High Throughput Sequencing and Assessing Disease Risk

Shannon M. Rego and Michael P. Snyder

Department of Genetics, Stanford University, Stanford, California 94305

Correspondence: srego@stanford.edu

High-throughput sequencing has dramatically improved our ability to determine and diagnose the underlying causes of human disease. The use of whole-genome and whole-exome sequencing has facilitated faster and more cost-effective identification of new genes implicated in Mendelian disease. It has also improved our ability to identify disease-causing mutations for Mendelian diseases whose associated genes are already known. These benefits apply not only in cases in which the objective is to assess genetic disease risk in adults and children, but also for prenatal genetic testing and embryonic testing. High-throughput sequencing has also impacted our ability to assess risk for complex diseases and will likely continue to influence this area of disease research as more and more individuals undergo sequencing and we better understand the significance of variation, both rare and common, across the genome. Through these activities, high-throughput sequencing has the potential to revolutionize medicine.

Since Watson and Crick first described the double helix, our understanding of the genome and ability to put that understanding to use in a medical setting has progressed at a staggering rate. In the intervening 60+ years, the genes underlying >3000 Mendelian phenotypes have been identified, the cost of whole-genome sequencing has dropped to $1000, and whole-exome sequencing has become a commonplace test in a clinical setting (see illumina.com/systems/hiseq-x-sequencing-system/system.html) (Chong et al. 2015).

High-throughput sequencing has played a notable role in these advances by drastically increasing the speed and reducing the cost with which sequencing can be performed. Despite this progress, genetic medicine remains a practice that is far more often reactive than proactive. To varying degrees, genetics can predict our future risk for a broad range of adverse medical outcomes, from cancer to diabetes and from coronary artery disease to Alzheimer's disease. For a number of highly penetrant Mendelian diseases, identifying a genetic mutation can equate with a near certain likelihood of developing or being born with a serious disease. Identifying this type of genetic risk can allow individuals to plan for future care needs, notify relatives of potential disease risk, make family planning decisions, increase disease surveillance, and, in some cases, to take preventative measures or make informed decisions about treatment options.

The potential benefits of genetic risk assessments are not limited to adults and children. For

S.M. Rego and M.P. Snyder

decades, newborn screening has identified at birth children with a variety of genetic diseases that often manifest suddenly and severely and which, when identified and treated early, can lead to much better outcomes. Genetic testing can tell us a great deal about a fetus, and even an embryo. Largely thanks to high-throughput sequencing, prenatal testing—long restricted to identifying chromosomal abnormalities visible on a karyotype—is now restricted more by ethical concerns than limitations of technology. When a known genetic disease exists in a family, parents routinely choose to use preimplantation genetic diagnoses in conjunction with in vitro fertilization to prevent the implantation of embryos with disease-causing mutations.

Many common medical conditions follow complex inheritance patterns and are caused by a combination of genetic risk spread across the genome plus environmental factors, and in these cases a genetic risk assessment can alert individuals to medical risks they may have some ability to influence; for example, an individual with increased genetic risk for developing type II diabetes may be able to mitigate this risk through dietary and exercise interventions. Improvements in genetic testing and our understanding of the genome will continue to make the process of assessing genetic disease risk faster and more accurate, which will, in turn, likely lead to better biological understanding of disease and better treatments and preventative measures. However, with these benefits come challenges, including both ethical and practical concerns, that will be important to address to realize the full benefits of these technologies.

PREDICTING MENDELIAN DISEASE RISK

Assessing Mendelian Disease Risk in Adults and Children

Mendelian disease—disease caused by a mutation in one or both copies of a single gene—is where genetic disease diagnoses and prediction began with Vernon Ingram's discovery of the genetic cause of sickle cell disease in 1956 (see genome.gov/pages/education/genetictimeline.pdf). Mendelian disease is the easiest type for which to

identify the underlying genetic cause, and, once identified, the easiest for which to predict disease risk. This is because, unlike complex disease, which is typically thought to involve multiple genes (see below), simple patterns of segregation of a single genetic mutation within a family can be used to pinpoint the underlying genetic cause of Mendelian disease.

The ability to predict risk for a Mendelian genetic condition begins with identifying the gene or genes in which mutations can lead to the phenotype in question. In some cases, this is simple, as the genetic cause of the phenotype has been previously identified and the inheritance pattern is known. In such cases, determining disease risk for a genetic condition begins with testing the gene or genes known to be associated with the disease presentation. Below is an example of such a situation.

