

# Individualized Medicine from Prewomb to Tomb

Eric J. Topol<sup>1,\*</sup>

<sup>1</sup>The Scripps Translational Science Institute, The Scripps Research Institute and Scripps Health, La Jolla, CA 92037, USA

\*Correspondence: [etopol@scripps.edu](mailto:etopol@scripps.edu)

<http://dx.doi.org/10.1016/j.cell.2014.02.012>

That each of us is truly biologically unique, extending to even monozygotic, “identical” twins, is not fully appreciated. Now that it is possible to perform a comprehensive “omic” assessment of an individual, including one’s DNA and RNA sequence and at least some characterization of one’s proteome, metabolome, microbiome, autoantibodies, and epigenome, it has become abundantly clear that each of us has truly one-of-a-kind biological content. Well beyond the allure of the matchless fingerprint or snowflake concept, these singular, individual data and information set up a remarkable and unprecedented opportunity to improve medical treatment and develop preventive strategies to preserve health.

## From Digital to Biological to Individualized Medicine

In 2010, Eric Schmidt of Google said “The power of individual targeting—the technology will be so good it will be very hard for people to watch or consume something that has not in some sense been tailored for them” (Jenkins, 2010). Although referring to the capability of digital technology, we have now reached a time of convergence of the digital and biologic domains. It has been well established that 0 and 1 are interchangeable with A, C, T, and G in books and Shakespeare sonnets and that DNA may represent the ultimate data storage system (Church et al., 2012; Goldman et al., 2013b). Biological transistors, also known as genetic logic gates, have now been developed that make a computer from a living cell (Bonnet et al., 2013). The convergence of biology and technology was further captured by one of the protagonists of the digital era, Steve Jobs, who said “I think the biggest innovations of the 21<sup>st</sup> century will be at the intersection of biology and technology. A new era is beginning” (Issacson, 2011).

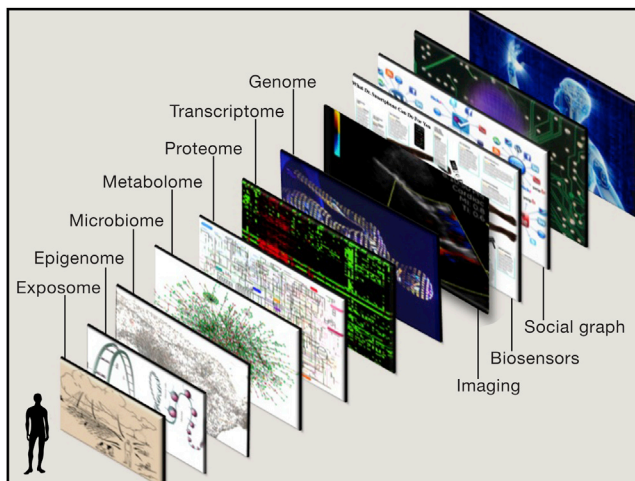
With whole-genome DNA sequencing and a variety of omic technologies to define aspects of each individual’s biology at many different levels, we have indeed embarked on a new era of medicine. The term “personalized medicine” has been used for many years but has engendered considerable confusion. A recent survey indicated that only 4% of the public understand what the term is intended to mean (Stanton, 2013), and the hackneyed, commercial use of “personalized” makes many people think that this refers to a concierge service of medical care. Whereas “person” refers to a human being, “personalized” can mean anything from having monogrammed stationary or luggage to ascribing personal qualities. Therefore, it was not surprising that a committee representing the National Academy of Sciences proposed using the term “precision medicine” as defined by “tailoring of medical treatment to the individual characteristics of each patient” (National Research Council, 2011). Although the term “precision” denotes the objective of exactness, ironically, it too can be viewed as ambiguous in this context

because it does not capture the sense that the information is derived from the individual. For example, many laboratory tests could be made more precise by assay methodology, and treatments could be made more precise by avoiding side effects—without having anything to do with a specific individual. Other terms that have been suggested include genomic, digital, and stratified medicine, but all of these have a similar problem or appear to be too narrowly focused.

The definition of individual is a single human being, derived from the Latin word *individu*, or indivisible. I propose individualized medicine as the preferred term because it has a useful double entendre. It relates not only to medicine that is particularized to a human being but also the future impact of digital technology on individuals driving their health care. There will increasingly be the flow of one’s biologic data and relevant medical information directly to the individual. Be it a genome sequence on a tablet or the results of a biosensor for blood pressure or another physiologic metric displayed on a smartphone, the digital convergence with biology will definitively anchor the individual as a source of salient data, the conduit of information flow, and a—if not the—principal driver of medicine in the future.

## The Human GIS

Perhaps the most commonly used geographic information systems (GIS) are Google maps, which provide a layered approach to data visualization, such as viewing a location via satellite overlaid with street names, landmarks, and real-time traffic data. This GIS exemplifies the concept of gathering and transforming large bodies of data to provide exquisite temporal and location information. With the multiple virtual views, it gives one the sense of physically being on site. Although Google has digitized and thus created a GIS for the Earth, it is now possible to digitize a human being. As shown in Figure 1, there are multiple layers of data that can now be obtained for any individual. This includes data from biosensors, scanners, electronic medical records, social media, and the various omics that include



**Figure 1. Geographic Information System of a Human Being**

The ability to digitize the medical essence of a human being is predicated on the integration of multiscale data, akin to a Google map, which consists of superimposed layers of data such as street, traffic, and satellite views. For a human being, these layers include demographics and the social graph, biosensors to capture the individual's physiome, imaging to depict the anatomy (often along with physiologic data), and the biology from the various omics (genome-DNA sequence, transcriptome, proteome, metabolome, microbiome, and epigenome). In addition to all of these layers, there is one's important environmental exposure data, known as the "exposome."

DNA sequence, transcriptome, proteome, metabolome, epigenome, microbiome, and exposome. Going forward, I will use the term "panoromic" to denote the multiple biologic omic technologies. This term closely resembles and is adopted from panoramic, which refers to a wide-angle view or comprehensive representation across multiple applications and repositories. Or more simply, according to the Merriam-Webster definition of panoramic, it "includes a lot of information and covers many topics." Thus the term panoromic may be well suited for portraying the concept of big biological data.

The first individual who had a human GIS-like construct was Michael Snyder. Not only was his whole genome sequenced, he also collected serial gene expression, autoantibody, proteomic, and metabolomic (Chen et al., 2012) samples. A portion of the data deluge that was generated is represented in the Circos plot of Figure 2 or an adoption of the London Tube map (Shendure and Lieberman Aiden, 2012). The integrated personal omics profiling (iPOP) or "Snyderome," as it became known, proved to be useful for connecting viral infections to markedly elevated glucose levels. With this integrated analysis in hand, Michael Snyder changed his lifestyle, eventually restoring normal glucose homeostasis. Since that report in 2012, Snyder and his team have proceeded to obtain further omic data, including whole-genome DNA methylation data at multiple time points, serial microbiome (gut, urine, nasal, skin, and tongue) sampling, and the use of biosensors for activity tracking and heart rhythm. Snyder also discovered that several extended family members had smoldering, unrecognized glucose intolerance, thereby changing medical care for multiple individuals.

