give 7 billion different combinations, one for every person alive. 270 such biomarkers with an appropriate distribution, a very small number compared to the millions of differences between different human genomes, would be sufficient to identify every single atom in the observable universe. Computer models like those we use for weather forecasting give increasingly better predictions the more (and the more different types of) information they can be based on, which is typically not the case for many statistical procedures. Every patient is different. Every tumor (and in fact almost every cell of every tumor) could be considered a different “orphan” disease. We will therefore probably need all the information we can get to address this individuality, integrated by “virtual patient/virtual individual” models of every patient (including every functionally distinct subset of tumor cells in every tumor and, at least for prevention, every individual), which we had proposed in our FET-flagship project IT future of medicine (ITFoM, see www.itfom.eu) as the only reasonable way to integrate the huge medical datasets generated by the wide range of technologies available now, and likely to become available in the future.

Abbreviations

CGH, comparative genome hybridization; ChIP, chromatin immunoprecipitation; CNVs, copy-number variations; ENCODE, Encyclopedia of DNA elements; FFPE, formalin fixed paraffin embedded; FRT-seq, on-flowcell reverse transcriptase sequencing; GO-ESP, Grand Opportunity Exome sequencing project; HiTS-FLIP, high throughput sequencing-fluorescent ligand profiling; ICGC, International Cancer Genome Consortium; NHLBI, National Heart, Lung and Blood Institute; PCR, polymerase chain reaction; PGP, Personal Genome Project; SNPs, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas.

Disclosures




I am founder of and advisor to Alacris Theranostics, which aims to develop ‘virtual patient’ models for use in therapy choice and drug development.

Acknowledgements

I am very grateful to many colleagues at the Max Planck Institute for Molecular Genetics, the Dahlem Centre for Genome Research and Medical Systems Biology and Alacris Theranostics, who have contributed to this review, in particular Sylvia Krobitch and Wilfried Nietfeld, who have helped me fight with the many references, and Sylvia Krobitch, Marie-Laure Yaspo, Michal Schweiger, Marc Sultan, Ralf Sudbrak, Margret Hoehe, Albert Poustka,

Tatjana Borodina, Jörn Glöckler and Bodo Lange for discussions and feedback. I am also very grateful to many colleagues outside, especially Ulf Landegren, Stefan Beck and George Church, for free exchange of ideas and to two anonymous referees for helpful suggestions.

References

- Cohen SN, Chang AC, Boyer HW, Helling RB: **Construction of biologically functional bacterial plasmids in vitro.** *Proc Natl Acad Sci USA.* 1973, **70**:3240-4.
 - Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci USA.* 1977, **74**:5463-7.
 - Maxam AM, Gilbert W: **A new method for sequencing DNA.** *Proc Natl Acad Sci USA.* 1977, **74**:560-4.
 - Church GM, Gilbert W: **Genomic sequencing.** *Proc Natl Acad Sci USA.* 1984, **81**:1991-5.
 - Hunkapiller T, Kaiser RJ, Koop BF, Hood L: **Large-scale and automated DNA sequence determination.** *Science* 1991, **254**:59-67.
 - Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
 - Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al.: **The sequence of the human genome.** *Science* 2001, **291**:1304-51.
- 
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers Y, Frazier ME, Scherer SW, Strausberg RL, et al.: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, **5**:e254.
- 
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song X, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452**:872-6.
 - Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, et al.: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456**:53-9.
 - Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, et al.: **The diploid genome sequence of an Asian individual.** *Nature* 2008, **456**:60-5.
- 
- Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S,

Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, et al.: **DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome.** *Nature* 2008, **456**:66-72.