Hypothetical Case Studies

Case 1. A 5-year-old boy presents to pediatric genetics clinic with café-au-lait spots (darkened patches of skin), freckling in the armpit and groin areas, and lisch nodules (characteristic spots in the iris of the eye). Based on these symptoms, he meets clinical diagnostic criteria for neurofibromatosis type 1, an autosomal dominant condition, and genetic testing can therefore be limited to the only gene in which mutations are known to cause this condition—the *NF1* gene. High-throughput sequencing of the *NF1* gene reveals that the patient is heterozygous for a stop-gain mutation, and loss-of-function mutations are the type of mutation previously associated with disease in this gene. There are now clear clinical and molecular diagnoses of neurofibromatosis type 1 in the patient, but what about his family members? To know the risk of other family members, we must first know the mutation status of the patient's parents. Neurofibromatosis type 1 is known to be caused by de novo mutations (new mutation in the affected individual) in 50% of cases, so there is a 50% chance that the patient inherited the mutation from one of his parents. It is noted in clinic that the father has two café-au-lait spots, but this alone is not enough for a clinical

diagnosis of neurofibromatosis and café-au-lait spots are fairly common in the general population, so both parents may need to be tested. If a mutation is found in one of the parents, then each of their children has a 50% risk of inheriting the same mutation and having neurofibromatosis type 1. If the mutation is not found in either parent, the most likely scenario is that the mutation was de novo and other family members are not at risk. However, in this case, there remains a very small risk that the patient's disease was caused by germline mosaicism, in which one of the parents has some normal gamete cells and some gamete cells with the mutation. Regardless of the source of the patient's mutation, if the patient should someday have children, each child will have a 50% chance of inheriting his mutation and developing neurofibromatosis type 1. Assessing disease risk for the patient and family members in this particular case is fairly easy because the disease phenotype has already been connected to a single gene, and the disease shows complete penetrance (individuals with a pathogenic mutation show symptoms 100% of the time). However, not all cases are so clear-cut.

Case 2. In a more complicated case, a 33-year-old woman is referred to a cancer genetics clinic because of a family history of breast and ovarian cancer (Fig. 1). She does not have a personal history of cancer. Her family history is indicative of hereditary breast and ovarian cancer (HBOC), an autosomal dominant condition in which loss-of-function mutations in one of several tumor suppressor genes lead to a high lifetime risk of breast and ovarian cancer. Because the patient does not herself have cancer, she is not the ideal candidate for genetic testing. The best person to test is the closest relative affected with cancer. In this case, it is the patient's father's side of the family that has the concerning history of cancer, but there are no affected relatives on that side of the family still living. However, the patient's sister died of breast cancer at age 35 and their father is still living. If her sister did have a mutation in an HBOC gene, she would most likely have inherited it from her father. Therefore, in this case, it makes the most sense to begin by testing the patient's father for mutations in the HBOC genes, the most commonly implicated of which are *BRCA1* and *BRCA2*. Genetic testing reveals
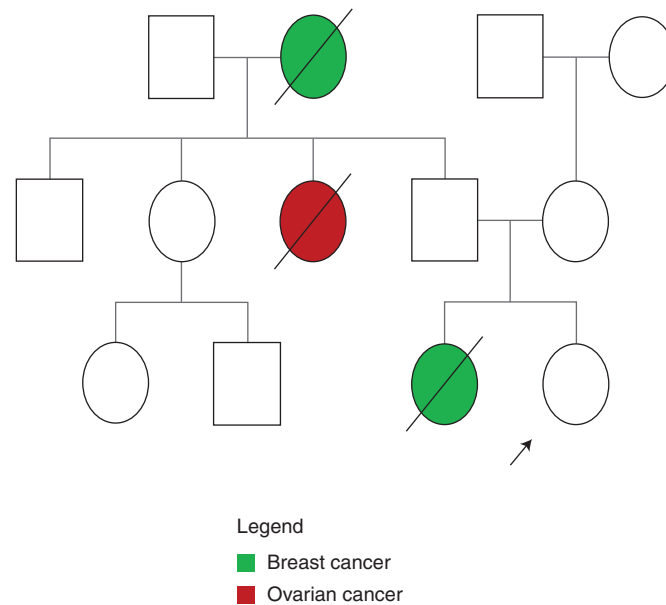

Legend
- 🟩 Breast cancer
- 🟥 Ovarian cancer

Figure 1. A pedigree showing a family history of breast and ovarian cancer. Circles represent women. Squares represent men. An arrow indicates the patient, the 33-year-old women described in case 2 in the text.