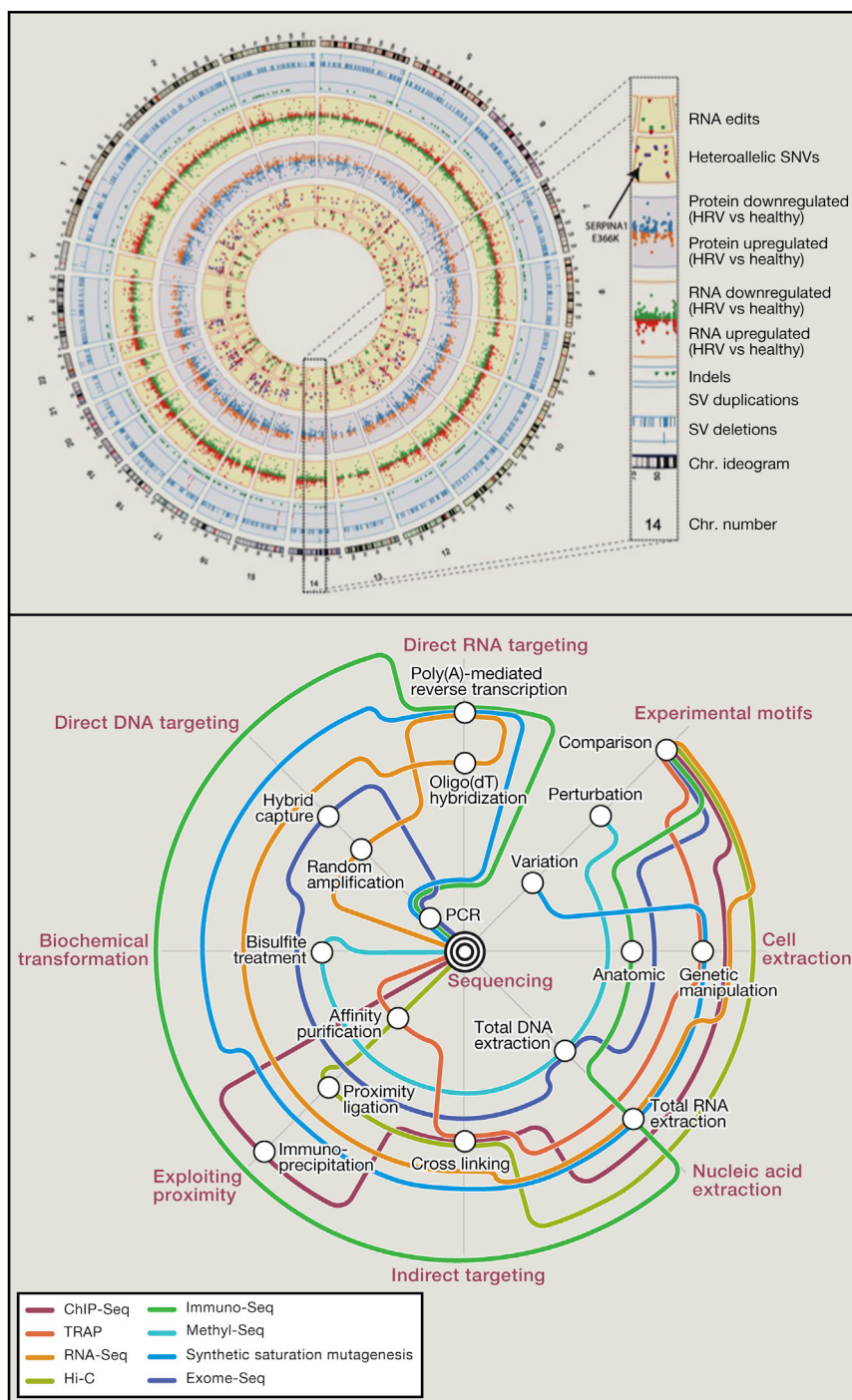
Of note, to obtain the data and process this first panoromic study, it required an armada of 40 experienced coauthors and

countless hours of bioinformatics and analytical work. To give context to the digital data burden, it took 1 terabyte (TB) for DNA sequence, 2 TB for the epigenomic data, 0.7 TB for the transcriptome, and 3 TB for the microbiome. Accordingly, this first human GIS can be considered a remarkable academic feat and yielded key diagnostic medical information for the individual. But, it can hardly be considered practical or scalable at this juncture. With the cost of storing information continuing to drop substantially, the bottleneck for scalability will likely be automating the analysis. On the other hand, each omic technology can readily be undertaken now and has the potential of providing meaningful medical information for an individual.

### The Omic Tools

#### Whole-Genome and Exome Sequencing

Perhaps the greatest technologic achievement in the biomedical domain has been the extraordinary progress in our ability to sequence a human genome over the past decade. Far exceeding the pace of Moore's Law for the relentless improvement in transistor capacity, there has been a  $>4$  log order (or  $0.00007^{\text{th}}$ ) reduction in cost of sequencing (Butte, 2013), with a cost in 2004 of ~\$28.8 million compared with the cost as low as \$1,000 in 2014 (Hayden, 2014). However, despite this incomparable progress, there are still major limitations to how rapid, accurate, and complete sequencing can be accomplished. High-throughput sequencing involves chopping the DNA into small fragments, which are then amplified by PCR. Currently, it takes 3 to 4 days in our lab to do the sample preparation and sequencing at 30x to 40x coverage of a human genome. The read length of the fragments is now ~250 base pairs for the most cost-effective sequencing methods, but this is still suboptimal in determining maternal versus paternal alleles, or what is known as phasing. Because so much of understanding diseases involves compound heterozygote mutations, *cis*-acting sequence variant combinations, and allele-specific effects, phasing the diploid genome, or what we have called "diplomics" (Tewhey et al., 2011), is quite important. Recently, Molecuola introduced a method for synthetically stitching together DNA sequencing reads yielding fragments as long as 10,000 base pairs. These synthetic long reads are well suited for phasing. Unfortunately, the term "whole-genome sequencing" is far from complete because ~900 genes, or 3%–4% of the genome, are not accessible (Marx, 2013). These regions are typically in centromeres or telomeres. Other technical issues that detract from accuracy include long sequences of repeated bases (homopolymers) and regions rich in guanine and cytosine. Furthermore, the accuracy for medical grade sequencing still needs to be improved. A missed call rate of 1 in 10,000, which may not seem high, translates into a substantial number of errors when considering the 6 billion bases in a diploid genome. These errors obfuscate rare but potentially functional variants. Beyond this issue, the accurate determination of insertions, deletions, and structural variants is impaired, in part due to the relatively short reads that are typically obtained. The Clinical Sequencing Exploratory Research (CSER) program at the National Institutes of Health is aimed at improving the accuracy of sequencing for medical applications (National Human Genome Research Institute, 2013).



**Figure 2. Plots of Panoramic Information**

Top: Circos plot of the Snyder genome. From outer to inner rings: chromosome ideogram; genomic data (pale blue ring), structural variants >50 bp (deletions [blue tiles], duplications [red tiles]), indels (green triangles); transcriptomic data (yellow ring), expression ratio of viral infection to healthy states; proteomic data (light purple ring), ratio of protein levels during human rhinovirus (HRV) infection to healthy states; transcriptomic data (yellow ring), differential heteroallelic expression ratio of alternative allele to reference allele for missense and synonymous variants (purple dots) and candidate RNA missense and synonymous edits (red triangles, purple dots, orange triangles, and green dots, respectively). From [Chen et al. \(2012\)](#) with permission.

Bottom: Adopted London Tube model of integrated omics from Shendure and Aiden Integration of the many applications of next-generation DNA sequencing, which include sites of DNA methylation (methyl-seq), protein-DNA interactions (ChIP-seq), 3D genome structure (Hi-C), genetically targeted purification of polysomal mRNAs (TRAP), the B cell and T cell repertoires (immuno-seq), and functional consequences of genetic variation (synthetic saturation mutagenesis) with a small set of core techniques, represented as open circles of “stations.” Like subway lines, individual sequencing experiments move from station to station until they ultimately arrive at a common terminal—DNA sequencing. From [Shendure and Lieberman Aiden \(2012\)](#) with permission.

From [Shendure and Lieberman Aiden \(2012\)](#) with permission.

have seen a remarkable surge of definition of their genomic underpinnings ([Boycott et al., 2013](#)). In the first half of 2010, the basis for four rare diseases was published, but in the first half of 2012, that number jumped to 68 ([Boycott et al., 2013](#)). With the power of sequencing, it is anticipated that the molecular basis for most of the 7,000 known Mendelian diseases will be unraveled in the next few years.

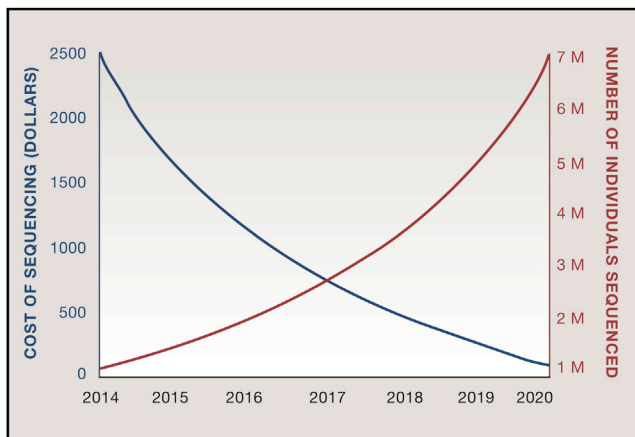
### The 1.5% versus 98.5% Genome Sequencing Dilemma

The exome consists of only ~40 Mb, or 1.5% of the human genome. There is continued debate over the use of whole-exome sequencing compared with whole-genome sequencing, given the lower cost of sequencing an exome, that can be readily captured via kits

Despite these shortcomings, the ability to identify rare or low-frequency variants that are pathogenic has been a major outgrowth of high-throughput sequencing. Well beyond the genome scans and genome-wide association studies that identified common variants associated with most complex, polygenic diseases and human traits, sequencing leads to high definition of the uncommon variants that typically have much higher penetrance. For example, rare Mendelian conditions

from a few different companies (Agilent SureSelect, Illumina TruSeq, and Roche NimbleGen). Exome sequencing is typically performed at much deeper coverage, >100× (as compared with 30×–40× for whole genome), which enhances accuracy, and the interpretation of variants that affect coding elements is far more advanced compared with the rest of the genome. However, the collective output from genome-wide association studies of complex traits has indicated that ~80% of the





