ANNUAL REVIEWS

Annual Review of Medicine

Precision Medicine: Functional Advancements

Thomas Caskey

Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030; email: tcaskey@bcm.edu

Annu. Rev. Med. 2018. 69:1–18

The *Annual Review of Medicine* is online at med.annualreviews.org

https://doi.org/10.1146/annurev-med-041316-090905

Keywords

whole-genome sequence, metabolomics, magnetic resonance imaging, molecular diagnosis, machine learning

Abstract

Precision medicine was conceptualized on the strength of genomic sequence analysis. High-throughput functional metrics have enhanced sequence interpretation and clinical precision. These technologies include metabolomics, magnetic resonance imaging, and I rhythm (cardiac monitoring), among others. These technologies are discussed and placed in clinical context for the medical specialties of internal medicine, pediatrics, obstetrics, and gynecology. Publications in these fields support the concept of a higher level of precision in identifying disease risk. Precise disease risk identification has the potential to enable intervention with greater specificity, resulting in disease prevention—an important goal of precision medicine.

INTRODUCTION

Late-onset adult diseases are the largest healthcare burden in both morbidity and care cost in the United States. Cardiovascular and neurologic disorders and cancer have the highest clinical burden (1). Given the aging population, new approaches for risk identification and disease prevention are needed. This review presents the technological and medical advancements for precision medical identification of disease risk and therapy intervention at the level of the individual patient ("N of 1").

Attractive approaches to the goal of precision medicine are use of noninvasive, high-throughput, and high-content technologies that provide actionable information. These parameters lend themselves to acceptance in medical practice. The goals of precision medicine are disease risk identification and pathology intervention. This disease-prevention strategy satisfies the goals of lower health cost and reduction of morbidity.

There are innovative technologies now available and in development that advance the purposes of precision medicine. The following are emphasized because they have been tested and found useful: genome sequencing, metabolomics, pharmacogenomics, proteomics, magnetic resonance imaging (MRI), bioinformatics, machine learning, and the electronic record of personal medical history and family medical history. Recent progress in transcriptome sequencing adding interpretative value has been reported (2) but is not included in this review. Function measurements are critical to subsequent therapeutic intervention. It is therefore essential to have functional tools for precision medicine that complement the prediction of aberrant function determined by DNA sequence information.

GENOMIC SEQUENCING

Genomic sequencing was enabled in clinical settings by advances in instrumentation with features of high throughput, lower cost, and cloud storage of data. These Illumina instruments differed from the ABI Technologies instruments that were used to obtain the first human genome frameworks (3, 4). The first approach to genome sequencing targeted the 180,000 expressed regions of the 22,000 known genes. The exome is 1.2% of the 3.2 gigabases in a human genome or about 30 million base pairs. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) are the common approaches for clinical settings. WES requires a hybridization capture technology prior to sequencing (5). The WES strategy entails lower cost and shorter analytical time than WGS. Clinical application of WES is practical, as reflected by the immediate success in genetic disease diagnosis and new disease gene/disease discoveries (6). The transition to WGS was enabled by advances in instrumentation, collaborative efforts of genome centers and industry, cloud storage organization, and the establishment of a consortium on human genomes for use in analytical validation. WGS is the most comprehensive method for analyzing human genomes; it does not require capture technology, and as stated, it sequences the entire genome. WGS allows the identification of disease-causative variants in exonic (protein coding), intronic (splice regions), and regulatory regions for identification of protein expression and gene regulation mechanisms (7). WGS therefore surpasses WES for medical sequence knowledge, providing not only the full sequence of expressed genes in the genome but also sequence information for genomic regulatory elements. The enormity of the WGS data set provides analytic challenges for medical interpretation by physicians and patients, as well as the challenge of controlling costs for whole-genome N-of-1 sequence analysis.

DNA sequence information is extensively used by the biotechnology and pharmaceutical industries for drug development, including new drug target identification and selection of genetic cohorts for clinical trials (8, 9). Each of these applications involves significant challenges, but once

analytical bottlenecks are navigated, information from WGS can be readily automated. Medical interpretation cannot.

Both WES and WGS are in clinical use today. Each sequencing strategy continues to improve and has advanced far beyond a "research" designation. Clinical utilities are illustrated in a later section (Fields of Application of Precision Medicine Advancements).

Interpretation

Interpretation of a genomic sequence is not a simple issue (10). Medical interpretation of DNA sequence variation has to consider not only the enormity of the genome (3.2 billion base pairs) (11) but also the high variability among general populations. An individual genome can vary at 4.1–5.0 million positions compared to a reference genome, thanks to ancestral population sequence variant biases (**http://www.gwascentral.org**) and the continued introduction of new mutations (12). The transcription–translation process has a built-in tolerance to genomic variation, as observed in the fact that although the genetic triplet codes for 64 codons, only 20 amino acids are encoded. The amino acids are encoded by one to six triplet codes (e.g., tryptophan and methionine arise from a single codon, whereas leucine can arise from six different codons). Genome sequence variation is tolerated by the permissiveness of the degenerative genetic code, and substitution of amino acids with similar biochemical properties can result in no deleterious change in a protein. The medical interpretation of sequence variations (damaging or tolerated) must take these features into account before rendering a precision diagnosis based on sequence alone. Informatics algorithms that take into account these features are used to rapidly estimate the impact of amino acid changes and regulatory sequence substitutions on gene function (13).

Disease-specific databases provide a strategy for improving precision. Disease databases are available that associate sequence variations (mutations) with diseases established to be causative by published patient studies. An example is the breast cancer database set up by Global Alliance (**https://www.patientpower.info/video/the-advanced-breast-cancer-global-alliance**), HGMD (**http://www.qiagenbioinformatics.com**), Clinvar (**http://www.ncbi.nlm.nih.gov/clinvar**), OMIM (**http://www.omim.org**), Leiden Open Variation Database, DECIPHER (14), and recent publications are encyclopedias for identifying sequence variants causative of disease. Using this knowledge, one can interpret the disease risk for many sequence variants in an individual's genome. It is estimated that each individual has 10,000 protein-coding variants in exons and many more noncoding variants in the whole genome. Despite these variations, risks of many cardiovascular diseases, cancers, and neurodegenerative diseases can be predicted by sequence alone.

Additive to use of databases, one can make a disease risk prediction if the mutation damages or eliminates synthesis of the protein. Variants creating deletions, insertions, canonical splice sites, and premature termination are examples of variants that are predicted to lead to loss of function (15). Comparative sequence analysis using informatics algorithms can evaluate additional variants' impacts on functions (13). It can identify highly conserved portions of a gene across species and among humans by the biochemical characters of amino acid substitutions. Examples of conserved sequence regions include sequences that encode gene regulatory regions and protein-protein interaction, transmembrane protein regions, catalytic active sites where sequence is critical to function of proteins, and low-frequency occurrences suggesting essential functions. Although it appears cumbersome, informatics automation has simplified interpretation of sequence variants for clinical use. WGS is identifying regulatory mutations and thereby building knowledge of the genetic basis of disease. Presently, we are aware of 246 inherited diseases caused by 612 noncoding variants (16). However, informatics cannot completely solve the challenges of interpretation.

