Quantitative Biology 2013, 1(1): 71–90
DOI 10.1007/s40484-013-0005-3

QB

REVIEW

Personal genomes, quantitative dynamic omics and personalized medicine

George I. Mias and Michael Snyder*

Department of Genetics, Stanford University School of Medicine, Stanford University, Stanford, CA 94305, USA
* Correspondence: mpsnyder@stanford.edu

Received November 1, 2012; Revised November 14, 2012; Accepted November 14, 2012

The rapid technological developments following the Human Genome Project have made possible the availability of personalized genomes. As the focus now shifts from characterizing genomes to making personalized disease associations, in combination with the availability of other omics technologies, the next big push will be not only to obtain a personalized genome, but to quantitatively follow other omics. This will include transcriptomes, proteomes, metabolomes, antibodyomes, and new emerging technologies, enabling the profiling of thousands of molecular components in individuals. Furthermore, omics profiling performed longitudinally can probe the temporal patterns associated with both molecular changes and associated physiological health and disease states. Such data necessitates the development of computational methodology to not only handle and descriptively assess such data, but also construct quantitative biological models. Here we describe the availability of personal genomes and developing omics technologies that can be brought together for personalized implementations and how these novel integrated approaches may effectively provide a precise personalized medicine that focuses on not only characterization and treatment but ultimately the prevention of disease.

INTRODUCTION

With the advent of high-throughput technologies genomic science has experienced great leaps, rapidly expanding its domain beyond the characterization of short genomic reads in the early days of sequencing to the possibility of obtaining personalized genomes, once considered the holy grail of genomic methodology and technology development. The value of personalized genomic analysis, and evaluation of variant associations to disease, is becoming more apparent, even spurring directly to consumer implementations. Further developments in the last few years now lead to a more ambitious goal: the longitudinal monitoring of multiple omics components in individuals and the characterization of the molecular changes associated with disease onset in individuals, at an unprecedented level. In this review we describe technological and methodological developments in personal genomics, and the new promise of multiple omics profiling, including transcriptomes, proteomes, metabolomes, autoantibodyomes and so forth, (sample omics analysis workflows shown in Figures 1–4). We then discuss a framework on how such data may be integrated with a view towards the application of a personalized precise and preventive medicine, and describe an implementation of this approach. The technological developments and methodology allow for inroads into the future of quantitative personal medicine, which we can now plan carefully by taking into account not only the scientific developments that need to be implemented, but also the social implications coupled to ethical and legal considerations.

GENOMIC SEQUENCING

In 2001 the completion of the Human Genome Project (HGP) was announced effectively with the publication of the first complete human genome sequence. The HGP came at a hefty $2.7 billion cost using the best technology of the time, making it seemingly prohibitive to expect personal genome sequences to be achieved shortly thereafter. Yet the immense technological advancement,

spurred by motivation by the National Institute of Health (NIH) and the National Human Genome Research Institute (NHGRI) to bring down genomic costs, led to an unprecedented growth in technology and methodology, enabling the drop in sequencing costs (http://www.genome.gov/sequencingcosts) to continue at a rate beyond the most optimistic projections of 2001 ($<$ \$4000 currently). While initially the human genome was a combination of multiple individual genomic data [1–3], the developments by 2008 had allowed the determination of genomic individual makeup [4–7]. It is now possible to personalize Whole Genome Sequencing (WGS), and the dwindling sequencing costs promise the possibility of affordability for all in the near future [8]. These developments encouraged efforts to characterize disease on a genomic level, towards the application of an all-encompassing genomic medicine, at the molecular level. The initial goals were the characterization of populations for large studies, now shifting to the individual.

Multiple technologies development/dropping costs

The HGP relied on technology using Sanger-based capillary sequencing [1] with an estimated production of 115k base pairs per day (kbp/day) [9]. The NHGRI spurred progress by encouragement through the \$1000 genome program (http://www.genome.gov/11008124-al-4), leading to the industry development of multiple massively parallel [10] sequencing platforms (e.g., Roche/454, based on pyrosequencing [11–13]; Life Technologies SOLiD [14–16]; Illumina [5,6]; Complete Genomics based on DNA nanoball sequencing [17]; Helicos Biosciences [18]; and recently single molecule real-time technology [19,20] by Pacific Biosciences). These next generation sequencing platforms are now being supplemented but what has been termed as third-generation sequencing, [21], including such nanopore technologies as announced early in 2012 by Oxford Nanopore Technologies [22]. The technological developments and competition resulted in a drastic and continuing drop in sequencing cost, processing times and exponential increases in number of reads produced.

An alternative to sequencing the whole genome has been whole exome sequencing (WES) [23]. This technology aims to study the exonic regions of the genome (~2%–3%), which are associated to several Mendelian disorders. It offers a lower cost option (e.g., Illumina, Agilent, and Niblegen platforms, see Clark et al. for a comparison of the latter two [24]) and has received immense attention, including the Exome Sequencing Project (ESP) (see the Exome Variant Server at http://evs.gs.washington.edu/EVS/), supported by the National Heart, Lung and Blood Institute (NHLBI).

Quantitating genomic variation

Concurrently with the technological developments, our understanding of the human genome has grown immensely since the publication of the reference genome in 2003. The aim was to determine the precise role of each base in the genome and identify genomic variants (Figure 1). Several collaborative large-scale efforts pursued such investigations. The International HapMap Consortium [25,26] tried to identify common population variants and led to the development of public databases, such as dbSNP [27] (http://www.ncbi.nlm.nih.gov/SNP/), which catalogues Single Nucleotide Polymorphisms (SNPs) (defined as occurring in $>1\%$ of the population to differentiate from Single Nucleotide Variants (SNVs)). This has revealed great genomic variation both in global populations [28,29] and populations of admixed ancestry [30–33].

Typically the technologies involve the assignment of reads to the reference genome to determine the structure of the underlying sequence, including variation (Figure 1). Beyond nucleotide variation, other genomic differences have been investigated, including small insertions and deletions (indels), copy number variations (CNVs) indicating varying numbers of segments and longer chromosomal segments that contribute to Structural Variation (SVs) — SVs are defined for segments of chromosomes larger than 1000 bp (Figure 1A). Such efforts have been based on microarray methodology [34–37] and even higher-resolution in structural variants may be achieved with other methods [38–41]. Structural variants have been publically made available in the database of Genomic Structural Variation (dbVAR; http://www.ncbi.nlm.nih.gov/dbvar/).

Furthermore, functional elements have been extensively catalogued by the Encyclopedia of DNA Elements consortium (ENCODE; http://genome.gov/encode ~10 production projects), with funding from the NHGRI. ENCODE data, including regulatory elements and RNA and protein level elements, have now been released and the project has received widespread attention [42–45]. The ENCODE project aims at a biochemical genomic characterization, with a thorough mapping of transcribed regions, transcription factor binding sites, open chromatin signatures, chromatin modification and DNA methylation. Such extensive data still needs to be annotated [46] interpreted in terms of biological significance, mechanisms and connections to phenotype and will likely prove invaluable in our interpretation of personalized genomic differences.

Though initially limited by the number of complete genomic sequences, such data are now continuously updated and expanded by information from other projects such as the 1000 Genomes Project [47] as discussed

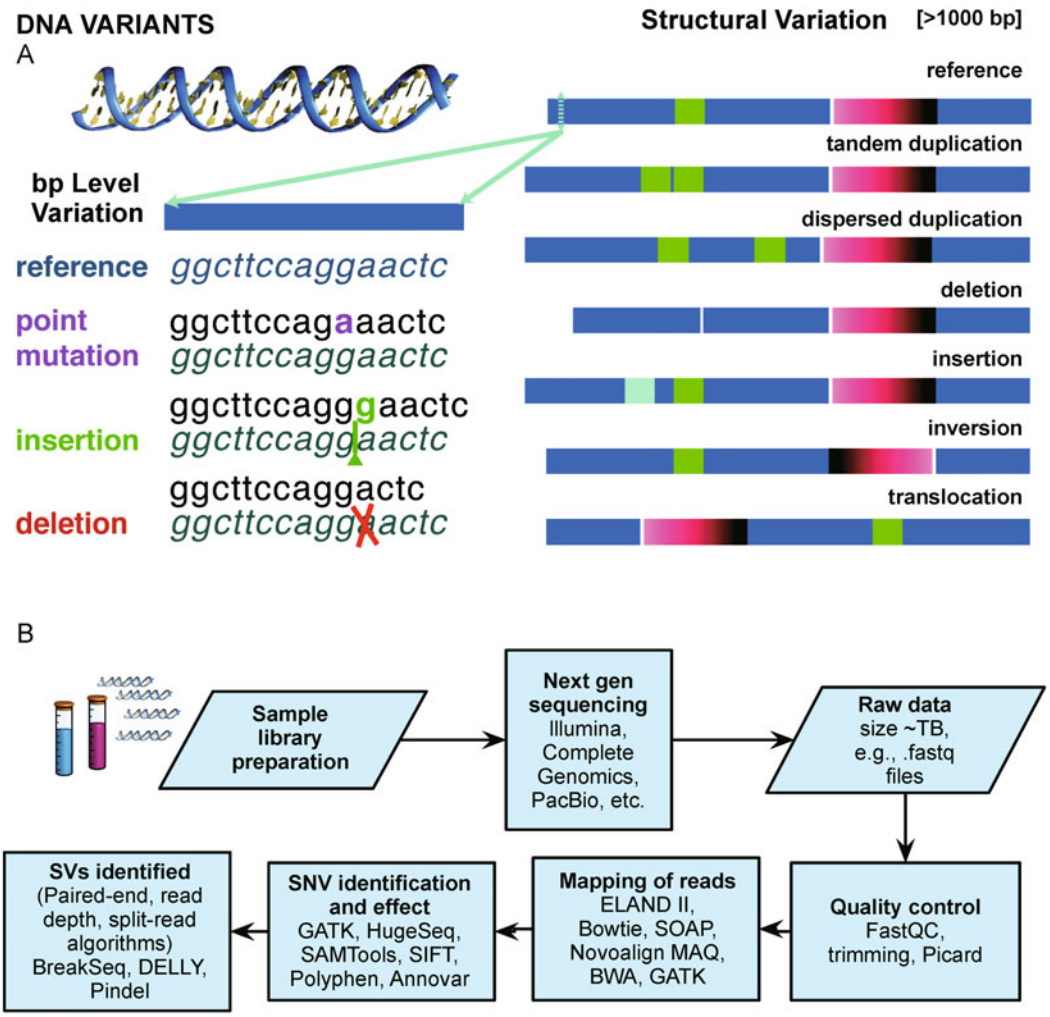

Figure 1. Genomic variants. (A) Variation in the human genome. The personal genomic code can differ from the published reference genome. Basic examples of variation are shown on a single or few base variants (e.g., point mutations, insertions and deletions), or a larger scale for structural variants (>1000 bp, e.g., large insertions, deletions, inversions, tandem repeats, translocations). (B) Sample variant analysis workflow. In a genomic variant analysis, for example, after sample preparation and sequencing the raw files can be passed through quality control (e.g., using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and removing PCR artifacts using tools as Picard (http://picard.sourceforge.net)). Reads are mapped to the genome and variants are assessed, e.g., mapping with several algorithms, including ELAND II (Illumina), SOAP [221], MAQ and Burrows-Wheeler Aligner (BWA) [222] and Novoalign by © Novocraft Technologies (http://www.novocraft.com). Read re-alignment can be performed, e.g., using Genome Analysis Toolkit (GATK) [223], or HugeSeq [211], to call variants, including implementations with Sequence Alignment Map format Tools (SAMtools) [224], annotation using Annovar [225], SIFT [226] and Polyphen [227] for determining variant effects on proteomic translation [228]. Furthermore, using a variety of methods the structural variants can be determined. For example the *paired-end mapping* method considers how paired-end reads mapped to the reference to assign deletions and insertions, from reads whose mapped span is longer or shorter than the average span; inversions, from position and relative orientations of the ends of reads [39,40]. The *read depth* method allows the possibility to identify the proportional genomic copy number variation. In the approach of Abyzov et al. [229] the read depth considered as an image is analyzed using image processing techniques, viz. mean-shift-theory [230]. Programs such as Pindel [231] and BreakSeq [232] consider *split-read* analysis to determine breakpoints of insertions and deletions. DELLY [233] by Rausch et al. takes into account paired-end and split-read methods for determining structural variants. Many packages for analysis are available through the Bioconductor [234] project as implemented in the freely available R statistical analysis platform (http://www.R-project.org).