**Figure 3. Hypothetical Plot of Cost of Sequencing and Number of Individuals Sequenced over the Next 6 Years**

As of early 2014, <100,000 individuals have had whole-genome sequencing, leaving the information difficult to fully interpret (of limited informativeness or value). When millions of people undergo sequencing, with the full gamut of diverse phenotypes and ancestries, and the cost for sequencing continues to drop, a virtuous cycle of informativeness is established. With the new capability in 2014 to have whole-genome sequencing at a cost of \$1,000, along with extremely high throughput, it is likely that millions of individuals will be sequenced in the next 3–4 years. The cost of sequencing will continue to drop throughout this time, as increasing numbers of individuals undergo sequencing. Projections suggest that at least 20,000 individuals with each phenotype may be necessary to reliably identify rare, functional genomic variants. Accordingly, once millions of individuals across all main phenotypes and ancestries are sequenced, there is a new set point, or threshold, of informativeness.

incriminated loci are in noncoding regions, outside the confines of genes (Koboldt et al., 2013). It is fair to say that we have long underestimated the importance of the rest of the genome, but its high density of key regulatory features provides intricate and extraordinarily tight control over how genes operate. Recent whole-genome sequencing studies have identified many critical variants in noncoding portions of the genome (Khurana et al., 2013). A typical whole human genome sequence will contain ~3.5 million variants compared with the reference genome, predominantly composed of single-nucleotide polymorphisms but also including insertion-deletions, copy number variants, and other types of structural variants (Frazer et al., 2009). Today, analysis of most of the 3.5 million variants is left with the “variant of unknown significant” (VUS) diagnosis. As more people get sequenced with the full range of disease phenotypes, the proportion of VUS will drop, and each sequence will become more informative. Figure 3 provides a theoretical plot of how further reduction of the cost of whole-genome sequencing will also be accompanied by large numbers of individuals undergoing sequencing. In 2014, still well under 100,000 people have had whole-genome sequencing with only a very limited number of phenotypes addressed. At some point in the future, sequence data get progressively more informative at a lower price point, thus establishing particular value of whole-genome sequencing. It is not just about getting a large number of people with diverse medical conditions and diverse ancestries sequenced. The drive to informativeness will clearly be enhanced by incorporating family genomic assessment, especially for determining whether

rare variants are meaningful. Here, too much focus on the individual can result in a loss of context, back to our analogy of the Google map of maximal zoom obscuring understanding. By anchoring the genomics of family members, such as was done with the important discovery of PCSK9 rare variants (Hall, 2013) in cholesterol metabolism, progress in genomic medicine will be catalyzed.

At this juncture, however, it appears that exome and whole-genome sequencing provide complementary information. As the cost of whole-genome sequencing is further reduced, along with the availability of enhanced analytical tools for the nongene 98.5% content interpretation, exome sequencing may ultimately become obsolete.

### Single-Cell Sequencing

The ability to perform sequencing of individual cells has provided remarkable new insights about human biology and disease (Shapiro et al., 2013; Battich et al., 2013; Owens, 2012). The unexpected heterogeneity in DNA sequence from one cell to another, such as has been well documented in tumor tissue and even in somatic cells in healthy individuals, has enlightened us about intraindividual genomic variation. The concept of “mosaicism” has gained rapid acceptance—with multiple mechanisms—ranging from gamete formation, embryonic development, to somatic mutation in cells in adulthood, that account for why each of us has cells with different DNA sequences (Lupski, 2013; Wang et al., 2012; Macosko and McCarroll, 2012; Poduri et al., 2013). It remains unclear whether mosaicism has functional significance beyond being tied to certain congenital conditions and cancers, but this is an active area of research that is capitalizing on single-cell sequencing technology. This is especially the case in neuroscience in order to explain the observed frequent finding of transposons, which appear to involve between 80 and 300 unique insertions for each neuron and are potentially associated with neurologic diseases (Poduri et al., 2013).

Sperm, which tend to swim solo, are particularly well suited for single-cell genomics. This work has quantified recombination rates of ~25 events per sperm, identified the hot spots where these events are most likely to occur, and determined genomic instability as reflected by the rate of de novo mutations (Wang et al., 2012; Poduri et al., 2013). Such de novo mutations, which increase in sperm with paternal age, are associated with autism, schizophrenia, and intellectual disability (Poduri et al., 2013; Kong et al., 2012; Veltman and Brunner, 2012; de Ligt et al., 2012).

Intriguing, and possibly revolutionary, single-cell methods using in situ sequencing protocols are set to offer precise spatial information in addition to linear sequence data. In situ sequencing holds the potential to resolve the spatial distribution of copy number variants, circular DNA, tumor heterogeneity, and RNA localization. A number of methods have been published in the last year, and progress is likely to accelerate in the near future.

### Transcriptomics, Proteomics, and Metabolomics

As opposed to the DNA sequence, which is relatively static, RNA reflects the dynamic state of the cell. Gene expression of a particular tissue of the whole genome has been available via microarrays for several years, but RNA sequencing (RNA-seq)

is a relatively new tool that transcends simple expression by capturing data on gene fusions, alternative spliced transcripts, and posttranscriptional changes, along with the whole gamut of RNAs (including microRNA [e.g., miRNAseq], small RNA, lincRNA, ribosomal RNA, and transfer RNA).

A particularly valuable metric related to RNA is the expression quantitative trait locus (eQTL). By having both genome-wide association study (GWAS) data and whole-genome gene expression at baseline with or without particular stimuli, functional genomic assessment has been enabled. For example, [Westra et al. \(2013\)](#) used eQTLs and loci derived from GWAS to provide functional genomic, mechanistic insights for multiple complex traits, including lupus and type 1 diabetes.

The proteome, metabolome, and autoantibody landscape can be assessed for an individual approaching the whole-genome level via recent advances in mass spectrometry and protein arrays. Using these techniques, posttranslational modifications of proteins, protein-protein interactions, or the small-molecule metabolites produced by these proteins can be revealed. Emerging technologies such as RNA-mediated oligonucleotide annealing, selection, and ligation sequencing (RASL-seq), barcoded small hairpin RNA (shRNA) libraries, and combinatorial antibody libraries provide inexpensive and efficient views of biology. Longer read sequencing provides the opportunity to sequence antibodies, which typically have variable and constant regions composed of ~2,000 nucleotides.

### **Microbiome**

Perhaps no area of biology has received more attention in recent years than the microbiome. Just the gut microbiome has orders of magnitude more DNA content than germline human DNA and has markedly heightened diversity. Our commensal bacterial flora has been shown to play an important role in various medical conditions ([Cho and Blaser, 2012](#)). From fecal samples using a 16S ribosomal amplicon sequencing method, the gut microbiome has been the subject of intensive prospective clinical assessment. It was determined that there were three major enterotypes of the intestinal microbiome based on the predominant bacterial species, such as *Bacteroides*, *Ruminococcus*, or *Prevotella* ([Arumugam et al., 2011](#)). The resident species appear to be quite stable over an extended period of time and to be initially transmitted via the mother at childbirth ([Faith et al., 2013](#)). As the interface between genomics and the host's environment, the microbiome clearly plays a pivotal role in defining each individual. The influence of the diet on the gut microbiome, such as the content of fiber, along with the underpinning of malnutrition, has been documented ([Gordon et al., 2012](#); [Ridaura et al., 2013](#)). For example, even an individual's response to medications, such as digoxin ([Haiser et al., 2013](#)), or multiple drugs used for cancer, has been shown to be linked to the bacterial flora of the gut microbiome ([Viaud et al., 2013](#); [Iida et al., 2013](#)).

### **Epigenome**

There has been extraordinary progress in our ability to map the human epigenome from DNA methylation to histone modifications and chromatin structure ([Ziller et al., 2013](#); [Rivera and Ren, 2013](#)). The prolific ENCODE project has provided troves of data detailing the role of regulatory elements such as enhancers and insulators and how they are tied to DNA methylation

