Applications of high-throughput DNA sequencing to benign hematology

Vijay G. Sankaran1,2 and Patrick G. Gallagher3

1Division of Hematology/Oncology, Boston Children’s Hospital, and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; 2Broad Institute, Cambridge, MA; and 3Division of Neonatal/Perinatal Medicine, Departments of Pediatrics, Pathology, and Genetics, Yale University School of Medicine, New Haven, CT

The development of novel technologies for high-throughput DNA sequencing is having a major impact on our ability to measure and define normal and pathologic variation in humans. This review discusses advances in DNA sequencing that have been applied to benign hematologic disorders, including those affecting the red blood cell, the neutrophil, and other white blood cell lineages. Relevant examples of how these approaches have been used for disease diagnosis, gene discovery, and studying complex traits are provided. High-throughput DNA sequencing technology holds significant promise for impacting clinical care. This includes development of improved disease detection and diagnosis, better understanding of disease progression and stratification of risk of disease-specific complications, and development of improved therapeutic strategies, particularly patient-specific pharmacogenomics-based therapy, with monitoring of therapy by genomic biomarkers. (Blood. 2013;122(22):3575-3582)

Introduction

Hematology has a long history of being a discipline at the forefront of applying novel technology to understanding and diagnosing disease. Maxwell Wintrobe pointed out in his classic text, “Blood, Pure and Eloquent,” that important developments in the field of hematology were often driven by technology. For instance, early observations of blood and marrow morphology were enabled by the advances in microscopy. Similarly, quantitation of the different cellular elements of blood were made possible by development of the hemocytometer.1

Hematology continued to lead the way into the molecular era, with the description of sickle cell anemia as the first molecular disease,2 solution of the hemoglobin molecule by x-ray crystallography (the first multisubunit protein to be understood at the molecular level3), and determination of the molecular basis of sickle cell disease at the amino acid level.4 As Wintrobe’s classic text was being published, breakthrough technologies in the fledgling field of molecular genetics were leading to groundbreaking developments. Studying sickle cell disease and the thalassemia syndromes, investigators linked genetic polymorphisms to human disease, identified disease-causing mutations at the DNA level, and developed strategies for prenatal diagnosis (reviewed in Sankaran and Nathan5). A few years later, the first human disease gene isolated by positional cloning, the chronic granulomatous disease gene CYBB, was identified.6 Recently, advances in genomic technologies have led to numerous discoveries in hematology, detailed below.

As we make our way through the 21st century, technology continues to rapidly evolve. The use of next-generation DNA sequencing has dramatically advanced the way we assess gene expression, protein-DNA interactions, long-range DNA interactions, and both normal and pathologic DNA variation.7,8 This latter area is the focus of this review, which is one of a series of reviews on the application of high-throughput sequencing approaches to hematology. In this review, we discuss the application of these approaches to benign hematology, including red blood cell, neutrophil, and other white blood cell disorders. Malignant hematologic disorders and bleeding disorders, including abnormalities in platelets, are covered in other reviews. A review from Deborah Nickerson provides an overview of the high-throughput DNA sequencing technology. Because a comprehensive review is impractical, we provide vignettes that illustrate how high-throughput DNA sequencing is impacting hematology. These highlight applications of these approaches for disease diagnosis, gene discovery, and a better understanding of the genetic basis of complex traits. Prognostic and therapeutic implications of advances driven by sequencing technology are discussed and future applications highlighted.

Disease diagnosis

Genome-wide targeted exon capture followed by high-throughput DNA sequencing (whole-exome sequencing [WES]) is a technique that provides an unbiased analysis of coding exons and their associated splice junctions.9,10 WES is an excellent tool for identifying disease-causing mutations in monogenic disease and it can be applied to disorders with recessive or dominant inheritance, or de novo mutations (Figure 1). The cost of DNA sequencing has markedly decreased and computational platforms have rapidly improved, making WES a viable alternative to traditional Sanger sequencing–based techniques for genetic diagnosis.11,12 Excellent candidate disorders for utilizing WES are those that have great genotypic variability, that is, mutations in numerous genes lead to the same clinical phenotype, and/or where the genetic loci of interest are very large, where traditional sequencing strategies...
suffer from slow throughput. Indeed, in some cases, the use of WES or targeted exome sequencing for molecular diagnosis is cheaper than traditional Sanger sequencing. Targeted exome sequencing has been applied to many disorders with genotypic variability including the long QT syndrome, cardiomyopathy, and severe combined immunodeficiency syndrome. There are many potential applications for targeted WES in clinical hematology (Table 1). Clinical syndromes where targeted exome sequencing has already been applied for diagnostic purposes include Fanconi anemia (FA), the neuroacanthocytosis (NA) syndromes, and Diamond-Blackfan anemia (DBA).