S.M. Rego and M.P. Snyder

that the patient's father does have a rare missense mutation in *BRCA1* that has not been previously reported. The laboratory classifies this variant as a variant of unknown significance; it is unclear whether this variant represents normal variation in the population or whether it is disease-causing and was therefore likely to have led to the patient's notable family history of cancer. In this case, because we do not know whether the mutation identified in the patient's father is causing the cancer phenotype in the family, testing the patient for this variant would not yield any practical benefit, and most genetics clinics would not test her for the variant. She would undergo high-risk HBOC screening with mammogram and/or magnetic resonance imaging (MRI) regardless of her mutation status and would be offered testing for her father's genetic variant only if in time it is reclassified as disease-causing or likely to be disease-causing. This patient's risk for developing breast and ovarian cancer is unclear, in part, because her mutation status is unclear, and it is further complicated by the fact that HBOC is a condition with reduced penetrance; not everyone who has a mutation shows symptoms of disease. This is particularly true of males with *BRCA1* gene mutations. They still have a small risk of developing male breast cancer and have risk for several other types of cancer that are elevated when compared to the general population risk. However, these risks are far less significant than the ∼70% lifetime risk of breast cancer and the 40% lifetime risk of ovarian cancer faced by females with *BRCA1* mutations (Brose et al. 2002). With reduced-penetrance conditions, estimating disease risk is not as simple as identifying a disease-causing mutation.

Clinical Impact of High-Throughput Sequencing

High-throughput sequencing has had a notable impact on how genetic disease is diagnosed in a clinical setting. Familial cancer syndromes, like the one described above, are an excellent example of a clinical setting in which high-throughput sequencing has dramatically changed the approach to testing for causative genes. High-throughput sequencing first made it possible to expand the known genes in which mutations cause familial breast cancer from *BRCA1* and *BRCA2* to a list that now includes dozens of genes. Before high-throughput sequencing, looking for a disease-causing mutation was a guessing game. If a patient had a striking family history of breast cancer, testing would begin with *BRCA1* and *BRCA2* and continue from there one gene at a time in the event that insurance would cover the sequencing cost or the patient was willing to pay out-of-pocket. Now, most genetic testing companies offer multigene panels for various types of cancer risk prediction, and dozens of genes may be sequenced in parallel to quickly identify a potential cause of the cancer phenotype in the family. In pediatric genetics settings, an increasing number of clinics are using whole-exome sequencing as a first-line molecular test, and in many cases this approach ends up costing less than ordering testing on multiple single genes or gene panels (Stark et al. 2016). The positive side of this development made possible by high-throughput sequencing is that gene discovery and diagnoses have become both much faster, and, in cases involving diseases like cancer and cardiomyopathy, identifying mutations faster can save lives. The down side is that testing more genes means higher chances for patients to receive variants of unknown significance—genetic variants where the significance is not clear because of lack of information. Often laboratories can reclassify such variants after a few years when more information becomes available, but, in the meantime, patients are often left with uncertainty that can cause a great deal of anxiety.

Another significant impact of high-throughput sequencing in Mendelian disease diagnoses is in the area of newborn screening. Since the 1960s, every state in the United States has implemented a screening program meant to identify children likely to have potentially deadly but treatable autosomal recessive genetic conditions at birth (Alexander 2003). Because autosomal recessive conditions often manifest in an infant with no previous family history of the disease, newborn screening provides a valuable opportunity to identify a condition early to minimize adverse outcomes. Though the conditions tested

for vary from state to state, the majority are metabolic genetic diseases in which dietary interventions, medications, or other treatments may mitigate outcomes that can range from progressive cognitive impairment to sudden metabolic decompensation and even death. Currently, most newborn screening is performed using a small sample of a newborn's blood to do biochemical screening. If a newborn tests positive for a condition, an important next step is high-throughput sequencing to identify potential disease-causing mutations. Mutations identified with high-throughput sequencing are then confirmed with Sanger sequencing, long considered the "gold standard" test for verifying genetic variants. Whether or not to change this process to forgo biochemical screening and begin with high-throughput sequencing of either a select list of genes or whole-exome or -genome sequencing is an active point of debate in the medical genetics and bioethics communities as high-throughput sequencing continues to get faster and cheaper (see blogs.cdc.gov/genomics/2014/12/30/newborn-screening).

