Genomics of Immune Diseases and New Therapies*

Michael Lenardo, Bernice Lo, and Carrie L. Lucas

Molecular Development of the Immune System Section, Laboratory of Immunology, and Clinical Genomics Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; email: lenardo@nih.gov

Annu. Rev. Immunol. 2016. 34:121–49

First published online as a Review in Advance on December 23, 2015

The *Annual Review of Immunology* is online at immunol.annualreviews.org

This article's doi: 10.1146/annurev-immunol-041015-055620

Keywords

genomics, biochemical, mechanism, immune disorder, targeted therapy

Abstract

Genomic DNA sequencing technologies have been one of the great advances of the 21st century, having decreased in cost by seven orders of magnitude and opening up new fields of investigation throughout research and clinical medicine. Genomics coupled with biochemical investigation has allowed the molecular definition of a growing number of new genetic diseases that reveal new concepts of immune regulation. Also, defining the genetic pathogenesis of these diseases has led to improved diagnosis, prognosis, genetic counseling, and, most importantly, new therapies. We highlight the investigational journey from patient phenotype to treatment using the newly defined XMEN disease, caused by the genetic loss of the MAGT1 magnesium transporter, as an example. This disease illustrates how genomics yields new fundamental immunoregulatory insights as well as how research genomics is integrated into clinical immunology. At the end, we discuss two other recently described diseases, CHAI/LATAIE (CTLA-4 deficiency) and PASLI (PI3K dysregulation), as additional examples of the journey from unknown immunological diseases to new precision medicine treatments using genomics.

> In almost all things, what they contain of useful or applicable nature is hardly perceived unless we are deprived of them, or they become deranged in some way.
>
> William Harvey (1657)

INTRODUCTION

Medical genetics became a medical subspecialty in the 1950s with the hope that understanding the genetic basis of disease would accelerate diagnosis, prognosis, and treatment (1). The first genetic (primary) immunodeficiency, Bruton's agammaglobulinemia, and the first successful etiological therapy, intravenous immunoglobulin replacement, were defined in 1952 (2). However, four decades passed before genetic and molecular technologies advanced enough to permit the identification of the causative gene, *Bruton tyrosine kinase* (*BTK*) (3, 4). New technologies for the investigation of genes have resulted in a cornucopia of newly defined immunodeficiencies and immunoregulatory diseases and an explosion of knowledge of human immune regulation (5, 6).

Genomics is the investigation of nucleotide sequence variants in the entire nuclear DNA content of an organism and their effects on phenotype. For most of the history of medical genetics, the molecular state of the genome in a patient's cell was essentially unknowable. Next-generation sequencing (NGS) technologies have now made this knowledge routine. Molecular genetic analysis in humans has also been accelerated by new techniques that allow biochemical and molecular analysis of human genetic defects with a rigor approaching that previously possible only in model organisms. These include efficient cell cultivation, improved electroporation techniques to introduce DNA into cells, altering gene expression with interfering RNA, and genomic editing with tools such as the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system (7). Gene mapping and candidate gene analyses for primary immunodeficiencies have been largely replaced with NGS (5, 6). Computational approaches have streamlined the identification and validation of causal gene variants. Although cross validation of human disease genes with genetically altered mice (a species that shares 99% of its genes with humans) is a mainstay of immunology, many human diseases cannot be modeled in the mouse, such as caspase-8 deficiency and CTLA-4 haploinsufficiency (8, 9). Thus, we are now capable of diving into an ocean of 3 billion nucleotides with a reasonable hope of swimming back to the surface with pearls of knowledge about the immune system. This essay is about our experience making such deep-sequence explorations.

APPROACHES TO STUDYING THE GENETIC CONTRIBUTION TO DISEASE

Uncovering the sequence of the human genome introduced the possibility that genomics could yield new insights into cellular regulation and disease mechanisms (10). However, knowledge about function required comparative genomics investigating DNA sequence variations between individuals, especially those with specific diseases. Early on, cost prevented wide deployment of DNA sequencing, so alternative genomic screening strategies were developed. The hypothesis emerged that common single-nucleotide polymorphisms (SNPs) would correlate with common diseases (the common SNP–common disease hypothesis) (11). In a procedure called genome-wide association (GWA), relatively inexpensive hybridization chips were used to probe common, well-defined SNPs to correlate predominant allele frequencies with a variety of diseases, including most common autoimmune diseases. For the SNPs on these chips, MHC class II alleles consistently showed significant and reproducible linkage disequilibrium with immunological diseases, but other associations showed only slight linkage (12). Specifically, very few SNPs in the coding portion of the genome had a large effect for any disease (11–13). Another complication of GWA studies was

that many SNPs correlated with disease were in noncoding regions, had modest contribution to disease, and were not medically actionable (11–13). This was puzzling and did not explain inheritance inferred from other analyses, such as twin studies. Also, these data were hard to interpret because they were based on statistical correlation; determining causation, though accomplished in some cases, was difficult (11, 12). For certain complex diseases, GWA studies suggest that a group of gene variants, each with a small effect, act in concert to produce the disease phenotype (11–13).

As the cost of sequencing DNA dropped precipitously and access to NGS technology became widely available, emphasis shifted to direct gene sequence investigation. Currently genomic investigation uses NGS of exonic DNA captured by hybridization chips, the exome, or the entire nuclear DNA of a cell from a human subject, the genome. Study of individual exomes and genomes, now increasing exponentially and including a panoply of disease cohorts, yielded a big surprise. Every human genome contains a surprisingly immense number of rare variants compared to the reference genome. These are usually private, meaning that they are restricted to the person's immediate blood relatives: parents, siblings, and children (14). Hence, the genomic "software" for every newborn is unique and has many variants—generally defects but occasionally improvements—that affect cellular programs. Furthermore, in contrast to common SNPs with weak effects, rare single-nucleotide variants (SNVs) and copy number variants (CNVs) can have highly penetrant and deleterious effects on phenotype (14). Single-nucleotide changes found at frequencies greater than 1% in a population are called SNPs, and those less frequent, usually exceedingly rare or unique to an individual, are called SNVs. How can we interpret the difference in disease heritability between common SNPs and rare SNVs? Any DNA change arising in a family with strong deleterious effects would be eliminated from the population as it is passed on to successive generations because of decreased reproductive fitness. Thus, common SNPs would have been selected for weak disease effects in order to be successfully transmitted within the population. A caveat to this assertion is that common SNPs could not have been selected against by recent specific environmental influences such as drugs and toxins, and therefore some may still have unexpectedly potent deleterious effects. On the other hand, most rare variants would be limited to one or a small number of generations in a family, depending on how severely they affect reproductive fitness. The concept of rare, family-restricted variants that have severe phenotypes has been called clan genomics, and pursuit of these variants has had a major impact on genomic investigation of diseases (14). Rare variation with large effect sizes is likely contributing a significant proportion to the missing heritability of complex traits and disease (14, 15).