Table 1 "N of 1" genome report[a]

Mode of inheritance	Gene	Variant	Zygosity	Associated conditions	Clinical significance
autosomal recessive	*CFTR*	NM_000492.3:c.1521_1523delCTT NP_000483.3:p.Phe508del	heterozygote	cystic fibrosis	pathogenic
autosomal recessive	*GBA*	NM_000157.3:c.1226A>G NP_000148.2:p.Asn409Ser	heterozygote	Gaucher's disease	pathogenic
autosomal dominant	*BRCA1*	NM_007294.3: c.213-2A>C	heterozygote	hereditary breast and ovarian cancers	pathogenic
autosomal dominant	*LDLR*	NM_000527.4: c.654_656delTGG NP_000518.1: p.Gly219del	heterozygote	familial hypercholesterolemia	pathogenic
unknown	*MSH6*	NM_000179.2:c.3558_3565delTGAAAGTA NP_000170.1:p.Glu1187Ilefs	heterozygote	Lynch syndrome	likely pathogenic
unknown	*MC1R*	NM_002386.3:c.86dupA NP_002377.4:p.Asn29Lysfs	heterozygote	cutaneous malignant melanoma	VUS

[a]This report is synthesized from several individuals from the 209 study to protect privacy (17). Omitted from this abbreviated illustration is the description of the disease, risk of occurrence of disease, frequency of the variant, and medical references supporting the diagnosis. The report typically does not make medical recommendations.

At this time, genomic sequence alone is not sufficient for precision genomic diagnosis of disease risk, nor for identifying genetic diseases in affected patients. Recognition of this has led to a cautious classification of genomic variants as pathogenic, likely pathogenic, and "variant of uncertain significance" (VUS) to assist physicians and patients with diagnostic precision reporting. An example of an N-of-1 genome report, actually a composite of findings from several patients, is shown in **Table 1** and discussed in the following subsections. The genomic information in this report would enable a medical geneticist, genetic counselor, and board-certified specialist to counsel a patient on disease risk.

Autosomal Recessive Conditions

Autosomal recessive conditions are commonly detected in families with no history of genetic disease. Cystic fibrosis variants are commonly detected because the population carrier (heterozygote) frequency of this disease is 2.5% in North Americans (OMIM #219700). Gaucher's disease variants are found commonly in Ashkenazi Jews (1.5%) but not exclusively (OMIM #230800). These two recessively inherited disorders illustrate a feature of disease penetrance. Cystic fibrosis carriers generally have normal adult health. Adults who are carriers of a Gaucher's mutant gene have increased risk of Parkinson's disease (18). The medical community is recognizing more examples of penetrance features associated with recessive disorders as we determine the health impact of a single gene mutation over a life span.

Autosomal Dominant Conditions

Autosomal dominant conditions are recognized in families with a multigenerational history of a specific disease. They can also emerge by de novo mutation; achondroplasia, for example, occurs predominantly by de novo mutation in the *FGFR-3* gene (19). In the case of dominant disorders, the second, normal copy of the gene does not protect from disease.

Cardiovascular patients commonly have a multigenerational family history of disease. Type II hypercholesterolemia is due to mutations in the low-density lipoprotein cholesterol (LDL) receptor. The issue of penetrance is illustrated by the fact that children with two copies of the

defective gene have early-onset vascular disease and those with a single copy develop cardiovascular disease prematurely in midlife. This genetic risk discovery opened the cardiovascular genetic field for risk identification (20). Because the incidence of LDL receptor mutations is 2% in the US population and statins reduce disease risk for these patients, screening has been proposed as a public health initiative (21).

Cancer provides another example of autosomal dominant risk identification. The pathogenic variants in *BRCA1* and *BRCA2* are found among individuals with family history for breast and ovarian cancer (2%) (22). Population-based screening has already been enacted in Israel, where only three variants account for 95% of risk of *BRCA1* and *BRCA2* cancers, thus simplifying screening and controlling cost of this public health strategy (23). Not all patients who carry pathogenic variants will develop either breast or ovarian cancer by the age of 65. For these *BRCA1* gene sequence variants, the disease risk is 26–55%, and for *BRCA2* variants it is 25%. The mechanism of avoidance of disease is not understood at this time, but suppressor genes are suspected. Suppressor mutations are well known in *Saccharomyces cerevisiae*, *Escherichia coli*, and *Drosophila* (24), though none have been discovered in humans to date. Genetic modifiers of penetrance of *BRCA1* and *BRCA2* have been studied by polygenic panels of SNPs in an effort to genetically localize "suppressors" (25).

Pathologic gene sequence variants are considered "risk variants." Their identification in an individual patient places considerable responsibility on the physician and counselor to translate this finding into a plan for care, surveillance, or management.

Variants of Uncertain Significance

The report in **Table 1** notes two variants of uncertain significance (VUS). The first example is a mutation in the mismatch repair gene, *MSH6* (OMIM 615083). This risk finding is within the list of genes (26) recommended for reporting by the American College of Medical Genetics and Genomics (ACMG) (27). A pathogenic variant in *MSH6* has been reported in ~10% of families with Lynch syndrome, also known as hereditary non-polyposis colorectal cancer. In this example, the causative variant is a premature chain termination mutation predicting a truncation of the protein. Although the specific mutation is not reported in the literature, the variant is predicted to cause loss of function. Given the lifetime cancer risk of 80% for *MSH6* pathogenic variants, the finding was reported to this individual, who had a positive family history of colon cancer. The patient can use colonoscopy surveillance with the knowledge of the identified sequence-based risk. The second VUS is a variant in a gene (*MC1R*, melanocortin receptor 1) that regulates melanin levels and thus skin color (OMIM # 155541). Individuals carrying this variant have a four- to eightfold increase in risk of squamous and basal cell cancer. Because the gene is known to be highly polymorphic (varying in sequence) and only a few variants are reported to reduce function of the gene, cautious interpretation of the cancer risk is warranted. The patient can utilize the sequence-derived risk information to reduce disease risk by reducing sun exposure, avoiding use of mutagenic UV-blocking agents, undergoing regular medical dermatologic surveillance for cancer lesions, and opting for early surgical removal of such lesions.

METABOLOMICS

Metabolomics has the potential to detect 3,000 small-molecule compounds (analytes) in serum (**Figure 1**). Advances in mass spectroscopy currently permit reliable quantitation of 1,500 analytes for each patient. Recent genetic twin studies (28) established a heritability factor of 70–80% for levels of 650 analytes in serum. Metabolomics provides a new functional diagnostic approach to evaluate disease risk for cardiovascular, metabolic (diabetes), and renal disease. Metabolomics quantitation and gene sequencing used in concert reveal the genetic basis of many metabolism differences among individuals. Metabolomics has the advantage of evaluating an entire pathway

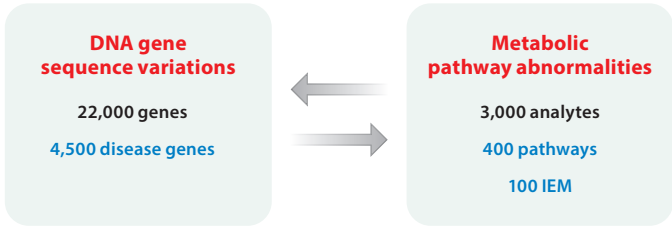


Figure 1

DNA sequence/metabolomic linkage strategy. The figure illustrates the complementarity of whole-genome sequencing of pathogenic variants and metabolic function.

(carbohydrate, urea cycle, lipids) in health and disease. In two recent twin and N-of-1 studies (17, 28), 54 associations of disease-causative gene mutations were made by detection of metabolic pathway abnormalities. The genetic use of those associations is illustrated in the next section (29, 30), which shows how pathway analysis can lead to resolution of an undiagnosed childhood disease. (A specific single-gene mutation identified in the pathway accounted for the metabolic abnormality.) We have previously reported the strategy to identify or refute a VUS as disease causative in a study of normal individuals (31).

