Published in final edited form as:

Curr Allergy Asthma Rep. 2011 October ; 11(5): 352–360. doi:10.1007/s11882-011-0211-x.

Novel Sequencing-based Strategies for High-Throughput Discovery of Genetic Mutations Underlying Inherited Antibody Deficiency Disorders

Hong-Ying Wang and Ashish Jain

Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, CRC, 5W-3840, 10 Center Drive, Bethesda, MD 20892, USA

Hong-Ying Wang: wanghongying@niaid.nih.gov; Ashish Jain: ajain@niaid.nih.gov

Abstract

Human inherited antibody deficiency disorders are generally caused by mutations in genes involved in the pathways regulating B-cell class switch recombination; DNA damage repair; and B-cell development, differentiation, and survival. Sequencing a large set of candidate genes involved in these pathways appears to be a highly efficient way to identify novel mutations. Herein we review several high-throughput sequencing approaches as well as recent improvements in target gene enrichment technologies. Systematic improvement of enrichment and sequencing methods, along with refinement of the experimental process is necessary to develop a cost-effective high-throughput resequencing assay for a large cohort of patient samples. The Hyper-IgM/CVID chip is one example of a resequencing platform that may be used to identify known or novel mutations in patients with various types of inherited antibody deficiency.

Keywords

High-throughput sequencing; Microarray; Next-generation sequencing; Target sequence enrichment; Sequence capture by hybridization; PCR; Cost-effective; Assay development

Introduction

The humoral immune response comprises a series of sequential events involving activation of hundreds of gene products to orchestrate B-cell activation, survival, differentiation, and antibody production [1]. Genetic defects affecting B-cell development, T-cell function, class switch recombination (CSR) and somatic hypermutation (SHM), DNA damage repair, plasma cell differentiation, or memory B-cell survival can all lead to some kind of inherited antibody deficiency disorders, including agammaglobulinemia (no B cells and therefore no antibodies, such as Bruton's tyrosine kinase deficiency), hypogammaglobulinemia (low serum antibody levels), selective deficiencies of antibody subclasses (IgA deficiency, IgG deficiency), or selective deficiency of antibodies to certain protein antigens [2•, 3•, 4]. Among them, the common variable immune deficiency (CVID) and the hyper-IgM syndrome (HIGM) are two of the most widely investigated antibody deficiencies associated with marked hypogammaglobulinemia.

HIGM is a family of rare and severe primary immune deficiency disorders associated with defects in the B-cell maturation processes of CSR and SHM [5]. Several single-gene defects were identified in genes involved in CD40–CD40LG signaling and downstream DNA recombination/repair pathways, including *CD40LG*, *CD40*, *AICDA* (activation-induced cytidine deaminase), *UNG* (uracil-DNA glycosylase), and *NEMO* (nuclear factor- κ B essential modulator) [6–11]. In contrast, CVID is a group of relatively prevalent (1 in 25,000) and heterogeneous disorders with unknown genetic causes in the majority of cases [12]. Autosomal dominant mutations in *TACI* (transmembrane activator and CAML interactor) are suspected in approximately 10% to 15% cases of CVID, as well as selective IgA deficiency [13, 14], while recessive mutations in genes such as *ICOS* (inducible costimulator of activated T cells), *MSH5* (mutS homolog 5), *SH2D1A* (SH2 domain-containing protein 1A), *BAFFR* (B-cell-activating factor receptor), and B-cell receptor components CD19, CD20, and CD81 [15–21] were additionally identified in less than 1% of CVID cases. Because CVID patients are often complicated with other autoimmune, malignancy (cancer), or allergic conditions, it is possible that many cases of CVID have a polygenic etiology [22].

Despite diverse underlying genetic causes, most patients with antibody deficiency present with symptoms of low serum antibody levels and recurrent infection. As more and more disease susceptibility genes are identified, comprehensive molecular diagnosis is required to quickly pinpoint the exact genetic defect so that appropriate medical treatment can be administered [23]. Furthermore, identification of additional disease susceptibility genes or genetic modifiers that contribute to the disease severity may help us delineate the molecular mechanism of humoral responses as well as B-cell deficiency and B-cell malignancy disorders. At present, only a few single genes, such as *CD40LG*, *NEMO*, and *TACI*, are implemented in clinical genetic tests [24], which does not satisfy the needs of most patients. A high-throughput mutation screening strategy is therefore required to analyze a large number of disease susceptibility genes in large cohorts of patients.