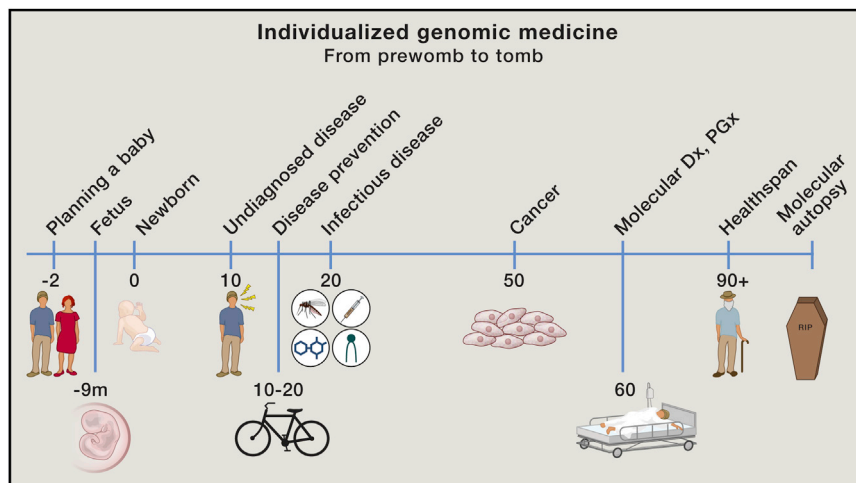
and histone changes ([Dawson and Kouzarides, 2012](#)). Like gene expression, epigenomic findings are highly cell-type specific, with more than 200 different cell types in the human body. For methylation, whole-genome bisulphite sequencing has recently been performed for 30 diverse human cell types ([Rivera and Ren, 2013](#)). Epigenomic reprogramming has a clear-cut role in cancer, be it via transcription factors or chromatin regulators ([Dawson and Kouzarides, 2012](#); [Suvà et al., 2013](#)). Although access to tissue to define epigenomic signatures is a limiting factor outside of the cancer space, it is apparent that many other diseases are affected by epigenomic dynamics, such as complications of diabetes, rheumatoid arthritis, or hypertension ([Pirola et al., 2010](#); [Liu et al., 2013](#); [Fratkin et al., 2012](#)). Furthermore, epigenomic changes affect susceptibility to diseases, as has been shown with open chromatin related to the TCF7L2 gene ([Groop, 2010](#)) and parental origin of sequence variants for Type 2 diabetes mellitus and breast and prostate cancer ([Kong et al., 2009](#)). This parent-of-origin issue may be tied to transgenerational epigenomic instability, as has been well documented in plants, and is certainly a key element of human biology and heritability ([Schmitz et al., 2011](#)).

### **Physiome and Exposome**

Understanding and quantifying an individual's physiology and environmental interactions are crucial to digitizing a human being. Through wearable biosensors and smartphones, this has become eminently practical. Continuous tracking is now obtainable for most key physiologic metrics, including blood pressure, heart rhythm, glucose, blood oxygen saturation, brain waves, intraocular eye pressure, and lung function indices. Similarly, there are environmental sensors that connect with smartphones to quantify such indices as air pollution, pollen count, radiation, water quality, ambient humidity, electromagnetic fields, and the presence of pesticides in food.

### **Bioinformatics**

Fundamental to individualized medicine is the ability to analyze the immense data sets and to extract all of the useful, salient information. This is exemplified by the task of sifting through a trio of whole-genome sequences to find a causative mutation in a proband with an undiagnosed disease. Typically, this translates to finding one critical nucleotide variant out of well over one to two million single base variants and simplifying the analysis by only considering variants that change amino acid sequence or lead to obvious splicing defects ([Maher, 2011](#)). Identifying the signal from the noise, with the vast majority of variants categorized as "unknown significance" (VUS), is the crux of the challenge. Moreover, the tools to assess structural variants and indels are not as extensively developed and validated. So there are considerably more data that come from the sequencer for an individual than can be fully and accurately mined. Beyond this, there is the need for better integration of the multiple GIS layers, such as panoramic and biosensor data, and the ability to provide an integrative multiscale approach to an individual's data set. Although not the comprehensive multilayer as depicted in [Figure 2](#), [Zhang et al. \(2013\)](#) recently used an integrated systems approach, including omics of both human and mice brains to discover genetic networks in Alzheimer's disease. Although in the past there were generally insufficient efforts to understand epistasis, gene-gene interactions have been



**Figure 4. Timeline of Sequencing Applications in Medicine from Prewomb to Tomb**

The medical application of genomics is relevant to many points during an individual's lifespan. Prior to conception, a couple can have genomic screening for important recessive alleles. An expectant mother, at 8–12 weeks of pregnancy, can now have single tube of blood used to accurately assess chromosomal abnormalities of the fetus, determine gender, and even have whole-genome sequencing of the fetus performed. At birth, sequencing the genome of the newborn can be used to rapidly diagnosis many critical conditions for which a time delay, which frequently can occur with the present heel stick screening methods, might lead to irrevocable damage. The molecular basis for serious, undiagnosed conditions can often be established by sequencing the individual with parents of siblings. Ultimately, omic information at a young age will be useful by providing susceptibility to various medical conditions that have actionable prevention strategies. Sequencing can be done to define a pathogen for

more rapid and accurate approaches to infectious diseases. The driver mutations and key biologic underpinning pathways of an individual's cancer can frequently be pinpointed by omics. The root causes of common polygenic conditions such as diabetes or coronary heart disease may ultimately be defined at the individual level. Specific sequence variants of germline DNA or the gut microbiome have relevance for response to prescription medication (both efficacy and safety). Defining the genomics of healthspan, rather than the traditional focus on diseases, may prove to be especially worthwhile to understand protective alleles and modifier genes. For an individual with sudden death, a molecular autopsy via sequencing can be performed, along with family survivors, to determine the cause of death and potentially prevent untimely or avoidable deaths of members of the family and subsequent generations.

upstaged by the complexity of a higher-order bioinformatics challenge.

### How the Omic Tools Reboot Medicine A Prewomb-to-Tomb Assessment

At many points in the span of a lifetime, the unique biology of the individual will play an increasing role. As depicted in the timeline of Figure 4, I will go through each topic sequentially.

#### Preconception

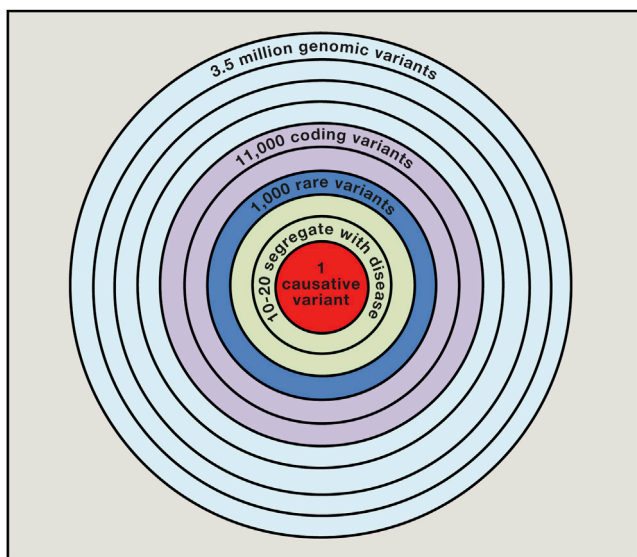
The ability to determine carrier mutations for each prospective parent has been greatly enhanced through multiple direct-to-consumer sources, including 23andMe, Counsyl, GenePeeks, and Good Start Genetics. This can be considered the ultimate form of prevention for major recessive conditions and has only received modest attention to date. Counsyl screens for more than 100 recessive Mendelian traits, and 23andMe screens 50 carrier conditions. The carrier rates for many serious conditions are higher than most people would suspect, such as 1 in 35 for spinal muscular atrophy, 1 in 40 for cystic fibrosis, and 1 in 125 individuals for Fragile X syndrome (Test, 2013). GenePeeks uses carrier data from ~100,000 DNA sequence variants for each prospective parent to perform a computer simulation of 10,000 "digital babies," determining the probabilistic odds of significant Mendelian disorders (Couzin-Frankel, 2012). They are already using their analytic methodology to screen sperm from the Manhattan cryobank; until now, sperm banks have been completely unregulated and without genomic assessment (Almeling, 2013; Rincon, 2013). The concept of higher DNA resolution preconception screening is attractive, given that there are many more pathogenic variants in the genes that are implicated in disease, such as cystic fibrosis (~2,000 variants), than are conventionally assessed (Sosnay et al., 2013).

#### Fetal Sequencing

Whereas the diagnosis of chromosomal aberrations such as trisomy-21, 18, 13 required amniocentesis or chorionic villi sampling, there are now four different maternal blood sampling assays to accomplish the same assessment with extremely high (>99%) accuracy (Morain et al., 2013). Relying on the plasma fetal DNA present in adequate quantity from a maternal blood sample at 8–10 weeks of pregnancy, such testing has been transformative, preempting the need for amniocentesis in all but the rare exception when results are ambiguous. Here is a great example of using plasma-free DNA sequencing to avoid an invasive test that carries a small but important risk of miscarriage. However, with more than 4 million births in the United States each year, only a tiny fraction (<2%) of prenatal maternal blood sampling has yet been performed in clinical practice. Multiple groups have demonstrated the ability to do a fetal whole-exome sequence from a maternal blood sample (Fan et al., 2012) or whole-genome sequencing from both parents' DNA along with the maternal plasma-free DNA (Kitzman et al., 2012; Lo et al., 2010), but this takes a rather extensive computing and bioinformatics effort that is not presently scalable. Undoubtedly, that will be resolved over time but will engender the serious bioethical issues of what constitutes the appropriate reasons for termination of pregnancy. But, at the same time, it will afford the opportunity to make the molecular diagnoses of conditions in utero and facilitate treatment then or at the earliest time after birth.