High-Throughput Sequencing and Mendelian Disease Discovery

The earlier case studies (cases 1 and 2) involve established genotype–phenotype correlations, but determining risk is much more difficult for phenotypes that are either previously undescribed or that have been described but for which no causative genes have yet been identified (Fig. 2). The reason this task is so difficult is that the typical human genome contains millions of identifiable genomic variants, the vast majority of which represent normal variation. Narrowing these variants down to a disease-causing variant is performed by filtering, often based on information available in a variety of databases, for example, information about how common variants are in a population, whether or not they are in protein-coding regions, what their impact is on the protein product of the gene, etc. However, even after these filters are applied, one can be left with thousands of candidate variants, many of which are in genes that are poorly understood from a biological point of
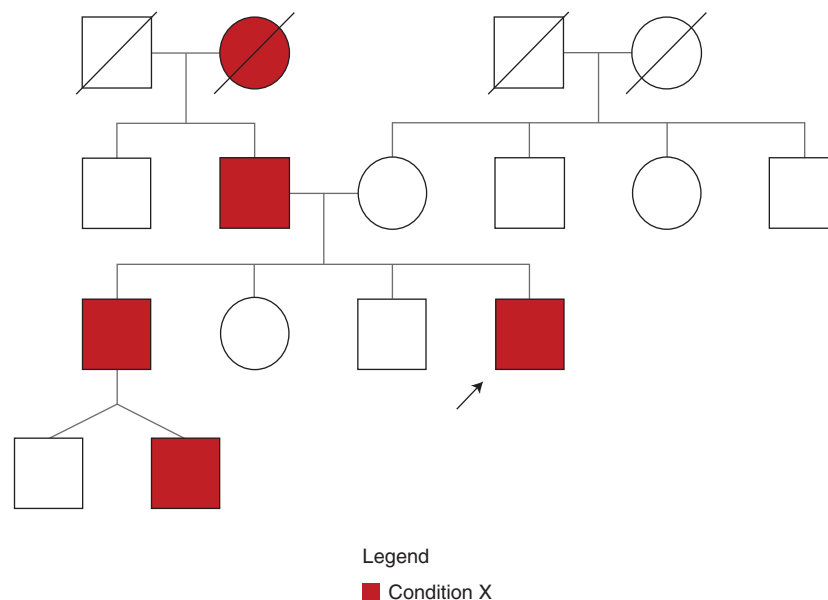

Legend
■ Condition X

Figure 2. A pedigree for a family showing autosomal dominant inheritance for a hypothetical unknown disease here called "condition X." Whole-exome or whole-genome sequencing can be performed on multiple affected and unaffected individuals to identify genetic changes present in all affected individuals but no unaffected individuals.

S.M. Rego and M.P. Snyder

view. In such cases, the most powerful tool available to assist in the identification of a disease-causing mutation is a pedigree, which can suggest likely patterns of inheritance. Combined with DNA samples from multiple affected and unaffected family members, this information about the disease status of each family member can be used to narrow the search for causative variants to those that segregate with the pattern of disease observed in the family (i.e., variants that exist in all affected individuals and no unaffected individuals). Below is a hypothetical example of just such a case in which a pedigree for undiagnosed disease X appears to show an autosomal dominant pattern of inheritance. If DNA samples can be provided from all living family members, it is possible to narrow down potentially disease-causing variants by searching only for variants that segregate with disease. In this case, because of the autosomal dominant pattern of inheritance suggested by the pedigree, we would search for heterozygous variants only. Without the ability to look for variants that segregate with disease in this way, in most cases it is nearly impossible to identify a disease-causing variant if it is located in a gene that has not been previously associated with the phenotype in medical literature. The reason for this is that even after filtering variants from an individual affected with disease based on other available information, there will be far too many variants left to investigate independently with the types of follow-up studies necessary to confirm the variants' significance.

High-throughput sequencing has dramatically impacted our ability to identify disease-causing mutations, and, consequently, to predict risk for Mendelian conditions. Prior to high-throughput sequencing, this was accomplished using linkage studies, which is a method by which a trait or disease is mapped to a location on the genome by demonstrating cosegregation of that trait or disease with genetic markers with known locations in the genome. This involved identification of a chromosome and broad region in which a genomic marker was segregating, and then a great deal more time and effort to identify the actual gene and variant responsible

for causing disease; this is a process that could take years if it was successful at all. The idea behind linkage studies is, like with disease X above, based in the use of a pedigree and DNA samples to use segregation to identify a causative variant. The major difference is that linkage studies entail sequencing very limited parts of the genome. Today, however, high-throughput sequencing has made it possible to sequence whole exomes and whole genomes quickly and affordably, providing a far better alternative. Rather than piecing segregation together one marker at a time, we can sequence whole exomes or whole genomes for multiple family members and, in the best-case scenarios, obtain an answer in a fraction of the time. This is especially important because many variants are rare or private and cannot be readily identified from large linkage studies. As a result of high-throughput sequencing, the pace at which genes are being linked to phenotypes has picked up.