High-Throughput Sequencing Strategy for Mutation Discovery and Identification of Novel Disease Susceptibility Genes

Unbiased Mutation Screening Approach on Whole Genome

Depending on the goal of the study and number of samples involved, we may choose whole genome sequencing, genome-wide association analysis (GWA), or whole exome sequencing [25]. Each approach has its own advantages and limitations. Whole genome sequencing is a comprehensive approach that can identify all types of mutation, but at the present time, it is very expensive to conduct. Alternatively, GWA can be used to identify quantitative trait loci in complex genetic diseases by statistically analyzing the co-segregation of each known single nucleotide polymorphism (SNP) or DNA marker under a special study design of family or population basis [26]. Because the trait loci identified under GWA are usually situated on genome intervals covering several megabases (Mb), a large sequencing and bioinformatics effort is still needed to localize the causative variants within those regions [27, 28]. Therefore, this approach is informative to identify disease susceptibility genes for rare genetic diseases but is not cost-effective.

In contrast, whole exome sequencing is a recently developed mutation-screening tool that shows strong potential in comprehensive genetic testing. As most known genetic diseases are caused by mutations in the coding sequences or intron–exon junction sites, whole exome sequencing may be used to quickly identify novel causative mutations affecting protein structure or function in any annotated exons [29, 30]. For the human genome, more than 30 Mb of unique exon sequences can be enriched for sequencing using the latest next-

generation sequencing (NGS) instruments. However, based on current technical limitations and associated cost, whole exome sequencing is only appropriate for studies involving a small number of patient samples.

Biased Mutation Screening on a Set of Candidate Genes

For large cohorts of samples, the candidate gene sequencing approach is much more popular and cost-effective. Candidate genes are generally selected based on their known biological functions and the developmental processes in which they may be involved. It is assumed that mutations in gene products associated with the same pathway will produce similar phenotypes [3•]. This seems to be true for HIGM and some cases of CVID because many known genetic defects are located in the CD40-CD40LG and APRIL (a proliferation-inducing ligand)/BAFF-TACI/BAFFR activation pathways. Therefore, direct mutation detection by sequencing a number of candidate genes responsible for B-cell development or neutralizing antibody response is clinically relevant and economically amenable [1].

To successfully develop a resequencing assay, each step, from patient enrollment, target gene selection, and enrichment to sequence detection and mutation analysis, must be carefully considered in terms of project scale, overall sequencing cost, mutation detection sensitivity and accuracy, experimental robustness, and labor time. With the advancement of modern target gene sequencing technology, a scalable mutation-screening assay covering tens to thousands of candidate genes can be established in a cost-effective way to sequence patients with various types of antibody deficiency. In the following sections, we focus on methods suitable for high-throughput target gene enrichment and resequencing and discuss current technological developments and possible ways to further improve these assays.

Current Approaches for Target Gene Sequence Enrichment

The most important features of sequence enrichment methods are their specificity and capture efficiency, irrespective of base composition and guanine–cytosine (GC) content, as the enrichment quality determines the final outcome of the whole mutation-screening project [31••].

In general, for a small number of candidate genes and a large number of samples, direct polymerase chain reaction (PCR) enrichment is the most effective and economical approach. For large-scale candidate gene sequencing and whole exome sequencing, sequence capture by hybridization is more appropriate due to its labor efficiency, regardless of target scale [32]. For midscale enrichment of 100 genes or less, PCR amplification or hybridization-based sequence capture may be used (Table 1). Additional enrichment methods such as molecular inversion probes that enrich target sequences from less than 1 µg of input DNA by highly multiplexing sequence capture are still in the developmental stage. Due to their relatively poor capture efficiency, reproducibility, and high probe cost [32, 33], we do not discuss them in detail here.

Target DNA Enrichment by Polymerase Chain Reaction and High-Throughput Optimization

PCR has long been used to amplify larger genomic regions (up to 20 kb with long-range PCR) and small exons with the use of specific primer pairs. However, to sequence hundreds of genes by PCR, a large number of primers must be designed and tested to ensure robust and uniform amplification of each amplicon. Moreover, thousands of single-plex PCR reactions will consume a lot of PCR reagents and DNA templates, which is sometimes not feasible when patient DNA samples are limited. To solve this problem, multiplex PCR involving two or more primer pairs can be set up between certain primer pairs of similar size, GC content, and amplification efficiency. To further increase the multiplexing capability, universal primer sequences can be attached to the 5' end of sequence-specific

primer pairs. After several cycles of sequence-specific amplification, a universal PCR is conducted to enrich all target sequences in the same tube [34]. Nonetheless, setting up many multiplex reactions in a uniform way is not easy, not to mention the incremental cost of special primer design.