below, which has allowed us to have a better view of the great variability in each individual genome (~3–4×10^6 SNPs, >200000 SVs of varying sizes, ~1500 SVs> 2 kbp), with much of the variation considered rare (1%–5%). Genome-Wide Association Studies (GWAS) try to associate the common variants to disease, by combining the now readily available extensive variant information and allelic variability, with linkage disequilibrium (a description of the correlation patterns between proximal variants). The NHGRI provides a publically available catalogue of published GWAS (http://www.genome.gov/gwastudies) [48]. The early expectations of finding common traits and genomic features unique to diseases have proven more complicated, as the genomic variability turns out to be higher than expected and additionally the genetic variants need further validation.

Use of WGS and WES has been successful in the identification of somatic mutations. Mendelian disorders including neurological disorders, and cancer have been characterized using WES [49–58], including some recent single-cell studies [59,60]. Genomics may help classifying cancer subtypes, and possible treatment, and such research is at the center of WGS, with projects such as the Cancer Genome Atlas [61] (http://cancer-genome.nih.gov/), and the International Cancer Genome Consortium (http://www.icgc.org). Additionally, cancer specific public databases already are available [62], including a cancer cell line encyclopedia [63], and genome characterization has been carried out, for example in ovarian cancer [61], melanoma [64], lymphocytic leukemia [65], breast cancer [66–69] and acute myeloid leukemia (AML) [70,71].

Personalized risk evaluation

One of the goals of personalized genome interpretation is the evaluation of disease risk factors based on an individual's variant and allelic distribution composition. Such information may be compared to similar individuals with known disease associations to assess whether an individual shows increased or decreased risk compared to the control group. A combination of know SNPs and personalized variants has been found to be effective [72–75] and has been used in clinical studies; more recently, a seminal study by Ashley et al. [76] evaluated disease risk for a patient with family history of vascular disease.

Personalized evaluation of potential drug responses can be based on the effects of variants [77,78], including drug selection, sensitivity and dosage estimation, e.g., cardiovascular drugs [79], schizophrenia related medications [80]. For example, PharmGKB (http://www.pharmgkb.org) provides a curated database of possible genomics information [81,82], exploring the impact of genomic variation on drug responses as these relate to expressed genes and associated pathways and disorders. The future applications are to include a precise drug dosage for an individual, avoiding trial and error methods and providing more effective treatment.

The evaluation of personalized risk based on genomes is now appearing in direct-to-consumer services. Companies like 23andMe, deCODEme, (and previously Navigenics), offer to assess individual genotypes and offer disease based interpretation services based on Mendelian disorder evaluation and including pharmacogenomics responses. These are mostly based on SNPs evaluation and the tests though limited in scope do offer interpretation attractive to multiple consumers.

Personal Genomes Project

Presently thousands of genomes have been completely sequenced. One of the first large scale projects has been the 1000 Genomes Project [47], that has made its data publically available, and has encouraged the development of streamlined bioinformatics tools to analyze the variation in the individual genomes (Figure 1). This project aims to combine data from 2500 individuals from multiple populations, at a 4× coverage.

Another grand scale effort driven by George Church's group at Harvard University is the Personal Genome Project (PGP) [83–85]. The project has been recruiting individuals who can share their medical and other information together with genomic information online (http://www.personalgenomes.org). The volunteers share full DNA sequences, RNA and protein profile information in addition to extensive phenotype information including medical records and environmental considerations, with all the data made publically available, and plans to expand to 100000 individuals [86]. One of the rather unique features of the PGP project is that it differs in consent of participants as compared to traditional studies. The ownership of the data is to be open and publically available without restrictions, not only for the initial perspective of the study, but open to follow-up or additional investigations. The scope is participatory, with the volunteers for the project interacting directly with the researchers. To address informed consent, participants pass a basic genetic literacy exam and must understand the project's scope. Additionally, they provide complete medical history, immunization and medications history, which becomes part of the publically available subject information. The access to the individual's data in the project can be either private to the participant and researchers only or completely public, depending on the participant's choice. The availability of extensive patient and omic information will be invaluable to researchers in developing robust analysis models for characterizing

genomes and disease and the PGP project, and its publically open structure model, will be at the forefront of such efforts.

BEYOND THE GENOME: OTHER OMICS

Transcriptomics

Though the genetic code in DNA is the almost identical (besides cellular variation), different cells have different gene expressions, corresponding to the kind of cell, developmental stage and physiological state. The collection of the transcripts in a cell (e.g., mRNA, non-coding RNA and small RNAs), the transcriptome, is essential in our understanding of cell function, and response to disease. Considerations must include start and end sites of genes, and coding, alternative splicing and post-transcriptional modifications.

Initially inroads were made using high-density oligo microarrays, and in-house custom made microarrays [87], with high-density arrays having resolutions up to 100 bp [88–91]. While relatively inexpensive, these methods suffered from relying on prior knowledge of the genome, and faced technical issues such as background and saturation effects [92]. Hybridization interactions between probe sets in short oligo microarrays lead to spurious correlations [92,93].

The development of RNA sequencing (RNA-Seq) brought higher coverage, better precision and quantitation, and higher resolution and sensitivity, bringing RNA-Seq technology and transcriptomics on par with genomic sequencing [94–98]. RNA-Seq considers reads that correspond to millions of transcriptomic fragments that are mapped to the reference genome, to provide information on transcripts that may not be in the existing genomic annotation, allowing the search for novel transcripts, and even identification of SNPs and other variants, while showing remarkable reproducibility (Figure 2). Transcriptome profiling has included looking at cancers [99–101], including breast cancer [102], gastrointestinal tumors [103] and prostate cancer [104].

Mass spectrometry, proteomics and metabolomics

Gene expression was expected to correlate with protein levels in a cell and it was thought that methods such as RNA-Seq would be enough to ascertain the proteomic expression corresponding to gene expression. Proteins are expected to be closer to phenotype, as they participate in every aspect of cellular biology, but their expression levels are difficult to quantitate, partly because of translational control in cells, possible degradation and sampling issues [105–107]. The development of electrospray ionization brought mass spectrometry (MS) to the field of proteomics and the possible identification of thousands of molecules based on mass [108–112]. This has enabled not only the cataloguing of proteins, but also querying post-translational modifications [113,114]. As the techniques matured, liquid chromatography tandem mass spectrometry (LC-MS/MS) has become standard, and novel instruments (e.g., Velos family [115] by Thermo Scientific; quadrupole time-of-flight mass spectrometers (QTOFs) by Agilent) allow unprecedented precision to enable the development of methods to

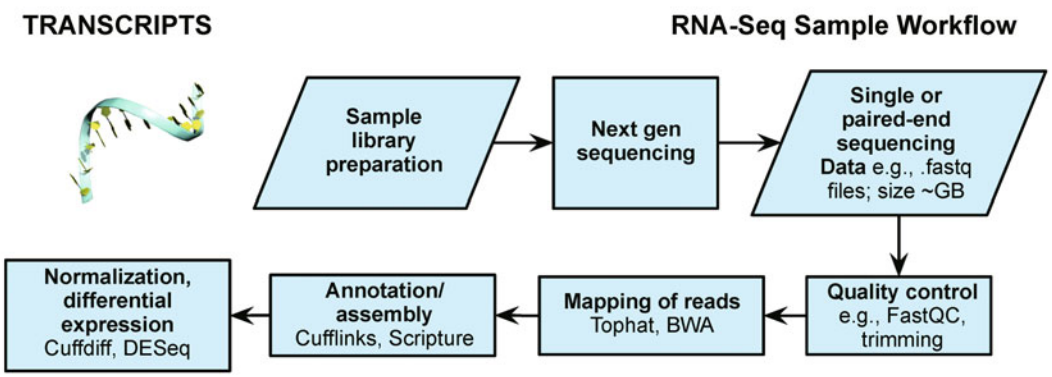

Figure 2. RNA-Seq analysis. In RNA-Seq analysis, short reads can be assembled and then mapped to the reference genome (with tools such as Illumina's ELAND, MAQ and BWA [222], Bowtie [235–237], SOAP [221], and others). A recent protocol by Trapnell et al. [238] describes in detail the use of dedicated RNA-Seq programs from the Tuxedo suite, such as TopHat [239], Cufflinks [240,241] and an R implementation called CummeRBund as a Bioconductor package (an alternative is to run these directly or using GenePattern [242,243], which also includes possible reconstruction by Scripture [244]). Other programs such as DESeq, another package in Bioconductor, can also help test for differential expression [245]. The numerous analyses availabilities are now publically discussed online, in a forum (http://SEQanswers.com/) that discusses many other examples and all aspects of the mapping process [246].

identify thousands of proteins (~4000–6000 over 2 days), and quantitate protein levels [73,116] (Figure 3). One set of methods uses stable isotopic labeling by amino acids in cell culture (SILAC) to label cell in light and heavy isotopes of amino acids providing double spectral peaks in MS for identification and quantitation [117–120] — this method is now supplemented by 'spike-in'/'super' SILAC which has been used to measure biopsy tumor proteomes [121]. Another possibility is to use isobaric tags for relative and absolute quantitation (iTRAQ) [122,123] or tandem mass tag (TMT) labeling [73,124,125], and other methods, including spiking in peptides for absolute quantitation. Finally, it is possible to employ label-free methods for quantitation, which do not rely on tags, including integrating signal methods and MS spectral counting [126–131].