13. Ahn S, Kim T, Lee S, Kim D, Ghang H, Kim D, Kim B, Kim S, Kim W, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha J, Kim K, Lee B, Bhak J, Kim S: **The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group.** *Genome Res* 2009, **19**:1622-9.
14. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, et al.: **Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding.** *Genome Res* 2009, **19**:1527-41.
15. Pushkarev D, Neff NF, Quake SR: **Single-molecule sequencing of an individual human genome.** *Nat Biotechnol* 2009, **27**:847-50.
-
16. International HapMap Consortium: **Integrating ethics and science in the International HapMap Project.** *Nat Rev Genet* 2004, **5**:467-75.
17. International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437**:1299-320.
-
18. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, et al.: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449**:851-61.
19. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsepas C, Xie X, Byrne EH, McCarrroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, et al.: **Genome-wide detection and characterization of positive selection in human populations.** *Nature* 2007, **449**:913-8.
-
20. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, et al.: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-80.
-
21. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermant BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borchering AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, et al.: **Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays.** *Science* 2010, **327**:78-81.
-
22. Ozsolak F, Ting DT, Wittner BS, Brannigan BW, Paul S, Bardeesy N, Ramaswamy S, Milos PM, Haber DA: **Amplification-free digital gene expression profiling from minute cell quantities.** *Nat Methods* 2010, **7**:619-21.
23. Schadt EE, Turner S, Kasarskis A: **A window into third-generation sequencing.** *Hum Mol Genet* 2010, **19**:R227-40.
24. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, et al.: **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature* 2011, **475**:348-52.
-
25. Metzker ML: **Sequencing technologies - the next generation.** *Nat Rev Genet* 2010, **11**:31-46.
-
26. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061-73.
-
27. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56-65.
-
28. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM: **Evolution and functional impact of rare coding variation from deep sequencing of human exomes.** *Science* 2012, **337**:64-9.
-
29. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, et al.: **International network of cancer genome projects.** *Nature* 2010, **464**:993-8.
30. Jones DTW, Jäger N, Kool M, Zichner T, Hutter B, Sultan M, Cho Y, Pugh TJ, Hovestadt V, Stütz AM, Rausch T, Warnatz H, Ryzhova M, Bender S, Sturm D, Pleier S, Cin H, Pfaff E, Sieber L, Wittmann A, Remke M, Witt H, Hutter S, Tzaridis T, Weischenfeldt J, Raeder B, Avci M, Amstislavskiy V, Zapatka M, Weber UD, et al.: **Dissecting the genomic complexity underlying medulloblastoma.** *Nature* 2012, **488**:100-5.
-
31. Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-8.
-
32. Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, Angrist M, Bhak J, Bobe J, Callow MJ, Cano C, Chou MF, Chung WK, Douglas SM, Estep PW, Gore A, Hulick P, Labarga A, Lee J, Lunshof JE, Kim BC, Kim J, Li Z, Murray MF, Nilsen GB, Peters BA, Raman AM, Rienhoff HY, Robasky K, Wheeler MT, et al.: **A public resource facilitating clinical use of genomes.** *Proc Natl Acad Sci USA*. 2012, **109**:11920-7.
-
33. Bashardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M: **Direct genomic selection.** *Nat Methods* 2005, **2**:63-9.

34. Turner EH, Ng SB, Nickerson DA, Shendure J: **Methods for genomic partitioning.** *Annu Rev Genomics Hum Genet* 2009, **10**:263-84.
35. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ: **Target-enrichment strategies for next-generation sequencing.** *Nat Methods* 2010, **7**:111-8.
- F1000Prime RECOMMENDED**
36. Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M: **Performance comparison of exome DNA sequencing technologies.** *Nat Biotechnol* 2011, **29**:908-14.
- F1000Prime RECOMMENDED**
37. Mertes F, Elsharawy A, Sauer S, van Helvoort JMLM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ: **Targeted enrichment of genomic DNA regions for next-generation sequencing.** *Brief Funct Genomics* 2011, **10**:374-86.
38. Myllykangas S, Buenrostro JD, Natsoulis G, Bell JM, Ji HP: **Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing.** *Nat Biotechnol* 2011, **29**:1024-7.
- F1000Prime RECOMMENDED**
39. Koretzky GA: **The legacy of the Philadelphia chromosome.** *J Clin Invest* 2007, **117**:2030-2.
40. Korbelt JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M: **Paired-end mapping reveals extensive structural variation in the human genome.** *Science* 2007, **318**:420-6.
- F1000Prime RECOMMENDED**
41. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM: **Transcriptome sequencing to detect gene fusions in cancer.** *Nature* 2009, **458**:97-101.
- F1000Prime RECOMMENDED**
42. McPherson A, Hormozdiani F, Zayed A, Giuliani R, Ha G, Sun MGF, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, Pacheco M, Marra MA, Hirst M, Nielsen TO, Sahinalp SC, Huntsman D, Shah SP: **deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data.** *PLoS Comput. Biol.* 2011, **7**:e1001138.
- F1000Prime RECOMMENDED**
43. Du Manoir S, Speicher MR, Joos S, Schröck E, Popp S, Döhner H, Kovacs G, Robert-Nicoud M, Lichter P, Cremer T: **Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization.** *Hum Genet* 1993, **90**:590-610.
44. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: **Large-scale copy number polymorphism in the human genome.** *Science* 2004, **305**:525-8.
- F1000Prime RECOMMENDED**
45. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**:949-51.
- F1000Prime RECOMMENDED**
46. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacós M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, et al.: **Global variation in copy number in the human genome.** *Nature* 2006, **444**:444-54.
- F1000Prime RECOMMENDED**
47. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, MacDonald JR, Onyiah I, Pang AWC, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2010, **464**:704-12.
- F1000Prime RECOMMENDED**
48. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE: **Fine-scale structural variation of the human genome.** *Nat Genet* 2005, **37**:727-32.
- F1000Prime RECOMMENDED**
49. Khaja R, Zhang J, MacDonald JR, He Y, Joseph-George AM, Wei J, Rafiq MA, Qian C, Shago M, Pantano L, Aburatani H, Jones K, Redon R, Hurles M, Armengol L, Estivill X, Mural RJ, Lee C, Scherer SW, Feuk L: **Genome assembly comparison identifies structural variants in the human genome.** *Nat Genet* 2006, **38**:1413-8.
- F1000Prime RECOMMENDED**
50. Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES: **High-resolution mapping of copy-number alterations with massively parallel sequencing.** *Nat Methods* 2009, **6**:99-103.
- F1000Prime RECOMMENDED**
51. Baker M: **Structural variation: the genome's hidden architecture.** *Nat Methods* 2012, **9**:133-7.
52. Xu Y, Cheng W, Nie P, Zhou F: **WinHAP: an efficient haplotype phasing algorithm based on scalable sliding windows.** *PLoS ONE* 2012, **7**:e43163.
- F1000Prime RECOMMENDED**
53. Suk E, McEwen GK, Duitama J, Nowick K, Schulz S, Palczewski S, Schreiber S, Holloway DT, McLaughlin S, Peckham H, Lee C, Huebsch T, Hoehe MR: **A comprehensively molecular haplotype-resolved genome of a European individual.** *Genome Res* 2011, **21**:1672-85.
54. Yang H, Chen X, Wong WH: **Completely phased genome sequencing through chromosome sorting.** *Proc Natl Acad Sci USA* 2011, **108**:12-7.
- F1000Prime RECOMMENDED**
55. Duitama J, McEwen GK, Huebsch T, Palczewski S, Schulz S, Verstrepen K, Suk E, Hoehe MR: **Fosmid-based whole genome haplotyping of a HapMap trio child: evaluation of Single Individual Haplotyping techniques.** *Nucleic Acids Res* 2012, **40**:2041-53.
- F1000Prime RECOMMENDED**
56. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, Jiang Y, Dahl F, Tang YT, Haas J, Robasky K, Zaranek AW, Lee J, Ball MP, Peterson JE, Perazich H, Yeung G, Liu J, Chen L, Kennemer MI, Pothuraju K, Konvicka K, Tsoumpko-Sitnikov M, Pant KP, Ebert JC, Nilsen GB, Baccash J, Halpern AL, Church GM, Drmanac R: **Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells.** *Nature* 2012, **487**:190-5.
57. Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, Bibikova M, Chuang H, Kruglyak S, Ronaghi M, Eberle MA, Fan J: **Whole-genome**