Prenatal Risk Assessment for Mendelian Disease

In addition to diagnosing disease and predicting risk for disease in adults and children, high-throughput sequencing has also had a significant impact on how and when disease risk can be identified in newborns, fetuses, and even embryos. In these types of cases, far more than those involving adults and children, testing options are often limited more by ethical concerns than by the limits of technology.

In a prenatal setting, the landscape of risk assessment has changed markedly during the past several years and is likely to change even more in the near future. Screening for prenatal genetic conditions has typically focused on aneuploidies, which are chromosomal genetic abnormalities, including Down syndrome. Pregnant women could be identified as high risk for having a fetus affected by an aneuploidy for one of several reasons: (1) age (risk for having a child with an aneuploidy increases with maternal age), (2) ultrasound findings indicative of an aneuploidy, (3) previous pregnancy affected by an aneuploidy or family history, and (4) routine biochemical screening indicating higher

Cite this article as *Cold Spring Harb Perspect Med* 2019;9:a026849

risk. Women determined to be high risk for any of these reasons are given the option of undergoing invasive tests, including chorionic villus testing or amniocentesis to procure cell samples either from the chorionic villi (part of the placenta) or the amniotic fluid surrounding the fetus. These cells are then used to produce a karyotype that can be used to diagnose large chromosomal aberrations including aneuploidies. Chorionic villus sampling (CVS) and amniocentesis can also be used to detect genomic copy number variations, or to identify known mutations in the family such as point mutations and insertions or deletions.

High-throughput sequencing has opened many more options for prenatal diagnostic testing. Now, when ultrasound findings or family history give reason to suspect a fetus could be affected with a Mendelian disease, the mother can undergo amniocentesis and the resulting fetal DNA can be used for single-gene sequencing, multigene panel testing, or even whole-exome or whole-genome sequencing. The increasing use of broad sequencing tests such as whole-exome and whole-genome sequencing in a prenatal setting can improve the chances for diagnoses, but also poses ethical and practical challenges. These include the increased likelihood of identifying genetic variants of uncertain significance, which can cause stress and uncertainty for families. Additionally, there are numerous ethical questions involved in the decision of whether or not to do broad genetic testing on fetuses (Yurkiewicz et al. 2014). Such practices raise questions about the possibility of individuals terminating pregnancies for reasons other than serious diseases, which is a concern often summed up in popular media by the term "designer babies." Another concern is that of ownership of genetic information, that is, do parents have a right to their own child's genetic information, or does it belong to the child alone? What if a mutation is identified on prenatal whole-exome testing that causes adult-onset neurodegeneration? Do the parents have a right to know this information before their baby is even born or old enough to comprehend the information for him-/herself? These difficult questions are only some of those

that are being addressed as broad genetic sequencing becomes more prevalent in prenatal clinics.

In addition to the advances made possible by high-throughput sequencing, another recent development has also made a game-changing impact on the landscape of prenatal testing. Noninvasive prenatal testing (NIPT) is a form of prenatal genetic testing that uses the cell-free fetal DNA circulating in a mother's blood during pregnancy. This type of testing is currently used in a clinical setting primarily for detecting the most common trisomies, including Down syndrome. The benefits of this type of test over traditional CVS and amniocentesis are significant; they are performed with blood drawn from the mother and are therefore noninvasive and pose no risk of miscarriage. NIPT can also be offered as early as 10 weeks of pregnancy, whereas CVS and amniocentesis are not offered until several weeks later. CVS and amniocentesis are still considered the "gold-standard" tests for prenatal diagnostics, but the accuracy of NIPT is not far behind, and in time this type of testing may replace the more invasive options. Although NIPT is currently used in clinical settings primarily to identify fetal chromosomal abnormalities, it has already been shown in research settings that this technology can be used in conjunction with high-throughput sequencing for single-gene analysis and even for fetal whole-genome sequencing (Fan et al. 2012; You et al. 2014).