In the newly identified 54 examples, the genes were frequently associated with inborn errors of metabolism (IEMs) in children (31). Two additional individual adults had an IEM detected by metabolite analysis. Children are screened for the IEM with the objective of early treatment and reduction of disease damage (32). At this time, it is clear adults have metabolic "penetrance" for these typically "recessive" traits. The analyte quantitative variation for adults can be as much as 4–5 standard deviations above or below population norms. It is unknown if adults could benefit from dietary or drug management as is standard practice for children with IEMs. We are in early days of such new discoveries of genetically determined metabolomic abnormalities. The situation is reminiscent of the introduction of automated analyte testing whose results sometimes revealed an unsuspected disease diagnosis. Mass spectrometry detects 150 times as many analytes as current laboratory tests and is far more sensitive. When metabolic variation is matched to the genetic causation, we gain greater precision and expand informed therapy options.

PHARMACOGENOMICS

Mass spectrometry analysis has improved pharmacogenomics by enabling detection of drug level, drug metabolites, and drug effectiveness, as well as sensitive detection of liver and kidney toxicity. The technology is widely used by the biotech and pharmaceutical industries to prepare for US Food and Drug Administration (FDA) approval of a new chemical entity. Genome studies have identified numerous drug metabolism genes and sequence variations affecting drug levels. Iatrogenic diseases caused by drugs are influenced by the variable expression of these genes, as well as inhibition (via drug/drug interaction) of metabolizing enzymes. Considerable research has focused on sequence variations of these metabolizing genes, which influence dosing levels. Despite this knowledge, predicting effective first dosage has been difficult to achieve. An extensively studied example is the anticlotting drug warfarin (33), whose activity is affected by *CYP2C9*, *VKORC1*, and *CYP4F2* gene variants. However, success has been realized for 6-mercaptopurine in leukemia therapy (34). Three to fourteen percent of *TPMT* variants lower drug metabolism and increase toxicity. Individuals with one nonfunctional *TPMT* allele experience moderate toxicity, and individuals with two nonfunctional *TPMT* alleles experience severe myelosuppression. Identification of the poor-metabolism *TPMT* variants is now a test for selecting initial dosage for leukemia patients. Research

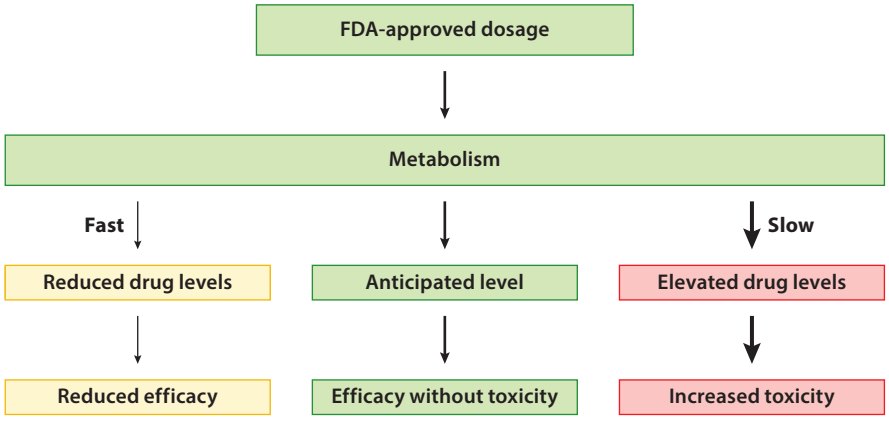


Figure 2

FDA-approved dosing is typically determined by testing for drug efficacy and toxicity on a genetically heterogeneous cohort (*green pathway*). Individuals may be underdosed (*yellow pathway*) owing to "fast" P450-metabolizing gene variants or elevated P450 levels caused by an increase in P450 gene transcription due to a second drug. Other patients (*red pathway*), having a "slow" P450-metabolizing allele or P450 enzyme inhibition by a second drug, may be overdosed and experience drug toxicity.

is now directed at the actual measurement of drug (and metabolite) level and earlier toxicity detection. Use of the genetic sequence of metabolizing genes together with mass spectrometry quantitation provides a dual and more accurate N-of-1 solution to the quandary of efficacy versus toxicity when determining the optimal drug dosage for a patient.

Figure 2 illustrates how genetic sequencing and mass spectrometry improve precision dosing of individual patients. FDA-approved dosing is typically determined by testing for drug efficacy and toxicity on a genetically heterogeneous cohort. At the approved dose, patients with faster-metabolizing gene variants will have lower than expected drug levels and efficacy; those with slower-metabolizing gene variants will have elevated drug levels and toxicity. Genetic sequencing and mass spectroscopy provide the drug levels as well as determining therapy efficacy and toxicity effects. We have previously reported toxicity metabolic effects for two commonly used drugs, statins and acetaminophen, linked to high drug level (31). **Figure 3** (29, 30) provides three specific illustrations of the pharmacogenomics principles shown in **Figure 2**.

Immunologically driven drug toxicities are associated with specific DNA variants in the HLA and HLB loci. Catastrophic Steven Johnson syndrome is linked to HLA sequence variants influencing responses to specific drugs. Examples include HLA-B*57:01 for abacavir (35) and flucloxacillin, and HLA-B*15:02 for carbamazepine and phenytoin (see Reference 36 for more examples). In the case of these HLA-B* variants, implementing pharmacogenomics and mass spectroscopy drug quantitation enables us to avoid iatrogenic drug injury as we strive for precision diagnosis and therapy.