haplotyping by dilution, amplification, and sequencing. *Proc Natl Acad Sci USA.* 2013, **110**:5552-7.



58. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA: **Detection of ultra-rare mutations by next-generation sequencing.** *Proc Natl Acad Sci USA.* 2012, **109**:14508-13.



59. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspina A, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, et al.: **A draft sequence of the Neandertal genome.** *Science* 2010, **328**:710-22.



60. Meyer M, Kircher M, Gansauge M, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, Filippo C de, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Dereviako AP, Patterson N, Andrés AM, Eichler EE, et al.: **A high-coverage genome sequence from an archaic Denisovan individual.** *Science* 2012, **338**:222-6.



61. Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, Parkhomchuk D, Zatloukal K, Lehrach H: **Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis.** *PLoS ONE* 2009, **4**:e5548.

62. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, Li F, Wu K, Liang J, Di Shao, Wu H, Ye X, Ye C, Wu R, Jian M, Chen Y, Xie W, Zhang R, Chen L, Liu X, Yao X, Zheng H, Yu C, Li Q, Gong Z, Mao M, Yang X, Yang L, Li J, Wang W, et al.: **Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm.** *Cell* 2012, **148**:873-85.



63. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM: **Methylome analysis using MeDIP-seq with low DNA concentrations.** *Nat Protoc* 2012, **7**:617-36.



64. ENCODE Project Consortium: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.

65. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Fritzsche S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altschuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, et al.: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57-74.



66. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, et al.: **Landscape of transcription in human cells.** *Nature* 2012, **489**:101-8.



67. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A,

Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fritzsche S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, et al.: **Architecture of the human regulatory network derived from ENCODE data.** *Nature* 2012, **489**:91-100.



68. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutayin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, et al.: **An expansive human regulatory lexicon encoded in transcription factor footprints.** *Nature* 2012, **489**:83-90.



69. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters.** *Nature* 2012, **489**:109-13.



70. Spivakov M, Akhtar J, Kheradpour P, Beal K, Girardot C, Koscielny G, Herrero J, Kellis M, Furlong EEM, Birney E: **Analysis of variation at transcription factor binding sites in Drosophila and humans.** *Genome Biol* 2012, **13**:R49.

71. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayin T, Lajoie B, Lee B, Lee K, London D, Lotakis D, Neph S, et al.: **The accessible chromatin landscape of the human genome.** *Nature* 2012, **489**:75-82.

72. Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M, Gerstein M: **Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors.** *Genome Biol* 2012, **13**:R48.



73. Bernstein BE, Meissner A, Lander ES: **The mammalian epigenome.** *Cell* 2007, **128**:669-81.

74. **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group.** *Cell* 1993, **72**:971-83.

75. Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG, Meissner A: **Quantitative comparison of genome-wide DNA methylation mapping technologies.** *Nat Biotechnol* 2010, **28**:1106-14.

76. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung W, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, et al.: **Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications.** *Nat Biotechnol* 2010, **28**:1097-105.

77. Hirst M, Marra MA: **Next generation sequencing based approaches to epigenomics.** *Brief Funct Genomics* 2010, **9**:455-65.



78. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A: **Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling.** *Nat Protoc* 2011, **6**:468-81.



79. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ: **A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the**

- Infinium HumanMethylation450 BeadChip® for methylome profiling.** *PLoS ONE* 2012, **7**:e50233.
80. Diep D, Plongthongkum N, Gore A, Fung H, Shoemaker R, Zhang K: **Library-free methylation sequencing with bisulfite padlock probes.** *Nat Methods* 2012, **9**:270-2.
81. Krueger F, Kreck B, Franke A, Andrews SR: **DNA methylome analysis using short bisulfite sequencing data.** *Nat Methods* 2012, **9**:145-51.
82. Pastor WA, Huang Y, Henderson HR, Agarwal S, Rao A: **The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine.** *Nat Protoc* 2012, **7**:1909-17.
83. Dedeurwaerder S, DeFrance M, Calonne E, Denis H, Sotiriou C, Fuks F: **Evaluation of the Infinium Methylation 450K technology.** *Epigenomics* 2011, **3**:771-84.
- F1000Prime RECOMMENDED**
84. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A: **The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing.** *PLoS ONE* 2010, **5**:e8888.
85. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324**:929-30.
- F1000Prime RECOMMENDED**
86. Clarke J, Wu H, Jayasinghe L, Patel A, Reid S, Bayley H: **Continuous base identification for single-molecule nanopore DNA sequencing.** *Nat Nanotechnol* 2009, **4**:265-70.
- F1000Prime RECOMMENDED**
87. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SV: **Direct detection of DNA methylation during single-molecule, real-time sequencing.** *Nat Methods* 2010, **7**:461-5.
- F1000Prime RECOMMENDED**
88. Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128**:693-705.
89. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of in vivo protein-DNA interactions.** *Science* 2007, **316**:1497-502.
- F1000Prime RECOMMENDED**
90. Adli M, Zhu J, Bernstein BE: **Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors.** *Nat Methods* 2010, **7**:615-8.
- F1000Prime RECOMMENDED**
91. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome.** *Cell* 2008, **132**:311-22.
- F1000Prime RECOMMENDED**
92. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J: **A map of open chromatin in human pancreatic islets.** *Nat Genet* 2010, **42**:255-9.
- F1000Prime RECOMMENDED**
93. Taussig MJ, Stoevesandt O, Borrebaeck CAK, Bradbury AR, Cahill D, Cambillau C, Daruvar A de, Dübel S, Eichler J, Frank R, Gibson TJ, Gloriam D, Gold L, Herberg FW, Hermjakob H, Hoheisel JD, Joos TO, Kallioniemi O, Koegl M, Koegl M, Konthur Z, Korn B, Kremmer E, Krobitch S, Landegren U, van der Maarel S, McCafferty J, Muyldermans S, Nygren P, Palcy S, et al.: **ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome.** *Nat Methods* 2007, **4**:13-7.
94. Song K, Lee S, Ban C: **Aptamers and their biological applications.** *Sensors (Basel)* 2012, **12**:612-31.
95. Dreier B, Plückthun A: **Rapid selection of high-affinity binders using ribosome display.** *Methods Mol Biol* 2012, **805**:261-86.
96. Agnew HD, Rohde RD, Millward SW, Nag A, Yeo W, Hein JE, Pitram SM, Tariq AA, Burns VM, Krom RJ, Fokin VV, Sharpless KB, Heath JR: **Iterative in situ click chemistry creates antibody-like protein-capture agents.** *Angew Chem Int Ed Engl* 2009, **48**:4944-8.
- F1000Prime RECOMMENDED**
97. van Steensel B, Henikoff S: **Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase.** *Nat Biotechnol* 2000, **18**:424-8.
- F1000Prime RECOMMENDED**
98. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science* 2013, **339**:1074-7.
- F1000Prime RECOMMENDED**
99. Nutiu R, Friedman RC, Luo S, Khrebtkova I, Silva D, Li R, Zhang L, Schroth GP, Burge CB: **Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument.** *Nat Biotechnol* 2011, **29**:659-64.
- F1000Prime RECOMMENDED**
100. Stormo GD, Zhao Y: **Determining the specificity of protein-DNA interactions.** *Nat Rev Genet* 2010, **11**:751-60.
101. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**:289-93.
- F1000Prime RECOMMENDED**
102. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G: **Three-dimensional folding and functional organization principles of the Drosophila genome.** *Cell* 2012, **148**:458-72.
- F1000Prime RECOMMENDED**
103. Tanizawa H, Noma K: **Unravelling global genome organization by 3C-seq.** *Semin Cell Dev Biol* 2012, **23**:213-21.
104. van Steensel B, Dekker J: **Genomics tools for unraveling chromosome architecture.** *Nat Biotechnol* 2010, **28**:1089-95.
105. McGuffee SR, Smith DJ, Whitehouse I: **Quantitative, genome-wide analysis of eukaryotic replication initiation and termination.** *Mol Cell* 2013, **50**:123-35.
- F1000Prime RECOMMENDED**
106. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo M: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science* 2008, **321**:956-60.
- F1000Prime RECOMMENDED**
107. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
- F1000Prime RECOMMENDED**