In comparison to whole transcriptome profiling, the numbers of proteins identified in proteome profiling tend to be less in comparison, particularly since low peptide levels cannot be amplified (*cf.* polymerase chain reaction methods for sequencing methods). Additionally, the current bottom-up (shotgun) proteomics methodology uses digestion with endopeptidases such as trypsin to obtain peptides of small enough mass to be identified by MS/MS, resulting in many fragments that cannot be identified in MS, which may possibly be alleviated by top down approaches that do not employ a digestion step [132–136]. However, proteomics provides insights that are missing from transcriptomic analysis, especially given the low correlations between protein and transcriptome differential gene expressions [73,137–142].

Multiple proteomes have been quantitatively profiled, including characterization of ovarian cancer [143], an integrated approach that combines transcriptome and proteome information in a human cancer cell line by Nagaraj et al. [144], integrative gastric cancer characterization and effects of post-translational modifications [145], and looking for biomarkers in other cancers [146,147].

In addition to developments in proteomics, MS has encouraged the study of small molecules. The behavior of small molecules in cells though difficult to track provides insight into many common disorders. The set of all cellular small molecules is collectively called the metabolome. Metabolic processes are vital in biological pathways and a systems analysis of molecular cell complexity might lead to biomarker discovery, and possibly disease risk assessment, diagnosis and treatment [148]. Similar to proteomics, metabolomics can employ mass spectrometry to identify compounds [149] (Figure

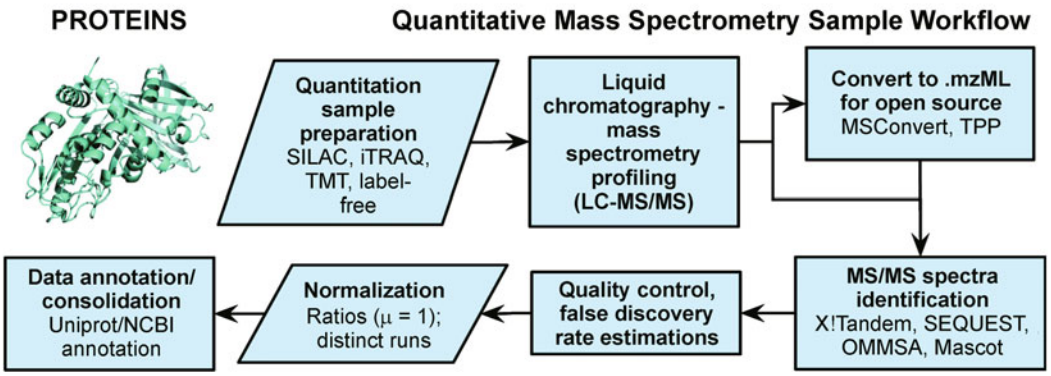

Figure 3. Proteome analysis. In quantitative proteomics using mass spectrometry typical approaches employ trypsin digestion coupled with tagging methods — non label-free methods include use of isotopic labeling (SILAC) or isobaric tagging (iTRAQ, TMT). One typical bottom-up-approach setup uses a combination of high affinity liquid chromatography coupled with two rounds of mass spectrometry (LC-MS/MS) to fractionate peptides for identification and obtain their mass spectra. Raw files may be analyzed using vendor software or converted to open formats (such as .mzXML, .mzData or the current standard .mzML [247–249], e.g., using MSConvert [250]). The mass spectra can be mapped to known protein using a protein library, or less frequently *de novo* assembled, using an array of programs (e.g., X!Tandem [251], SEQUEST [252], Mascot [253], Open Mass Spectrometry Search Algorithm (OMSSA) [254], Proteome Discoverer by Thermo Scientific, or MassHunter Workstation by Agilent). Quality control includes estimation of false discovery rates (FDR), often using a reverse database search [105,255,256]. Quantitation can be carried out to estimate relative levels of proteins in different samples (employing standardization and normalization of average sample ratios to a unit mean). Finally annotation is made using databases such as UniProt or NCBI. Some of the analysis can be performed using suites and programs, such as PEAKS [257], the Trans-Proteomic Pipeline (TPP) [258–261], multiple tools from ProteoWizard [250], OpenMS [262–264] or vendor complete solutions Proteome Discoverer and MassHunter Workstation mentioned above. Multiple other programs for mass spectrometry are available (e.g., see http://www.msutils.org).

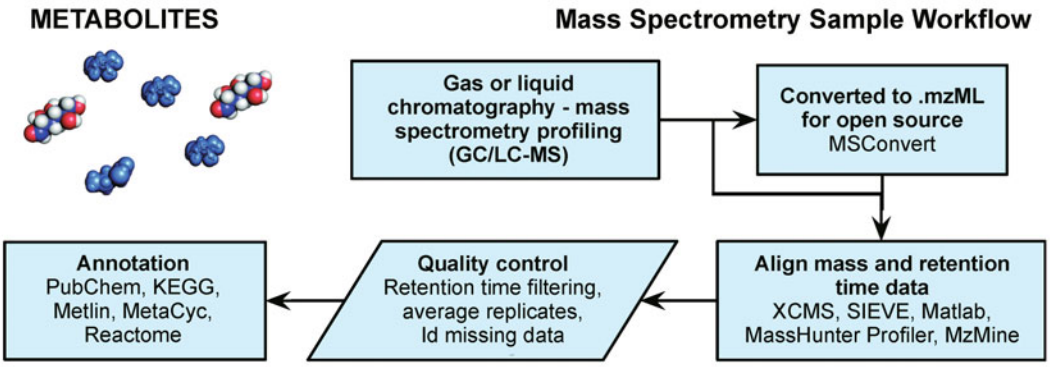

Figure 4. Metabolome analysis. In metabolomics analysis chromatography columns are used for purification and preparation of samples coupled to mass spectrometry (gas chromatography (GC) or liquid chromatography (LC)-MS); standards for specific compounds may also be used in parallel for positive identification. Raw files may be analyzed using vendor software or converted to open formats (such as .mzXML, .mzData or the current standard .mzML [247–249], e.g., using MSConvert). The spectral data may be aligned for retention time and mass intensity calibration, e.g., using XCMS [265–267], SIEVE by Thermo Scientific, Matlab toolboxes by MathWorks, MassHunterProfiler by Agilent, MzMine [268,269]. After quality control and statistical analysis, masses of interest can be annotated using databases, e.g., Metlin [155,156], KEGG [151], MetaCyc [153,270,271], Reactome [157–161].

4) and cataloguing is under way, with thousands of metabolites identified by structure, mass and occasionally associated biological processes [150–161]. The identification of compounds can be based on MS/MS application and use of known compound spectra, or via use of standards against which mass spectra are compared. The profiling of metabolic components on an individualized basis can provide insights into pharmacogenomics and personalized medications, in addition to potential biomarkers, for example cholesterol levels and coronary artery [162,163]. The metabolomics of cancer has been extensively studied [164–166] and Type 2 Diabetes has been investigated [167], and *in vivo* interactions with proteins are being evaluated [168].

Other omics

Genomes, transcriptomes and metabolomes have received widespread attention and currently offer the most quantitative data, provided by robust and comprehensive omics technologies, both in terms of experimental, as well as computational methodology. However multiple other omics are available, and these numbers are increasing, with a few notable technologies mentioned below:

- Autoantibodyomes: In addition to profiling of proteins directly, the reactivity of proteins to autoantibodies may be profiled on a large scale. Spotted protein arrays [169–173] have been implemented to study for example effects in cancer [174], immune response [175] and recently diabetes [176]. Another approach is the Nucleic Acid Programmable Protein Array (NAPPA) constructed by spotting plasmid DNA to effectively express and code the proteins on the array and used for immunoprofiling [177,178]. Furthermore functional peptide arrays have also been constructed [179,180]. Complementary technologies such as bead-based immunoassays are also being actively developed, such as the Luminex xMAP assay [181].
- Microbiomes: Omics profiling could also include mapping of the personal microbiome, the complete set of microbes in an individual (e.g., found mainly on the skin or in the gut, conjunctiva, saliva and mucosa) using possibly a combined omics approach to look at genetic makeup and metabolic components [182–187]. The human microbiota (http://www.human-microbiome.org) have been associated to obesity [188] and diabetes [189,190] and have also been suspected to play an active role in the development of immunity [191]. The dynamic monitoring of microbiome-related changes can help identify the specific microbiota involved in disease responses, elucidate microbiome-host interactions and how the individual variability in components impacts developmental and metabolic processes.
- Methylomes: In addition to genomics, epigenomic information, such probing the methylome, i.e., identifying all genomic sites of cytosine methylation [192,193], might provide information about differentiation and regulation of gene expression. Methylation analysis and data interpretation can be challenging [194,195] but methods are improving as more data becomes available. Methylome analysis has now been carried out in blood components [196], stem cells [197] and ovarian cancer [61], and it might prove invaluable in assessing epigenomic effects on individual development and

health.

PERSONALIZED MEDICINE

The developments of the many different omics technologies outlined above have given us tremendous insight into the human genome and associations to diseases, especially with the rise of the personal genome. The NHGRI recognizing the importance of these developments and the directions necessary to enhance health care, outlined in 2011 a vision for the future of personalized medicine [198] encompassing five domains of development that included understanding the structure of genomes, their biology, improving our understanding of the biology of disease, advancing medicine and improving the effectiveness of healthcare. The aims had been set to a shift towards personalized medicine within two decades, but the availability of the technology and constant decreasing costs have made pilot investigations of personalized medicine a current possibility [73]. Genetic variation has proven adequate for understanding group differences in disorders, but a truly personalized implementation needs to consider an individual. Clinicians are already considering molecular markers in their evaluation of patients, and particularly cancer [199–203]. The typical clinical diagnosis involves the observation of symptoms traditionally confirmed utilizing a small set of molecular markers. In diseases that share a common set of symptoms, some rare, such diagnosis is often complicated and prolonged, especially for heterogeneous disorders that need additional information to enable classification and subsequent specific treatments. Genetic and environmental factors create additional variability in disease severity, progression and treatment responses. Thus, traditional assays together with the aforementioned current omics technologies, that allow monitoring of thousands of molecular components, will facilitate and accelerate differential diagnostics and sub-classification through utilizing a more complete set of disease markers. A personalized approach will result in better targeting of diseases, introduce higher precision through measurement of larger sets of molecular components and ideally implemented at an early age to assess disease risk and have a preventative rather than retrospective treatment focus.

A personal approach is by its nature an $n = 1$ study, which helps eliminate variation between individuals that are treated as a group, but still requires some verification and establishment of a baseline for comparison. As such, the profiling of healthy physiological states in a longitudinal approach may provide such a basis, if multiple time points with similar physiological state makeup are sampled. Multiple omics can supply multiple supporting datasets at each time point, with each complementary technology providing additional supporting information for a baseline establishment. This introduces the concept of complete omics monitoring of individuals over time, making personalized medicine a more dynamic proposition. The dynamic changes of molecular components may be associated to the individual's changing physiological states, and mapped onto pathways to identify the onset and progression of disease, including possible preventive measures. In our suggested implementation, termed integrative Personal Omics Profiling (iPOP) which we followed in the study discussed below [73] we integrate the omics components discussed above in a longitudinal approach with three essential steps (Figure 5):

I) *Risk estimation*: As discussed above the personal and common genomic variants determined in an individual genome can be associated to disease [76], with pharmacogenomic evaluation to determine possible drug response. An early age whole genome sequencing, possibly at birth, can provide a list of possible increased risk disorders and lead to taking preventive measures. This may be done in combination with a complete medical and family history, as for example implemented in the PGP project, and in conjunction with classical clinical risk factor profiling.