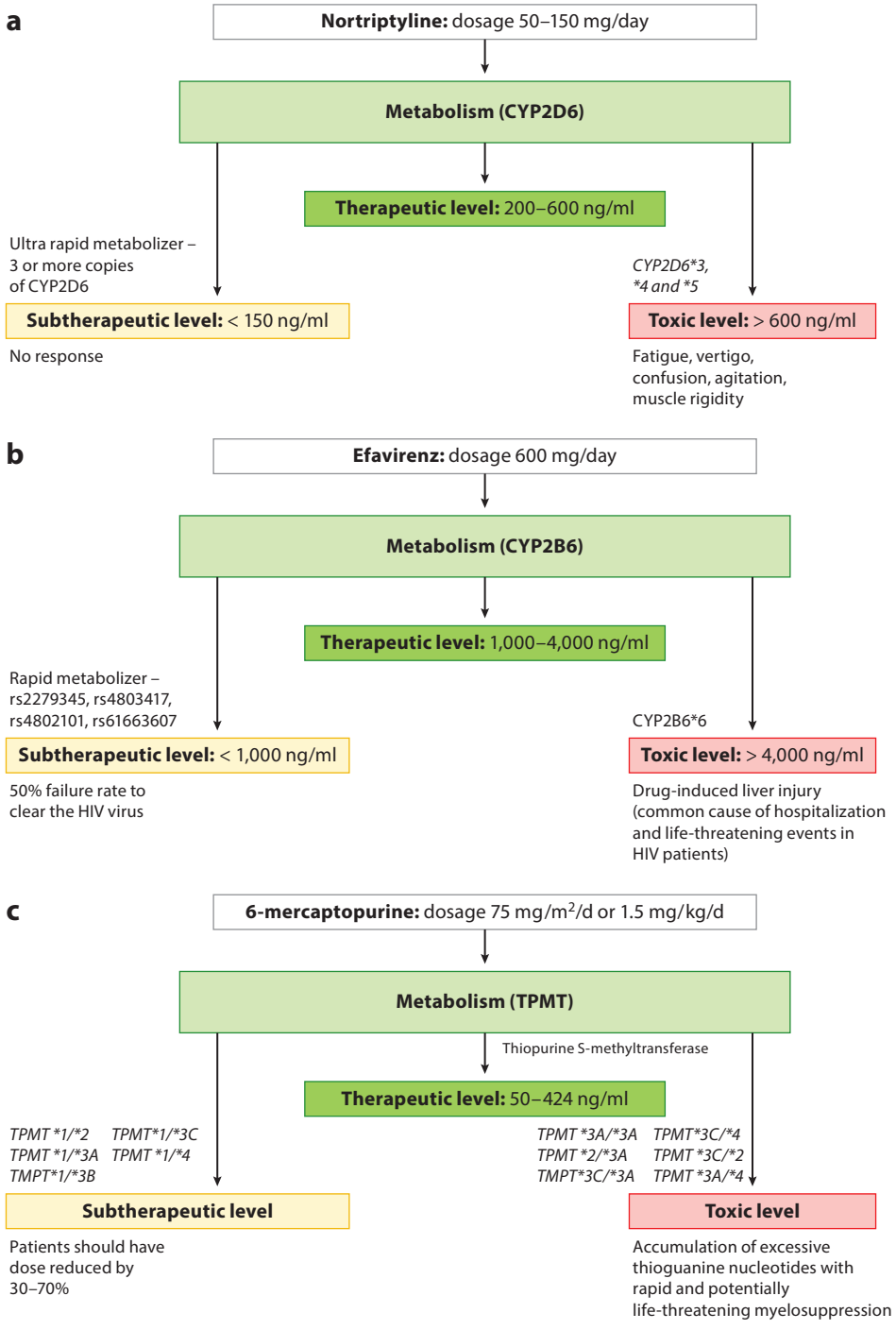
The recent report of a severe neurologic disorder due to fatty acid amide hydrolase emphasizes the need for new strategies to detect iatrogenic disease (37). Metabolomics is an obvious consideration for early detection of injury.

PROTEOMICS

Proteomics has been difficult to develop because there is wide variation in levels of specific proteins in sera (albumin and human growth hormone illustrate the challenge). Elisa protein quantitation

Figure 3

Three specific illustrations of the pharmacogenomics principles shown in **Figure 2**. (*a*) The clinical effectiveness and toxicity of nortriptyline are illustrated by *CYP2D6* genetic variants. (*b*) For efavirenz, both failure of efficacy and toxicity are determined by *CYP2B6* genetic variants. (*c*) Both excessive and toxic levels of 6-mercaptopurine are caused by *TPMT* genetic variants.



and flow sorting are examples of current clinically accepted technologies. A novel use of conjugates of monoclonal antibodies and metals (38) has increased the number of proteins (~50) that can be quantitatively detected by mass spectroscopy but falls short of the large numbers of proteins desired for genome-wide detection. There is a critical need for this functional technology to measure both protein levels determined by gene coding variants and their regulation. SomaLogic (39) recently claimed to have a technology capable of measuring 5,000 proteins with specificity and qualitative data by utilizing a "sandwich" aptamer technique. This functional tool would be a welcome addition, correlating protein level to genome sequence variants.

IMAGING

Magnetic resonance imaging (MRI) has advanced in resolution, speed of scan, and imaging of blood flow dynamics in heart, veins, and arteries. New software for diffusion wave imaging has been excellent for imaging cardiac and vascular system anomalies without contrast media, conducted noninvasively and without radiation risk (40). 3D time-of-flight magnetic resonance has the ability to detect unsuspected high-risk disease phenomena such as aneurysms (**Figure 4**),

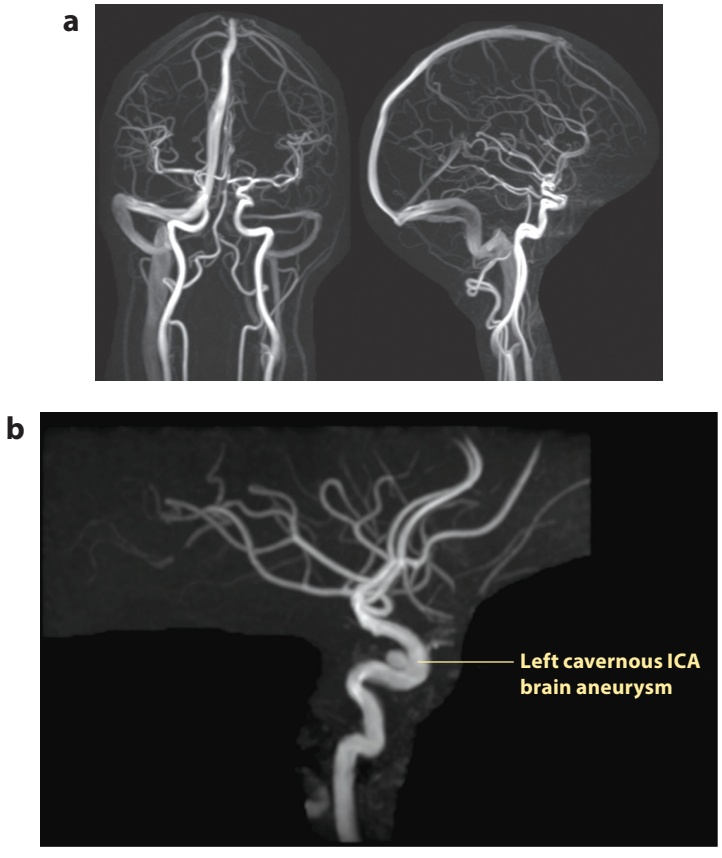


Figure 4

An example of the use of interference magnetic resonance imaging and visual detail obtained by this noninvasive technology: intracranial arteries and veins from noncontrast magnetic resonance angiography in (*a*) a normal case and (*b*) a positive HLI case [a 47-year-old female with an unexpected, newly identified left cavernous internal carotid artery (ICA) aneurysm].

vascular narrowing, and other abnormalities prior to a catastrophic event. Thus, MRI can reduce disease risk. Early detection of cancer lesions of the lung, pancreas, colon, prostate, and ovaries using T1 and diffusion weighted imaging provides precision diagnosis for early surgical intervention and cure (41). In contrast to identifying risk via DNA sequencing, as is done for colon and breast cancer, MRI imaging identifies the transition from genetic risk to actual disease. The two approaches, genetic and imaging, are linked for greater success in precision medicine directed at reducing cancer morbidity. One publication has suggested MRI does not decrease mortality for any cancer (42). These competing approaches to early cancer intervention should be considered as phenotypic/genetic detection methods, each with merit.

A recent report has suggested stem cell replication error rate, not genetic predisposition, is the major driver of neoplasia (43). This new information may influence MRI becoming the primary strategy for early cancer intervention in precision medicine. In addition to its value in noninvasively following patients with an established genetic risk, MRI serves as a precision medicine approach for neoplasias where there is no known genetic predisposition. MRI is already accepted for follow-up of patients treated for cancer either surgically or by medications.