FA is a heterogeneous bone marrow failure syndrome associated with defective DNA repair. Affected individuals exhibit cancer predisposition and frequently suffer from various congenital anomalies. Inherited primarily in an autosomal-recessive manner, over a dozen FA genes have been described. Application of exome sequencing to FA patients has identified a variety of mutations in FA-associated genes, several of which were novel. New gene discovery in FA genes has also been carried out with exome sequencing. A truncating mutation of the XRCC2 gene was discovered in a male child from a consanguineous Saudi family with an FA phenotype. XRCC2 is one of 5 RAD51 paralogs that act non-redundantly in the pathway of homologous recombination repair.

The NA syndromes are a group of heterogeneous neurodegenerative disorders that share the feature of having acanthocytes present on peripheral blood smear (Figure 1). NA syndromes include chorea-acanthocytosis (ChAc), X-linked McLeod syndrome (MLS), Huntington disease-like 2 (HDL2), and pantothenate kinase-associated neurodegeneration (PKAN). Diagnosis is difficult, particularly in the early stages of disease or when the presentation is atypical. Multiple genetic loci are involved and include mutations in chorein (VPS13A) in ChAc, XK (XK) in MLS, junctophilin-3 (JPH3) in HDL2, and pantothenate kinase 2 (PANK2) in PKAN. Most NA mutations are private, that is, each kindred has a unique mutation, making mutation detection difficult, and several of the NA genes are very large, making traditional Sanger sequencing cumbersome. Walker and colleagues used exome sequencing to identify compound heterogeneous mutations of the VPS13A gene in 2 NA patients, allowing precise genetic diagnosis and providing information for genetic counseling of affected patients and their family members.

Obtaining a precise molecular diagnosis when a patient presents with complex phenotypic features is another application of exome sequencing. Cullinane and colleagues studied a woman with oculocutaneous albinism, recurrent infections, bleeding diathesis, and neutropenia with the working clinical diagnosis of Hermansky-Pudlak syndrome. However, homozygosity mapping and exome sequencing identified mutations in 2 disease loci: the SLC45A2 gene locus associated with oculocutaneous albinism and the G6PC3 gene locus associated with congenital neutropenia. Additional findings of this woman and her sibling were described by Fernandez and coworkers.

Extending disease phenotype-genotype relationships and disease gene discovery

Making diagnoses in patients with hematologic disorders has proven valuable, as illustrated by the examples discussed in the prior section. In many hematologic disorders, much of the genetic etiology remains undefined. WES gives an opportunity to define and extend the spectrum of mutations causing a particular disease. DBA is a hypoplastic anemia characterized by a specific reduction in both mature red blood cells and their progenitors (Figure 2A). Approximately 50% to 70% of cases are attributable to mutations in ~10 different ribosomal proteins (RPs), the most frequent of which is RPS19, mutated in 25% of cases. Targeted WES has been used to study RP genes in DBA patients, identifying mutations in 15 of 17 patients in 1 study.

WES in a family with 2 affected male siblings with a clinical diagnosis of DBA without RP gene mutations identified mutations

Table 1. Potential applications of targeted WES in clinical hematology

<table>
<thead>
<tr>
<th>Potential applications</th>
<th>NA syndromes</th>
<th>Congenital neutropenia</th>
<th>Bone marrow failure syndromes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown hemolytic anemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfusion-dependent anemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congenital neutropenia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aplastic anemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyskeratosis congenita</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
in the critical X-linked hematopoietic transcription factor GATA1. An additional DBA patient was found to have similar mutations in GATA1. These mutations favor production of a short form of GATA1 that lacks the first 83 amino acids (Figure 2B). Further work is needed to understand how these mutations impair erythropoiesis and to explore whether any connection exists between these mutations and the more common RP gene mutations found in DBA. It is interesting to note that other GATA1 missense mutations found in the N-terminal zinc finger of this transcription factor result in very different phenotypes involving dyserythropoietic anemia, thalassemia, erythropoietic porphyria, and/or macrothrombocytopenia. These differences have been suggested to be due to variable effects on different GATA1 binding partner proteins and are distinct from the DBA-associated GATA1 mutations.