An astronomical number of private variants (thousands or more per genome) with low minor allele frequencies (MAFs) have been uncovered in human genomic sequences around the world. This is likely because of our rapid evolutionary success and the dramatic increase of human populations in a comparatively short span of evolutionary time (16). There may also be increasing retention of disease susceptibility variants because of better health due to hygiene and medicine. The other feature of clan genomics is homozygosity of private variants through consanguinity. Early in human history, certain isolated hunter-gatherer clans likely practiced a high level of endogamy, causing homozygous retention of private variants through consanguinity (17). As small populations moved or suffered from disease or starvation, the gene pool would have been subject to genetic drift and founder effects through population bottlenecks. Even today, this tendency to consanguinity has persisted through social, economic, and religious forces. In large regions of the world, accounting for roughly 10% of humanity, consanguineous reproduction retains and promotes homozygosity of private SNVs, especially deleterious loss-of-function (LOF) mutations, without dispersing them throughout the regional or national population (16, 17).

Powerful and evolving NGS technologies now provide the principal approach to identifying disease-associated SNVs and short (<50 bp) insertions and deletions (indels) (5, 6, 15, 18, 19).

Whole-genome sequencing (WGS) of paired ends, comparative genomic hybridization (CGH) arrays, and chromosome analysis can identify structural variations (SVs, defined as >50 bp), including CNVs and more complex genomic rearrangements (18, 19). Individual genomes may harbor nearly 10,000 SNVs, affecting almost 2% of the genome, and these are an important source of disease variants (18–20). These new genomic technologies provide an unbiased method to identify gene mutations in Mendelian, monogenic, and de novo inheritance patterns. Not all monogenic disorders can be considered Mendelian; de novo gene variants that are congenital but not inherited, and by definition private, cause important disease phenotypes (18, 20). Family-based WGS studies have estimated that each individual's genome contains many germline de novo mutations (DNMs) and that these are an important source of disease mutations (20–22). These mutations, found in the affected proband but not in the parents or siblings, are potentially more deleterious, because they have not been subjected to natural selection. DNA sequence analysis of selected cell populations can also uncover somatic mutations (SM) such as *FAS* gene mutations that are present in a subpopulation of CD4 and CD8 double negative T cells but not in the germline of certain autoimmune lymphoproliferative syndrome (ALPS) patients (23). Somatic reversion can also ameliorate disease phenotype by correcting the genetic defect in appropriate cell lineages (24).

In consanguineous families, NGS of genomes of affected individuals together with parents and other affected or unaffected relatives combined with homozygosity mapping can efficiently identify disease-associated variants. Nonconsanguineous; large, multigenerational; and multiplex pedigrees can also be used to identify rare inherited variants. NGS has caused immunology research to emphasize forward genetics, in which a gene(s) variant is sought to explain a specific phenotype, usually a disease. Forward genetics using NGS is the best unbiased approach to correlating genotypes with phenotypes and has already added a substantial number of new entities to over 250 genetic diseases of the immune system (25). The new technology has also uncovered thousands of disease-associated human gene variants (14, 18, 19, 25). As more comprehensive data sets are assembled, correlating gene variants with various immunological pathologies and infection susceptibility will reveal the genetic landscape of immune diseases.

NGS data coupled with bioinformatics analyses have become increasingly valuable for diagnosing known genetic disorders but are generally not adequate on their own to conclusively implicate new genetic variants in disease pathogenesis (26). When computational approaches do succeed, it is usually when there is a clearly defined clinical phenotype that occurs in multiple unrelated individuals and maps to the same gene in which nucleotide changes cause similar biochemical effects (27, 28). Additional biochemical and molecular validation is essential. This is because the number of variants is large but the availability of affected samples for rare diseases is often limited. Also, there are surprisingly many severe LOF variants found in apparently healthy people—almost 100 genes per individual, with 20 being homozygous (29). The problem of inadequate validation has been documented in population studies showing that hundreds of variants reported in the human gene mutation database as disease alleles and lacking adequate biochemical validation can be found in apparently healthy individuals (15, 26, 30). Although statistical and bioinformatics tools will improve, the most convincing evidence of associations between gene variants and disease will be based on molecular and biochemical validation in human cell lines or experimental organisms such as mice and zebrafish (26, 31). Validation requires addressing three major issues: How does the nucleotide alteration affect the protein biochemically? How does the gene mutation fit a specific model of inheritance? This can be complicated by reduced penetrance or expressivity of the genetic change. Finally, how does the specific biochemical change derange the cellular pathways in a manner that potentially explains the observed phenotype/disease? As described below, validation is greatly simplified when a cellular phenotype that clearly relates to disease manifestations can be defined.

CLINICAL PHENOTYPE: XMEN

The first two XMEN patients were brothers who had a clinical history of mild immunodeficiency with repeated but generally self-limited respiratory and gastrointestinal infections (32). Overall, their development was normal and they otherwise appeared healthy on clinical exam. Clinical laboratory investigation revealed the fact that the patients had strikingly elevated EBV levels that ranged between 100,000 and 1 million copies per milliliter of blood; normal results are often completely negative but can occasionally reach up to 800 copies per milliliter (34). These analyses defined the clinical phenotype: mild infectious history and astronomical levels of EBV in developmentally normal males. The occurrence of this phenotype in two affected males in the same family suggested a genetic basis.