108. Ozsolak F, Platt AR, Jones DR, Reifenger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM: **Direct RNA sequencing.** *Nature* 2009, **461**:814-8.
- F1000Prime RECOMMENDED**
109. Mamanova L, Turner DJ: **Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq).** *Nat Protoc* 2011, **6**:1736-47.
- F1000Prime RECOMMENDED**
110. Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR: **A High-Throughput Method for Illumina RNA-Seq Library Preparation.** *Front Plant Sci* 2012, **3**:202.
111. Valencia A, Hidalgo M: **Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics.** *Genome Med* 2012, **4**:61.
112. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: **The antisense transcriptomes of human cells.** *Science* 2008, **322**:1855-7.
113. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, Lehrach H, Soldatov A: **Transcriptome analysis by strand-specific sequencing of complementary DNA.** *Nucleic Acids Res* 2009, **37**:e123.
114. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A: **Comprehensive comparative analysis of strand-specific RNA sequencing methods.** *Nat Methods* 2010, **7**:709-15.
- F1000Prime RECOMMENDED**
115. Saxena A, Carninci P: **Whole transcriptome analysis: what are we still missing?** *Wiley Interdiscip Rev Syst Biol Med* 2011, **3**:527-43.
116. Armour CD, Castle JC, Chen R, Babak T, Loerch P, Jackson S, Shah JK, Dey J, Rohl CA, Johnson JM, Raymond CK: **Digital transcriptome profiling using selective hexamer priming for cDNA synthesis.** *Nat Methods* 2009, **6**:647-9.
- F1000Prime RECOMMENDED**
117. Plessy C, Bertin N, Takahashi H, Simone R, Salimullah M, Lassmann T, Vitezic M, Severin J, Olivarius S, Lazarevic D, Hornig N, Orlando V, Bell I, Gao H, Dumais J, Kapranov P, Wang H, Davis CA, Gingeras TR, Kawai J, Daub CO, Hayashizaki Y, Gustinich S, Carninci P: **Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan.** *Nat Methods* 2010, **7**:528-34.
- F1000Prime RECOMMENDED**
118. Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M, Hayashizaki Y, Carninci P: **CAGE: cap analysis of gene expression.** *Nat Methods* 2006, **3**:211-22.
119. Takahashi H, Lassmann T, Murata M, Carninci P: **5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing.** *Nat Protoc* 2012, **7**:542-61.
120. Fullwood MJ, Wei C, Liu ET, Ruan Y: **Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses.** *Genome Res* 2009, **19**:521-32.
121. Blencowe BJ: **Alternative splicing: new insights from global analyses.** *Cell* 2006, **126**:37-47.
122. Skotheim RI, Nees M: **Alternative splicing in cancer: noise, functional, or systematic?** *Int J Biochem Cell Biol* 2007, **39**:1432-49.
123. Blehman R, Marioni JC, Zumbo P, Stephens M, Gilad Y: **Sex-specific and lineage-specific alternative splicing in primates.** *Genome Res* 2010, **20**:180-9.
- F1000Prime RECOMMENDED**
124. Auboeuf D, Carmo-Fonseca M, Valcarcel J, Biamonti G: **Alternative splicing and cancer.** *J Nucleic Acids* 2012, **2012**:363809.
125. Germann S, Gratadou L, Dutertre M, Auboeuf D: **Splicing programs and cancer.** *J Nucleic Acids* 2012, **2012**:269570.
126. Wang ET, Cody NAL, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, Lécuyer E, Burge CB: **Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins.** *Cell* 2012, **150**:710-24.
- F1000Prime RECOMMENDED**
127. Taggart AJ, DeSimone AM, Shih JS, Filloux ME, Fairbrother WG: **Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo.** *Nat Struct Mol Biol* 2012, **19**:719-21.
- F1000Prime RECOMMENDED**
128. DeVeale B, van der Kooy D, Babak T: **Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective.** *PLoS Genet* 2012, **8**:e1002600.
- F1000Prime RECOMMENDED**
129. Li JB, Levanon EY, Yoon J, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM: **Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing.** *Science* 2009, **324**:1210-3.
- F1000Prime RECOMMENDED**
130. Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ, Greengard P, Heintz N: **A translational profiling approach for the molecular characterization of CNS cell types.** *Cell* 2008, **135**:738-48.
- F1000Prime RECOMMENDED**
131. Chi SW, Zang JB, Mele A, Darnell RB: **Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.** *Nature* 2009, **460**:479-86.
- F1000Prime RECOMMENDED**
132. Budhu A, Ji J, Wang XW: **The clinical potential of microRNAs.** *J Hematol Oncol* 2010, **3**:37.
133. Su W, Kleinhanz RR, Schadt EE: **Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques.** *Mol Syst Biol* 2011, **7**:490.
- F1000Prime RECOMMENDED**
134. Röhr C, Kerick M, Fischer A, Kühn A, Kashofer K, Timmermann B, Daskalaki A, Meinel T, Drichel D, Börno ST, et al.: **High-throughput miRNA and mRNA sequencing of paired colorectal normal, tumor and metastasis tissues and bioinformatics modeling of miRNA-I therapeutic applications.** *PLoS ONE* 2013, in press
Rinn JL, Chang HY: **Genome regulation by long noncoding RNAs.** *Annu Rev Biochem* 2012, **81**:145-66.
135. Rinn JL, Chang HY: **Genome regulation by long noncoding RNAs.** *Annu Rev Biochem* 2012, **81**:145-66.
136. Tang Y, Luo X, Cui H, Ni X, Yuan M, Guo Y, Huang X, Zhou H, Vries N de, Tak PP, Chen S, Shen N: **MicroRNA-146A contributes to abnormal activation of the type I interferon pathway in**