Assessing Mendelian Disease Risk in Embryos

When a genetic disease runs in a family and the causative mutation has already been identified in a family member, prospective parents may choose to initiate genetic testing before pregnancy. Preimplantation genetic diagnosis (PGD) is an option for parents who wish to ensure that their child does not have the previously identified mutation. This process involves producing several embryos in vitro, testing the embryos for the familial mutation, and selecting only embryos that do not have the mutation for implantation. The role of high-throughput sequencing in embryonic diagnosis is still limited for technical

S.M. Rego and M.P. Snyder

reasons that mainly have to do with turnaround time and, for the purpose of diagnosing embryos, the difficulty of performing high-throughput sequencing and getting accurate results with a very limited amount of DNA (Martín et al. 2013; Manegold-Brauer et al. 2014). However, it seems likely that these challenges will be overcome, and that high-throughput sequencing will eventually be used on a regular basis for these types of diagnoses, just as they are in the context of finding genetic causes of disease in adults and children. Assuming that time comes, we will be faced with many of the same ethical challenges and questions brought up by the use of high-throughput sequencing of fetuses, which is a perfect example of a situation in which advancements in technology have outpaced the development of the necessary ethical frameworks.

COMPLEX DISEASE

Whereas Mendelian disease is caused by mutations in one or both copies of a single gene, complex disease is, as per its name, much more complicated and therefore much more difficult to accurately predict risk for. Complex disease—disease caused by a combination of environmental factors (including diet and exercise, among other things) and the cumulative risk of variants spanning the genome—includes coronary artery disease, type II diabetes, osteoporosis, and asthma, to name a few.

We often see the evidence of genetic risk for complex diseases in families (individuals over multiple generations are affected), but there is no indication of a Mendelian inheritance pattern. Confounding efforts to assess genetic risk in these cases is the fact that many environmental factors will be the same or similar for family members. For example, if a child grows up in a family that eats a diet high in sugar, his diet may well be high in sugar as an adult. This influences his risk for a condition like type II diabetes, as well as genetic variants he shares with his parents. This confounding makes it difficult to determine how much of an individual's risk for a given disease is genetic and how much is environmental.

Heritability

In assessing risk for complex disease, one important question we often ask is "How much of an individual's risk for developing a given disease has to do with genetic factors and how much is environmental?" Fortunately, it is possible to quantify the heritability of a trait using twin studies. This type of study can and has been performed for all different types of traits and diseases showing complex inheritance, including height, sexual orientation, intelligence, autism, and diabetes, to name a few. Twin studies entail comparing the frequency with which monozygotic (identical) twins share a given trait to the frequency with which dizygotic (fraternal) twins share the same trait. Monozygotic twins share ∼100% of their genetic material and dizygotic twins share ∼50% of their genetic material (just like typical nontwin siblings). However, because dizygotic twins, like monozygotic twins, develop in the same womb and as siblings will generally be raised in very similar if not identical circumstances, most of the confounding factors are removed and we can ascribe much the difference in their traits and disease predisposition to genetics, especially for traits or conditions that are present at birth or at a very young age (before divergent environments have much of an impact). Heritability for a given trait or disease can range from 0 to 1, with 0 indicating the disease has no genetic component (for example, a viral illness) at all, and 1 indicating the disease is completely caused by genetics with no environmental impact (for example, cystic fibrosis). This number can be useful in considering disease risk. An early twin study of type II diabetes estimated the heritability at 25% (Poulsen et al. 1999). For an individual with a strong family history of type II diabetes, the practical import of this information is that their risk of developing the condition is strongly impacted by genetics but is much more strongly impacted by environmental factors, many of which are in that person's control.

Using Genome-Wide Association Studies to Identify Genetic Correlates of Risk

Estimates of heritability give us important information about the extent to which risk for a given

disease is determined by genetics, but they do not tell us how one person's genetic risk compares with another. For this, we have genome-wide association studies (GWAS). GWAS studies attempt to identify common genomic variants that are associated with increased or decreased risk for a particular disease or trait. These studies typically involve dividing a large cohort into two groups: one control group and one group that has the phenotype in question. Sites of common genetic variation called single-nucleotide polymorphisms (SNPs) are compared between these two cohorts with the goal of identifying SNPs that are significantly more common in one cohort than the other. In doing so, we can identify SNPs that are associated with either an increased or decreased risk for a disease or trait. This type of study has been performed for an incredible range of diseases and traits, including common diseases like hypertension and high cholesterol, traits like eye color and facility with math, and reaction and sensitivity to various types of drugs.

Once an SNP has been associated with increased or decreased risk for a given disease or trait, that information can be used to predict the likelihood of disease in individuals. If there are numerous SNPs that correlate with increased or decreased risk for a disease, the impact of those SNPs can be combined to produce a risk assessment incorporating the cumulative risk of all the associated SNPs. A number of companies, including 23andMe, have taken advantage of this approach to produce risk assessments for common conditions (Figs. 3 and 4).