To overcome this bottleneck, we developed duplex and short-/long-range PCR conditions to amplify 1,576 coding exons from 148 genes in our midscale mutation- screening study [35•]. To reduce assay time and workload, approximately 800 PCR primer pairs were pre-aliquoted into designated wells of PCR plates by a robot and air-dried. Additional primer plates were also prepared for any primer pairs showing weak amplification on agarose gel electrophoresis during the quality-control process. Therefore, all weak amplicons were amplified twice to increase their overall quantity (Fig. 1). As a result, many PCR plates can be quickly set up when patient samples are ready.

Because the PCR amplification efficiency of different target sequences is highly variable (largely due to sequence secondary structure, primer design, or template quality), to increase the evenness of coverage for all target sequences, many high-throughput sequencers require quantitation of each PCR product and pooling of equimolar amounts of individual amplicons before sequence detection. Equimolar pooling is relatively easy to conduct with a small number of large amplicons but is more laborious and costly when a large number of small amplicons are involved. For microarray-based sequencing in our mutation-screening studies [35•], additional normalization was omitted because we had amplified weak amplicons twice [36]. The entire volume in each well of the PCR plate can be pooled together in one tube for DNA purification and subsequently used for array hybridization after fragmentation and labeling, which dramatically reduces the overall labor time and cost [35•].

For NGS, because additional PCR amplification steps are involved in the preparation of sequencing library [37•], pooling of equimolar amounts of individual amplicons is essential for direct PCR-enriched targets; otherwise, certain amplicons may be overrepresented or underrepresented in the massive sequence output [35•].

Recently, a new type of multiplex PCR based on RainDance microdroplet (RainDance Technologies, Lexington, MA) PCR technology was developed in which a single primer pair is embedded in a water-in-oil droplet. The primer droplet is then merged at a 1:1 ratio with template droplet premixed with PCR reagents to create an amplification-ready droplet in which a single amplicon is produced. In this way, the amplification interference between different primer pairs is eliminated and the specificity and uniformity of target enrichment are dramatically improved. As reported, greater than 99% of all amplicons were successfully amplified, and 95% of targeted regions showed uniform coverage (read difference less than 10-fold) [38]. Currently, up to 4,000 amplicons (300–600 bp at 40%–60% GC content, maximum of 2 kb) can be simultaneously processed, and the capacity is expected to increase to 20,000 [31•], which is sufficient for many large-scale candidate gene sequencing projects. However, one major disadvantage of this microdroplet PCR technology is that expensive equipment such as the RDT100 is required for droplet production, and only one sample can be processed at a time in a several-hour time frame, which limits its application in high-throughput mutation-screening studies.

Target Sequence Enrichment by Hybridization

Hybridization-based target sequence enrichment technologies use tens of thousands of synthesized oligonucleotide probes on array surface or in solution or soluble biotinylated RNA probes, together with streptavidin-coated beads to capture target sequences from a pool of fragmented genomic DNA that is already ligated with linker or universal primer sequences. After washing and elution, the captured sequences are amplified by universal PCR for

downstream sequence detection [31••, 39]. Generally, array-based sequence capture is easy to conduct but has limited sample input and requires longer hybridization time, whereas soluble probe-based enrichment is more flexible in terms of the scale of capture content and number of DNA samples, although the experimental procedures are slightly more complicated [33, 39]. Unlike direct PCR enrichment, the same experimental effort is needed for sequence capture by hybridization regardless of the target scale, and there is no need to design and test a large number of primers; therefore, target sequences at the Mb level can be enriched easily in a short time period by hybridization-based sequence capture.

Sequence capture array was first applied to next-generation resequencing [40] and resequencing arrays [41] in 2007. Since then, several enrichment systems have become commercially available, such as the sequence capture microarrays from Roche NimbleGen (Basel, Switzerland) and the SureSelect solution-hybridization-based Target Enrichment System from Agilent Technologies (Santa Clara, CA). Both NimbleGen and Agilent provide custom format arrays or probes of 3- to 5-Mb sequence capture capacity, and whole human exome array or probes with 30-Mb or greater sequence coverage.

For hybridization-based sequence capture, shorter input DNA fragments at about 100 to 200 bp seem to perform better than longer fragments in terms of capture specificity and evenness of sequence coverage [42]. The design of capture probes is also important for capture efficiency and reproducibility across different samples. Shorter capture probes (≤ 120 -mer) together with overlapping tiling strategy usually perform better than longer probes with an end-to-end tiling strategy [43]. Generally, capture probes may not be used to target repetitive sequences or highly homologous sequences due to the problems of cross-hybridization and downstream sequence alignment, and target sequences that have very high adenine–thymine (AT) or GC content or are located near repeat-rich regions may not be fully captured due to poor hybridization or secondary structure [44]. The unique coding sequences of most exons can be captured efficiently (97%) compared with genomic intervals with many repetitive sequences (50%) [43], which makes the hybridization capture assay extremely suitable for whole exome sequencing.