In the remainder of this review, we walk through the genomic process of disease gene discovery, starting with patient phenotype and ending with the molecular definition of a new disease and its therapy. To illustrate the gene discovery process, we use the example of mutations in the *MAGT1* gene that cause XMEN [X-linked immunodeficiency with magnesium defect, Epstein-Barr virus (EBV) infection, and neoplasia] disease (32). These examples will be in sidebars. Identification of the XMEN disease gene offered new theoretical and basic insights into magnesium regulation as well as an etiological therapy for the most important consequence of the disease, uncontrolled active EBV infection (33, 34). At the end of the review, we describe two additional examples of gene disease discovery that led directly to improved therapies for life-threatening immune diseases.

CLINICAL AND LABORATORY PHENOTYPE

The process begins with the chief complaint. It can be severe or, as was the case with XMEN disease, a mild increase in susceptibility to childhood infections (32) (see sidebar "Clinical Phenotype: XMEN"). In the realm of immune disease, the phenotype often relates to deficiencies in immune responses to viral, bacterial, fungal, or parasitic infections (2, 8, 35, 36). Conversely, it can involve overactive immune responses leading to cellular or antibody-mediated autoimmunity or atopic disease, such as asthma (37, 38). Immune homeostasis can also be deranged with the accumulation of immune cells and/or the distortion of immune subset frequencies. Combinations of these abnormalities often defy obvious explanation due to the complex web of regulatory interactions in immune responses (e.g., severe immunodeficiency with excessive IgE) (37). Reasons to believe an immune abnormality is primarily genetic include early onset, occurrence in a multiplex family, marked severity of disease, and a characteristic syndrome that emerges in unrelated families and cannot be attributed to environmental causes. Clinical lab examination, depending on the depth of analysis available, can uncover abnormal lymphocyte subsets and basic functional defects in B and T cell activation and proliferation. The collective clinical data become the foundation of the phenotype.

The functional cellular defects that are characteristic of an immune disease can be investigated with elaborate tests because patient blood samples provide abundant cells to study. A key goal is to define a cellular phenotype as a benchmark of the disease. This clarifies the genetic model and provides a molecular confirmation of candidate genetic variants. Later, it can be used as a diagnostic test for individuals suspected of having the disease. Immunophenotyping by flow cytometry for lymphocyte subsets can document derangements in lymphoid development or survival, proliferation, and differentiation (see sidebar "Laboratory Phenotype: XMEN"). Though this cannot survey tissue-resident immune cells, immunophenotyping can be used to assess many distinct cell types.

LABORATORY PHENOTYPE: XMEN

In XMEN cases, flow cytometry revealed a modest decrease in peripheral naive CD4 T cells that inverted the CD4:CD8 ratio and was accompanied by a decreased number of $CD31^+$ recent thymic emigrants (32). These alterations were mild, however, and would not have been expected to have severe infectious consequences, especially not chronic active EBV infection. Functional analyses were more revealing: Both CD4 and CD8 T cells were impaired in upregulating activation markers, including CD69, CD25, and CD95, after T cell receptor (TCR) cross-linking (32). Interestingly, this deficit was overcome by using a combination of second messenger mimics (phorbol ester and ionomycin), implying a defect in proximal TCR signaling. By contrast, the patient B cells were normal in number and response to B cell receptor and Toll-like receptor stimulation (32).

MODES OF INHERITANCE

With the cellular and clinical phenotype in hand, NGS was a logical approach to deciphering XMEN disease, but evaluating SNVs depends on the presumptive mode of inheritance determined from the family pedigree (18, 19, 25, 26). The principal modes are derived from classic Mendelian recessive (an allele that must be present on both chromosomes to affect phenotype; the associated trait skips generations) and dominant (an allele that affects phenotype even if present on only one chromosome; the trait occurs in successive generations) concepts as well as the possibility of haploinsufficiency, X-linked disorders, and de novo occurrence of mutations.

In current databases of Mendelian monogenic immune diseases, autosomal recessive (AR) disorders are 3–4 times more common than autosomal dominant (AD) diseases, partly because they were the easiest to solve using classic approaches (25). Immunodeficiencies due to AR LOF mutations can be caused by homozygous alleles or two different LOF alleles (biallelic or compound heterozygous) (37, 38). In countries with high consanguinity, AR disorders involve LOF alleles that are usually homozygous.

AD disorders are increasingly being reported and arise mainly in two genetic forms: activating, in which the variant allele has hyperactive function, and dominant-interfering, in which the variant allele impairs the function of the normal allele. AD disease, the other side of the Mendelian coin, have been increasingly identified because NGS can uncover causative mutations in small pedigrees or large numbers of unrelated people even with marked variations in penetrance. PASLI disease (p110δ-activating mutation causing senescent T cells, lymphadenopathy, and immunodeficiency; also known as activated PI3K-δ syndrome, or APDS) is an AD disease caused by heterozygous mutations that upregulate phosphatidylinositol 3-kinase activity due to changes either in the gene encoding the leukocyte-restricted p110δ enzymatic subunit or the ubiquitous p85α regulatory subunit (27, 28, 39, 40). How the enzymatic gain of function (GOF) causes lymphocyte proliferation, differentiation, and senescence with metabolic changes that prevent adequate immune responses is described in greater depth below.

The second type of AD mutations is dominant interfering (DI) (41, 42). These mutations usually produce an anomalous protein that interferes with the function of the normal protein. For example, in ALPS, apoptosis signaling is inhibited by heterozygous alleles encoding defective FAS proteins that bind to the normal protein and hold it hostage in defective receptor complexes (41, 42). Similarly, in AD autoimmune polyendocrine syndrome (APS), mutations in the plant homology domain 1 of the *AIRE* gene generate defective proteins that sequester the normal AIRE protein in abortive transcriptional complexes in nuclear speckles (43). Rarely, genetic GOF mutations lead to immunodeficiency, which some have called autosomal dominance of the third kind (44).