An example of MRI surveillance for cancer development is the Tanner Project (**http://www.tannerproject.org/**). Mutations in the p53 gene can give rise to different cancers in different individuals within a family where the p53 risk mutation occurs. Frequent whole-body scans of an individual at risk can detect the transition from risk to disease for the multiple organs where cancer can emerge. This approach has succeeded in early detection of prostate cancer where prostate-specific antigen testing and biopsy have frequently failed (44). This detection is made possible by use of differential imaging technology (diffusion wave technology), which can distinguish hypertrophy from cancer. Differential imaging of the pancreas can distinguish cysts, which are common, from cancer, which is rare (45).

MACHINE LEARNING

Machine learning has demonstrated its potential in medical practice, including digital pathology, disease prognosis, and diagnosis. The first targets for machine learning are now identified for precision medicine. The challenge is to train for diagnosis from multiple data sets derived from the technologies discussed above as well as future technologies. Interpretation of genetic sequence data has benefited from computer condensation of sequence variants (~10,000 for exons) into a handful of genetic disease risk candidates. When further validated, machine learning will provide trained geneticists and physicians with a translational tool for clinical reporting of sequence data. Exon sequencing prior to machine learning took 12–14 weeks of personnel curation time. An advanced genomic search engine (HLI Search) by Human Longevity, Inc., has reduced the analysis time for identification of disease-causing variants in N-of-1 or trio cases from 12–14 weeks to less than a second (30). The HLI Search engine is not formally machine learning but is the initial step toward that objective. It encompasses traditional components of a genomic search engine but more importantly enables users to query multiple data types simultaneously and is highly interactive for specific queries, responding in <500 ms. Two successful projects, in obesity and coronary artery disease, have illustrated N-of-1 precision medicine utility beyond the 209 study (17). This improvement utilizes a proprietary index technology that allows retrieval of genomic, annotation, and phenotypic data from 0.10000 individuals simultaneously. Analysis of N-of-1 or trio cases through genotype and phenotype integration further improves precision of diagnosis. An example of this new initiative links disease-causative mutations and abnormal metabolite levels. This program uses biochemical and gene pathway analysis to establish a precision diagnosis in children and adults. A reduction of the current 2–3 h analysis time per case to less than a second

would be highly utilized. This feature is particularly important as we expand into large population studies. These achievements will additionally enhance rapid precision diagnosis of childhood IEMs (acute disease) and adult metabolic disease (chronic disease).

A goal of machine learning is to enhance physician access to new knowledge provided by precision diagnosis and interpretation of integrated health data, so as to ultimately guide safe and effective therapy choices. Current data sets are composed of electronic family and personal history, DNA sequence, metabolomics, proteomics, imaging, and pharmacogenomics. These data sets frequently reside in silos rather than being integrated for precision diagnosis and optimal therapy selection. The initial goals for information technology have been realized in rapid simplification of data analysis. Machine learning has a greater challenge. Analysis of different data sets in a logic-based strategy directed toward individual diagnosis would need to satisfy medical diagnosis criteria and be self-correcting. This is similar to the training of an individual physician, who considers several possible diagnoses, tests each by functional assays and imaging, and learns from successes and errors. Machine learning for medical practice will be critical to the utility of our new technologies for precision medicine and has the advantage of taking into account the larger medical practice experience.

FIELDS OF APPLICATION OF PRECISION MEDICINE ADVANCEMENTS

Pediatrics

DNA sequencing paired with metabolomics has enabled precision diagnosis of numerous pediatric disorders (6). Given an undiagnosed case that defies conventional medical resolution, the likelihood of making a gene-based diagnosis is now 25% (6). The approach has been so successful that a national program for undiagnosed disorders is now supported by the National Institutes of Health, philanthropists, the biotech industry, and some insurance carriers. Gene-based diagnosis is fast becoming the first option for a complex undiagnosed case, avoiding the high cost and time of the "odyssey" of less informative older standard diagnostics. Speed of precision diagnosis is critical in pediatrics, since corrective therapies have their greatest efficacy in protecting the central nervous system when used early. Newborn screening, driven predominantly by mass spectroscopy, is applicable to ~50 IEMs that have therapy options. The majority of gene-based discoveries are from undiagnosed pediatric cohorts. A recent report on newborn WGS has questioned pediatric WGS utility over mass spectroscopy, noting a 20% false negative detection rate (46) of IEM by WGS.

Thanks to the combination of DNA sequencing and metabolomics quantitation, the number of new disease/gene associations from the aggregate research community is predicted to grow to 500 per year. This discovery rate is estimated from reporting from exon sequencing and predictably will accelerate with application of WGS, which can detect regulatory sequence mutations now missed by exon sequencing. A recent review reports over 254 published regulatory mutations (16) that affect the function of single disease genes. Two disease categories where regulatory mutations have been extensively studied over many years are hemoglobinopathies and clotting disorders.

Two examples of undiagnosed pediatric cases are illustrated in **Table 2**. In both examples, the very high levels of a metabolite prior to the enzyme defect in the pathway targeted the focus on downstream DNA sequence analysis of pathway genes. In **Figure 5**, ADSL acts at two sites in the metabolism of purines. The finding of markedly elevated levels of SAICA and S-ADO focused attention on this gene where the disease causative mutation was identified by exon sequencing (47).

Table 2 Metabolome/gene precision diagnosis

Disease	Genetic status	Metabolite	Mutations
aromatic amino acid decarboxylase deficiency	compound heterozygote	3-methoxytyrosine	DDC mutations, C286G>A (pG96R) and 260C>T (pP87L)
adenylosuccinate lyase deficiency	homozygote	succinyladenosine	ADSL mutation C1277G>A (pArg426Itis)

Obstetrics

DNA sequencing has dramatically altered the field of prenatal diagnosis. Newborn chromosomal aneuploidy (trisomy 21, 18, 13, X, Y) is common. The invasive nature of amniocentesis and chorionic villus biopsy for prenatal diagnosis has limited these diagnostic approaches to mothers at increased risk (age >35). This left undiagnosed the majority (80%) of newborns with these aneuploidies. A genetic technology breakthrough was the ability to make fetal aneuploidy diagnosis by sequencing mother's serum (48). The technique determined both maternal and fetal DNA genome sequence (~5–10%) in the mother's circulation. Informatic subtractive methods accurately made the diagnosis of the fetal aneuploidy. Noninvasive prenatal diagnosis (NIPD) has been rapidly accepted by expectant mothers and obstetrics/gynecology specialists (49). The