As the sequences captured are a very small portion of the input DNA, prior- or post-enrichment PCR with universal primers is still necessary to increase the amount of captured sequences, which could introduce additional mutations or lead to uneven sequence coverage due to differential target sequence amplification efficiency [31••]. Therefore, compared with direct PCR enrichment, hybridization-based sequence capture methods usually have lower capture efficiency, specificity, and uniformity, resulting in lower target sequence coverage and higher error rate.

Another disadvantage of hybridization enrichment assay is that the reagents/arrays are expensive, and the workflow is usually complicated, with multiple steps of enzymatic processing and PCR amplification. To increase sample throughput and reduce per sample assay cost, tens of patient DNA samples may be barcoded by ligating the fragmented DNA with a unique 6-bp barcode sequence in the middle of the universal primer. Equimolar amounts of fragmented samples are then combined and enriched together in one assay [45, 46]. The efficiency of this multisample enrichment was found to be compatible with that of a single-sample enrichment assay if a set of barcode-targeted adaptors are present [46]. Interestingly, the sample barcode strategy is particularly suitable for analyzing a large number of samples on NGS platforms in one sequencing run [39], which dramatically reduces per sample cost for target sequencing of hundreds of genes. As the refinement of probe design and experimental processes, it is possible that sequence capture by hybridization systems will dominate the market for sequence enrichment in midscale to large-scale target gene sequencing and whole exome sequencing.

High-Throughput Platforms for Target Sequence Detection

Currently, the enriched target sequences are detected by three approaches: traditional Sanger sequencing, NGS, and high-density oligonucleotide microarray (Table 2).

Traditional Sanger Sequencing

Sanger sequencing uses four-color dideoxynucleotides to randomly terminate synthesis of newly elongated DNA chains at every base position. The resulting DNA molecules are separated by capillary electrophoresis. Although about 96 high-quality reads are generated per run, each with reading length of 500 to 1,000 bases, the throughput of Sanger sequencing cannot match the gigabase output of the latest next-generation sequencers. In addition, the DNA to be sequenced must be individually cloned or amplified and purified, which is labor intensive, costly, and time consuming, even for a limited number of genes [47]. Therefore, Sanger sequencing is generally used for confirming mutations detected by other high-throughput sequencing platforms in large mutation-screening projects.

Next-Generation Resequencing Technology

Currently, three major commercial NGS platforms are widely used in large sequencing centers: Roche 454 GS FLX (454 Life Sciences, Branford, CT), Illumina Solexa (San Diego, CA) genome analyzer, and Applied Biosystems SOLiD system (Life Technologies Corp., Carlsbad, CA), each with its own advantages and disadvantages (Table 2). The first two platforms are based on pyrosequencing or sequencing by DNA synthesis, whereas the SOLiD system is based on sequence-specific DNA ligation with a collection of two-base encoding octamers. During the process, fragmented and cloned DNA is immobilized on an array or beads, the 4-base nucleotides or 16 dinucleotide adaptors are sequentially passed through, and incorporated bases/dinucleotides in each sequencing reaction are recorded [37•, 48, 49]. Compared with Sanger sequencing, millions of parallel sequencing reactions can be performed simultaneously on NGS platforms, generating a huge amount of sequence data (Table 2).

At the present time, all three NGS platforms present a short read length and high base-call error rate due to their sequencing chemistries, and also have highly variable sequence coverage that is dependent on target sequence GC content, composition, and library preparation methods. At least 15-fold sequence coverage is required for reliable detection of variants and alignment of unique reads [35•, 44]. Because PCR is involved in target gene enrichment as well as generation of cloned sequence libraries, all NGS platforms are highly susceptible to experimental errors such as DNA contamination, chimera library between different samples, and an increased mutation rate from PCR. The end result is that multiple sequencing runs may be needed, which is very costly and time consuming [50]. An additional problem is that no efficient system exists to store and manage the huge image and raw sequence data generated per NGS run, not to mention that a large bioinformatics effort is required for downstream data analysis [51]. Therefore, to develop a truly cost-effective mutation-screening study on NGS platform, its sequencing chemistry and experiment robustness, as well as coverage of target sequences and per base coverage uniformity must be improved to increase the accuracy and sensitivity of variant detection and to reduce the overall sequencing depth required. In addition, to improve the extraction of high-quality data from the massive NGS output, more advanced sequence alignment tools for short-read sequences and paired-end sequencing are needed [52].