human lupus by targeting the key signaling proteins. *Arthritis Rheum* 2009, **60**:1065-75.



137. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J, Lönnerberg P, Linnarsson S: **Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq.** *Genome Res* 2011, **21**:1160-7.



138. Goetz JJ, Trimarchi JM: **Transcriptome sequencing of single cells with Smart-Seq.** *Nat Biotechnol* 2012, **30**:763-5.

139. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen S, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RV, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, et al.: **The genetic basis of early T-cell precursor acute lymphoblastic leukaemia.** *Nature* 2012, **481**:157-63.



140. Plessy C, Desbois L, Fujii T, Carninci P: **Population transcriptomics with single-cell resolution: a new field made possible by microfluidics: a technology for high throughput transcript counting and data-driven definition of cell types.** *Bioessays* 2013, **35**:131-40.

141. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC: **Biosynthetic labeling of RNA with uracil phosphoribosyl-transferase allows cell-specific microarray analysis of mRNA synthesis and decay.** *Nat Biotechnol* 2005, **23**:232-7.



142. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322**:1845-8.



143. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest ARR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS: **Tiny RNAs associated with transcription start sites in animals.** *Nat Genet* 2009, **41**:572-8.



144. Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, Malterer G, L'Hernault A, Schilhabel M, Schreiber S, Rosenstiel P, Zimmer R, Eick D, Friedel CC, Dölken L: **Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution.** *Genome Res* 2012, **22**:2031-42.



145. Duan J, Shi J, Ge X, Dölken L, Moy W, He D, Shi S, Sanders AR, Ross J, Gejman PV: **Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines.** *Sci Rep* 2013, **3**:1318.

146. Deal RB, Henikoff S: **A simple method for gene expression and chromatin profiling of individual cell types within a tissue.** *Dev Cell* 2010, **18**:1030-40.



147. Gay L, Miller MR, Ventura PB, Devasthali V, Vue Z, Thompson HL, Temple S, Zong H, Cleary MD, Stankunas K, Doe CQ: **Mouse TU**

tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. *Genes Dev* 2013, **27**:98-115.



148. Mitra RD, Shendure J, Olejnik J, Edyta-Krzyszanska-Olejnik, Church GM: **Fluorescent in situ sequencing on polymerase colonies.** *Anal Biochem* 2003, **320**:55-65.

149. Larsson C, Grundberg I, Söderberg O, Nilsson M: **In situ detection and genotyping of individual mRNA molecules.** *Nat Methods* 2010, **7**:395-7.

150. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M: **Deep proteome and transcriptome mapping of a human cancer cell line.** *Mol Syst Biol* 2011, **7**:548.



151. Stingle S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z: **Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells.** *Mol Syst Biol* 2012, **8**:608.



152. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.** *Mol Cell Proteomics* 2012, **11**:O111.016717.

153. Zanivan S, Meves A, Behrendt K, Schoof EM, Neilson LJ, Cox J, Tang HR, Kalna G, van Ree JH, van Deursen JM, Trempp CS, Machesky LM, Linding R, Wickström SA, Fässler R, Mann M: **In vivo SILAC-based proteomics reveals phosphoproteome changes during mouse skin carcinogenesis.** *Cell Rep* 2013, **3**:552-66.



154. Walter G, Büssov K, Cahill D, Lueking A, Lehrach H: **Protein arrays for gene expression and molecular interaction screening.** *Curr Opin Microbiol* 2000, **3**:298-302.

155. Charboneau L, Tory H, Scott H, Chen T, Winters M, Petricoin EF, Liotta LA, Paweletz CP: **Utility of reverse phase protein arrays: applications to signalling pathways and human body arrays.** *Brief Funct Genomic Proteomic* 2002, **1**:305-15.