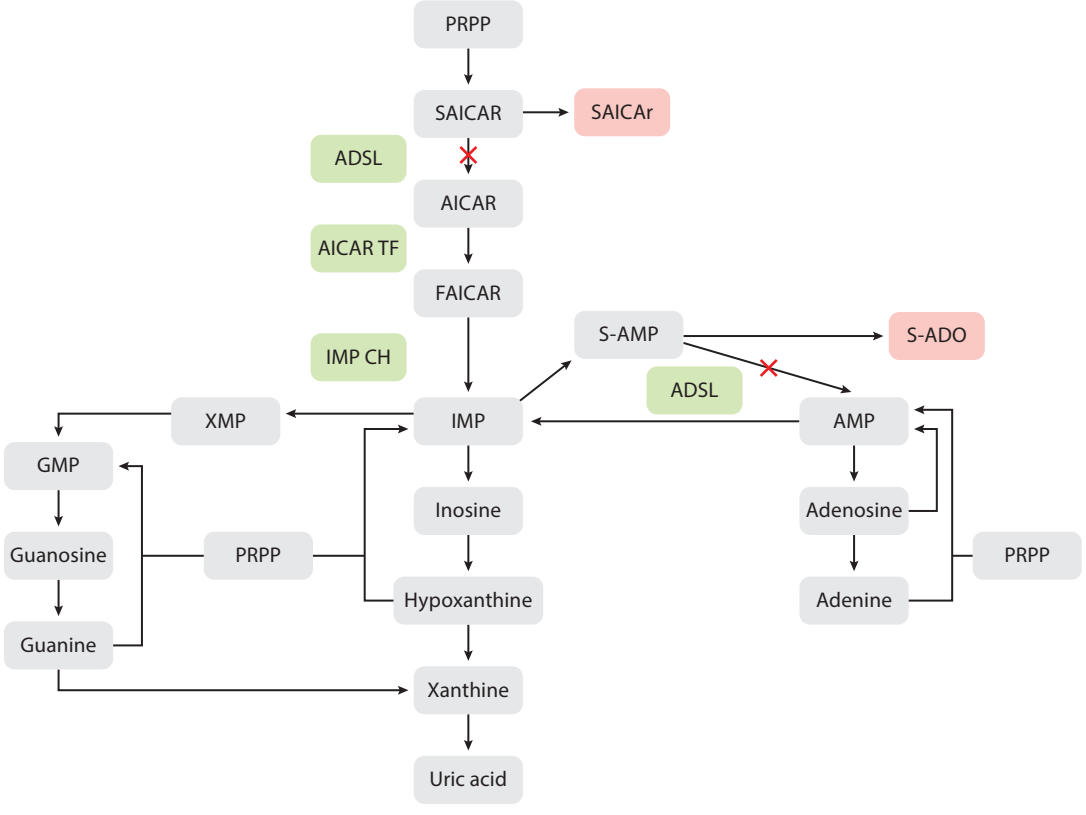


Figure 5

Elevated analytes SAICAr and S-ADO (*red*); identification of gene variants in the disease-causing gene ADSL (*green*) in the purine pathway.

feasibility of prenatal whole-genome and exon sequencing is under study in the wake of a report of successful isolation of fetal cells from a mother's blood (50). Only 2–10 cells are needed for sequencing. Despite the promise of fetal cell isolation, an elegant study (51) has demonstrated the ability to diagnose a fetal gene defect from maternal serum sequence analysis. The diagnosis of adrenal hyperplasia led to in utero fetal hormonal replacement therapy (51). This is the first example I am aware of in which precision in utero diagnosis directed precision in utero therapy. The case illustrates a future direction for in utero precision medicine.

Genetic screening of expectant parents for autosomal recessive disorders is now standard of practice. The American College of Obstetrics and Gynecology (ACOG) has set standards for disease inclusion (26). Disease examples include Tay Sachs, cystic fibrosis, Gaucher's disease, and mucolipidosis IV. Corporations offer disease lists for the screen that exceed ACOG's. The objective is to identify parents who both carry the recessive disease variant, thus offering the option of prenatal diagnosis and avoidance of disease in the newborn.

Detection of dominantly inherited triplet-repeat diseases from the family history is challenging because of "anticipation." The mechanism of clinical anticipation is now understood at a molecular level as a triplet nucleotide repeat expansion, generation to generation (52, 53). New DNA sequence analytic programs have enabled detection of triplet repeat risk from WGS analysis (54). This development improves precision diagnosis of reproductive risk for a normal individual. A healthy at-risk individual may be carrying unstable triplet expansions that can give rise to affected offspring with markedly expanded repeats (replication errors). Fragile X and myotonic dystrophy (53) were first to be identified among the 42 repeat sequence diseases now known. Although family history can identify this risk in some families, DNA sequence is the preferred method to precisely identify or rule out the risk of an individual or family member for a repeat sequence disease in offspring.

Internal Medicine

Many adult-onset disorders are inherited in an autosomal dominant manner, which places special importance on pedigree analysis for risk identification. Our clinical diagnoses of cardiovascular disease, cardiomyopathy, cancers, and neurodegenerative disorders provide a "hunting license" for family genetics disease risk for the patient and among younger asymptomatic family members. Software programs such as PhenoTips (**https://phenotips.org/about**) not only provide an electronic record of family illness but also suggest a genetic diagnosis to be considered for the asymptomatic family member. Family medical information is used to focus on genetic risk genes for the N-of-1 study.

There are a limited number of reports on sequencing "healthy adults." In the first of these reports (H. Tang, E.F. Kirkness, C. Lippert, et al., unpublished data), use of family and personal medical history together with exon sequence established a gene-based disease diagnosis for 28% of N-of-1 adult volunteers. Family and personal history focused sequence studies for disease/gene associations in 25% of cases.

A separate study (55) of individuals sequenced for a pediatric diagnosis found that 17.5% had risk factors for adult disorders on sequence alone, without knowledge of medical or family history. A study (56) of 55,685 hospital admissions found lifestyle and genetic factors independently associated with coronary artery disease. Unsuspected cardiovascular disease risk based on exon sequencing of three risk genes (*LDLR*, *ABOB*, and *PCSK9*) was found at an incidence of 1:222 (57) in an adult cohort and 1:273 for a pediatric cohort (58).

A recent study of 209 adult volunteers used an extensive set of functional measurements to identify disease risk (17). These added measurements were metabolomics, whole-body MRI scanning

and cardiovascular rhythm monitoring, cognition testing, and metabiome testing, in addition to WGS and personal and family history. In 8% of these volunteers, whole-body MRI found previously unrecognized diseases (cancer and cardiovascular) requiring prompt attention. The majority would not have been considered at risk on the basis of currently recognized germline genetic risks. Genomic disease associations were found in 25% of individual participants, in keeping with earlier studies. There were 12 associations of sequence variants and metabolome analytes, including two adult IEMs. None of these disease risk mutations would have been recognized without metabolomics data.

Collectively, the studies make the case for adult genetic disease precision diagnosis as a strategy for accurate diagnosis and disease prevention when DNA sequencing is complemented by functional data. Editorials have also supported this initiative (59).

In addition to these reported N-of-1 results, the American College of Medical Genetics and Genomics panel of experts has suggested 59 genes on which, if a disease risk allele is identified, there is obligation to report the finding (27). Each gene/disease selection has established therapy, surveillance, or management recommendations created by medical experts in the specialty. Examples in this list of 59 include long QT syndrome (*KCNH2*), familial hypercholesterolemia (*APOB*, *LDLR*, and *AC5K9*), and familial adenoma polyposis (*MUTYH*). In the experience of the 209 study of N-of-1 normal volunteers, the obligation to report was 2.4%. Undoubtedly, this list of genes for obligatory reporting will expand. Presently, the identification of disease risk by personal and family history, WGS, metabolomics, proteomics, and MRI reaches 35–50% for adults whose median age is 56.

SUMMARY

Precision medicine was conceptualized based on the power of genetics, but it has been strengthened and enabled by the electronic medical record with family history and by the functional and structural technologies of MRI, metabolomics, and information technology. **Figure 6** summarizes applications of the new technologies that enable precision medicine. The first objectives in precision diagnosis are well advanced. Accurate diagnosis is a significant advancement toward precise disease therapy and prevention.

Precision medicine has to pass the test of efficacy, utility, and expansion beyond the current pilot studies. Challenges in our efforts toward precision diagnosis and therapy exist. One involves the potentially expanding cost of evaluating ambiguous test results and the creation of anxiety in patients and study volunteers due to uncertainty. Examples include VUS in sequencing, incidentalomas on MRI, and unexplained metabolite variations. Medicine has met such challenges previously with the careful adoption of new technologies, including new surgical techniques, new prosthetic and implantable devices, and new drug classes.