MODE OF INHERITANCE: XMEN

For XMEN disease, the occurrence of disease in two boys suggested the possibility of an XLR mode of inheritance. This hypothesis was supported by the skewing of X chromosome inactivation in the unaffected mother such that nearly all of her hematopoietic cells showed inactivation of the X chromosome that she transmitted to both boys (32). This was strong prima facie evidence for a mutation on the X chromosome that was deleterious to her hematopoietic cells (hence, only cells with the normal X chromosome could be found) and responsible for X-linked disease in her sons. Solution hybridization reagents targeting the X chromosome were used to prepare exons from the boys' and the mother's DNA for NGS. Sequencing of the captured DNA revealed a 10-bp deletion in a magnesium transporter gene called *MAGT1* located on the X chromosome (32).

One specialized Mendelian mode of inheritance, combining heterozygosity and LOF, is called haploinsufficiency because the normal phenotype depends on full diploid gene dosage and a single defective allele reduces the amount of protein by half (45). These behave genetically like AD alleles. For example, patients with CTLA-4 haploinsufficiency with autoimmune infiltration (CHAI) disease have heterozygous, germline LOF mutations in *CTLA4* (9, 46). These decrease CTLA-4, resulting in abnormal T cell infiltration and damage to organs such as the lung, brain, and bowel (9).

X-linked recessive (XLR) mutations involve the X chromosome and predominantly affect males who inherit hemizygous defective alleles from their mothers. There are several highly penetrant immunological XLR disorders, including X-linked SCID, IPEX, and XMEN diseases (32, 47, 48) (see sidebar "Mode of Inheritance: XMEN"). These resemble recessive mutations because they are inherited through an unaffected mother. Finally, other important classes of mutations that are not, strictly speaking, Mendelian are DNMs, found in the proband but absent in parents and siblings. The DNA replication error rate is estimated to be 1 per 10^8 base pairs, generating 30–100 DNMs in each generation and serving as the source of unique private SNVs. These increase with paternal age and can also be mosaic by occurring postzygotically (20, 49, 50). DNMs can be experimentally identified by NGS of trios comprising the parents and affected child (50, 51). A variant type of DNM occurs in somatic cells. SMs usually contribute to genetic disease by conferring a proliferative or survival advantage as well as an abnormal phenotype in the affected cells. Cancer is the chief example of this.

PHENOTYPE VARIATION

In reverse genetics, inbred mouse and rat strains aid the genetic engineering of disease models because they reduce background variation and generate consistent phenotypes. Although 10% of the world is consanguineous, modern human populations are generally highly outbred (16). Thus, in humans, the phenotypes of gene mutations can be highly variable because they shine through heterogeneous genomic atmospheres. This genetic admixing, together with environmental influences, makes it difficult, particularly with small numbers of patients with a given gene variant, to precisely define disease phenotype. Coupled to this vast tumult in gene mixing, human genetics has a powerful phenotype detection mechanism provided by the medical and research professions that globally detect rare and variable phenotypes. This has increased the pace of gene discovery and validation but also revealed how disease associated with a specific gene can present in many different guises.

In human genetics, variation associated with a specific genotype, particularly for disease genes, was classically described in two ways: penetrance and expressivity (52–55). Penetrance refers to

PENETRANCE: XMEN

There are approximately two dozen XMEN patients from different countries and ethnic backgrounds, and the clinical and immunological phenotype is remarkably consistent. Although females can be carriers, there are no known males harboring silent *MAGT1* mutations, perhaps indicating that the disease is 100% penetrant on the cellular level with variable age of apparent symptom onset (34). As noted above, patients show a decrease in $CD4^+$ T cells, mild immunodeficiency, and a severe loss of immunity against EBV. Accompanying lab findings include a reduction in activation markers such as CD25, and a reduced level of NKG2D on $CD8^+$ T cells and natural killer (NK) cells. Yet, how these features fit together to explain a mechanism of disease was a puzzle.

whether or not a phenotype, in a sort of all-or-none fashion, is present in individuals with the same gene mutation (see sidebar "Penetrance: XMEN"). Expressivity was used to describe pleiotropy associated with specific gene mutations. With the dizzying array of human conditions associated with genomic variations, the difference between these two concepts has been blurred. Penetrance probably best describes syndromes with a highly stereotypical set of disease features where not everyone with a specific gene mutation gets the disease. For example, a healthy parent harboring a dominant-interfering *FAS* mutation can pass the mutation to a child who develops clinically significant ALPS (41). In this situation, where unaffected individuals harbor gene mutations identical to those with full-blown clinical disease, one would say the penetrance is incomplete. Mathematically, penetrance measures the proportion of those affected in a given population of individuals with a specific genotype (55). Penetrance of ALPS (AD) and CHAI disease (haploinsufficiency) ranges approximately from 50% to 60% (9, 41, 45). No molecular explanation is known for this surprisingly low penetrance. It is also important to distinguish between disease-causing variants with incomplete penetrance and risk alleles for complex diseases that are identified in GWA studies and cannot, on their own, serve as a primary cause of disease. Variations in expressivity are ubiquitous; essentially every genetic disease associated with a specific gene shows variation in the severity and type of clinical manifestation. As a consequence, this term is used less often in contemporary genomics.

There are a multitude of causes for variation in penetrance and expressivity (**Figure 1**) (52–55). Identified causes intrinsic to the genome include the molecular nature of the gene mutation and the degree to which it is damaging to protein function; additional mutations in either the same gene (*cis*) or other genes (*trans*); interindividual gene expression differences even between alleles; gene dosage; digenic disorders in which there are actually two obligatory causative mutations; sex dependence; age; imprinting; somatic mosaicism, and epigenetic modification. Of special significance are modifier genes (see below) that are often unlinked to the primary disease-causing gene but are critical determinants of the genetic atmosphere that influences penetrance and expressivity. For example, low expression levels of the affected gene can lead to a more severe abnormal phenotype, and this can be determined by genetic background (55). In addition there are extrinsic causes due to environmental influences on the affected individual (52–55). A good example of this is genetic susceptibility to viral pathogens or mycobacteria that is silent until exposure to the specific pathogen causes severe disease; individuals who never encounter the pathogen will have zero disease penetrance (56, 57). Differences in penetrance and expressivity create disease disguises that make it difficult to implicate a specific gene variant as the cause of a genetic disease. Violations of the mode of inheritance, inconsistencies in the pedigree, uncertainty about which disease characteristics are the most relevant, and confounding effects from disorders that are actually due to a different gene variant are significant challenges for genomic detective work. Only with further investigation of the