II) *Dynamic profiling of multiple omics*: Starting with a healthy or 'steady state' baseline, by monitoring changes in the molecular components over multiple time points, drastic or gradual changes in physiological states might be assessed and the dynamic onset of disease profiled, and possibly prevented. Such profiling may be done on blood components, which are easily obtainable currently in the clinic. The individual blood components are excellent reflectors of generalized physiological state of an individual, as the blood circulates and receives inputs from multiple tissues throughout the body. The components may be processed to track multiple omics, such as transcriptome, proteome, metabolome and autoantibodyome, etc., which as mentioned offer complementary information, especially given the modest correlation observed between transcriptomic and proteomic components [137–142]. A recent study of profiles of tumors changing over time also employed an integrative approach on genomic and transcriptomic components [204]. Implementing this monitoring on healthy individuals will allow the monitoring of disease onset and physiological changes from various healthy, disease and recovery states, and following thousands of molecular component levels and responses at corresponding physiological states.

III) *Data integration and biological impact assessment*: The multiple omics data can be analyzed individually to characterize their temporal response profile. This may be done using standard statistical time-series analysis, extensively used in all quantitative disciplines, such as

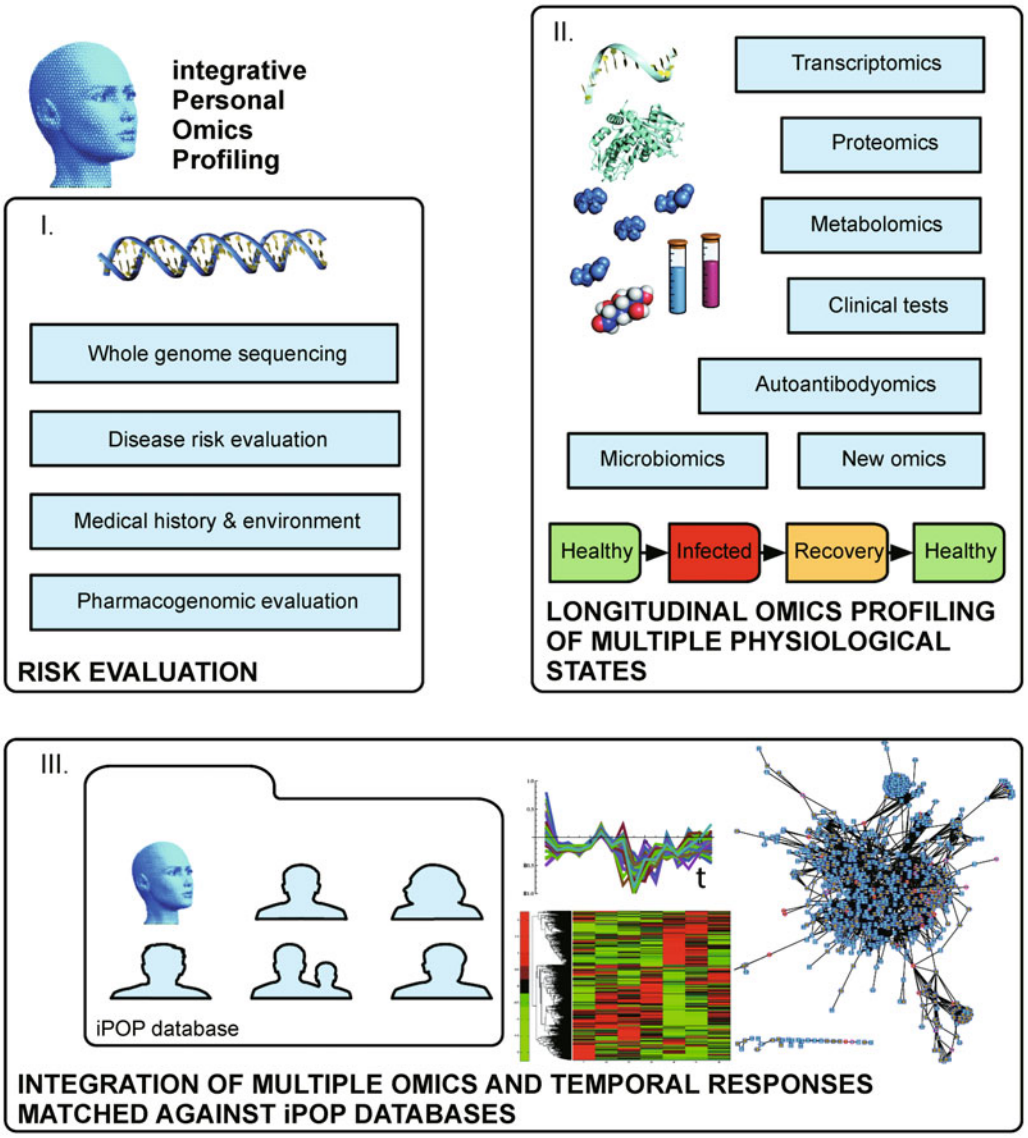

Figure 5. iPOP for personalized medicine. The framework described in the text employs multi-omics analyses (see above and Figures 1–4) that may be implemented for individuals. In step I) *Risk estimation* for disease is carried out using a whole genome sequencing to perform variant analysis coupled to medical history, environmental considerations and pharmacogenomics evaluations. In step II) *Dynamic profiling of multiple omics* using an array of technologies follows multiple omics longitudinally in a subject as they progress through their different physiological states, including healthy, disease, and recovery states. Thus thousands of molecular components are collected over time for III) *Data integration and biological impact assessment*, using temporal patterns to obtain matched omics information, correlate and classify responses, compare against pathway databases and visualize components, e.g., current pathway tools include DAVID [206,272], KEGG [151], Reactome [157–161], Ingenuity Pathway Analysis (IPA); networks can be visualized using Cytoscape [207], various R packages through Bioconductor [234], Matlab by MathWorks and several others. The future iPOP implementations may be gathered into a curated database of iPOP-disease associations that may help in categorizing an omics dynamic response to a catalogued physiological state and disease onset, with potential diagnostic capabilities.

physics, economics and finance, as discussed by Bar-Joseph et al. [205]. The dynamic signature of the signals for each molecular component can be studied for autocorrelation, periodicity or spikey behavior, corre-

sponding to causal changes or abnormal physiological state conditions resulting from the onset of disease, infections, or environmental effects. The different classes of temporal response can be checked for biological pathway and gene ontology enrichment [151,157–161,206–210], and corresponding disease associations in comparison to a database of other longitudinal profiles (coupled to complete electronic records of omic and medical histories). Such a database is a necessary and powerful resource towards the realization of personalized medicine based on omics data profiling.

Example implementation of personalized medicine: iPOP

To show the feasibility and practical applicability of iPOP we profiled a healthy individual, 54, over a period of initially 14 (now 33) months [73]. This initial time series covered healthy states, and two viral states, including a human rhinovirus (HRV) infection at the initiation of the study and a respiratory syncytial virus (RSV) infection 289 days later. The iPOP used blood samples to extract omic components from peripheral blood mononuclear cells (PBMCs) and serum, which were analyzed to obtain a complete DNA, RNA, protein, metabolite and autoantibody profile. Initially a complete medical exam was performed with standard clinical tests before time-point profiling began. In a first step, WGS with two platforms was carried out (Complete Genomics and Illumina, at 150- and 120-fold coverage respectively) and WES with three platforms (Nimblegen, Illumina and Agilent) and helped identify a large number of variants ($>3\times10^6$ SNPs; $>2\times10^5$ indels; >2000 SVs). Using multiple platforms allowed us to determine high-confidence and novel variants (using HugeSeq [211]). Evaluation of genetic disease risks based on variants was carried out, both by looking for known disease associations using dbSNP and the Online Mendelian Inheritance in Man (OMIM, http://omim.org/) database and using the RiskO-Gram algorithm [76] which integrates information from multiple alleles to assess risk against a similarly matched data cohort. This revealed significantly increased risk for various disorders, including open angle glaucoma, dyslipidemia, coronary artery disease, basal cell carcinoma, type 2 diabetes (T2D), age related macular degeneration and psoriasis. This encouraged the subject to follow up on these disorders, and also start monitoring glucose and glycated hemoglobin (HbA1c) levels, which surprisingly increased beyond normal levels following the RSV infection, and the subject was diagnosed by his physician for T2D 369 days into the study. Related to T2D, pharmacogenomic considerations revealed a possibly favorable (glucose lowering) response to diabetic drugs rosiglitazone and metformin, should treatment become necessary. Furthermore, the autoantibodyome profiling of the subject (Invitrogen ProtoArrays profiling of 9483 protein reactivities to Immunoglobulin G (IgG)) revealed increased reactivity in multiple proteins, including DOK6 (related to insulin receptors), and GOSR1, BTK and ASPA, previously reported to show high reactivity by Winer et al. in insulin resistant patients [176]. The subject initiated and still maintains a strict dietary and exercise regiment supplemented with low doses of acetylsalicylic acid, which helped him control his glucose and HbA1c levels, which after a considerable time period (~months) have now returned to normal levels.

In addition a range of omics were profiled over time for up to 20 different timepoints over the span of the study including high coverage transcriptome (RNA-Seq of PBMCs, 2.67 billion reads mapped to 19714 isoforms corresponding to 12659 genes), proteome (MS of PBMCs, identifying a total of 6280 proteins; 3731 consistently across most timepoints), metabolome (MS of serum, profiling 6862 and 4228 metabolites during periods of HRV and RSV infections respectively, with ~20% identified based on mass and retention times alone). The dynamic transcriptome, proteome and metabolome profiles were analyzed in a novel integrated framework based on spectral analysis of the time series. This allowed the identification of temporal patterns in the combined data, corresponding to biological processes that varied with physiological state changes, including the onset of T2D seen in multiple omics components, and common signatures of HRV and RSV infections. While several gene associations to pathways were known, multiple genes showed similar patterns that had not been reported before and merit further investigation.

OTHER CONSIDERATIONS AND FUTURE DIRECTIONS

The iPOP study discussed above revealed the complexities and characteristics of personal genomes, transcriptomes, proteomes and metabolomes and showed the feasibility of personalized longitudinal profiling that can provide actionable health information. Multiple omics data integration still presents a formidable challenge and merits further development. Each omics technology produces different kinds of data, including multiple formats (e.g., data files range from simple text, and extensible markup, e.g., .xml, to vendor closed-source formats). Additionally, each omics set requires its own quality control analysis, further confounded by different error and noise levels associated to the different technologies. As each of the data sets also presents different signal and noise distributions, this makes uniform normalization approaches across omics challenging, especially if considering multimodal dynamic data.