#### Neonatal Sequencing

Monogenic diseases, many of which present in the first month of life, are a major cause of neonatal fatality and morbidity (Saunders et al., 2012). Despite routine heel sticks for blood sent out for analysis, with attendant delays of several days to weeks in obtaining results, there has not been any improvement in



**Figure 5. Legend from Whole-Genome Sequencing to Identification of a Causative Variant**

Following sequencing, alignment, and annotation, the 3.5 million variants—all genetic variants in the family (unaffected and affected individuals)—are analyzed to identify known disease variants. Then inheritance-based and population-based filters are applied. Phenotype-informed ranking and functional filters are used to then determine the root-cause variant. The timeline for this involves (1) sample preparation of 2 days, (2) sequencing of 2 days in fast output mode, the (3) preliminary analysis in ~24 hr (6 hr for variant calling, 1 hr annotation, 1 hr for each candidate variant), and (4) literature review to exclude and include genes hit by potential candidate variants—an additional ~5 days. The cost for analysis is predominantly personnel and compute time on the cloud, estimated to be ~\$300.

reducing neonatal mortality related to genetic disorders in the past 20 years (Kaiser, 2013; Sosnay et al., 2013). Now, it has been shown that whole-genome sequencing of newborns can be accomplished in <48 hr and can lead to highly actionable information for managing a neonate's condition, such as in the classic example of phenylketonuria or galactosemia, whereby irreversible damage might otherwise occur (Saunders et al., 2012; Kaiser, 2013; Sosnay et al., 2013).

#### **Undiagnosed, Idiopathic, and Rare Diseases**

The diagnosis of an XIAP mutation in a child with fulminant pancolitis with successful, curative treatment is often cited as the first case of sequencing to save an individual's life (Worthey et al., 2011). Since that report, there have been several other cases that used whole-genome or exome sequencing, along with other omic tools, for making the molecular diagnosis of idiopathic conditions (Jacob et al., 2013). For example, the National Institutes of Health Undiagnosed Disease Program uses exome sequencing to facilitate the diagnosis (Gahl and Tift, 2011). Recently, the group at Baylor College of Medicine published a series of 250 individuals, of whom 80% were pediatric and largely affected by neurologic conditions, who underwent whole-exome sequencing. In that cohort, there were many affected patients with a known Mendelian trait but without a specific root cause established. A molecular diagnosis was made in 25% of the cohort (Yang et al., 2013). At the Scripps Health and Scripps Research Institute, we have screened more

than 100 individuals for the potential of having an idiopathic disease. This requires review by a multidisciplinary physician panel to assure that a comprehensive evaluation of the patient has been performed before turning to DNA sequencing. The first of 15 individuals who we enrolled into our protocol was 16 years old and had an incapacitating neurologic condition. She and her parents were sequenced, and in Figure 5, the bioinformatics challenge of interpreting the three whole-genome sequences is presented, along with the molecular diagnosis of an ADCY5 mutation that had not been previously described. In our 15 probands at Scripps, we have had a successful molecular diagnosis (using criteria as described by the Baylor group) in 8 individuals. However, establishing the diagnosis represents only the first step of the desired strategy, as providing an effective treatment is the fundamental goal. Unfortunately, the number of individuals for whom that has been achieved is quite limited to date, but strategies using repurposing of existing drugs, drugs that were partially developed but not commercialized, or acceleration of the development of genomically guided therapies are all actively being pursued. With an estimate of at least one million individuals in the United States with a serious medical condition left without a diagnosis, such progress is encouraging (Jacob et al., 2013).

#### **Disease Prevention**

At some point in the future, it is hoped that having DNA sequence information will pave the way for prevention of an individual's predisposed conditions. To date, however, that concept has not been actualized for a few principal reasons. First, most of the complex, polygenic traits have not had much more than 10% of their heritability explained by the common variants assessed by GWAS. The "missing" or unsolved heritability detracts from the ability to assign an individual any certainty or risk or protection from a particular condition. There are some notable exceptions, such as age-related macular degeneration or type 1 diabetes mellitus, in which combinations of common and rare variants can provide a well-characterized, quantified risk profile. Second, there is the appropriate question of whether the knowledge of a risk allele is actionable. Prototypic here is the apoE4 allele, which carries an unequivocal high risk for Alzheimer's disease, yet there is no proven strategy to prevent the disease. So, even armed with known risk, there is a lack of knowledge for how to mitigate it. Third, the way data for genomic susceptibility are analyzed via a population approach makes it difficult to extrapolate such average findings to a particular individual. For example, someone may have low-frequency modifier genes for a condition at risk or unusual environmental interactions that markedly affect susceptibility. Notwithstanding these issues, as millions of individuals with diverse phenotypes undergo whole-genome sequencing (Figure 3), the ability to provide meaningful risk data will increase. Having the full GIS of each individual will further enrich the probabilistic approach of providing vulnerability data early in one's life. For example, if one's sequence data indicate a risk for hypertension, that risk may be further modulated by knowledge of his/her proteins, metabolites, microbiome, and epigenomics. The use of a biosensor watch to passively collect continuous blood pressure measurements could make diagnosis at the earliest possible time, avoiding any end-organ damage to the heart or kidney. Specific



treatments could be used that are biologically based from one's GIS. Similarly, for asthma, the panoramic information, coupled with biosensors that track air quality, pollution, forced expiratory volume, and other relevant physiologic metrics, could prove useful to prevent an attack.

A futuristic way in which genomics and biosensors will ultimately converge is through injectable nanosensors that put the blood in continuous surveillance mode (Ferguson et al., 2013). Such sensors have the ability to detect a DNA, RNA, autoantibody, or protein signal and to wirelessly transmit the signal to the individual's smartphone. This sets up the potential for detecting endothelial sloughing from an artery before a heart attack (Damani et al., 2012), plasma tumor DNA in a patient being treated for or in remission from cancer, or a child with known genomic risk of autoimmune diabetes that is developing autoantibodies to pancreatic  $\beta$ -islet cells long before there has been destruction. The blood under continuous surveillance concept highlights the potential ability to temporally detect a risk signal for a major clinical event and to implement true preventive therapy. That could be intensive antiplatelet medication to prevent a heart attack, genomic-guided treatment of cancer recurrence at the earliest possible juncture, or immunomodulation therapy for autoimmune diabetes.

### **Infectious Diseases**

Whole-genome sequencing has proven to be particularly useful for tracking pathogen outbreaks, such as for tuberculosis (Gardy et al., 2011), methicillin-resistant *Staphylococcus aureus*, antibiotic-resistant *Klebsiella pneumoniae*, and *Clostridium difficile* (Eyre et al., 2013; Harris et al., 2013; Snitkin et al., 2012). Beyond identifying the particular bacterial or viral strain that accounts for an outbreak's origin and spread, sequencing is likely to prove to be quite useful for rapid, early characterization of the cause of infection and specific, effective antibiotic therapy. For sepsis, the current standard of care is to take blood or other body fluids for culture, which typically takes 2 days to grow out. Additionally, there is at least another day required to determine sensitivities to a range of antibiotics. In the future, with lab-on-a-chip sequencing platforms that attach to or are integrated with a smartphone (Biomeme and QuantuMD), it may be feasible to do rapid sequencing of the pathogen and determination of the optimal treatment. Such a strategy would preempt the need for broad-spectrum antibiotic use, and the rapid diagnosis and targeted treatment would likely have a favorable impact on prognosis in these very high-risk, critically ill patients.

### **Cancer**

With cancer's basis in genomics, there have been extensive efforts to characterize the principal driver mutations and biologic pathways, especially through The Cancer Genome Atlas (TCGA) (Kandoth et al., 2013a, 2013b; Alexandrov et al., 2013). Our understanding of the biology of cancer has expanded exponentially and, with it, so has the appreciation for its extreme complexity. Perhaps the two classical "Hallmarks of Cancer" Reviews in *Cell*, one in 2000 and the sequel in 2011, best exemplify this (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). The diagram to explain the principal mechanisms of cancer was already exceptionally complex in 2000 and became at least a log order more intricate a decade later. In a more recent review of the cancer genome landscape, the Johns Hopkins group pro-

vided perspective for the 84 known oncogenes and 54 tumor suppressor genes that have been fully validated (Vogelstein et al., 2013). There will unquestionably be more, but estimates of the total number of genes involved in pivotal mutations may wind up being ~200. Beyond this, the principal pathways involving cell survival, cell fate, and DNA damage repair are recognized. Certain cancers have a relative low burden of mutations per megabase of the tumor genome, such as acute myelogenous leukemia (<1), whereas others are quite high, like lung adenocarcinoma or squamous cell carcinoma (~50) (Kandoth et al., 2013b; Vogelstein et al., 2013). Mutations of certain genes, such as the P53 tumor suppressor, are found in some patients with any of 12 common forms of cancer (Kandoth et al., 2013b). Our old taxonomy of cancer based upon the organ of origin may be considered inapt, for knowledge of the driver mutation(s) and pathway could be more useful for individualizing treatment. "N of 1" case reports with whole-genome sequencing have been particularly illuminating for the clonal origin of an individual's cancer (Brannon and Sawyers, 2013; Haffner et al., 2013). In the past 2 years, the Food and Drug Administration has approved almost 20 new drugs that target a specific mutation for cancer.