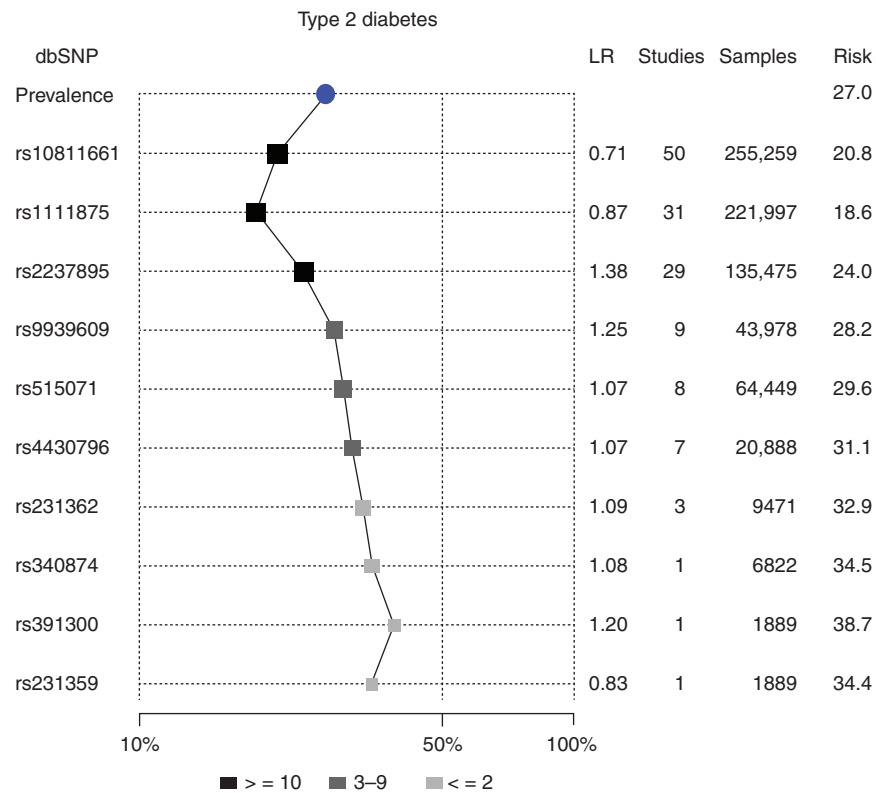| dbSNP	LR	Studies	Samples	Risk
Prevalence				27.0
rs10811661	0.71	50	255,259	20.8
rs1111875	0.87	31	221,997	18.6
rs2237895	1.38	29	135,475	24.0
rs9939609	1.25	9	43,978	28.2
rs515071	1.07	8	64,449	29.6
rs4430796	1.07	7	20,888	31.1
rs231362	1.09	3	9471	32.9
rs340874	1.08	1	6822	34.5
rs391300	1.20	1	1889	38.7
rs231359	0.83	1	1889	34.4

Figure 3. A sample risk assessment for an individual for type II diabetes. The blue dot indicates a baseline general population risk for a Caucasian male to develop type II diabetes in his lifetime. Each black and gray dot represents the cumulative impact on that risk of individual single-nucleotide polymorphisms (SNPs) (represented by the rsIDs on the *left*). The final risk number at the *bottom*, a 34.4% lifetime risk of developing type II diabetes for this individual, is based on the cumulative impact of 10 SNPs (Chen et al. 2012).
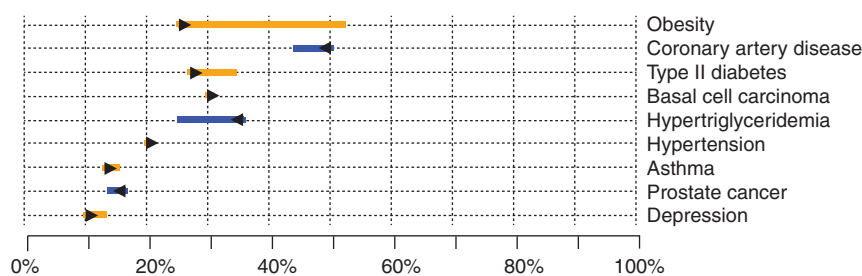
S.M. Rego and M.P. Snyder

Figure 4. A sample risk assessment for multiple conditions for the same individual in Figure 3. The arrow represents the baseline risk for developing the disease for a Caucasian male. The end of the colored line indicates the final risk taking into account genetic single-nucleotide polymorphisms (SNPs) identified in the individual. Orange lines indicate increased risk and blue indicate decreased risk (Chen et al. 2012).