Furthermore, the amounts of information per omics set can vary, e.g., ~5000 proteins, ~20000 transcript isoforms, ~6000–10000 metabolites, ~9000 autoantibody-protein reactivities and so forth. Hence, gene-centric approaches, that integrate data corresponding to, associated or interacting with the same genes, will not always work, as the different components may not match. The integration of information per component is made more difficult with multiple existing gene and protein annotations, often resulting in a many-to-many map in the gene-protein integration, and correspondingly lacking metabolite-protein/gene annotations and associations. Finally, if considering dynamic datasets, this also results in multiple instances where time points might be missing data for some of the molecular components (especially evident in mass spectrometry and shotgun proteomics, where proteins are identified through different peptides). These complications of omics data integration necessitate that each individual omics data set is analyzed independently up to normalization, and then integrated with the other information. New integrative methodology has to account for such different normalizations, missing data, and also integration that is not gene-based, but rather incorporates time-series analyses, as for example was carried out in the iPOP study [73]. Classification of changes by temporal response, and possibly interaction data leads to an interpretation of components based on shared similar dynamics and avoids some of the issues of insufficient annotations and missing information. Such an interpretation lends itself to a clinical setting where dynamic changes are associated to varying personalized physiological states, and may be adopted by the medical community.

To facilitate the wide adoption of the methods into personalized medicine, the integrated data analysis will require optimization of current computational tools to rapidly and efficiently handle as well as visualize the multiple omics data. As a first step, the amount of computation time for different analyses must be reduced from days (in the case of mapping sequence data and quantitative proteomics in current omics analyses presented above) to hours or less to have immediate relevance to active medical examinations. Secondly, better visualizations of omics data, though difficult, are also necessary, as multidimensional information is difficult to collate, present, and interpret (many efforts are addressing this, e.g., Circos plots that allow multiple sequence information to be displayed together are now widely adopted [212]). Incorporating such information with clinical data and phenotypes presents a new challenge, requiring browsers that combine temporal information with multi-dimensional omics sets. We believe network analysis [213–217] presents an excellent visualization and integration possibility, allowing the combinations of multiple levels of networks, dynamically changing, that will include cellular information, component and corresponding disease temporal progressions, as well as medical assay data in a modularized approach. The computational analyses and visualization of omics data integration also reveal the known need to manage large amounts of data [218,219], both in terms of processing power, as well as storage capacity and maintaining easy accessibility, especially for the practicing clinician — with the recent advent of cloud computing providing one possible solution. Finally, the combination of omics data with medical records presents another challenge, with privacy and ethical issues that must be considered. Such improvements and standardization of approaches will help make the analysis available in a clinical setting and an increasingly larger set of patients, while encouraging the early adaptation of the integrated approaches by the scientific community towards personalized medicine applications.

As technology improves we expect to see advancements in each omics implementation discussed above. In terms of sequencing, continual improvements in depth and read length will allow unambiguous precise sequence mapping and additionally the querying of lower gene expression, coupled to higher accuracy in variant calling. With sequencing times becoming faster (e.g., whole genome sequencing in ~5–30 hours depending on platform at deep, ~100× coverage), and hardware more compact, eventually such technology will be available in the clinic, enabling the incorporation of all genomic, transcriptomic, microbiomic and autoantibodyomic profiling as parts of regular medical examinations. Correspondingly, mass spectrometry improvements (including table-top hardware now available) will improve mass accuracy, and higher sensitivity, allowing increases in the number of proteins identified and better quantitation, which can already be implemented in a clinical setting. The MS improvements in combination with better metabolite cataloguing will also improve the identification of small molecules. The protocol and methodology advancements will allow using a smaller volume of patient sample needed for iPOP (decreasing from ~80 mL to drops of blood) making it feasible to probe the omics on more regular basis for each patient, even providing home kits to send in self-collected samples (akin to what is already implemented to some degree by companies, e.g., 23andMe, that collect saliva samples for phenotyping).

The technological and methodological advancements will allow for effective iPOP implementations with multiple patients, but it will still take some time to evaluate what constitutes actionable information and which components will be most informative. Once these relevant components are identified monitoring technologies can be further developed to help possible clinical

implementations. This will certainly be alleviated by multiple iPOP studies providing the necessary aggregated information. However, clinical and psychological concerns need to be addressed and the possible impact to patient health being of paramount importance, in a medical process in which the patient is actively participating [220]. Such active participation requires the training of the public and health professionals to an understanding of genomic information, and how this omics knowledge impacts their health, and their families. Genetic counseling is a necessity, and the number of trained genetic counselors is steadily increasing. Informed consent will be necessary, but this requires an understanding of basic genomic terms that are not apparent to non-experts. To facilitate this, probably school curriculum adjustments will be needed to enable early education of the public.

The emergence of quantitative Personal Omics, including genomes transcriptomes, proteomes, metabolomes and other omics allows us to now combine them to yield personalized actionable health care information. Such research is at the forefront of medical science, and may help the characterization of disorders and the implementation of precise personal medicine aimed towards prevention rather than treatment. Careful forward planning, coupled to the continuing interest and participation of the public, government agencies and researchers, assures that the development of personalized omics will proceed beyond possible hurtles into a novel approach for the 21st century health care implementations.

ACKNOWLEDGEMENTS

We would like to thank the Stanford Genetics Department and the NIH for support through grant P50HG02357. GIM would also like to thank the NIH for support through training grant T32HG000044. We also thank Drs. Rui Chen, Jennifer Li Pook Than and Hogune Im for useful discussions.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing and analysis of the human genome. Nature, 409, 860–921.
2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001) The sequence of the human genome. Science, 291, 1304–1351.
3. International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. Nature, 431, 931–945.
4. Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J., et al. (2008) The diploid genome sequence of an Asian individual. Nature, 456, 60–65.
5. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456, 53–59.
6. Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature, 452, 872–876.
7. Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., et al. (2007) The diploid genome sequence of an individual human. PLoS Biol., 5, e254.
8. Snyder, M., Du, J. and Gerstein, M. (2010) Personal genome sequencing: current approaches and challenges. Genes Dev., 24, 423–431.
9. Mardis, E. R. (2011) A decade's perspective on DNA sequencing technology. Nature, 470, 198–203.
10. Tucker, T., Marra, M. and Friedman, J. M. (2009) Massively parallel sequencing: the next big thing in genetic medicine. Am. J. Hum. Genet., 85, 142–154.
11. Ronaghi, M., Uhlén, M. and Nyrén, P. (1998) A sequencing method based on real-time pyrophosphate. Science, 281, 363, 365.
12. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. and Nyrén, P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem., 242, 84–89.
13. Nyrén, P. (2007) The history of pyrosequencing. Methods Mol. Biol., 373, 1–14..
14. Nutter, R. C. (2008) New frontiers in plant functional genomics using next generation sequencing technologies. In Kahl, G. and Meksem, K. (eds.), The Handbook of Plant Functional Genomics: Concepts and Protocels. Wiley-VCH Verlag GmbH & Co. KGaA, Chapter 21, 431–446.
15. Dai, M., Thompson, R. C., Maher, C., Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M., Omenn, G. and Meng, F. (2010) NGSQC: cross-platform quality analysis pipeline for deep sequencing data. BMC Genomics, 11, S7.
16. Pandey, V., Nutter, R. C. and Prediger, E. (2008) Applied biosystems SOLiD™ system: ligation-based sequencing. In Janitz, M. (ed.), Next Generation Genome Sequencing: Towards Personalized Medicine. Wiley-VCH Verlag GmbH & Co. KGaA, Chapter 3, 29–42.
17. Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., Carnevali, P., Nazarenko, I., Nilsen, G. B., Yeung, G., et al. (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science, 327, 78–81.
18. Braslavsky, I., Hebert, B., Kartalov, E. and Quake, S. R. (2003) Sequence information can be obtained from single DNA molecules. Proc. Natl. Acad. Sci. USA, 100, 3960–3964.
19. Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., Holden, D., Saxena, R., Wegener, J. and Turner, S. W. (2010) Real-time DNA sequencing from single polymerase molecules. Methods Enzymol., 472, 431–455.
20. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009) Real-time DNA sequencing from single polymerase molecules. Science, 323, 133–138.
21. Schadt, E. E., Turner, S. and Kasarskis, A. (2010) A window into third-generation sequencing. Hum. Mol. Genet., 19, R227–R240.
22. Hayden, E. (2012) Nanopore genome sequencer makes its debut. Nature, 10.1038/nature.2012.10051.

23. Bainbridge, M. N., Wang, M., Burgess, D. L., Kovar, C., Rodesch, M. J., D'Ascenzo, M., Kitzman, J., Wu, Y. Q., Newsham, I., Richmond, T. A., et al. (2010) Whole exome capture in solution with 3 Gbp of data. Genome Biol., 11, R62.
24. Clark, M. J., Chen, R., Lam, H. Y., Karczewski, K. J., Chen, R., Euskirchen, G., Butte, A. J. and Snyder, M. (2011) Performance comparison of exome DNA sequencing technologies. Nat. Biotechnol., 29, 908–914.
25. The International HapMap Consortium. (2005) A haplotype map of the human genome. Nature, 437, 1299–1320.
26. The International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature, 449, 851–861.
27. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res., 29, 308–311.
28. Altshuler, D. and Clark, A. G. (2005) Genetics. Harvesting medical information from the human family tree. Science, 307, 1052–1053.
29. Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature, 451, 998–1003.
30. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., et al. (2008) Genes mirror geography within Europe. Nature, 456, 98–101.
31. Kidd, J. M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., Bryc, K., Degenhardt, J. D., Brisbin, A., Sheth, V., Chen, R., et al. (2012) Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. Am. J. Hum. Genet., 91, 660–671.
32. Galanter, J. M., Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A. V., Figueroa, L. U., Raska, P., et al. (2012) Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet., 8, e1002554.
33. Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., Froment, A., Bodo, J. M., Wambebe, C., Tishkoff, S. A., et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. USA, 107, 786–791.
34. Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., et al. (2006) Global variation in copy number in the human genome. Nature, 444, 444–454.
35. Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., et al. (2010) Origins and functional impact of copy number variation in the human genome. Nature, 464, 704–712.
36. Alkan, C., Coe, B. P. and Eichler, E. E. (2011) Genome structural variation discovery and genotyping. Nat. Rev. Genet., 12, 363–376.
37. Haraksingh, R. R., Abyzov, A., Gerstein, M., Urban, A. E. and Snyder, M. (2011) Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms. PLoS ONE, 6, e27859.
38. Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., Kim, P. M., Palejev, D., Carriero, N. J., Du, L., et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science, 318, 420–426.
39. Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., et al. (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods, 6, 677–681.
40. Korbel, J. O., Abyzov, A., Mu, X. J., Carriero, N., Cayting, P., Zhang, Z., Snyder, M. and Gerstein, M. B. (2009) PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genome Biol., 10, R23.
41. Quinlan, A. R. and Hall, I. M. (2012) Characterizing complex structural variation in germline and somatic genomes. Trends Genet., 28, 43–53.
42. The ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57–74.
43. Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012) Architecture of the human regulatory network derived from ENCODE data. Nature, 489, 91–100.
44. Ecker, J. R., Bickmore, W. A., Barroso, I., Pritchard, J. K., Gilad, Y. and Segal, E. (2012) Genomics: ENCODE explained. Nature, 489, 52–55.
45. Birney, E. (2012) The making of ENCODE: lessons for big-data projects. Nature, 489, 49–51.
46. Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res., 22, 1790–1797.
47. 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. Nature, 467, 1061–1073.
48. Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. and Manolio, T. A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA, 106, 9362–9367.
49. Haack, T. B., Danhauser, K., Haberberger, B., Hoser, J., Strecker, V., Boehm, D., Uziel, G., Lamantea, E., Invernizzi, F., Poulton, J., et al. (2010) Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. Nat. Genet., 42, 1131–1134.
50. Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010) A *de novo* paradigm for mental retardation. Nat. Genet., 42, 1109–1112.
51. Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., Gibbs, J. R., Brunetti, M., Gronka, S., Wuu, J., et al. (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron, 68, 857–864.
52. Bilgüvar, K., Oztürk, A. K., Louvi, A., Kwan, K. Y., Choi, M., Tatli, B., Yalnizoğlu, D., Tüysüz, B., Cağlayan, A. O., Gökben, S., et al. (2010) Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature, 467, 207–210.
53. Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet., 42, 30–35.
54. Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C.,

McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., et al. (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat. Genet., 42, 790–793.

55. Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C., Garimella, K. V., Fisher, S., Abreu, J., Barry, A. J., et al. (2010) Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N. Engl. J. Med., 363, 2220–2227.
56. Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., et al. (2012) *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. Nature, 485, 237–241.
57. Pugh, T. J., Weeraratne, S. D., Archer, T. C., Pomeranz Krummel, D. A., Auclair, D., Bochicchio, J., Carneiro, M. O., Carter, S. L., Cibulskis, K., Erlich, R. L., et al. (2012) Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature, 488, 106–110.
58. Agrawal, N., Frederick, M. J., Pickering, C. R., Bettegowda, C., Chang, K., Li, R. J., Fakhry, C., Xie, T. X., Zhang, J., Wang, J., et al. (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science, 333, 1154–1157.
59. Xu, X., Hou, Y., Yin, X., Bao, L., Tang, A., Song, L., Li, F., Tsang, S., Wu, K., Wu, H., et al. (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell, 148, 886–895.
60. Hou, Y., Song, L., Zhu, P., Zhang, B., Tao, Y., Xu, X., Li, F., Wu, K., Liang, J., Shao, D., et al. (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell, 148, 873–885.
61. The Cancer Genome Atlas Research Network. (2011) Integrated genomic analyses of ovarian carcinoma. Nature, 474, 609–615.
62. Küntzer, J., Maisel, D., Lenhof, H. P., Klostermann, S. and Burtscher, H. (2011) The Roche Cancer Genome Database 2.0. BMC Med. Genomics, 4, 43.
63. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature, 483, 603–607.
64. Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., Varela, I., Lin, M. L., Ordóñez, G. R., Bignell, G. R., et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. Nature, 463, 191–196.
65. Puente, X. S., Pinyol, M., Quesada, V., Conde, L., Ordóñez, G. R., Villamor, N., Escaramis, G., Jares, P., Beà, S., González-Díaz, M., et al. (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature, 475, 101–105.
66. Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W., Van Tine, B. A., Hoog, J., Goiffon, R. J., Goldstein, T. C., et al. (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature, 486, 353–360.
67. Ding, L., Ellis, M. J., Li, S., Larson, D. E., Chen, K., Wallis, J. W., Harris, C. C., McLellan, M. D., Fulton, R. S., Fulton, L. L., et al. (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature, 464, 999–1005.
68. Yost, S. E., Smith, E. N., Schwab, R. B., Bao, L., Jung, H., Wang, X., Voest, E., Pierce, J. P., Messer, K., Parker, B. A., et al. (2012) Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. Nucleic Acids Res., 40, e107.
69. Natrajan, R., Mackay, A., Lambros, M. B., Weigelt, B., Wilkerson, P. M., Manie, E., Grigoriadis, A., A'hern, R., van der Groep, P., Kozarewa, I., et al. (2012) A whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. J. Pathol., 227, 29–41.
70. Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., McGrath, S., Hickenbotham, M., et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature, 456, 66–72.
71. Link, D. C., Schuettpelz, L. G., Shen, D., Wang, J., Walter, M. J., Kulkarni, S., Payton, J. E., Ivanovich, J., Goodfellow, P. J., Le Beau, M., et al. (2011) Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. JAMA, 305, 1568–1576.
72. Dewey, F. E., Chen, R., Cordero, S. P., Ormond, K. E., Caleshu, C., Karczewski, K. J., Whirl-Carrillo, M., Wheeler, M. T., Dudley, J. T., Byrnes, J. K., et al. (2011) Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet., 7, e1002280.
73. Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L., Lam, H. Y., Chen, R., Miriami, E., Karczewski, K. J., Hariharan, M., Dewey, F. E., et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell, 148, 1293–1307.
74. Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science, 328, 636–639.
75. Bainbridge, M. N., Wiszniewski, W., Murdock, D. R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., Reid, J. G., Fink, J. K., Morgan, M. B., Gingras, M. C., et al. (2011) Whole-genome sequencing for optimized patient management. Sci. Transl. Med., 3, 87re3.
76. Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., Dudley, J. T., Ormond, K. E., Pavlovic, A., Morgan, A. A., et al. (2010) Clinical assessment incorporating a personal genome. Lancet, 375, 1525–1535.
77. Lesko, L. J. and Schmidt, S. (2012) Individualization of drug therapy: history, present state, and opportunities for the future. Clin. Pharmacol. Ther., 92, 458–466.
78. Evans, W. E. and Relling, M. V. (2004) Moving towards individualized medicine with pharmacogenomics. Nature, 429, 464–468.
79. Zineh, I. and Johnson, J. A. (2006) Pharmacogenetics of chronic cardiovascular drugs: applications and implications. Expert Opin. Pharmacother., 7, 1417–1427.
80. Gupta, S., Jain, S., Brahmachari, S. K. and Kukreti, R. (2006) Pharmacogenomics: a path to predictive medicine for schizophrenia. Pharmacogenomics, 7, 31–47.
81. Thorn, C. F., Klein, T. E. and Altman, R. B. (2010) Pharmacogenomics and bioinformatics: PharmGKB. Pharmacogenomics, 11, 501–505.
82. McDonagh, E. M., Whirl-Carrillo, M., Garten, Y., Altman, R. B. and Klein, T. E. (2011) From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. Biomark. Med., 5, 795–806.
83. Lunshof, J. E., Bobe, J., Aach, J., Angrist, M., Thakuria, J. V., Vorhaus, D. B., Hoehe, M. R. and Church, G. M. (2010) Personal genomes in progress: from the human genome project to the personal

genome project. Dialogues Clin. Neurosci., 12, 47–60.
84. Ball, M. P., Thakuria, J. V., Zaranek, A. W., Clegg, T., Rosenbaum, A. M., Wu, X., Angrist, M., Bhak, J., Bobe, J., Callow, M. J., et al. (2012) A public resource facilitating clinical use of genomes. Proc. Natl. Acad. Sci. USA, 109, 11920–11927.
85. Church, G. M. (2005) The personal genome project. Mol. Syst. Biol., 1, 2005.0030.
86. Jones, B. (2012) Genomics: personal genome project. Nat. Rev. Genet., 13, 599.
87. Clark, T. A., Sugnet, C. W. and Ares, M. Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science, 296, 907–910.
88. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science, 308, 1149–1154.
89. Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004) Global identification of human transcribed sequences with genome tiling arrays. Science, 306, 2242–2246.
90. Yamada, K., Lim, J., Dale, J. M., Chen, H., Shinn, P., Palm, C. J., Southwick, A. M., Wu, H. C., Kim, C., Nguyen, M., et al. (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science, 302, 842–846.
91. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W. and Steinmetz, L. M. (2006) A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. USA, 103, 5320–5325.
92. Okoniewski, M. J. and Miller, C. J. (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics, 7, 276.
93. Royce, T. E., Rozowsky, J. S. and Gerstein, M. B. (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic Acids Res., 35, e99.
94. Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet., 10, 57–63.
95. Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood,V., Goodhead, I., Penkett, C. J., Rogers, J. and Bähler, J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 453, 1239–1243.
96. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 320, 1344–1349.
97. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods, 5, 621–628.
98. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. and Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res., 18, 1509–1517.
99. Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. and Chinnaiyan, A. M. (2009) Transcriptome sequencing to detect gene fusions in cancer. Nature, 458, 97–101.
100. Mayr, C. and Bartel, D. P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell, 138, 673–684.
101. Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., Stebbings, L. A., Leroy, C., Edkins, S., Hardy, C., et al. (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet., 40, 722–729.
102. Shah, S. P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature, 486, 395–399.
103. Delahaye, N. F., Rusakiewicz, S., Martins, I., Ménard, C., Roux, S., Lyonnet, L., Paul, P., Sarabi, M., Chaput, N., Semeraro, M., et al. (2011) Alternatively spliced NKp30 isoforms affect the prognosis of gastrointestinal stromal tumors. Nat. Med., 17, 700–707.
104. Rajan, P., Elliott, D. J., Robson, C. N. and Leung, H. Y. (2009) Alternative splicing and biological heterogeneity in prostate cancer. Nat. Rev. Urol., 6, 454–460.
105. Gygi, S. P., Rochon, Y., Franza, B. R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol., 19, 1720–1730.
106. Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat. Biotechnol., 25, 117–124.
107. Cravatt, B. F., Simon, G. M. and Yates, J. R. 3rd. (2007) The biological impact of mass-spectrometry-based proteomics. Nature, 450, 991–1000.
108. Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. Nature, 422, 198–207.
109. Aebersold, R. (2003) Quantitative proteome analysis: methods and applications. J. Infect. Dis., 187, S315–S320.
110. Aebersold, R. (2003) A mass spectrometric journey into protein and proteome research. J. Am. Soc. Mass Spectrom., 14, 685–695.
111. Yates, J. R. 3rd, Gilchrist, A., Howell, K. E. and Bergeron, J. J. (2005) Proteomics of organelles and large cellular structures. Nat. Rev. Mol. Cell Biol., 6, 702–714.
112. Cox, J. and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem., 80, 273–299.
113. Mann, M. and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. Nat. Biotechnol., 21, 255–261.
114. Allmer, J. (2012) Existing bioinformatics tools for the quantitation of post-translational modifications. Amino Acids, 42, 129–138.
115. Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M. and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell. Proteomics, 10, M111.011015.
116. Ong, S. E. and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. Nat. Chem. Biol., 1, 252–262.
117. Ong, S. E. and Mann, M. (2007) Stable isotope labeling by amino acids in cell culture for quantitative proteomics. Methods Mol. Biol., 359, 37–52.
118. Ong, S. E. and Mann, M. (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat. Protoc., 1, 2650–2660.
119. Ong, S. E., Kratchmarova, I. and Mann, M. (2003) Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J. Proteome Res., 2, 173–181.
120. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino

acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics, 1, 376–386.

121. Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y. and Mann, M. (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. Nat. Protoc., 6, 147–157.
122. Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P. and Lee, K. H. (2007) 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics, 7, 3651–3660.
123. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics, 3, 1154–1169.
124. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem., 75, 1895–1904.
125. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. and Sanchez, J. C. (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal. Chem., 80, 2921–2931.
126. Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. Science, 312, 212–217.
127. Zybailov, B. L., Florens, L. and Washburn, M. P. (2007) Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. Mol. Biosyst., 3, 354–360.
128. Mueller, L. N., Rinner, O., Schmidt, A., Letarte, S., Bodenmiller, B., Brusniak, M. Y., Vitek, O., Aebersold, R. and Müller, M. (2007) SuperHirn — a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics, 7, 3470–3480.
129. May, D., Fitzgibbon, M., Liu, Y., Holzman, T., Eng, J., Kemp, C. J., Whiteaker, J., Paulovich, A. and McIntosh, M. (2007) A platform for accurate mass and time analyses of mass spectrometry data. J. Proteome Res., 6, 2685–2694.
130. Lundgren, D. H., Hwang, S. I., Wu, L. and Han, D. K. (2010) Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics, 7, 39–53.
131. Liu, H., Sadygov, R. G. and Yates, J. R. 3rd. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem., 76, 4193–4201.
132. Kusunoki, M., Tsutsumi, K., Nakayama, M., Kurokawa, T., Nakamura, T., Ogawa, H., Fukuzawa, Y., Morishita, M., Koide, T. and Miyata, T. (2007) Relationship between serum concentrations of saturated fatty acids and unsaturated fatty acids and the homeostasis model insulin resistance index in Japanese patients with type 2 diabetes mellitus. J. Med. Invest., 54, 243–247.
133. Shaffer, J. P. (2007) Controlling the false discovery rate with constraints: the Newman-Keuls test revisited. Biom. J., 49, 136–143.
134. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D. and Gygi, S. P. (2003) A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol., 21, 921–926.
135. Ahdesmäki, M., Lähdesmäki, H., Pearson, R., Huttunen, H. and Yli-Harja, O. (2005) Robust detection of periodic time series measured from biological systems. BMC Bioinformatics, 6, 117.
136. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol., 17, 994–999.
137. Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E. and Yates, J. R. 3rd. (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA, 100, 3107–3112.
138. Ning, K., Fermin, D. and Nesvizhskii, A. I. (2012) Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J. Proteome Res., 11, 2261–2271.
139. Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., Algenäs, C., Lundeberg, J., Mann, M. and Uhlen, M. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. Mol. Syst. Biol., 6, 450.
140. Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M. S., Gramolini, A. O., Morris, Q., Hallett, M. T., et al. (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. Cell, 125, 173–186.
141. Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M. and Nilsson, P. (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics, 10, 365.
142. Greenbaum, D., Jansen, R. and Gerstein, M. (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. Bioinformatics, 18, 585–596.
143. Petricoin, E. F. III, Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., et al. (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet, 359, 572–577.
144. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S. and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol., 7, 548.
145. Guo, T., Fan, L., Ng, W. H., Zhu, Y., Ho, M., Wan, W. K., Lim, K. H., Ong, W. S., Lee, S. S., Huang, S., et al. (2012) Multidimensional identification of tissue biomarkers of gastric cancer. J. Proteome Res., 11, 3405–3413.
146. Woolfson, A., Ellmark, P., Chrisp, J. S., Scott, M. A. and Christopherson, R. I. (2006) The application of CD antigen proteomics to pharmacogenomics. Pharmacogenomics, 7, 759–771.
147. Griffin, N. M. and Schnitzer, J. E. (2011) Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues. Mol. Cell. Proteomics, 10, R110.000935.
148. Suhre, K. and Gieger, C. (2012) Genetic variation in metabolic phenotypes: study designs and applications. Nat. Rev. Genet., 13, 759–769.
149. Theodoridis, G., Gika, H. G. and Wilson, I. D. (2011) Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. Mass Spectrom. Rev., 30, 884–906.
150. Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., et al. (2011) The human serum metabolome. PLoS ONE, 6, e16957.
151. Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28, 27–30.

152. Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., Zhou, Z., Han, L., Karapetyan, K., Dracheva, S., Shoemaker, B. A., et al. (2012) PubChem's BioAssay Database. Nucleic Acids Res., 40, D400–D412.
153. Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., et al. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res., 40, D742–D753.
154. Tautenhahn, R., Cho, K., Uritboonthai, W., Zhu, Z., Patti, G. J. and Siuzdak, G. (2012) An accelerated workflow for untargeted metabolomics using the METLIN database. Nat. Biotechnol., 30, 826–828.
155. Sana, T. R., Roark, J. C., Li, X., Waddell, K. and Fischer, S. M. (2008) Molecular formula and METLIN Personal Metabolite Database matching applied to the identification of compounds generated by LC/TOF-MS. J. Biomol. Tech., 19, 258–266.
156. Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R. and Siuzdak, G. (2005) METLIN: a metabolite mass spectral database. Ther. Drug Monit., 27, 747–751.
157. Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., et al. (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol., 8, R39.
158. Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., et al. (2009) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res., 37, D619–D622.
159. Matthews, L., D'Eustachio, P., Gillespie, M., Croft, D., de Bono, B., Gopinath, G., Jassal, B., Lewis, S., Schmidt, E., Vastrik, I., et al. (2007) An introduction to the reactome knowledgebase of human biological pathways and processes. Bioinformatics Primer, NCI/ Nature Pathway Interaction Database.
160. Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G. R., Wu, G. R., Matthews, L., et al. (2005) Reactome: a knowledgebase of biological pathways. Nucleic Acids Res., 33, D428–D432.
161. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res., 39, D691–D697.
162. Dumont, J., Huybrechts, I., Spinneker, A., Gottrand, F., Grammatikaki, E., Bevilacqua, N., Vyncke, K., Widhalm, K., Kafatos, A., Molnar, D., et al. (2011) FADS1 genetic variability interacts with dietary α-linolenic acid intake to affect serum non-HDL-cholesterol concentrations in European adolescents. J. Nutr., 141, 1247–1253.
163. Lu, Y., Feskens, E. J., Dollé, M. E., Imholz, S., Verschuren, W. M., Müller, M. and Boer, J. M. (2010) Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study. Am. J. Clin. Nutr., 92, 258–265.
164. Serkova, N. J. and Glunde, K. (2009) Metabolomics of cancer. Methods Mol. Biol., 520, 273–295.
165. Griffin, J. L. and Shockcor, J. P. (2004) Metabolic profiles of cancer cells. Nat. Rev. Cancer, 4, 551–561.
166. Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A. L., Kafri, R., Kirschner, M. W., Clish, C. B. and Mootha, V. K. (2012) Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science, 336, 1040–1044.
167. Newgard, C. B. (2012) Interplay between lipids and branched-chain amino acids in development of insulin resistance. Cell Metab., 15, 606–614.
168. Li, X., Gianoulis, T. A., Yip, K. Y., Gerstein, M. and Snyder, M. (2010) Extensive *in vivo* metabolite-protein interactions revealed by large-scale systematic analyses. Cell, 143, 639–650.
169. MacBeath, G. and Schreiber, S. L. (2000) Printing proteins as microarrays for high-throughput function determination. Science, 289, 1760–1763.
170. Haab, B. B., Dunham, M. J. and Brown, P. O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol., 2, RESEARCH0004.
171. Robinson, W. H., Steinman, L. and Utz, P. J. (2003) Protein arrays for autoantibody profiling and fine-specificity mapping. Proteomics, 3, 2077–2084.
172. Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B., Kamachi, M., Dean, E. J., Fournel, S., Fong, D., Genovese, M. C., de Vegvar, H. E., et al. (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat. Med., 8, 295–301.
173. Sharon, D., Chen, R. and Snyder, M. (2010) Systems biology approaches to disease marker discovery. Dis. Markers, 28, 209–224.
174. Hudson, M. E., Pozdnyakova, I., Haines, K., Mor, G. and Snyder, M. (2007) Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proc. Natl. Acad. Sci. USA, 104, 17494–17499.
175. Zhu, H., Hu, S., Jona, G., Zhu, X., Kreiswirth, N., Willey, B. M., Mazzulli, T., Liu, G., Song, Q., Chen,P., et al. (2006) Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. Proc. Natl. Acad. Sci. USA, 103, 4011–4016.
176. Winer, D. A., Winer, S., Shen, L., Wadia, P. P., Yantha, J., Paltser, G., Tsui, H., Wu, P., Davidson, M. G., Alonso, M. N., et al. (2011) B cells promote insulin resistance through modulation of T cells and production of pathogenic IgG antibodies. Nat. Med., 17, 610–617.
177. Miersch, S. and LaBaer, J. (2011) Nucleic Acid programmable protein arrays: versatile tools for array-based functional protein studies. Curr. Protoc. Protein Sci., Chapter 27, Unit27.2.
178. Sibani, S. and LaBaer, J. (2011) Immunoprofiling using NAPPA protein microarrays. Methods Mol. Biol., 723, 149–161.
179. Andresen, H. and Bier, F. F. (2009) Peptide microarrays for serum antibody diagnostics. Methods Mol. Biol., 509, 123–134.
180. Andresen, H., Grötzinger, C., Zarse, K., Kreuzer, O. J., Ehrentreich-Förster, E. and Bier, F. F. (2006) Functional peptide microarrays for specific and sensitive antibody diagnostics. Proteomics, 6, 1376–1384.
181. Wong, S. J., Demarest, V. L., Boyle, R. H., Wang, T., Ledizet, M., Kar, K., Kramer, L. D., Fikrig, E. and Koski, R. A. (2004) Detection of human anti-flavivirus antibodies with a West Nile virus recombinant antigen microsphere immunoassay. J. Clin. Microbiol., 42, 65–72.
182. Weinstock, G. M. (2012) Genomic approaches to studying the human microbiota. Nature, 489, 250–256.
183. Clemente, J. C., Ursell, L. K., Parfrey, L. W. and Knight, R. (2012) The impact of the gut microbiota on human health: an integrative view. Cell, 148, 1258–1270.
184. Grice, E. A. and Segre, J. A. (2012) The human microbiome: our second genome. Annu. Rev. Genomics Hum. Genet., 13, 151–170.
185. Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D. and Knight, R. (2012) Experimental and analytical

tools for studying the human microbiome. Nat. Rev. Genet., 13, 47–58.