So, with these leaps in understanding biology and introduction of new therapies, why has there been relatively little impact in the clinic to date? One major barrier is that we do not have drugs that can target tumor suppressor genes, making up ~40% of mix of principal, driver mutations. Sometimes there are workarounds for this issue, such as the tumor suppressor gene PTEN, which results in PI3-kinase activation, but more often, this is not the case. Even for oncogenes, less than 40% have a specific drug antagonist, as most are part of protein complexes (Vogelstein et al., 2013). A second critical issue is that there is marked heterogeneity in tumors, both within an individual's primary tumor and certainly intermetastatic. This appears to be a foundation for the common occurrence of relapse after an initial marked response, reflecting success directed to an oncogene but also that other undetected mutations become capable of propagating the tumor. The BRAF mutations, which are drivers in a variety of tumors, notably melanoma, thyroid, and colon, can be treated with a specific BRAF inhibitor. In the first 2 weeks of oral therapy, there is usually a marked response, but at 9–12 months, a relapse is quite typical (Sosman et al., 2012). Interestingly, when targeting BRAF for colon cancer, there appears to be primary resistance to these inhibitors (Prahallad et al., 2012) related to EGFR expression and emphasizing that the stroma, microenvironment of the tumor can still exert an important role. The issues of heterogeneity and resistance lend credence to the use of combinations of targeted drugs in the future, but that has yet to be explored at scale in prospective trials. A third largely unaddressed issue in the clinic is the involvement of the epigenome in tumorigenesis. At least 40 epigenome regulator genes are known that have highly recurrent somatic mutations in tumors across a variety of cancers, affecting multiple target genes simultaneously (Kandoth et al., 2013b; Vogelstein et al., 2013; Garraway and Lander, 2013; Shen and Laird, 2013). These are not screened for clinically, nor are there drugs available to modulate their effect.



In the clinic today, the bare bones of mutation screening are typically used, such as HER2 for breast cancer or KRAS for colon cancer. Recently, Foundation Medicine commercialized a targeted gene panel of 287 genes that have an established role in cancer (Frampton et al., 2013). Using predominantly fixed-formalin, paraffin-embedded samples, mutation cell concordance was established compared with mass spectrometric methodology (Sequenom), and the typical driver mutations were identified in a cohort of more than 2,000 individuals, such as TP53, KRAS, CDKN2A, and PIK3CA (Frampton et al., 2013). However, there was a long tail of uncommon mutations that was identified, reflecting the profound diversity of cancers.

This panel represents a step forward compared with a very limited gene mutation screen for commonly occurring drivers, which might even miss other pathogenic mutations within incriminated genes. The 287 genes assessed represent only <15% of genes and only the coding elements. This is in contrast to research studies of whole-genome and whole-exome sequencing, with paired germline DNA for each individual, to more precisely determine driver mutations (Kandoth et al., 2013b). Further, multiple recent studies have highlighted the role of noncoding elements of the genome to play a prominent role, such as TERT promoters in melanoma (Huang et al., 2013), a long, noncoding RNA SChLAP1 for aggressive prostate cancer (Prensner et al., 2013), and identification of ~100 non-coding driver variants for cancer using a new bioinformatics tool known as FunSeq (Khurana et al., 2013). Clearly, even a comprehensive exome would only represent a limited swath of sampling for root causes of cancer in an individual.

Cancer genomic medicine of the future will likely involve a GIS of the tumor with assessment of DNA sequence, gene expression, RNA-seq, microRNAs, proteins, copy number variations, and DNA methylation cross-referenced with the individual's germline DNA. But the issue of addressing heterogeneity still looms (Vogelstein et al., 2013; Bedard et al., 2013), and for that, there are a few possible steps, including deep sequencing of the tumor at multiple locations, single-cell sequencing, or the use of the "liquid biopsy" of cancer (Schwarzenbach et al., 2011; Leary et al., 2010; Forshew et al., 2012).

Cell-free tumor DNA in plasma, which is present in the vast majority of patients with cancer, has been shown to be a useful biomarker for following patients (Schwarzenbach et al., 2011; Leary et al., 2010; Forshew et al., 2012) and appears to have independent prognostic significance (Dawson et al., 2013). It may be that plasma tumor DNA is the best representative of the cancer for targeted treatment because avoidance of metastasis is of utmost concern. The ease of isolating and sequencing cell-free tumor DNA is likely to make this a very attractive and routine in the future. Especially appealing is the ability to sample on a more frequent and even continuous basis using biosensors. Whether the "liquid biopsy" will help override the challenges related to intratumor heterogeneity and recurrence of cancer awaits prospective evaluation. Also of particular interest, at some point, will be screening healthy people for cell-free tumor DNA to determine whether we are constantly facing microscopic tumor burden but are able to effectively keep the disease in check by a variety of homeostatic mechanisms.

### Molecular Diagnosis

When a patient receives a diagnosis of a chronic illness today, it is nonspecific and is based on clinical and not molecular features. Take, for example, diabetes mellitus type 2, which could reflect anything from insulin resistance, failure of  $\beta$ -islet cells, or a variety of subtypes, including  $\alpha$ -adrenergic receptor (ADRA2A) diabetes (Gribble, 2010) or a zinc transporter subtype (SLC30A8) (Sladek et al., 2007). Common genomic variants have been identified in pathways involving signal transduction, cell proliferation, glucose sensing, and circadian rhythm (Dupuis et al., 2010). Some individuals with a high fasting glucose have a G6PC2 variant that is associated with protection from diabetes (Bouatia-Naji et al., 2008). A genotype score, amalgamating the number of risk variants, has been shown to be helpful for identifying high susceptibility (Meigs et al., 2008). Moreover, there are 13 classes of drugs to treat diabetes, and the treatment could be made considerably more rational with knowledge of the individual's underlying mechanism(s).

This brief summary of the diabetes example reflects the need for a new molecular taxonomy across all diseases. When an individual is diagnosed (or at some point when risk can be defined), the molecular basis will be assessed and, ideally, when possible, the root cause will be established. Clearly, for many common diseases, there are multiple pathways implicated, and this may prove to be difficult. But there has yet to be a systematic attempt of providing such a molecular diagnosis in clinical care. Despite multiple reports of molecular subtypes of asthma (Wenzel, 2012), multiple sclerosis (Ottoboni et al., 2012), and colon (Sadanandam et al., 2013) and uterine cancer (Kandoth et al., 2013a), which appear to be linked with therapy and prognosis, this has yet to be made part of medical practice.

### Pharmacogenomics

Just as molecular subtyping of chronic disease is not part of medical practice, pharmacogenomics screening for either assurance of efficacy or avoidance of major side effects is predominantly ignored. With the use of GWAS, there has been an avalanche of discovery of alleles that are pivotal to individual drug response. Unlike polygenic disease, for which the penetrance for a common sequence variant is quite low (approximate odds ratio of 1.15), the typical genotype odds ratio for prescription drugs can be as high as 80, and for many, the range is 3- to 40-fold (reviewed in depth in Harper and Topol [2012]). The likely explanation for this pronounced impact of common variants on individual drug response is based on selection—as compared with diseases, the human genome has had very limited time to adapt to medication exposure. Despite there being more than 100 drugs that carry a genomic "label" by the Food and Drug Administration, meaning that there is a recommendation for genotype assessment before the drug is used, there is rarely any pharmacogenomic assessment in clinical practice. This needs to improve, and perhaps the availability of point-of-care testing will help, along with reduced cost, to eventually promote routine use. Beyond this barrier, there needs to be more genomic sequencing for commonly used drugs, with associated phenotypic determination of efficacy and side effects, along with systematic omic assessment for drugs in development. From the marked success of discovering genomic-drug interactions found to date, there is certainly the sense that, the more you look, the

more you will find. The potential here is to reduce the waste of pharmaceuticals, not just by avoiding drugs that will not provide efficacy for particular individuals but also by avoiding serious toxicity that can be either fatal or life threatening.

### Healthspan

The human reference genome is based upon multiple young individuals who had no phenotypic characterization. Accordingly, we know nothing about the reference human's natural history of disease, and one can consider this as a flawed standard for comparison. Ideally, we should have a reference genome that has had rigorous phenotyping. This is especially the case in an era of using sequencing in medical practice but with an inadequate comparator. Perhaps the optimal phenotype would be healthspan. At Scripps, we have defined healthy elderly as age >80 years with no history of chronic illness or use of medications. The cohort (known as "Wellderly") that we have assembled over the past 7 years of 1,400 individuals has an average age of 88, and we have completed whole-genome sequencing for 500 of these individuals. The intent is to provide a more useful reference genome with a clearly defined, uniform, and relevant phenotype.