156. Nam J, Thaxton CS, Mirkin CA: **Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins.** *Science* 2003, **301**:1884-6.



157. Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szgyarto CA, Persson A, Ottosson J, Wernérus H, Nilsson P, Lundberg E, Sivertsson A, Navani S, Wester K, Kampf C, Hober S, Pontén F, Uhlen M: **A genecentric Human Protein Atlas for expression profiles based on antibodies.** *Mol Cell Proteomics* 2008, **7**:2019-27.

158. Kurien BT, Scofield RH: **A brief review of other notable protein detection methods on acrylamide gels.** *Methods Mol Biol* 2012, **869**:617-20.

159. Zheng G, Lieber CM: **Nanowire biosensors for label-free, real-time, ultrasensitive protein detection.** *Methods Mol Biol* 2011, **790**:223-37.

160. Olek M: **Determination of diphenylamine residues in apples, and 4-aminobiphenyl residues in diphenylamine, by high-performance liquid chromatography and electrochemical detection.** *J Chromatogr* 1988, **447**:421-5.

161. Weibrecht I, Leuchowius K, Clausson C, Conze T, Jarvius M, Howell WM, Kamali-Moghaddam M, Söderberg O: **Proximity ligation assays: a recent addition to the proteomics toolbox.** *Expert Rev Proteomics* 2010, **7**:401-9.

162. Hammond M, Nong RY, Ericsson O, Pardali K, Landegren U: **Profiling cellular protein complexes by proximity ligation with dual tag microarray readout.** *PLoS ONE* 2012, **7**:e40405.
- F1000Prime RECOMMENDED**
163. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrzikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M: **Next-generation sequencing to generate interactome datasets.** *Nat Methods* 2011, **8**:478-80.
- F1000Prime RECOMMENDED**
164. Fujimori S, Hirai N, Ohashi H, Masuoka K, Nishikimi A, Fukui Y, Washio T, Oshikubo T, Yamashita T, Miyamoto-Sato E: **Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data.** *Sci Rep* 2012, **2**:691.
165. Weimann M, Grossmann A, Woodsmith J, Özkan Z, Birth P, Meierhofer D, Benlasfer N, Valovka T, Timmermann B, Wanker EE, Sauer S, Stelzl U: **A Y2H-seq approach defines the human protein methyltransferase interactome.** *Nat Methods* 2013, **10**:339-42.
- F1000Prime RECOMMENDED**
166. Zhu J, Larman HB, Gao G, Somwar R, Zhang Z, Laserson U, Ciccio A, Pavlova N, Church G, Zhang W, Kesari S, Elledge SJ: **Protein interaction discovery using parallel analysis of translated ORFs (PLATO).** *Nat Biotechnol* 2013, **31**:331-4.
- F1000Prime RECOMMENDED**
167. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, et al.: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**:59-65.
- F1000Prime RECOMMENDED**
168. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P: **Genomic variation landscape of the human gut microbiome.** *Nature* 2013, **493**:45-50.
- F1000Prime RECOMMENDED**
169. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI: **Human gut microbiome viewed across age and geography.** *Nature* 2012, **486**:222-7.
- F1000Prime RECOMMENDED**
170. Bochner BR: **New technologies to assess genotype-phenotype relationships.** *Nat Rev Genet* 2003, **4**:309-14.
171. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT: **Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides.** *Nat Biotechnol* 2010, **28**:856-62.
- F1000Prime RECOMMENDED**
172. Carette JE, Guimaraes CP, Wuethrich I, Blomen VA, Varadarajan M, Sun C, Bell G, Yuan B, Muellner MK, Nijman SM, Ploegh HL, Brummelkamp TR: **Global gene disruption in human cells to assign genes to phenotypes by deep sequencing.** *Nat Biotechnol* 2011, **29**:542-6.
- F1000Prime RECOMMENDED**
173. Ralser M, Kuhl H, Ralser M, Werber M, Lehrach H, Breitenbach M, Timmermann B: **The *Saccharomyces cerevisiae* W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt.** *Open Biol* 2012, **2**:120093.
174. Zador AM, Dubnau J, Oyibo HK, Zhan H, Cao G, Peikon ID: **Sequencing the connectome.** *PLoS Biol.* 2012, **10**:e1001411.
- F1000Prime RECOMMENDED**
175. Mardis ER: **The \$1,000 genome, the \$100,000 analysis?** *Genome Med* 2010, **2**:84.
- F1000Prime RECOMMENDED**
176. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB: **Clinical application of exome sequencing in undiagnosed genetic conditions.** *J Med Genet* 2012, **49**:353-61.
- F1000Prime RECOMMENDED**
177. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature* 2009, **461**:272-6.
- F1000Prime RECOMMENDED**
178. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP: **Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.** *Proc Natl Acad Sci USA.* 2009, **106**:19096-101.
- F1000Prime RECOMMENDED**
179. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: **Exome sequencing identifies the cause of a mendelian disorder.** *Nat Genet* 2010, **42**:30-5.
- F1000Prime RECOMMENDED**
180. Ku C, Naidoo N, Pawitan Y: **Revisiting Mendelian disorders through exome sequencing.** *Hum Genet* 2011, **129**:351-70.
181. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: **Exome sequencing as a tool for Mendelian disease gene discovery.** *Nat Rev Genet* 2011, **12**:745-55.
- F1000Prime RECOMMENDED**
182. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I: **Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders.** *J Hum Genet* 2012, **57**:621-32.
183. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widada S, et al.: **A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation.** *Nat Genet* 2009, **41**:535-43.
- F1000Prime RECOMMENDED**
184. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M: **Exome sequencing supports a de novo**

mutational paradigm for schizophrenia. *Nat Genet* 2011, **43**:864-8.