GWAS-based risk assessments for complex diseases and traits are usually performed with microarray chip-based technology that makes it possible to look for variants at millions of sites of common variation across the genome. This methodology preceded high-throughput sequencing by several years, and the first GWAS studies came out long before whole-exome and whole-genome sequencing were fast and cheap enough to be used with any regularity for identifying genomic sites associated with disease risk. However, now that whole-exome and whole-genome sequencing are more cost and time effective than ever before, high-throughput sequencing is likely to play a much more important role in assessing risk for complex diseases and traits going forward. In addition to identifying sites of common genetic variation that influence risk, whole-genome and whole-exome sequencing will also make it possible to better understand the role of rare genetic variants in complex disease.

The most significant difference between GWAS-based risk assessment for complex diseases versus risk assessments for Mendelian conditions is the difference between correlation and causation. Assessing risk for Mendelian conditions involves identifying a mutation that is known or likely to "cause" disease. Risk assessments for complex conditions are generally based on identifying genetic variants that "correlate" with disease. Complex disease risk assessment is still an imprecise science in that different algorithms based on different GWAS studies can produce very different results. In time, more GWAS studies with more individuals may decrease the disparities between different algorithms and lead to a better, more standardized method of predicting complex disease risk. Understanding environmental contributions and incorporating this information with genetic risk will also facilitate better risk prediction for complex disease.

CONCLUSIONS

The ever-increasing speed and accuracy of next-generation sequencing has had a notable impact on the speed and scope of our ability to identify genes implicated in disease risk and diagnose genetic disease, and the pace of improvement has not slowed. Higher-throughput technologies are being developed and tested with the goal of decreasing processing time and cost of sequencing yet further. Overall, our development of ethical frameworks and standards for different applications of high-throughput sequencing and our understanding of the impact of various genetic alterations on an individual's health are far more limiting factors in assessing genetic disease risk than limitations of sequencing technology. As improvements in sequencing technology seem likely to continue to outpace developments in these other areas, we will face interesting times as we see how this disparity manifests in future efforts to assess genetic disease risk.

Cite this article as *Cold Spring Harb Perspect Med* 2019;9:a026849

REFERENCES

Alexander D. 2003. The National Institute of Child Health and Human Development and phenylketonuria. *Pediatrics* **112:** 1514–1515.

Brose MS, Rebbeck TR, Calzone KA, Stopfer JE, Nathanson KL, Weber BL. 2002. Cancer risk estimates for *BRCA1* mutation carriers identified in a risk evaluation program. *J Natl Cancer Inst* **94:** 1365–1372.

Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, et al. 2012. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148:** 1293–1307.

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, et al. 2015. The genetic basis of Mendelian phenotypes: Discoveries, challenges, and opportunities. *Am J Hum Genet* **97:** 199–215.

Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. 2012. Non-invasive prenatal measurement of the fetal genome. *Nature* **487:** 320–324.

Manegold-Brauer G, Hahn S, Lapaire O. 2014. What does next-generation sequencing mean for prenatal diagnosis? *Biomark Med* **8:** 499–508.

Martín J, Cervero A, Mir P, Martinez-Conejero JA, Pellicer A, Simón C. 2013. The impact of next-generation sequencing technology on preimplantation genetic diagnosis and screening. *Fertil Steril* **99:** 1054–1061.e3.

Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. 1999. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—A population-based twin study. *Diabetologia* **42:** 139–145.

Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, Yeung A, Peters H, Mordaunt D, Cowie S, et al. 2016. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med* **18:** 1090–1096.

You Y, Sun Y, Li X, Li Y, Wei X, Chen F, Ge H, Lan Z, Zhu Q, Tang Y, et al. 2014. Integration of targeted sequencing and NIPT into clinical practice in a Chinese family with maple syrup urine disease. *Genet Med* **16:** 594–600.

Yurkiewicz IR, Korf BR, Lehmann LS. 2014. Prenatal whole-genome sequencing—Is the quest to know a fetus's future ethical? *N Engl J Med* **370:** 195–197.

Cold Spring Harbor Perspectives in Medicine

www.perspectivesinmedicine.org

Cold Spring Harbor
Perspectives in Medicine

High Throughput Sequencing and Assessing Disease Risk

Shannon M. Rego and Michael P. Snyder

Subject Collection   Next-Generation Sequencing in Medicine

For additional articles in this collection, see http://perspectivesinmedicine.cshlp.org/cgi/collection/