Moreover, there is another important application of healthspan genomics. Multiple studies have established that a research investment in understanding healthy aging would be more prudent than in any specific disease category (Goldman et al., 2013a). Because the cohort that we have enrolled carries a similar burden of common risk variants for chronic diseases compared with the general population, there are most likely a substantial number of modifier genes and protective alleles that may be ultimately identified. One example is from APP, a gene that has a variant for both high risk of Alzheimer's and another rare variant with marked protection from cognitive impairment of Alzheimer's (De Strooper and Voet, 2012). Unquestionably, there are many more such variants left to be discovered, and therein lies the potential for drug discovery efforts that can follow such findings from nature of particular genes and pathways that prevent diseases.

### Molecular Autopsies

Although physical autopsies have lost favor and become exceptionally rare, there is an opportunity to use sequencing to determine the cause of death, particularly when this occurs suddenly. Targeted or whole-genome sequencing for heritable heart disorders implicated in sudden death, including ion channel mutations and hypertrophic cardiomyopathy, can be performed in the deceased individual and family members. This approach is now actively being pursued in New York City for all sudden cardiac deaths (Erdmann, 2013) and may prove helpful in preventing this condition in family members.

### Future Directions

This prewomb-to-tomb review has emphasized that there is a disproportionate relationship between knowledge and implementation into clinical practice. For individualized medicine to take hold, it will require intensive, rigorous validation that these new approaches improve patient outcomes and are demonstrated to be cost effective. This proof will be essential for the medical community to embrace the opportunities but will also require educational programs that squarely address the knowledge chasm that currently exists for practicing physicians. A

second theme is that our efforts have been largely sequence centric and have not adequately taken into account or integrated the data from other omics, no less biosensors and imaging. Related to this deficiency, there is a profound shortage of data scientists in biomedicine, with unparalleled opportunities to process enormous, high-yield data sets. While we will increasingly rely on algorithms, artificial intelligence, and machine learning, the rate-limiting step necessitates talented bio-computing and bioinformatic human expertise.

One of the most attractive outgrowths of defining each individual's unique biology in an era with unprecedented digital infrastructure is to be able to share the data. By taking the deidentified data from each individual, including panoramic, biosensors, social graph, treatment, and outcomes, an extraordinary resource can now be developed. Such a massive open online medical information (MOOM) repository could provide matching capability to approximate a newly diagnosed individual's data as compared with all of those previously amalgamated. For a patient with cancer, for example, this could provide closest matches to the tumor GIF, demographics, treatment, and outcomes to select an optimal strategy; this would potentially take Bayesian principles to a new, enriched potential. Such a MOOM resource does not need to be confined to cancer, but the first to be announced was with the Leukemia & Lymphoma Society and Oregon Health Sciences University for 900 patients with liquid tumors (Winslow, 2013). Hopefully, this will be one of many data-sharing initiatives in medicine to go forward, now that such rich unique information can be captured at the individual level, and our computing infrastructure is so well suited to perform such functions. Although we are still at the nascent stages of individualized medicine, there has never been more promise and opportunity to reboot the way health care can be rendered. Only with systematic validation of these approaches at the intersection of biology and digital technology can we actualize this more precise, futuristic version of medicine.

### ACKNOWLEDGMENTS

I want to express my gratitude to my colleagues Ali Torkamani, PhD, and Erick Scott, MD, who reviewed the manuscript and offered editorial input and to Katrina Schreiber, who helped prepare the typescript and graphics. This work was supported by NIH/NCATS 1 UL1 TR001114. Dr. Topol is an advisor to Illumina, Genapsys, and Edico Genomics and is a cofounder of Cypher Genomics.

### REFERENCES

- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Almeling, R. (2013). The unregulated sperm industry. *New York Times*, November 30, 2013. <http://www.nytimes.com/2013/12/01/opinion/sunday/the-unregulated-sperm-industry.html>.
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., et al.; MetaHIT Consortium (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180.
- Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* 10, 1127–1133.

- Bedard, P.L., Hansen, A.R., Ratain, M.J., and Siu, L.L. (2013). Tumour heterogeneity in the clinic. *Nature* 501, 355–364.
- Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P., and Endy, D. (2013). Amplifying genetic logic gates. *Science* 340, 599–603.
- Bouatia-Naji, N., Rocheleau, G., Van Lommel, L., Lemaire, K., Schuit, F., Cavalcanti-Proença, C., Marchand, M., Hartikainen, A.L., Sovio, U., De Graeve, F., et al. (2008). A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. *Science* 320, 1085–1088.
- Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* 14, 681–691.
- Brannon, A.R., and Sawyers, C.L. (2013). “N of 1” case reports in the era of whole-genome sequencing. *J. Clin. Invest.* 123, 4568–4570.
- Butte, A.J. (2013). Should healthy people have their genomes sequenced at this time? *The Wall Street Journal*, February 15, 2013. <http://online.wsj.com/news/articles/SB10000872396390443884104577645783975993656?KEYWORDS=cost+of+whole+genome+sequencing>.
- 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.
- Cho, I., and Blaser, M.J. (2012). The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* 13, 260–270.
- Church, G.M., Gao, Y., and Kosuri, S. (2012). Next-generation digital information storage in DNA. *Science* 337, 1628.
- Couzin-Frankel, J. (2012). Genetics. New company pushes the envelope on pre-conception testing. *Science* 338, 315–316.
- Damani, S., Bacconi, A., Libiger, O., Chourasia, A.H., Serry, R., Gollapudi, R., Goldberg, R., Rapeport, K., Haaser, S., Topol, S., et al. (2012). Characterization of circulating endothelial cells in acute myocardial infarction. *Sci. Transl. Med.* 4, 126ra33.
- Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27.
- Dawson, S.J., Tsui, D.W., Murtaza, M., Biggs, H., Rueda, O.M., Chin, S.F., Dunning, M.J., Gale, D., Forshaw, T., Mahler-Araujo, B., et al. (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* 368, 1199–1209.
- de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* 367, 1921–1929.
- De Strooper, B., and Voet, T. (2012). Alzheimer’s disease: A protective mutation. *Nature* 488, 38–39.
- Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42, 105–116.
- Erdmann, J. (2013). Telltale hearts. *Nat. Med.* 19, 1361–1364.
- Eyre, D.W., Cule, M.L., Wilson, D.J., Griffiths, D., Vaughan, A., O’Connor, L., Ip, C.L., Golubchik, T., Batty, E.M., Finney, J.M., et al. (2013). Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N. Engl. J. Med.* 369, 1195–1205.
- Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The long-term stability of the human gut microbiota. *Science* 341, 1237439.
- Fan, H.C., Gu, W., Wang, J., Blumenfeld, Y.J., El-Sayed, Y.Y., and Quake, S.R. (2012). Non-invasive prenatal measurement of the fetal genome. *Nature* 487, 320–324.
- Ferguson, B.S., Hoggarth, D.A., Maliniak, D., Ploense, K., White, R.J., Woodward, N., Hsieh, K., Bonham, A.J., Eisenstein, M., Kippin, et al. (2013). Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci. Transl. Med.* 5, 213ra165.
- Forshaw, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D.W., Kaper, F., Dawson, S.J., Piskorz, A.M., Jimenez-Linan, M., Bentley, D., et al. (2012). Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* 4, 136ra68.
- Frampton, G.M., Fichtenholtz, A., Otto, G.A., Wang, K., Downing, S.R., He, J., Schnall-Levin, M., White, J., Sanford, E.M., An, P., et al. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31, 1023–1031.
- Fratkin, E., Bercovici, S., and Stephan, D.A. (2012). The implications of ENCODE for diagnostics. *Nat. Biotechnol.* 30, 1064–1065.
- Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* 10, 241–251.
- Gahl, W.A., and Tift, C.J. (2011). The NIH Undiagnosed Diseases Program: lessons learned. *JAMA* 305, 1904–1905.
- Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodtkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., et al. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364, 730–739.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Goldman, D.P., Cutler, D., Rowe, J.W., Michaud, P.C., Sullivan, J., Peneva, D., and Olshansky, S.J. (2013a). Substantial health and economic returns from delayed aging may warrant a new focus for medical research. *Health Aff. (Millwood)* 32, 1698–1705.
- Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E.M., Sipos, B., and Birney, E. (2013b). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494, 77–80.
- Gordon, J.I., Dewey, K.G., Mills, D.A., and Medzhitov, R.M. (2012). The human gut microbiota and undernutrition. *Sci. Transl. Med.* 4, 137ps12.
- Gribble, F.M. (2010). Alpha2A-adrenergic receptors and type 2 diabetes. *N. Engl. J. Med.* 362, 361–362.
- Groop, L. (2010). Open chromatin and diabetes risk. *Nat. Genet.* 42, 190–192.
- Haffner, M.C., Mosbrugger, T., Esopi, D.M., Fedor, H., Heaphy, C.M., Walker, D.A., Adejola, N., Gürel, M., Hicks, J., Meeker, A.K., et al. (2013). Tracking the clonal origin of lethal prostate cancer. *J. Clin. Invest.* 123, 4918–4922.
- Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., and Turnbaugh, P.J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Escherichia coli*. *Science* 341, 295–298.
- Hall, S.S. (2013). Genetics: a gene of rare effect. *Nature* 496, 152–155.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Harper, A.R., and Topol, E.J. (2012). Pharmacogenomics in clinical practice and drug development. *Nat. Biotechnol.* 30, 1117–1124.
- Harris, S.R., Cartwright, E.J., Török, M.E., Holden, M.T., Brown, N.M., Ogilvy-Stuart, A.L., Ellington, M.J., Quail, M.A., Bentley, S.D., Parkhill, J., and Peacock, S.J. (2013). Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* 13, 130–136.
- Hayden, E.C. (2014). Is the \$1,000 genome for real? *Nature News*, January 15, 2014. <http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530>.
- Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. *Science* 339, 957–959.
- Iida, N., Dzutsev, A., Stewart, C.A., Smith, L., Bouladoux, N., Weingarten, R.A., Molina, D.A., Salcedo, R., Back, T., Cramer, S., et al. (2013). Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* 342, 967–970.
- Issacson, W. (2011). *Steve Jobs* (New York: Simon & Schuster).