Precision medicine tools have a high potential for improved care and disease prevention. Though in its infancy, the field is not untried. One might describe its status in Winston Churchill's words: "Now this is not the end. It is not even the beginning of the end, but it is, perhaps, the end of the beginning."

The advances in precision medicine are driven by innovative technology and offer information to physicians. Integration of the advancements into standard of practice will require physician training and broadening from the tradition of organ system diagnosis now dominating medicine. The examples of p53 mutations leading to multi-organ neoplasias (60) and *AIRE* mutations (OMIM # 151623) leading to multiple endocrine gland failure illustrate the change. Physicians will need to consider fundamental germline DNA mutations and their disease effects on multiple organs in diagnosis and therapy selection.

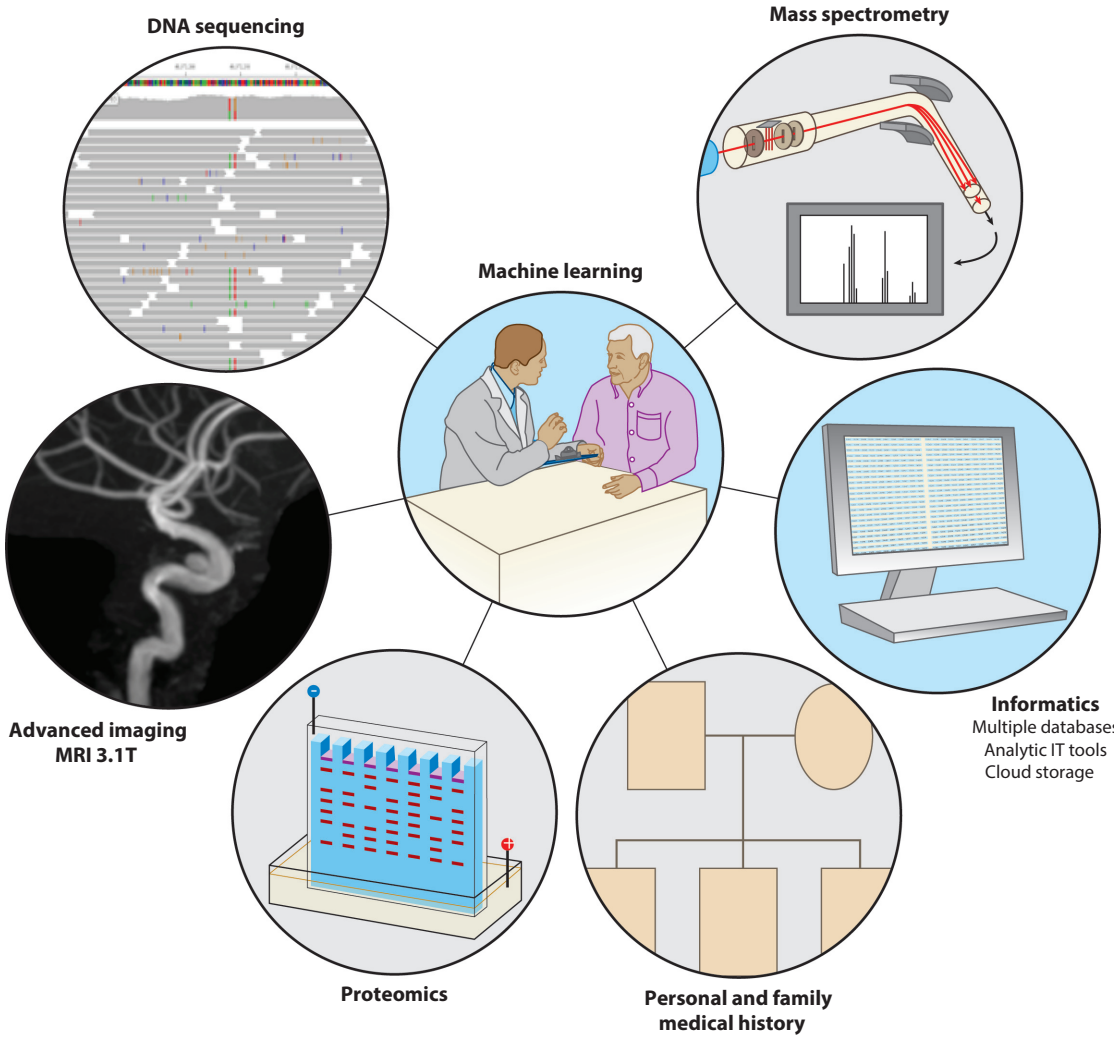


Figure 6

The functional data sets used to achieve a precision diagnosis of an N-of-1 case. Machine learning is illustrated as the method of data silo integration toward a clinical diagnosis.

Wide acceptance of precision medicine into practice will be driven by availability of precision therapy for the affected patient and prevention of disease for the unaffected but genetically at-risk client. We have illustrated the need for machine learning to "de-silo" data sets for precision diagnosis and ultimately therapy selection. Several recent publications extensively cover this need and opportunity (61–63).

DISCLOSURE STATEMENT

The author is a faculty member of the Department of Molecular and Human Genetics at Baylor College of Medicine, which owns a for-profit diagnostic laboratory, Maraca. He is an independent member of the Metabolomics Board of Directors and a consultant to Human Longevity Inc.

LITERATURE CITED

1. Bauer UE, Briss PA, Goodman RA, et al. 2014. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *Lancet* 384(9937):45–52
2. Cummings BB, Marshall JL, Tukiainen T, et al. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9(386):eaal5209
3. Lander ES, Linton LM, Birren B, et al., Intl. Hum. Genome Seq. Consort. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
4. Venter JC, Adams MD, Myers EW, et al. 2001. The sequence of the human genome. *Science* 291(5507):1304–51
5. Bainbridge MN, Wang M, Wu Y, et al. 2011. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol.* 12(7):R68
6. Yang Y, Muzny DM, Reid JG, et al. 2013. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N. Engl. J. Med.* 369:1502–11
7. Telenti A, Pierce LC, Biggs WH, et al. 2016. Deep sequencing of 10,000 human genomes. *PNAS* 113(42):11901–6
8. Druker BJ, Talpaz M, Resta DJ, et al. 2001. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* 344:1031–37
9. Sabatine MS, Giugliano RP, Keech AC, et al. 2017. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* 376:1713–22e
10. Nirenberg M. 1968. *The genetic code*. Nobel Lect., Dec. 12, Nobel Found., Stockholm, Sweden
11. The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526:68–74
12. Crow JF. 2000. The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* 1:40–47
13. Landrum MJ, Lee JM, Benson M, et al. 2016. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44(D1):D862–68
14. Smedley D, Robinson PN. 2015. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. *Genome Med.* 7(1):81
15. Caskey CT, Gonzalez-Garay ML, Pereira S, McGuire AL. 2014. Adult genetic risk screening. *Annu. Rev. Med.* 65:1–17
16. Smedley D, Schubach M, Jacobsen JO, et al. 2016. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am. J. Hum. Genet.* 99(3):595–606
17. Perkins BA, Caskey CT, Brar P, et al. 2017. Precision medicine screening using whole genome sequencing and advanced imaging to identify disease risk in adults. *bioRxiv*, May 3. doi: **https://doi.org/10.1101/133538**
18. Sidransky E1, Nalls MA, Aasly JO, et al. 2009. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* 361(17):1651–61
19. Rousseau F1, Bonaventure J, Legeai-Mallet L, et al. 1994. Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. *Nature* 371(6494):252–54
20. Brown MS, Golstein JL. 1985. *A receptor-mediated pathway for cholesterol homeostasis*. Nobel Lect., Dec. 9, Nobel Found., Stockholm, Sweden
21. Versmissen J, Oosterveer DM, Yazdanpanah M, et al. 2008. Efficacy of statins in familial hypercholesterolaemia: a long term cohort study. *BMJ* 337:a2423
22. Levy-Lahad E, Lahad A, King MC. 2014. Precision medicine meets public health: population screening for BRCA1 and BRCA2. *J. Natl. Cancer Inst.* 107(1):420
23. Gabai-Kapara E, Lahad A, Kaufman B, et al. 2014. Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. *PNAS* 111(39):14205–10
24. Forsburg SL. 2001. The art and design of genetic screens: yeast. *Nat. Rev. Genet.* 2:659–68
25. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. 2017. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl. Cancer Inst.* 109(7): In press. doi: 10.1093/jnci/djw302