186. Sonnenburg, J. L. and Fischbach, M. A. (2011) Community health care: therapeutic opportunities in the human microbiome. Sci. Transl. Med., 3, 78ps12.
187. Cho, I. and Blaser, M. J. (2012) The human microbiome: at the interface of health and disease. Nat. Rev. Genet., 13, 260–270.
188. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. and Gordon, J. I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature, 444, 1027–1031.
189. Wen, L., Ley, R. E., Volchkov, P. Y., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., Hu, C., Wong, F. S., Szot, G. L., Bluestone, J. A., et al. (2008) Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature, 455, 1109–1113.
190. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature, 490, 55–60.
191. Littman, D. R. and Pamer, E. G. (2011) Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host Microbe, 10, 311–323.
192. Pelizzola, M. and Ecker, J. R. (2011) The DNA methylome. FEBS Lett., 585, 1994–2000.
193. Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet., 13, 484–492.
194. Bock, C. (2012) Analysing and interpreting DNA methylation data. Nat. Rev. Genet., 13, 705–719.
195. Laird, P. W. (2010) Principles and challenges of genomewide DNA methylation analysis. Nat. Rev. Genet., 11, 191–203.
196. Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol., 8, e1000533.
197. Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature, 471, 68–73.
198. Green, E. D., Guyer, M. S.. and National Human Genome Research Institute. (2011) Charting a course for genomic medicine from base pairs to bedside. Nature, 470, 204–213.
199. Moch, H., Blank, P. R., Dietel, M., Elmberger, G., Kerr, K. M., Palacios, J., Penault-Llorca, F., Rossi, G. and Szucs, T. D. (2012) Personalized cancer medicine and the future of pathology. Virchows Arch., 460, 3–8.
200. Tsimberidou, A. M., Iskander, N. G., Hong, D. S., Wheler, J. J., Falchook, G. S., Fu, S., Piha-Paul, S. A., Naing, A., Janku, F., Luthra, R., et al. (2012) Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center Initiative. Clin. Cancer Res., 18, 6373–6383.
201. Parkinson, D. R., Johnson, B. E. and Sledge, G. W. (2012) Making personalized cancer medicine a reality: challenges and opportunities in the development of biomarkers and companion diagnostics. Clin. Cancer Res., 18, 619–624.
202. Modugno, F. and Edwards, R. P. (2012) Ovarian cancer: prevention, detection, and treatment of the disease and its recurrence. Molecular mechanisms and personalized medicine meeting report. Int. J. Gynecol. Cancer, 22, S45–S57.
203. Cho, S. H., Jeon, J. and Kim, S. I. (2012) Personalized medicine in breast cancer: a systematic review. J. Breast Cancer, 15, 265–272.
204. Roychowdhury, S., Iyer, M. K., Robinson, D. R., Lonigro, R. J., Wu,Y. M., Cao, X., Kalyana-Sundaram, S., Sam, L., Balbin, O. A., Quist, M. J., et al. (2011) Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci. Transl. Med., 3, 111ra121.
205. Bar-Joseph, Z., Gitter, A. and Simon, I. (2012) Studying and modelling dynamic biological processes using time-series gene expression data. Nat. Rev. Genet., 13, 552–564.
206. Dennis, G. Jr, Sherman, B. T., Hosack, D. A., Yang, J., Gao,W., Lane, H. C. and Lempicki, R. A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol., 4, P3.
207. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 27, 431–432.
208. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res., 13, 2498–2504.
209. Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics, 21, 3448–3449.
210. Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., et al. (2007) Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc., 2, 2366–2382.
211. Lam, H. Y., Pan, C., Clark, M. J., Lacroute, P., Chen, R., Haraksingh, R., O'Huallachain, M., Gerstein, M. B., Kidd, J. M., Bustamante, C. D., et al. (2012) Detecting and annotating genetic variations using the HugeSeq pipeline. Nat. Biotechnol., 30, 226–229.
212. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009) Circos: an information aesthetic for comparative genomics. Genome Res., 19, 1639–1645.
213. Dorogovtsev, S. N., Goltsev, A. V. and Mendes, J. F. F. (2008) Critical phenomena in complex networks. Rev. Mod. Phys., 80, 1275–1335.
214. Albert, R. and Barabasi, A. L. (2002) Statistical mechanics of complex networks. Rev. Mod. Phys., 74, 47–97.
215. Alon, U. (2003) Biological networks: the tinkerer as an engineer. Science, 301, 1866–1867.
216. Costa, L. F., Rodrigues, F. A. and Cristino, A. S. (2008) Complex networks: the key to systems biology. Genet. Mol. Biol., 31, 591–601.
217. Levy, E. D. and Pereira-Leal, J. B. (2008) Evolution and dynamics of protein interactions and networks. Curr. Opin. Struct. Biol., 18, 349–357.
218. Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L. and Nolan, G. P. (2011) Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology. Nat. Rev. Genet., 12, 224.
219. Trelles, O., Prins, P., Snir, M. and Jansen, R. C. (2011) Big data, but are we ready? Nat. Rev. Genet., 12, 224.
220. Biesecker, L. G. (2012) Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. Genet. Med., 14, 393–398.
221. Li, R., Li, Y., Kristiansen, K. and Wang, J. (2008) SOAP: short oligonucleotide alignment program. Bioinformatics, 24, 713–714.
222. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754–1760.
223. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for

analyzing next-generation DNA sequencing data. Genome Res., 20, 1297–1303.
224. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 2078–2079.
225. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res., 38, e164.
226. Ng, P. C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res., 31, 3812–3814.
227. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. and Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. Nat. Methods, 7, 248–249.
228. Flanagan, S. E., Patch, A. M. and Ellard, S. (2010) Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet. Test. Mol. Biomarkers, 14, 533–537.
229. Abyzov, A., Urban, A. E., Snyder, M. and Gerstein, M. (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res., 21, 974–984.
230. Wang, L. Y., Abyzov, A., Korbel, J. O., Snyder, M. and Gerstein, M. (2009) MSB: a mean-shift-based approach for the analysis of structural variation in the genome. Genome Res., 19, 106–117.
231. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics, 25, 2865–2871.
232. Lam, H. Y., Mu, X. J., Stütz, A. M., Tanzer, A., Cayting, P. D., Snyder, M., Kim, P. M., Korbel, J. O. and Gerstein, M. B. (2010) Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nat. Biotechnol., 28, 47–55.
233. Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V. and Korbel, J. O. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics, 28, i333–i339.
234. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol., 5, R80.
235. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol., 10, R25.
236. Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods, 9, 357–359.
237. Langmead, B. (2010) Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics, Chapter 11, Unit 11.7.
238. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc., 7, 562–578.
239. Trapnell, C., Pachter, L. and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25, 1105–1111.
240. Roberts, A., Pimentel, H., Trapnell, C. and Pachter, L. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics, 27, 2325–2329.
241. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan,G., van Baren, M. J., Salzberg, S. L., Wold, B. J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol., 28, 511–515.
242. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. and Mesirov, J. P. (2006) GenePattern 2.0. Nat. Genet., 38, 500–501.
243. Kuehn, H., Liberzon, A., Reich, M. and Mesirov, J. P. (2008) Using GenePattern for gene expression analysis. Curr. Protoc. Bioinformatics, Chapter 7, Unit 7.12.
244. Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M. J., Gnirke, A., Nusbaum, C., et al. (2010) *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol., 28, 503–510.
245. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol., 11, R106.
246. Li, J. W., Schmieder, R., Ward, R. M., Delenick, J., Olivares, E. C. and Mittelman, D. (2012) SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics, 28, 1272–1273.
247. Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W. H., Rompp, A., Neumann, S., Pizarro, A. D., et al. (2011) mzML — a community standard for mass spectrometry data. Mol. Cell. Proteomics, 10, R110.000133.
248. Deutsch, E. W. (2010) Mass spectrometer output file format mzML. Methods Mol. Biol., 604, 319–331.
249. Deutsch, E. (2008) mzML: a single, unifying data format for mass spectrometer output. Proteomics, 8, 2776–2777.
250. Kessner, D., Chambers, M., Burke, R., Agus, D. and Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics, 24, 2534–2536.
251. Craig, R. and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics, 20, 1466–1467.
252. Eng, J. K., McCormack, A. L. and Yates, J. R. III. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom., 5, 976–989.
253. Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 20, 3551–3567.
254. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. and Bryant, S. H. (2004) Open mass spectrometry search algorithm. J. Proteome Res., 3, 958–964.
255. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res., 2, 43–50.
256. Elias, J. E., Gibbons, F. D., King, O. D., Roth, F. P. and Gygi, S. P. (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nat. Biotechnol., 22, 214–219.
257. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A. and Ma, B. (2012) PEAKS DB: *de novo* sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics, 11, M111.010587.
258. Pedrioli, P. G. (2010) Trans-proteomic pipeline: a pipeline for proteomic analysis. Methods Mol. Biol., 604, 213–238.

259. Keller, A. and Shteynberg, D. (2011) Software pipeline and data analysis for MS/MS proteomics: the trans-proteomic pipeline. Methods Mol. Biol., 694, 169–189.
260. Deutsch, E. W., Shteynberg, D., Lam, H., Sun, Z., Eng, J. K., Carapito, C., von Haller, P. D., Tasman, N., Mendoza, L., Farrah, T., et al. (2010) Trans-Proteomic Pipeline supports and improves analysis of electron transfer dissociation data sets. Proteomics, 10, 1190–1195.
261. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., et al. (2010) A guided tour of the Trans-Proteomic Pipeline. Proteomics, 10, 1150–1159.
262. Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., Zerck, A., Reinert, K., et al. (2008) OpenMS — an open-source software framework for mass spectrometry. BMC Bioinformatics, 9, 163.
263. Kohlbacher, O., Reinert, K., Gröpl, C., Lange, E., Pfeifer, N., Schulz-Trieglaff, O. and Sturm, M. (2007) TOPP — the OpenMS proteomics pipeline. Bioinformatics, 23, e191–e197.
264. Bertsch, A., Gröpl, C., Reinert, K. and Kohlbacher, O. (2011) OpenMS and TOPP: open source software for LC-MS data analysis. Methods Mol. Biol., 696, 353–367.
265. Tautenhahn, R., Patti, G. J., Kalisiak, E., Miyamoto, T., Schmidt, M., Lo, F. Y., McBee, J., Baliga, N. S. and Siuzdak, G. (2011) metaXCMS: second-order analysis of untargeted metabolomics data. Anal. Chem., 83, 696–700.
266. Tautenhahn, R., Patti, G. J., Rinehart, D. and Siuzdak, G. (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. Anal. Chem., 84, 5035–5039.
267. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
268. Pluskal, T., Castillo, S., Villar-Briones, A. and Oresic, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics, 11, 395.
269. Katajamaa, M., Miettinen, J. and Oresic, M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics, 22, 634–636.
270. Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., et al. (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res., 36, D623–D631.
271. Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res., 38, D473–D479.
272. Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4, 44–57.