- Jacob, H.J., Abrams, K., Bick, D.P., Brodie, K., Dimmock, D.P., Farrell, M., Geurts, J., Harris, J., Helbling, D., Joers, B.J., et al. (2013). Genomics in clinical practice: lessons from the front lines. *Sci. Transl. Med.* 5, 194cm5.
- Jenkins, H.W. (2010). Google and the search for the future. *Wall Street Journal*, August 14, 2010. <http://ow.ly/qglOu>.
- Kaiser, J. (2013). Genomics. Researchers to explore promise, risks of sequencing newborns' DNA. *Science* 341, 1163.
- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., et al.; Cancer Genome Atlas Research Network (2013a). Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67–73.
- Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013b). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339.
- Khurana, E., Fu, Y., Colonna, V., Mu, X.J., Kang, H.M., Lappalainen, T., Sboner, A., Lochovsky, L., Chen, J., Harman, A., et al.; 1000 Genomes Project Consortium (2013). Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* 342, 1235587.
- Kitzman, J.O., Snyder, M.W., Ventura, M., Lewis, A.P., Qiu, R., Simmons, L.E., Gammill, H.S., Rubens, C.E., Santillan, D.A., Murray, J.C., et al. (2012). Noninvasive whole-genome sequencing of a human fetus. *Sci. Transl. Med.* 4, 137ra76.
- Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38.
- Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A., et al.; DIAGRAM Consortium (2009). Parental origin of sequence variants associated with complex diseases. *Nature* 462, 868–874.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471–475.
- Leary, R.J., Kinde, I., Diehl, F., Schmidt, K., Clouser, C., Duncan, C., Antipova, A., Lee, C., McKernan, K., De La Vega, F.M., et al. (2010). Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl. Med.* 2, 20ra14.
- Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31, 142–147.
- Lo, Y.M., Chan, K.C., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R., and Chiu, R.W. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* 2, 61ra91.
- Lupski, J.R. (2013). Genetics. Genome mosaicism—one human, multiple genomes. *Science* 341, 358–359.
- Macosko, E.Z., and McCarroll, S.A. (2012). Exploring the variation within. *Nat. Genet.* 44, 614–616.
- Maher, B. (2011). Human genetics: Genomes on prescription. *Nature* 478, 22–24.
- Marx, V. (2013). Next-generation sequencing: The genome jigsaw. *Nature* 501, 263–268.
- Meigs, J.B., Shrader, P., Sullivan, L.M., McAteer, J.B., Fox, C.S., Dupuis, J., Manning, A.K., Florez, J.C., Wilson, P.W., D'Agostino, R.B., Sr., and Cupples, L.A. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* 359, 2208–2219.
- Morain, S., Greene, M.F., and Mello, M.M. (2013). A new era in noninvasive prenatal testing. *N. Engl. J. Med.* 369, 499–501.
- National Human Genome Research Institute. (2013). Clinical sequencing exploratory research (CSER). <http://www.genome.gov/27546194>.
- National Research Council (2011). Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease (Washington, D.C.: The National Academies Press).
- Ottoboni, L., Keenan, B.T., Tamayo, P., Kuchroo, M., Mesirov, J.P., Buckle, G.J., Khoury, S.J., Hafler, D.A., Weiner, H.L., and De Jager, P.L. (2012). An RNA profile identifies two subsets of multiple sclerosis patients differing in disease activity. *Sci. Transl. Med.* 4, 153ra131.
- Owens, B. (2012). Genomics: The single life. *Nature* 491, 27–29.
- Pirola, L., Balcerczyk, A., Okabe, J., and El-Osta, A. (2010). Epigenetic phenomena linked to diabetic complications. *Nat. Rev. Endocrinol.* 6, 665–675.
- Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic mutation, genomic variation, and neurological disease. *Science* 341, 1237758.
- Prahalad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012). Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* 483, 100–103.
- Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., Davicioni, E., Erho, N., Ghadessi, M., et al. (2013). The long noncoding RNA SCHLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* 45, 1392–1398.
- Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henricsson, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341, 1241214.
- Rincon, P. (2013). GenePeaks firm to offer 'digital baby' screen for sperm donors. *BBC News*, October 4, 2014. <http://ow.ly/qj9wm>.
- Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. *Cell* 155, 39–55.
- Sadanandam, A., Lyssiotis, C.A., Homicsko, K., Collisson, E.A., Gibb, W.J., Wulschleger, S., Ostos, L.C., Lannon, W.A., Grotzinger, C., Del Rio, M., et al. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* 19, 619–625.
- Saunders, C.J., Miller, N.A., Soden, S.E., Dinwiddie, D.L., Noll, A., Alnadi, N.A., Andrews, N., Patterson, M.L., Krivohlavik, L.A., Fellis, J., et al. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.* 4, 154ra135.
- Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Ulrich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334, 369–373.
- Schwarzenbach, H., Hoon, D.S., and Pantel, K. (2011). Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* 11, 426–437.
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630.
- Shen, H., and Laird, P.W. (2013). Interplay between the cancer genome and epigenome. *Cell* 153, 38–55.
- Shendure, J., and Lieberman Aiden, E. (2012). The expanding scope of DNA sequencing. *Nat. Biotechnol.* 30, 1084–1094.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881–885.
- Snitkin, E.S., Zelazny, A.M., Thomas, P.J., Stock, F., Henderson, D.K., Palmore, T.N., and SEGREG, J.A.; NISC Comparative Sequencing Program Group (2012). Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* 4, 148ra116.
- Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., et al. (2012). Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* 366, 707–714.
- Sosnay, P.R., Siklosi, K.R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., Ramalho, A.S., Amaral, M.D., Dorfman, R., Zielinski, J., et al. (2013). Defining

the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat. Genet.* 45, 1160–1167.

Stanton, D. (2013). GfK survey. GfK, August 22, 2013. <http://www.gfk.com/us/news-and-events/press-room/press-releases/Pages/Only-27-of-US-Consumers-Have-Heard-of-Personalized-Medicine.aspx>.

Suvà, M.L., Riggi, N., and Bernstein, B.E. (2013). Epigenetic reprogramming in cancer. *Science* 339, 1567–1570.

Test, T.C. (2013). The Counsyl Test. <https://www.counsyl.com/>.

Tewhey, R., Bansal, V., Torkamani, A., Topol, E.J., and Schork, N.J. (2011). The importance of phase information for human genomics. *Nat. Rev. Genet.* 12, 215–223.

Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. *Nat. Rev. Genet.* 13, 565–575.

Viaud, S., Saccheri, F., Mignot, G., Yamazaki, T., Daillère, R., Hannani, D., Enot, D.P., Pfirschke, C., Engblom, C., Pittet, M.J., et al. (2013). The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* 342, 971–976.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558.

Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150, 402–412.

Wenzel, S.E. (2012). Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat. Med.* 18, 716–725.

Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243.

Winslow, R. (2013). Patients share DNA for cures. *Wall Street Journal*, September 16, 2013. <http://online.wsj.com/news/articles/SB10001424127887323342404579079453190552312?KEYWORDS=Patients+Share+DNA+for+Cures>.

Worthey, E.A., Mayer, A.N., Syverson, G.D., Helbling, D., Bonacci, B.B., Decker, B., Serpe, J.M., Dasu, T., Tschannen, M.R., Veith, R.L., et al. (2011). Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* 13, 255–262.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* 369, 1502–1511.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707–720.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.