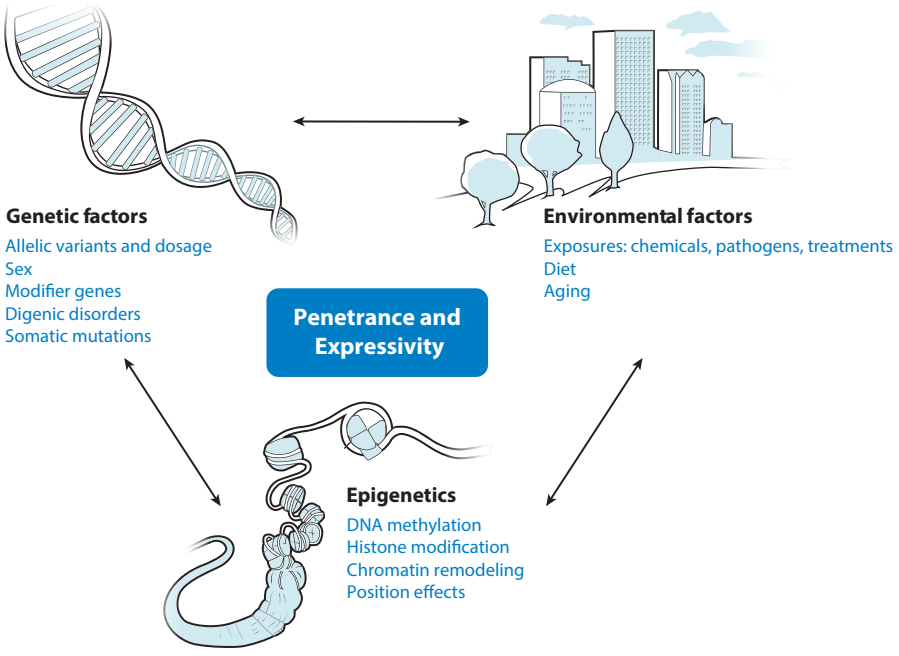

Figure 1

The causes of variable penetrance and expressivity of disease gene mutations.

specific genotype and phenotype, especially the accumulation of data from more people with the same genotype and phenotype, does the empiric association between the two become convincing.

GENOTYPE VARIATION AND BIOINFORMATICS

Once the genome has been interrogated by CGH arrays or NGS, variations can be defined by comparing the patient sequence to the standard reference assembly as well as family member genomes (5, 6, 15, 25, 51, 58–60). As alluded to above, a large number of SNVs will be identified, leading to the problem of determining which variants are potentially disease causing. Exonic sequences will reveal alterations in the coding sequences of particular genes. WGS can also be used to detect alterations in noncoding regions in human disease (58–61). However, the definition of disease-causing alterations in noncoding regions is still in its infancy because it involves casting a searchlight across immense stretches of the genome with complicated regulatory structures that are still only dimly perceived. Even within coding regions, analysis is challenging because there are thousands of variants, many private, to be sifted through (14, 26). The problem was documented in the first analyses of the 1000 Genomes Project (26, 30). Even considering only severe LOF variants such as nonsense, indel, and splice site–disrupting SNVs, it was estimated that genomes from healthy humans could contain up to 100 LOF variants, of which 20 would be homozygous (29). A similar analysis of missense mutations from this data set predicted that 281–515 missense substitutions are protein damaging and LOF (see below) and between 40 and 85 would be homozygous (30). Further compounding the problem was the fact that a shockingly large number (hundreds) of variants in the human disease mutation database were found in healthy people and that the original publication of these variants had no biochemical validation of disease involvement (30). Thus, NGS

interpretation is prone to false-positive disease correlations (26, 30). To surmount these problems, recommendations for variant interpretation emphasize weighing multiple lines of evidence. When available, statistical arguments across many families are certainly strongly supportive, but difficult to deploy (26). Most disease phenotypes are very rare, and it is difficult, without having identified a gene, to assemble a well-defined patient group. For more common conditions, general and less-specific phenotypes are binned together (e.g., common variable immunodeficiency) and later prove to have multiple genes involved, thereby defying a uniform approach (62, 63). Indeed, in defining over a dozen new diseases, our program has not been able to use a statistical study design.

The approaches we prefer to use are (*a*) judicious choice of candidate variants by a stratification process (51, 58, 60) and (*b*) biochemical and molecular validation. These have the virtues of proven success and providing a molecular understanding of disease that benefits treatment and prognosis. A variety of software tools are continuously evolving to stratify gene candidates. We discuss the guiding principles for variant evaluation and leave the reader to consult with bioinformatics experts for specific recommendations for genomics projects. The stratification of candidate variants is essentially a guesstimate of the probability that a given variant is disease causing, generally using a population of affected individuals that is too small to allow a statistical correlation. First, it is crucial to ensure that the variant is not a sequencing or annotation error (51). NGS technologies are high-throughput approaches designed to generate bulk but not complete sequence information, and a sequence variant can be most accurately described if it has been redundantly sequenced in the data set. This redundancy is usually expressed as the read number at that site or -fold coverage: 30 is a good number; 100 is better, but beyond that is probably unnecessary (51). Medium-sized indels and CNVs can be hard to judge with confidence. Large CNVs or rearrangements require CGH arrays or chromosome banding analyses. Even with good coverage, once a variant becomes highly suspicious, it should be confirmed by conventional Sanger sequencing on a separate sample from the patient.