Iron-refractory iron-deficiency anemia is an autosomal-recessive hereditary xerocytosis (HX) is an autosomal-dominant hereditary xerocytosis (HX) characterized by primary erythrocyte dehydration (Figure 2C). Although a locus for HX had been identified at
is due to mutation in a conserved intronic element of the GATA2 gene.50

WES of patients from the French Severe Chronic Neutropenia Registry revealed GATA2 gene mutations in 7 kindreds with unexplained neutropenia, which in several cases evolved to MonoMAC or MDS/AML.51 These studies indicate that GATA2 gene mutations should be added to the list of congenital neutropenia genes associated with susceptibility to infection including the ELANE, HAX1, CXCR4, WAS, SBDS, GF11, and G6PC3 loci.

**Genetic basis of complex traits**

Genome-wide association studies (GWAS) have consistently demonstrated, with a few exceptions, that common variants have small to modest phenotypic effects, often requiring tens to hundreds of thousands of patient samples to have sufficient power to detect statistically-significant associations.9 When studying complex diseases or traits, a priori, it is unknown whether the likelihood of developing a specific disease complication or trait is more closely linked to common or rare variants. To address the former, many GWAS have been performed, attempting to identify common variants contributing to complex disease.52-56 These studies have revealed that in most complex diseases, common variants explain only a small fraction of genetic risk.

Recent studies suggest that rare, independent mutations may also contribute to phenotypic variation in complex diseases such as hypertension, hypertriglyceridemia, and cholesterol level variation.57-62 Thus, for genes subject solely to purifying selection, rare, independent mutations and not common variants predominate.58 Sequencing of linked genes identified by functional, genetic, or other methods, provide information to our understanding of the genetic contribution to complex disease. This has led to the studies that combine WES with genome-wide linkage or association analyses for identification of complex trait-associated variants.

These strategies have been successful leveraged to identify genes that may underlie common variation in hematologic traits found from GWAS.63-67 Galaneau and colleagues performed one of the first studies showing that rare variants in genes implicated from GWAS studies can help identify causal genes in these loci for hematologic traits.68 By sequencing candidate genes near loci implicated in fetal hemoglobin (HbF) level variation from GWAS, they were able to show that rare variants in MYB were associated with HbF levels. This suggests that MYB is likely, at least in part, responsible for the effect seen from GWAS signals at the 6q23 locus.

Auer and colleagues performed WES on 761 African Americans and then imputed newly discovered variants into a larger sample of 13,000 African Americans for association with traits for hemoglobin, hematocrit, white blood cell count, and platelet count.69 This led to discovery of association between coding variants in MPL and higher platelet count, CD36 and lower platelet counts, LCT and higher white blood cell count, and α-globin gene variants with lower hemoglobin. This was one of the first studies to demonstrate that imputation of low-frequency missense variants identified by WES onto GWAS data are a powerful approach to dissecting complex, genetically heterogeneous traits in population-based studies.69

Although promising, the appropriate numbers of patients needed to power these types of studies have not been determined and statistical approaches to test for such associations are rapidly being developed and are evolving.53,58,70

16q23-q24 by traditional linkage analysis, a number of factors complicated identifying the disease gene including a paucity of large, informative kindreds, and large blocks of repetitive, recombinant DNA sequence in the region containing the HX locus.34 Zarychanski and colleagues used WES to study individuals from one of the original HX kindreds from Rochester, NY and additional kindred from Winnipeg, MB.35 This led to discovery of mutations in PIEZO1, encoded by the FAM38A gene, in both HX kindreds (Figure 2D). These findings were confirmed in additional HX kindreds by 2 other groups who also used WES to identify the HX disease gene.36,37 PIEZO proteins are recently identified channels that mediate mechanotransduction in mammalian cells.38 In a large genome-wide association study, a single-nucleotide polymorphism near the PIEZO1 gene was strongly associated with cellular volume, as determined by mean corpuscular hemoglobin concentration.39 These findings indicate that this newly discovered protein plays an important role in erythrocyte volume homeostasis.

Con genital neutropenia and primary myelofibrosis are both very rare conditions in infancy. Five infants with neutropenia, recurrent and severe infections, defective platelet aggregation, myelofibrosis, and progressive bone marrow failure were studied by homozygosity mapping and WES.40 Independently, another group examined 7 patients with similar phenotypes in 5 families by linkage mapping and WES.41 These studies revealed missense mutations in VPS45, which encodes a protein that participates in trafficking in the endosomal pathway. Interestingly, fibroblasts from affected patients lacked lysosomes, suggesting a role for VPS45 in biogenesis of the endosomal-lysosomal pathway.40 This adds VPS45 to the growing list of lysosomal-related proteins associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.