Sequence Detection by Array Hybridization

A resequencing array uses millions of unique 25-mer probes synthesized on the chip to detect sequence variations in up to 300 kb of known sequence [53]. Generally, eight probes

corresponding to forward and reverse DNA strands are used per base position, each varying in the middle position with A, G, C, or T. By analyzing the hybridization intensity of each probe set to target sequence, a base call corresponding to the probe(s) with highest hybridization intensity is made.

To date, resequencing arrays have been evaluated as diagnostics tools in many complex genetic diseases, such as severe combined immunodeficiency [54]. We have developed a custom resequencing array called the Hyper-IgM/CVID chip to screen single nucleotide polymorphism changes in patients with HIGM or CVID [35•]. Approximately 300 kb sequences are interrogated on the chip, which covers the coding sequences and 6-bp intron–exon junction sequences of 148 candidate genes implicated in T-cell– B-cell interaction, nuclear factor- κ B activation, CSR and SHM, and a gene set enriched in microarray studies of *NEMO*-deficient HIGM patients [55]. Several steps of the resequencing assay were optimized to make it faster and more robust (Fig. 1), such as employment of two-round enzymatic fragmentation reactions by DNase I and YM-100 column separation to ensure a uniform fragmentation of target DNA to the required hybridization size of approximately 100 bp (Fig. 1). In addition, we developed a Web-based data analysis tool called SNP Explorer that allows nonexperienced researchers to analyze resequencing array data and report known or novel mutations from all or selected samples in one simple step. Using this Hyper-IgM/CVID chip, a mean nucleotide call rate of 98.65% (range, 94.3%–99.6%) was achieved for DNA samples from diverse sources, and many known mutations were identified in addition to several novel mutations that are possibly deleterious [35•].

Further improvements to the array-based resequencing assay are possible. We found that shorter amplicons are better than longer amplicons for array hybridization, as they eliminate most of the nontargeted sequences that interfere with labeling, fragmentation, and hybridization of effective target sequences (personal observation). Zheng et al. [56] also developed a fully multiplexed enrichment pipeline in which 1,500 genes were captured and amplified using common primers, followed by allele enrichment with mismatch repair detection to generate separate pools of homogenous variant or nonvariant DNA for highly accurate array hybridization. By comparing the ratio of hybridization intensity between the variant pool and nonvariant pool, the false-positive rate ($\sim 10^{-5}$) and false-negative rate ($\sim 10\%$) on all heterozygotes and certain small indels were significantly reduced.

Nevertheless, microarray-based resequencing applications have several intrinsic limitations. First, the microarray can only be used to detect known sequences and single-base changes, not indels, unless they are known and presynthesized on the array. Second, because the microarray determines base composition by hybridization intensity, a high signal-to-noise ratio is required for accurate base calling; thus, a large amount of labeled target DNA (1 to ~ 50 μ g, depending on the target size and enrichment method) is needed for hybridization [57]. Third, only unique sequences can be interrogated on the chip to avoid cross-hybridization from repetitive and homologous sequences, and a higher false-positive rate is found for heterozygote detection in diploid genome sequences. Therefore, array-based mutation-screening assays can only be applied in specific cases [58].

Conclusions

Taken together, the recent rapid advancement of NGS technology represents the technology trend in future mutation discovery applications. Compared with resequencing arrays, NGS-based approaches do not need prior knowledge of genome annotation, and nanogram quantities of material are sufficient [58]. In combination with advanced target sequence enrichment technology, the massively parallel NGS may revolutionize large-scale clinical applications with its low cost, high specificity, and high throughput [52]. Currently, several

proof-of principle studies have been conducted in disease models such as retinitis pigmentosa and neurofibromatosis [59, 60], and all demonstrated sensitive and accurate mutation detection under sufficient sequence coverage ($>30\times$). As the cost and error rates of NGS continue to drop while the read length, sequence throughput, and data management and analysis tools continue to improve [52], we foresee their implementation for clinical diagnostics in the near future.

Although it may be inevitable that NGS platforms ultimately will replace many microarray-based resequencing applications, with their established experimental protocols combined with a fast and straightforward data analysis pipeline, microarrays may still fill a niche in certain applications, such as resequencing of small virus, bacteria, or mitochondria genomes [57]. For the human genome, the first US Food and Drug Administration–approved AmpliChip CYP450 from Roche (Basel, Switzerland) will still be used in clinical testing for some time [58]. In addition, as new mutations continue to be found in existing disease susceptibility genes, repeated sequencing of those genes on a standardized microarray platform is more robust and economical than Sanger sequencing or NGS. For example, the Hyper-IgM/CVID chip that we developed may provide rapid and cost-effective mutation screening of hundreds of disease-relevant genes in many antibody-deficient patients, as well as patients with certain other B-cell malignancy diseases.