Another initial consideration is whether the mutated gene is expressed in cell types with an abnormal phenotype relevant to the disease. It is generally safe to assume that the altered gene must be expressed in the affected cells, especially if there is a strong cellular phenotype. Further filtering of gene candidates can then be prioritized considering the presumptive mode of inheritance (see sidebar "Finding the Mutation: XMEN"). Depending on the clarity of the clinical phenotype, the pedigree provides the road map for navigating the mutant allele among affected and unaffected individuals (58). Extensive NGS has shown that nucleotide variation is not uniformly distributed in the genome and that a variant further gains significance if it falls in a conserved region of the genome, indicating that the function of that region does not tolerate much variation (20). Conversely, certain types of genes (a random example of this would be olfactory receptors) frequently harbor LOF mutations and are an unlikely cause of immune disease (58). Also important is whether the MAF shows that the variant is unique or very rare in the variant databases (ExAC, dbSNP, etc.). There is no precise cutoff, but <0.1% could be allowed, given that the reported SNV universe is so vast and covers so many groups of individuals with and without disease, there is a chance that the variant would have been reported. For example, heterozygous carriers of AR mutant alleles will be present in healthy control populations because they will have not have an associated disease. Nonetheless, a high MAF decreases the chance a variant is disease causing. Together these strategies can drastically shrink the list of possibilities from thousands to less than a hundred (47, 51, 56, 58–60).

The common assumption for a disease-causing variant is that it will have a severe deleterious effect on gene function (26, 31, 51, 58, 64–67). Largely owing to the wide use of exome sequencing, private variants with big effects have been located mainly in protein-coding regions of the genome. These currently account for about 85% of all Mendelian disease variants (63).

FINDING THE MUTATION: XMEN

The search for a pathogenic variant causing XMEN disease was simplified to a great extent by the hypothesis that the gene was X-linked, which was based on the mother's skewed X chromosome inactivation (32). However, when the variants were prioritized, no apparently protein-damaging candidate variant emerged that was present in both boys but heterozygous in the mother. The absence of a suitable variant raised a suspicion about the coverage of genes on the X chromosome by the sequencing analysis. In fact, the software was designed to exclude mismatches over 2 bp as erroneous sequences and would miss larger deletions, which would be consigned to the "junk pile" of sequences from the sequencing run. If such an exclusion had occurred, then there would be genes covered in the mother that had no coverage in the children. In fact, two genes met this criterion. When the reads were examined carefully, two nonoverlapping partial sequences of the *MAGT1* gene were found that were separated by a 10-bp gap. If this 10-bp region was removed from the X chromosome reference sequence and the boys' sequences were reanalyzed, then a large number of sequences would be recovered from the software's junk pile of discarded sequences, showing a 10-bp deletion in the genomic DNA of both boys (and heterozygous in the mother) that spanned the gap between sequences that had passed the quality control filter. Thus, by examining the coverage of genes, a problem with the software filters was detected and overcome. Further analysis of the deletion showed that it spanned an intron-exon junction, causing a frameshift and premature stop codon in the protein (32). This variant was predicted by bioinformatics to be a deleterious mutation.

Generally, nonsense mutations together with deletions (large > small) and splice variants, especially if they are frameshift mutations, are more likely to be deleterious than missense mutations (26, 29). Base changes that simply substitute an alternate codon for the same amino acid, termed synonymous, are usually ignored. The problem then becomes how to determine if a nonsynonymous change is likely to be deleterious just by looking at it. Fortunately, ingenious software has been developed to consider various parameters of the nonsynonymous change and determine if it is likely to be damaging to the protein (64–67). These include whether the mutation is in an evolutionarily conserved residue; whether it is a marked chemical change, e.g., substituting an acidic residue for a basic residue; whether it affects buried residues; whether it induces a steric clash between amino acids; and how it affects surface electrostatics, among other considerations. Generally, if the functional or conserved parts of the protein structure are drastically compromised, then it is judged to be likely damaging to the protein. Examples of widely used programs, which are constantly evolving, are SIFT, PROVEAN, CUPSAT, PolyPhen2, MutPred, MutationTaster, GERP and PhyloP, CADD, and variations thereof (64–67). Studies have shown that these programs have great value in prioritizing candidate disease variants (64, 66). Also facilitating disease variant identification are databases that aggregate reference exome data (e.g., ExAc Browser, **http://exac.broadinstitute.org/**). These techniques can reduce the candidate variant pool usually to one or a handful of possibilities.

BIOCHEMICAL VALIDATION

The critical final step in the journey (and often the longest and most labor intensive) is validation (26, 68, 69). This is the bottleneck in genomics research that connects DNA sequencing and computational analysis to clinical application (**Figure 2**). It includes both making an argument for causality and developing a molecular understanding of the disease mechanism. If a strong cellular phenotype has been defined, then cellular knockouts or knockdowns (e.g., using CRISPR, siRNA, or shRNA) to decrease expression of the gene or ectopic expression of the mutant allele might