Inherited aplastic anemia syndromes include FA and dyskeratosis congenita. However, the primary cause in a subset of familial aplastic anemia patients is unknown. Walne and coworkers studied 2 children from a consanguineous Tunisian family affected with familial aplastic anemia without a known genetic diagnosis.42 WES identified homozygous nonsense mutations in the thrombopoietin receptor gene, MPL. Previously, biallelic mutations in the MPL gene have been associated with congenital neutropenia, including those associated with Chediak-Higashi syndrome, Hermansky-Pudlak syndrome type 2, Griscelli syndrome, Cohen syndrome, and variants in the endosomal adaptor protein p14 associated with primary immunodeficiency.
Using findings from high-throughput sequencing to guide treatment

WES and other high-throughput sequencing approaches have already had and will continue to have a major impact upon disease diagnosis at the molecular level and disease gene discovery, but a major question that remains is whether these findings are actionable for therapeutic purposes. A few examples exist in the literature where targeted therapies resulted from the findings of WES. An excellent example of this involves the case of a child who presented at a young age with severe inflammatory bowel disease. However, a definitive diagnosis could not be made based upon clinical findings alone. WES helped to identify an XIAP mutation in this child, which is associated with a unique X-linked immunodeficiency syndrome. As a result, this child received an allogeneic hematopoietic stem cell transplant and had complete resolution of their previously intractable inflammatory bowel disease. In another example, WGS of an individual who was followed using a variety of genomic tools over a 14-month period predicted an increased risk of diabetes from the additive effects of multiple genetic variants present. The individual was found to have signs of glucose intolerance, and resultant lifestyle modifications led to improved glucose tolerance.

These examples are only single case reports and larger scale studies have yet to demonstrate whether such high-throughput sequencing approaches can have a major impact on therapy. The immediate utility of identifying new disease and quantitative trait loci is an improved understanding of disease pathogenesis. From these findings, the goal is the elucidation of novel, targeted therapeutic strategies. However, as is evident from study of numerous other Mendelian diseases, the road from genetic discovery, to understanding underlying biology, and ultimately to patient therapy is lengthy and plagued by a numerous hurdles along the way.

Ethical concerns and other considerations

Identification of actionable, incidental findings during genome-wide DNA sequencing genetic studies is a major concern of many patients, as well as health care providers. A well-known case is that of a patient who underwent WES in a search for autism-associated genetic variants. In the course of these studies, he was found to have pyruvate kinase deficiency. Known to suffer from an undiagnosed, lifelong hemolytic anemia, the results were conveyed to the patient’s hematologist, confirming the clinical diagnosis.

Many ethical and practical questions remain unanswered. How frequent are actionable findings found when performing WES or whole-genome sequencing (WGS)? Data are conflicting even answering this simple question. When WES or WGS data are obtained, should the data be curated for a set of specific variants? When potential deleterious variants are identified, should they be handled? Who should notify and counsel the patient? In 2012, the American College of Medical Genetics (ACMG), stated that for “results that are generated in the course of screening asymptomatic individuals, it is critical that the standards for what is reported be high to avoid burdening the health care system and consumers with what could be large numbers of false positive results.” In 2013, the Working Group on Incidental Findings in Clinical Exome and Genome Sequencing of the ACMG provided specific recommendations, focusing on incidental findings of clinical import with actionable results, identifying a subset of variants they felt laboratories have an obligation to report. Already concerns have been voiced about these recommendations. These discussions are beyond the scope of this review. Interested readers can consult recent discussion of these topics.

Whole-genome sequencing

WGS holds the promise of revealing the critical deleterious and at-risk alleles in an individual genome wide (or at least in those parts of the genome that can be sequenced, as discussed in the next paragraph below). Although the cost of WGS has dramatically fallen, issues of data storage, workflow and analysis, and clinical and ethical concerns persist. A recent report of patients who underwent WGS at the Medical College of Wisconsin revealed that a definitive diagnosis was obtained in 7 of 26 (27%) patients. Although initial concerns revolved around cost and data accuracy, the major challenges faced were logistics of delivering the data to clinicians, how clinicians used the genetic data, and how patients and their families dealt with incidental findings. Another major challenge that plagues the interpretation of WGS is the fact that alterations in most parts of the genome are still not interpretable, in contrast to the modifications in coding regions that cause clearly interpretable changes in amino acid sequence or splicing sites. As a result, even in cases where WGS is performed, often the analysis is limited to coding regions of the genome that can more readily be analyzed.

It is also important to remember that some diseases are caused by mutations located in regions of the genome that cannot be sequenced using even the latest WGS approaches, as was recently described for medullary cystic kidney disease. It is likely that some unidentified hematologic diseases may lie in such regions of the genome, which are refractory to current high-throughput sequencing approaches.