185. Calvo SE, Compton AG, Hershman SG, Lim SC, Lieber DS, Tucker EJ, Laskowski A, Garone C, Liu S, Jaffe DB, Christodoulou J, Fletcher JM, Bruno DL, Goldblatt J, Dimauro S, Thorburn DR, Mootha VK: **Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing.** *Sci Transl Med* 2012, **4**:118ra10.



186. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery.** *Am J Hum Genet* 2012, **90**:7-24.



187. Vissers LELM, Ligt J de, Gilissen C, Janssen I, Steehouwer M, Vries P de, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BWM, Hoischen A, Vries BBA de, Brunner HG, Veltman JA: **A de novo paradigm for mental retardation.** *Nat Genet* 2010, **42**:1109-12.



188. Bodmer W, Bonilla C: **Common and rare variants in multifactorial susceptibility to common diseases.** *Nat Genet* 2008, **40**:695-701.

189. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale BM, Daly MJ, Sklar P, Sullivan PF, Bergen S, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Purcell SM, Haas DW, Liang L, Sunyaev S, Patterson N, Bakker PIW de, Reich D, Price AL: **Extremely low-coverage sequencing and imputation increases power for genome-wide association studies.** *Nat Genet* 2012, **44**:631-5.



190. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, et al.: **Clinical assessment incorporating a personal genome.** *Lancet* 2010, **375**:1525-35.



191. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, Cheng Y, Clark MJ, Im H, Habegger L, Balasubramanian S, O'Huallachain M, Dudley JT, Hillenmeyer S, Haraksingh R, Sharon D, Euskirchen G, Lacroute P, Bettinger K, Boyle AP, Kasowski M, Grubert F, Seki S, Garcia M, Whirl-Carrillo M, Gallardo M, et al.: **Personal omics profiling reveals dynamic molecular and medical phenotypes.** *Cell* 2012, **148**:1293-307.



192. Treff NR, Fedick A, Tao X, Devkota B, Taylor D, Scott RT: **Evaluation of targeted next-generation sequencing-based**

preimplantation genetic diagnosis of monogenic disease. *Fertil Steril* 2013, **99**:1377-1384.e6.



193. Lun FMF, Tsui NBY, Chan KCA, Leung TY, Lau TK, Charoenkwan P, Chow KCK, Lo WYW, Wanapirak C, Sanguanserm Sri T, Cantor CR, Chiu RWK, Lo YMD: **Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma.** *Proc Natl Acad Sci USA*. 2008, **105**:19920-5.

194. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, Gammill HS, Rubens CE, Santillan DA, Murray JC, Tabor HK, Bamshad MJ, Eichler EE, Shendure J: **Noninvasive whole-genome sequencing of a human fetus.** *Sci Transl Med* 2012, **4**:137ra76.

195. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, Silva M de, et al.: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature* 2012, **483**:603-7.



196. Feero WG, Guttmacher AE, Collins FS: **Genomic medicine—an updated primer.** *N Engl J Med* 2010, **362**:2001-11.

197. Stratton MR: **Exploring the genomes of cancer cells: progress and promise.** *Science* 2011, **331**:1553-8.



198. Mardis ER: **Genome sequencing and cancer.** *Curr Opin Genet Dev* 2012, **22**:245-50.

199. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW: **A whole-cell computational model predicts phenotype from genotype.** *Cell* 2012, **150**:389-401.



200. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, et al.: **A community-driven global reconstruction of human metabolism.** *Nat Biotechnol* 2013, **31**:419-25.

201. Wierling C, Kühn A, Hache H, Daskalaki A, Maschke-Dutz E, Peycheva S, Li J, Herwig R, Lehrach H: **Prediction in the face of uncertainty: a Monte Carlo-based approach for systems biology of cancer treatment.** *Mutat Res* 2012, **746**:163-70.

202. Lennon GG, Lehrach H: **Hybridization analyses of arrayed cDNA libraries.** *Trends Genet* 1991, **7**:314-7.

203. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-70.

204. Pang AWW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, Park H, Hurler ME, Lee C, Venter JC, Kirkness EF, Levy S, Feuk L, Scherer SW: **Towards a comprehensive structural variation map of an individual human genome.** *Genome Biol* 2010, **11**:R52.