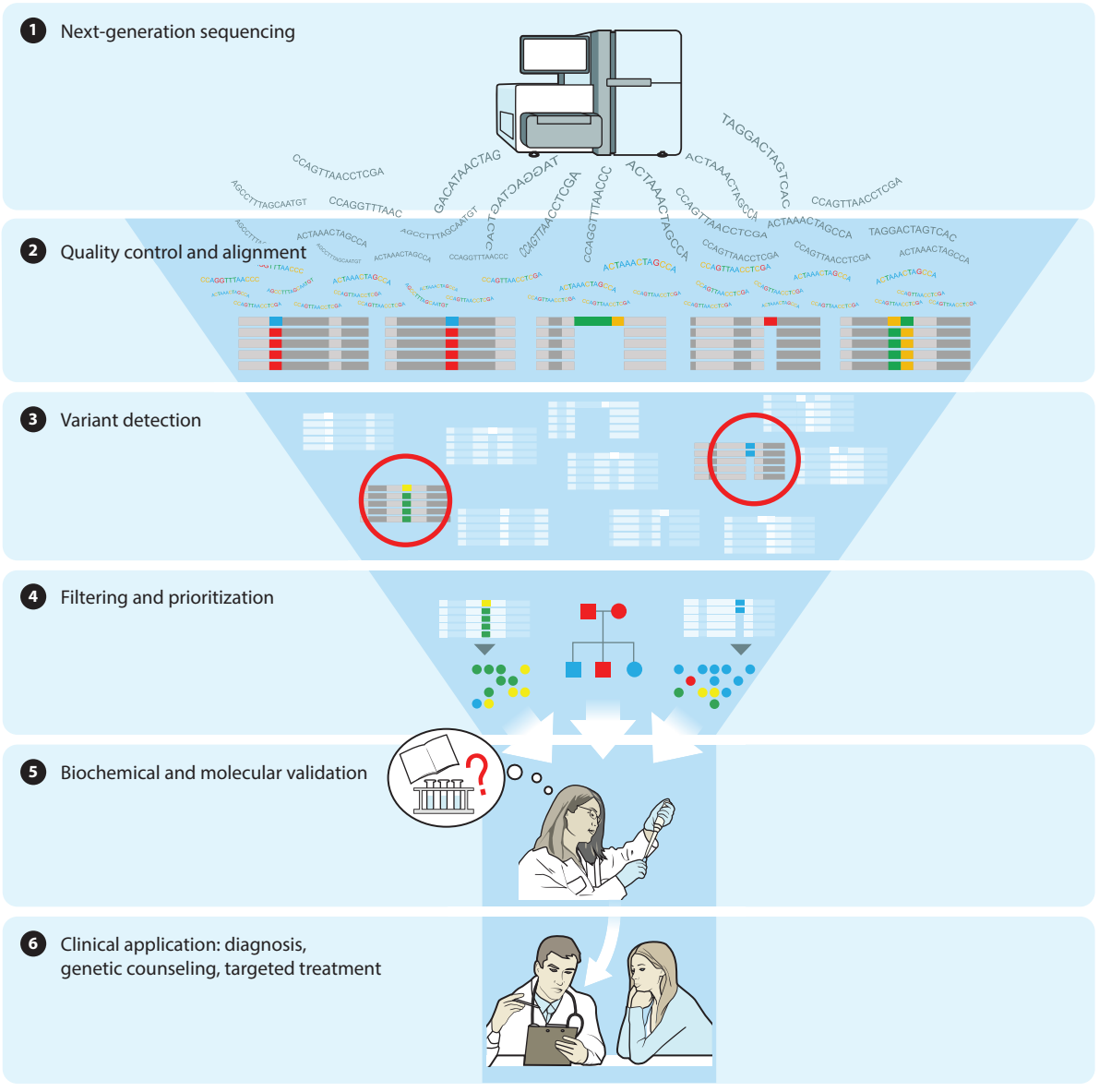

Figure 2

The bottleneck to genetic disease discovery is the biochemical and molecular validation of candidate disease variants obtained through next-generation DNA sequencing. Although computational methods can reduce and prioritize variants, further validation through direct experiments is required. The benefits of clinical application can only be achieved once validation has been carried out.

yield data demonstrating that the cellular phenotype associated with disease can be recapitulated in healthy cells (7, 8, 28, 32, 40, 68, 70). Also, introducing a normal gene version or knocking out a dominant variant might be sufficient to correct the abnormal phenotype in the patient cells (28, 32, 71). It is crucial to carry out biochemical investigations of the putative mutant protein because protein levels correlate poorly with mRNA in human cells (72). If a cellular phenotype

BIOCHEMICAL VALIDATION: XMEN

Validation of the first XMEN mutation was guided by a predicted premature stop codon in exon 7 (out of 10), a situation that typically leads to nonsense-mediated decay of the mRNA and accounts for loss of protein in roughly 30% of genetic diseases (75). Indeed, in both of the boys with XMEN disease, the mRNA was found to be depleted by polymerase chain reaction of cDNA, and a protein blot showed absence of the MAGT1 protein (32). A strong cellular phenotype in the form of defective TCR but not B cell receptor (BCR) signaling was observed in the patient cells and could be recapitulated by knocking down the MAGT1 protein in normal T cells.

Furthermore, artificial expression of a normal *MAGT1* gene transferred into patient T cells was able to correct the activation defect. Also, background literature on the gene indicated that it was a magnesium transporter, so Mg^{2+} levels and mobilization in immune cells were evaluated after stimulating T cells through the TCR. These investigations showed that there was a rapid TCR-gated Mg^{2+} flux that was required for a rapid and optimal Ca^{2+} flux. Further biochemical experiments revealed that the Mg^{2+} influx rapidly activated phospholipase C, γ 1 (PLCγ1), which cleaves phosphoinositide lipids into inositol triphosphate (IP_3) and diacylglycerol (DAG) (32). The reduction of IP_3 and DAG generation explained the blunted Ca^{2+} flux and impaired protein kinase C θ phosphorylation, respectively. Together, these defects short-circuited the induction of nuclear factors such as NFAT and NF-κB that control the transcriptional program underpinning normal T cell activation. Other pathways of TCR signaling involving p38 and ERK were Mg^{2+} independent, indicating an uncoordinated but not completely deficient molecular orchestration of signaling circuits needed for a healthy T cell response. Correspondingly, the patients have a relatively minor immunodeficiency. Experiments also showed that BCR activation of the orthologous enzyme PLCγ2 does not involve any Mg^{2+} flux, explaining why the number of B cells and their function in XMEN patients was the same as in healthy subjects (32).

Thus, the molecular validation of the *MAGT1* gene was consistent with the patients' cellular phenotype.

is not apparent or the disease involves a developmental process (e.g., differentiation of T cells), then more elaborate investigations are needed (26, 69, 73). Mouse models may reflect the cardinal features of human genetic diseases but are often incomplete (74). Moreover, they do not always reflect complicated human genetic disease mechanisms such as haploinsufficiency (9). Researchers can also use emerging new technology to induce pluripotent stem cells as a renewable source of cells to test gene variants. These cells can be differentiated into relevant cell populations in vitro, allowing gene experimentation in a physiological context (57, 68). These approaches can build the argument that the variant causes disease. The discoveries may then ultimately be confirmed by data from more patients and research as well as a corresponding phenotype in a relevant animal model (26, 68, 73) (see sidebar "Biochemical Validation: XMEN").