26. New MI, Tong YK, Yuen T, et al. 2014. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. *J. Clin. Endocrinol. Metab.* 99(6):E1022–E1030
27. Kalia SS, Adelman K, Bale SJ, et al. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* 19:249–55
28. Long T, Hicks M, Yu H-C, et al. 2017. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* 49:568–78
29. Donti TR, Cappuccio G, Hubert L, et al. 2016. Diagnosis of adenylosuccinate lyase deficiency by metabolomic profiling in plasma reveals a phenotypic spectrum. *Mol. Genet. Metab.* 8:61–66
30. Hou YC, Lavrenko V, Bartha I, et al. 2017. *An interactive search engine of massive genomic and phenotypic data*. Presented at Am. Coll. Med. Genet. Genom. Annu. Clin. Genet. Meet., Mar. 21, Phoenix, AZ. Abstr. No. 446
31. Guo L, Milburn MV, Ryals JAA, et al. 2015. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *PNAS* 112:E4901–E4910
32. Miller MJ, Kennedy AD, Eckhart AD, et al. 2015. Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. *J. Inherit. Metab. Dis.* 38(6):1029–39
33. Pirmohamed M, Burnside G, Eriksson N, et al. 2013. A randomized trial of genotype-guided dosing of warfarin. *N. Engl. J. Med.* 369:2294–303
34. Nguyen CM, Mendes MAS, Ma JD. 2011. Thiopurine methyltransferase (TPMT) genotyping to predict myelosuppression risk. *PLOS Curr*. 3:RRN1236
35. Cargnin S, Jommi C, Canonico PL, et al. 2014. Diagnostic accuracy of HLA-B*57:01 screening for the prediction of abacavir hypersensitivity and clinical utility of the test: a meta-analytic review. *Pharmacogenomics* 15(7):963–76
36. Thorn CF, Klein TE, Altman RB. 2013. PharmGKB: the pharmacogenomics knowledge base. *Methods Mol. Biol.* 1015:311–20
37. Kerbrat A, Ferré JC, Fillatre P, et al. 2016. Acute neurologic disorder from an inhibitor of fatty acid amide hydrolase. *N. Engl. J. Med.* 375(18):1717–25
38. Angelo M, Bendall SC, Finck R, et al. 2014. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* 20:436–42
39. Gawande BN, Rohloff JC, Carter JD, et al. 2016. Selection of DNA aptamers with two modified bases. *PNAS* 114(11):2898–903
40. Brunsing RL, Schenker-Ahmed NM, White NS, et al. 2016. Restriction spectrum imaging: an evolving imaging biomarker in prostate MRI. *J. Magn. Reson. Imag.* 45(2):323–36
41. McCammack KC, Kane CJ, Parsons JK, et al. 2016. In vivo prostate cancer detection and grading using restriction spectrum imaging–MRI. *Prostate Cancer Prostatic Dis.* 19(2):168–73
42. Bojadzieva J, Amini B, Day SF, et al. 2017. Whole body magnetic resonance imaging (WB-MRI) and brain MRI baseline surveillance in TP53 germline mutation carriers: experience from the Li-Fraumeni Syndrome Education and Early Detection (LEAD) clinic. *Fam. Cancer.* In press
43. Tomasetti C, Li L, Vogelstein B, et al. 2017. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 355(6331):1330–34
44. Sternberg IA, Vela I, Scardino PT. 2016. Molecular profiles of prostate cancer: to treat or not to treat. *Annu. Rev. Med.* 67:119–35
45. Megibow AJ, Baker ME, Morgan DE. 2017. Management of incidental pancreatic cysts: a white paper of the ACR Incidental Findings Committee. *J. Am. Coll. Radiol.* 14(7):911–23
46. Kaiser J. 2016. Baby genome screening needs more time to gestate. *Science* 354(6311):398–99
47. Donti TR, Cappuccio G, Hubert L, et al. 2016. Diagnosis of adenylosuccinate lyase deficiency by metabolomic profiling in plasma reveals a phenotypic spectrum. *Mol. Genet. Metab.* 8:61–66
48. Atwal PS, Donti TR, Cardon AL, et al. 2015. Aromatic L-amino acid decarboxylase deficiency diagnosed by clinical metabolomic profiling of plasma. *Mol. Genet. Metab.* 115(2–3):91–94
49. Lo YM, Chan KC, Sun H, et al. 2010. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* 2(61):61ra91
50. Hui L, Bianchi DW. 2017. Noninvasive prenatal DNA testing: the vanguard of genomic medicine. *Annu. Rev. Med.* 68:459–72

51. Kølvraa S, Singh R, Norman EA, et al. 2016. Genome-wide copy number analysis on DNA from fetal cells isolated from the blood of pregnant women. *Prenat. Diagn.* 36(12):1127–34
52. Edwards JG, Feldman G, Goldberg J, et al. 2015. Expanded carrier screening in reproductive medicine—points to consider: a joint statement of the American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Quality Foundation, and Society for Maternal-Fetal Medicine. *Obstet. Gynecol.* 125(3):653–62
53. Pieretti M, Zhang FP, Fu YH, et al. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66(4):817–22
54. Fu YH1, Pizzuti A, Fenwick RG Jr., et al. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255(5049):1256–58
55. Gonzalez-Garaya ML, McGuireb AL, Pereirab S, Caskey CT. 2013. Personalized genomic disease risk of volunteers. *PNAS* 110(42):16957–62
56. Posey JE, Rosenfeld JA, James RA, et al. 2016. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet. Med.* 18(7):678–85
57. Khera AV, Emdin CA, Drake I, et al. 2016. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* 375:2349–58
58. Abul-Husn NS, Manickam K, Jones LK, et al. 2016. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* 354(6319):aaf7000
59. Wald DS, Bestwick JP, Morris JK, et al. 2016. Child-parent familial hypercholesterolemia screening in primary care. *N. Engl. J. Med.* 375(17):1628–37
60. Evans JP, Berg JS, Olshan AF, et al. 2013. We screen newborns, don't we? Realizing the promise of public health genomics. *Genet. Med.* 15:332–34
61. Anderson MS, Su MA. 2011. AIRE and T cell development. *Curr. Opin. Immunol.* 23(2):198–206
62. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. 2017. Opportunities and obstacles for deep learning in biology and medicine. bioRxiv preprint 10:1101/142760
63. Obermeyer Z, Emanuel EJ. 2016. Predicting the future—big data, machine learning, and clinical medicine. *N. Engl. J. Med.* 375:13