Implications for clinical hematology

From our discussion here, it is clear that there is a great deal of uncertainty in how best to interpret and apply the findings from WES and WGS. Nevertheless, these approaches are already showing promise for clinical applications and in some centers sequencing of patients by these approaches has already been initiated. As such, it is important for practicing hematologists to be aware of both the type of information produced from such approaches and the limitations of these findings.

In general, having both unaffected and affected family members undergo sequencing (at least confirmatory sequencing for any potential mutations identified) helps to delineate causal mutations from those that are likely to be uninvolved in the clinical phenotype of interest. The information gleaned from the clinical differential diagnosis can be useful to narrow potential candidate mutations, particularly when the high-throughput approaches are used for disease diagnosis. For example, if one was examining a child suspected of having a congenital neutropenia, then one could initially focus the analysis on genes already implicated as having a role in this condition, including ELANE, HAX1, CXCR4, WAS, SBDS, GFI1, G6PC3, GATA2, and VPS45. This set of genes should be assessed first if the goal is identifying mutations likely to cause the disease observed in the patient, just as individual gene sequencing using Sanger methods would be sent as clinical tests. In
cases where no mutations are identified, then it is possible that the individual has a new genetic cause of their disease, but without a sufficient number of other family members or without other confirmatory information, such findings can rarely be of immediate clinical utility without performing a variety of research tests. We recommend that when information from clinical WES or WGS is going to be reported to patients or their families, that the known clinical information be used to determine the likely differential a priori to narrow the search for potential causal variants and then any identified mutations be evaluated in light of the known clinical information. When there is discrepancy between the clinical findings and the identified mutations, appropriate evaluation is necessary prior to reporting such information to patients and their families.

In addition, as with any other diagnostic test, there can be a variety of false-positive results. Geneticists have traditionally not had large healthy control populations to examine and with the deluge of high-throughput sequencing data, the certainty of presumed “validated” mutations in a variety of human diseases is coming into question. This may be attributable to the finding that many of the disease gene mutation databases include a large number of false positives (potentially as high as 25% of the reported mutations) and also may reflect the concept that many human diseases show substantial incomplete penetrance that was previously unappreciated due to ascertainment bias.55,56

Conclusions and future directions

High-throughput DNA sequencing technology, which has already had an impact upon hematologic disease diagnosis and gene discovery, holds significant promise for the future. In the coming years, genomic studies will permit discovery of new disease genes and modifier alleles and provide important insights into disease pathobiology. These findings can be leveraged into better understanding of disease progression and allow stratification of risk of disease-specific complications. Novel, specific diagnostic approaches can be developed using genomic-based datasets. Improved therapeutic strategies can be created, particularly patient-specific pharmacogenomics-based therapy, with better monitoring of therapy by genomic biomarkers.

Much more work needs to be done to realize the lofty goals outlined here. Limitations of current technologies, for example, coverage of regions of the genome that are refractory to high-throughput sequencing methods, need to be addressed. Improving understanding of the relevance of disease-associated variants is needed, as are efficient strategies for functional validation of results obtained from sequencing studies. At the same time, development of approaches to handle incidental findings, from dealing with uncertainty in causality of variants of unknown significance, to reporting and counseling of actionable variants, is needed. The clinical validity and diagnostic utility for genetic testing for hematologic-associated disease need to be defined. The need for education of both patients and clinicians to embrace and fully understand and use genomic data is great.

Thus, as Wintrobe referred to historical technologic developments in his classic text, there is little doubt that hematology continues to be dramatically influenced by the use of newer technologies, as illustrated by high-throughput DNA sequencing.

Acknowledgments

This work was supported in part by RO1HL065449 (P.G.G.) and HL007574-30 (V.G.S.) from the National Institutes of Health, National Heart, Lung, and Blood Institute, and a grant from the Doris Duke Research Foundation (P.G.G.).

Authorship

Contribution: V.G.S. and P.G.G. designed, organized, researched, and wrote this review.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Vijay G. Sankaran, Division of Hematology/Oncology, Boston Children’s Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 3 Blackfan Circle, CL 03001, Boston, MA 02115; e-mail: sankaran@broadinstitute.org; and Patrick G. Gallagher, Division of Neonatal/Perinatal Medicine, Departments of Pediatrics, Pathology, and Genetics, Yale University School of Medicine, 333 Cedar St, PO Box 208064, New Haven, CT 06520-8064; e-mail: patrick.gallagher@yale.edu.

References


Applications of high-throughput DNA sequencing to benign hematology

Vijay G. Sankaran and Patrick G. Gallagher