We expect that the rapid identification of causative mutations and genetic modifiers by high-throughput sequencing tools will not only advance studies on gene function and gene interaction with respect to phenotype change but also facilitate personalized genome therapeutics for each participating patient.

Acknowledgments

Preparation of this manuscript was supported by the Intramural Research Program of the National Institutes of Health/National Institute of Allergy and Infectious Diseases.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
  - Of major importance
1. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood*. 2008; 112:1570–1580. [PubMed: 18725575]
  2. Conley ME, Dobbs AK, Farmer DM, et al. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu Rev Immunol*. 2009; 27:199–227. [PubMed: 19302039] This is an in-depth review of B-cell immunodeficiencies.
  3. Vale AM, Schroeder HW Jr. Clinical consequences of defects in B-cell development. *J Allergy Clin Immunol*. 2010; 125:778–787. [PubMed: 20371392] This is an in-depth review of B-cell development and B-cell immunodeficiencies.
  4. Wood PM. Primary antibody deficiency syndromes. *Curr Opin Hematol*. 2010; 17:356–361. [PubMed: 20442656]
  5. Davies EG, Thrasher AJ. Update on the hyper immunoglobulin M syndromes. *Br J Haematol*. 2010; 149:167–180. [PubMed: 20180797]
  6. Allen RC, Armitage RJ, Conley ME, et al. CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science*. 1993; 259:990–993. [PubMed: 7679801]
  7. Aruffo A, Farrington M, Hollenbaugh D, et al. The CD40 ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. *Cell*. 1993; 72:291–300. [PubMed: 7678782]

8. Ferrari S, Giliani S, Insalaco A, et al. Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc Natl Acad Sci USA*. 2001; 98:12614–12619. [PubMed: 11675497]
9. Revy P, Muto T, Levy Y, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*. 2000; 102:565–575. [PubMed: 11007475]
10. Imai K, Slupphaug G, Lee WI, et al. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol*. 2003; 4:1023–1028. [PubMed: 12958596]
11. Jain A, Ma CA, Liu S, et al. Specific missense mutations in NEMO result in hyper-IgM syndrome with hypohydrotic ectodermal dysplasia. *Nat Immunol*. 2001; 2:223–228. [PubMed: 11224521]
12. Eibel H, Salzer U, Warnatz K. Common variable immunodeficiency at the end of a prospering decade: towards novel gene defects and beyond. *Curr Opin Allergy Clin Immunol*. 2010; 10:526–533. [PubMed: 20864887]
13. Castigli E, Wilson SA, Garibyan L, et al. TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat Genet*. 2005; 37:829–834. [PubMed: 16007086]
14. Salzer U, Chapel HM, Webster AD, et al. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet*. 2005; 37:820–828. [PubMed: 16007087]
15. Grimbacher B, Hutloff A, Schlesier M, et al. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat Immunol*. 2003; 4:261–268. [PubMed: 12577056]
16. Sekine H, Ferreira RC, Pan-Hammarstrom Q, et al. Role for Msh5 in the regulation of Ig class switch recombination. *Proc Natl Acad Sci USA*. 2007; 104:7193–7198. [PubMed: 17409188]
17. Morra M, Silander O, Calpe S, et al. Alterations of the X-linked lymphoproliferative disease gene SH2D1A in common variable immunodeficiency syndrome. *Blood*. 2001; 98:1321–1325. [PubMed: 11520777]
18. Warnatz K, Salzer U, Rizzi M, et al. B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. *Proc Natl Acad Sci USA*. 2009; 106:13945–13950. [PubMed: 19666484]
19. van Zelm MC, Reisli I, van der Burg M, et al. An antibody-deficiency syndrome due to mutations in the CD19 gene. *N Engl J Med*. 2006; 354:1901–1912. [PubMed: 16672701]
20. Kuijpers TW, Bende RJ, Baars PA, et al. CD20 deficiency in humans results in impaired Tcell-independent antibody responses. *J Clin Invest*. 2010; 120:214–222. [PubMed: 20038800]
21. van Zelm MC, Smet J, Adams B, et al. CD81 gene defect in humans disrupts CD19 complex formation and leads to antibody deficiency. *J Clin Invest*. 2010; 120:1265–1274. [PubMed: 20237408]
22. Agarwal S, Cunningham-Rundles C. Autoimmunity in common variable immunodeficiency. *Curr Allergy Asthma Rep*. 2009; 9:347–352. [PubMed: 19671377]
23. Hsu AP, Fleisher TA, Niemela JE. Mutation analysis in primary immunodeficiency diseases: case studies. *Curr Opin Allergy Clin Immunol*. 2009; 9:517–524. [PubMed: 19841577]
24. Ameratunga R, Woon ST, Neas K, et al. The clinical utility of molecular diagnostic testing for primary immune deficiency disorders: a case based review. *Allergy Asthma Clin Immunol*. 2010; 6:12. [PubMed: 20529312]
25. Zhu M, Zhao S. Candidate gene identification approach: progress and challenges. *Int J Biol Sci*. 2007; 3:420–427. [PubMed: 17998950]
26. Smith JG, Newton-Cheh C. Genome-wide association study in humans. *Meth Mol Biol*. 2009; 573:231–258.
27. Pare G. Genome-wide association studies—data generation, storage, interpretation, and bioinformatics. *J Cardiovasc Transl Res*. 2010; 3:183–188. [PubMed: 20560038]
28. Roberts R, Wells GA, Stewart AF, et al. The genome-wide association study—a new era for common polygenic disorders. *J Cardiovasc Transl Res*. 2010; 3:173–182. [PubMed: 20560037]
29. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA*. 2009; 106:19096–19101. [PubMed: 19861545]

30. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet.* 2010; 19:R145–R151. [PubMed: 20705737]
31. Turner EH, Ng SB, Nickerson DA, et al. Methods for genomic partitioning. *Annu Rev Genom Hum Genet.* 2009; 10:263–284. This is a comprehensive review of all sequence enrichment methods.
32. Turner EH, Lee C, Ng SB, et al. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods.* 2009; 6:315–316. [PubMed: 19349981]
33. Teer JK, Bonnycastle LL, Chines PS, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res.* 2010; 20:1420–1431. [PubMed: 20810667]
34. Meuzelaar LS, Lancaster O, Pasche JP, et al. MegaPlex PCR: a strategy for multiplex amplification. *Nat Methods.* 2007; 4:835–837. [PubMed: 17873887]
35. Wang HY, Gopalan V, Aksentijevich I, et al. A custom 148 gene-based resequencing chip and the SNP explorer software: new tools to study antibody deficiency. *Hum Mutat.* 2010; 31:1080–1088. [PubMed: 20652909] This method paper used an optimized resequencing array assay and bioinformatics tool to screen mutations in patients with antibody deficiency.
36. Kothiyal P, Cox S, Ebert J, et al. An overview of custom array sequencing. *Curr Protoc Hum Genet.* 2009; Chapter 7:Unit 7 17. [PubMed: 19360699]
37. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet.* 2010; 11:31–46. [PubMed: 19997069] This is a comprehensive review of NGS technology.
38. Tewhey R, Warner JB, Nakano M, et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol.* 2009; 27:1025–1031. [PubMed: 19881494]
39. Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods.* 2010; 7:111–118. [PubMed: 20111037]
40. Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods.* 2007; 4:903–905. [PubMed: 17934467]
41. Okou DT, Steinberg KM, Middle C, et al. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods.* 2007; 4:907–909. [PubMed: 17934469]
42. Mokry M, Feitsma H, Nijman IJ, et al. Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res.* 2010; 38:e116. [PubMed: 20164091]
43. Tewhey R, Nakano M, Wang X, et al. Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol.* 2009; 10:R116. [PubMed: 19835619]
44. Hoppman-Chaney N, Peterson LM, Klee EW, et al. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. *Clin Chem.* 2010; 56:1297–1306. [PubMed: 20562348]
45. Kenny EM, Cormican P, Gilks WP, et al. Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection. *DNA Res.* 2011; 18:31–38. [PubMed: 21163834]
46. Nijman IJ, Mokry M, van Bostel R, et al. Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. *Nat Methods.* 2010; 7:913–915. [PubMed: 20953175]
47. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics.* 2009; 93:105–111. [PubMed: 18992322]
48. Nowrousian M. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell.* 2010; 9:1300–1310. [PubMed: 20601439]
49. Smith DR, Quinlan AR, Peckham HE, et al. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* 2008; 18:1638–1642. [PubMed: 18775913]
50. Taub MA, Corrada Bravo H, Irizarry RA. Overcoming bias and systematic errors in next generation sequencing data. *Genome Med.* 2010; 2:87. [PubMed: 21144010]
51. Koboldt DC, Ding L, Mardis ER, et al. Challenges of sequencing human genomes. *Brief Bioinform.* 2010; 11:484–498. [PubMed: 20519329]
52. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem.* 2009; 55:641–658. [PubMed: 19246620]

53. Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet.* 1999; 21:42–47. [PubMed: 9915500]
54. Lebet T, Chiles R, Hsu AP, et al. Mutations causing severe combined immunodeficiency: detection with a custom resequencing microarray. *Genet Med.* 2008; 10:575–585. [PubMed: 18641513]
55. Jain A, Ma CA, Lopez-Granados E, et al. Specific NEMO mutations impair CD40-mediated c-Rel activation and B cell terminal differentiation. *J Clin Invest.* 2004; 114:1593–1602. [PubMed: 15578091]
56. Zheng J, Moorhead M, Weng L, et al. High-throughput, high-accuracy array-based resequencing. *Proc Natl Acad Sci USA.* 2009; 106:6712–6717. [PubMed: 19342489]
57. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic.* 2009; 8:174–183. [PubMed: 19535508]
58. Teng X, Xiao H. Perspectives of DNA microarray and next-generation DNA sequencing technologies. *Sci China C Life Sci.* 2009; 52:7–16. [PubMed: 19152079]
59. Chou LS, Liu CS, Boese B, et al. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem.* 2010; 56:62–72. [PubMed: 19910506]
60. Simpson DA, Clark GR, Alexander S, et al. Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. *J Med Genet.* 2011; 48:145–151. [PubMed: 21147909]
61. Fisher S, Barry A, Abreu J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 2011; 12:R1. [PubMed: 21205303]
62. Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics—principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn.* 2010; 12:539–551. [PubMed: 20805560]

Fig. 1. Overview of the resequencing microarray assay on the Hyper-IgM/CVID chip and high-throughput assay optimization. PCR—polymerase chain reaction. (Data from Wang et al. [35].)

Table 1

Performance and cost comparison of major candidate gene enrichment methods

	Direct PCR	Microdroplet PCR	Solution hybridization–based sequence capture	Array hybridization–based sequence capture
Major providers	–	RainDance Technologies	Agilent, NimbleGen	Agilent, NimbleGen, Febit ^a
Probes	Sequence-specific oligo primer pairs	1 primer pair per droplet	50- to 170-mer biotinylated RNA or oligo probes, no repeats	50- to 100-mer oligo probes, no repeat regions
Input DNA amount	Based on target scale	2–8 µg	<3 µg	1–20 µg
Special equipment	–	RDT100	–	Hybridization and elution systems
Sample throughput	Low	Low	Up to thousands with automation and sample barcode	Low: up to tens with sample barcode strategy
Cost per sample	Increases with target scale	Moderate	Same cost for large-scale target sequence capture	
Capture specificity, %	>93	>80	50–80	
Targets covered, %	>99.5	>95	>90	
Uniformity of coverage	Highest after normalization	High	Variable	
Fold enrichment	High	High	Moderate to high based on target or sample scale	
Target scale	Low to mid	Mid: up to 10 Mb	Mid to high: 125 k–1 Mb for HybSelect ^a ; 3–5 Mb custom capture, 30+ Mb for whole exome capture	

^aFebit Holding GmbH, Heidelberg, Germany*Mb* megabyte; *PCR* polymerase chain reaction

(Data from Turner et al. [31••], Mamanova et al. [39], and Fisher et al. [61].)

Table 2

Performance comparison of common sequencing detection platforms

	Sanger sequencing	Resequencing microarray	Next-generation sequencing
Latest platform	ABI 3730xl DNA Analyzer ^a	Affymetrix Custom Resequencing Array ^b	Roche 454 GS FLX Titanium
Sequencing chemistry	Dideoxynucleotide chain termination	Sequencing by hybridization	Pyrosequencing
Preferred template DNA	Cloned or PCR-amplified DNA	Mixture of PCR-amplified targets	Mixture of DNA fragments captured by hybridization and amplified by universal primers
Run time	2 h	16 h for hybridization	10 h
Read length	500–1,100 bases	Not limited	Up to 500 bases
Data output	96 kb/run	300 kb/array	400 Mb/run
Estimated cost ^c	~\$1–\$2 per read	\$300/array	\$60/Mb
Preferred mutations	SNP and indel	SNP	SNP
Raw base call accuracy, %	>99.99	Up to 99.99 for haploid genomes	>99
Limitation	1 target per reaction	Only unique known sequence or known indels targeted	Short sequence reads, indel detection and homologous sequence alignment problems; need strong central processing unit power and bioinformatics support
Sequencing applications	Low-throughput, mutation confirmation	Mid-throughput resequencing	High-throughput resequencing
			High-throughput, highly accurate resequencing

^aLife Technologies Corp., Carlsbad, CA

^bSanta Clara, CA

^cDoes not include cost of target sequence enrichment and capital equipment for sequencing

Mb megabyte; PCR polymerase chain reaction; SNP single nucleotide polymorphism

(Data from Kothiyal et al. [36], Metzker [37•], Nowrousian [48], Voelkerding et al. [52], Hacia [53], and Voelkerding et al. [62].)