Determining the disease mechanism for a specific variant or the gene itself may be straightforward or may involve substantial investigation. This is the art of experimental immunology, creatively combining cellular, molecular, and biochemical analyses with an understanding of disease manifestations (see sidebar "Molecular Pathophysiology: XMEN"). The central guidepost is the patient phenotype as defined by astute clinicians. It is very helpful if a mouse model is available because many cell populations can be interrogated, and well-established autoimmune disease models are available. Also, viruses, bacteria, and parasites can be used to test immune responses. If one can connect the mutant to well-studied molecular pathways, then there will be more reagents and fundamental knowledge to speed up the investigation. However, if the disease-implicating evidence from the genetics and the cellular phenotype is strong, then it would be a mistake to shy away from gene mutations that occur in little known pathways or that do not seem to make sense. This may yield immunological insights with the greatest novelty and impact, but

MOLECULAR PATHOPHYSIOLOGY: XMEN

The hallmark of XMEN disease is uncontrolled EBV infection that creates a strong susceptibility to B cell lymphoma even in childhood (33, 34). This selective pathogen susceptibility was difficult to explain on the basis of a relatively mild TCR signaling defect. Further analyses of antiviral mechanisms revealed a severe reduction of the NK cell–activating receptor, NKG2D, which was observed in numerous XMEN patients who had uncontrolled EBV infection. Moreover, B cells infected with EBV showed increased expression of specific ligands that mark the cells for cytolytic destruction via NKG2D. With a deficit of NKG2D, these infected cells became invisible to antiviral immunity and escaped elimination. MAGT1 deficiency reduced intracellular basal free Mg^{2+}, which impaired proper surface expression of NKG2D. Remarkably, adding excess Mg^{2+} to XMEN lymphocytes reversed the intracellular deficiency, resulting in rescue of NKG2D surface expression and cytolytic function against EBV-infected targets (34).

the investigative journey could be long and frustrating. Genomics is an unbiased exploration of genotype-phenotype interactions—so it has extraordinary potential to take the investigator into new, uncharted territory.

If we stop for a moment and ponder our journey through the incredibly powerful new genomic technologies, we realize that human genetic variation in disease is a gold mine for insight into molecular function. The depredations of immune function caused by genetic mutations have the potential to instruct us about how the immune system operates. As pointed out above, careful work must be done to be sure that the variant under investigation is not fool's gold. However, the full value of gold ore is not realized until it is refined and polished into a lustrous understanding of an immune mechanism. In medical genetics there is, we believe, an overemphasis on generating DNA sequences from patients and insufficient emphasis on biochemical work to understand the molecular and pathophysiological importance of gene variants that are detected. We have argued that if one assumes that 200 families with unknown genetic immunological diseases (for instance parent-child trios) will generate 200 terabytes of DNA sequence data and this will require 1–2 bioinformatics specialists to generate candidate variants, then it will likely require 10 to 12 biochemists and immunologists to biochemically validate the causative variants (69). However, many clinical sequencing groups have not prioritized biochemical efforts so that the mutant proteins can be convincingly validated or investigated in sufficient depth to understand the molecular mechanisms of disease and possibly identify therapeutic strategies. Hence, in-depth functional validation and biochemical studies are essential goals for human genetic studies. Further, immunology has the unique and powerful advantage of the system of interest being accessible for study and treatment—much more so than in cardiology, neurology, ophthalmology, etc. This level of effort will fashion the genetic gold first into nuggets of understanding and then, hopefully, into better diagnosis and treatment.

GENETIC ETIOLOGY OF DISEASE AND COMPLEX ADAPTIVE SYSTEM THEORY

In the pre-NGS era of medical genomics, diseases were divided into Mendelian, meaning monogenic (or possibly digenic, but not more), and complex, meaning polygenic (13, 63, 76). As we pointed out above, even Mendelian diseases reflect the contribution of multiple genes. Conversely, NGS has revealed that broad genetic diseases heretofore thought to be polygenic—autism, mental retardation, and others—are actually often caused by one of a collection of heterogeneous

mutations that are private, highly penetrant, and deleterious (20, 77). Immunogeneticists have discovered that common variable immunodeficiency (CVID) can be explained in the same way (63). It is likely that rare, highly penetrant mutations account for other immunological diseases, including the common autoimmune conditions. This is where the nascent field of systems biology can be influential. By assembling the affected genes in a common phenotype into a regulatory network, systems analyses may be able to place new variants (private, highly deleterious, and penetrant) into a logical framework of disease (78).

Understanding the effect of genes on cellular phenotype requires a model for understanding how the linear, one-dimensional information in DNA generates a three-dimensional dynamic system. The protein product of a disease gene does not work in isolation but is part of a large interactive system (76). To understand genetic interactions, it is useful to view the cell as a complex adaptive system (CAS) (79). A CAS is a collection of diverse elements that self-organize into a system according to simple rules. Although analogous to many interactome models, a CAS focuses on defining features that enable the system to adapt to new and unpredictable circumstances. Modeling has shown that CAS can generate large intricate networks with spectacularly complex behavior that surpasses the action of individual parts in surprising and unexpected ways (80). Intuitively, this seems an appropriate way to model how molecules form the structures and systems of immune cells.

CAS are known to exhibit three key properties: robustness, adaptability, and tipping. Robustness refers to stability in the face of unpredictable perturbations, such as genetic alterations, through a network of interacting elements (81). Functionally speaking, robustness and adaptability (internal, real-time changes to preserve function) are two sides of the same coin. To be clear, adaptability here does not refer to Darwinian adaptations through natural selection but immediate adjustments in biochemical responses during the life of a single cell or organism to the effects of mutations under different cellular conditions. NGS has shown that genomes of individuals who appear healthy tolerate a large number of private deleterious mutations (29, 30). The key to understanding disease pathogenesis is uncovering how the CAS in different individuals adapts to LOF or otherwise disease-causing changes to robustly preserve function. Examples of adaptability to genetic loss or severe environmental stress on cells have been reported (82, 83). In these instances, the random occurrence of new mutations and the unpredictability of pathogens and the environment mean that there can be no predetermined plan of adaptations to achieve robustness in the system. These same issues underlie any explanation of how penetrance and expressivity can vary so greatly even to the extent that relatives of people harboring the same deleterious genetic variant can remain healthy (41). From the perspective of a CAS, wide variation in the parts of a genetic system—gene families, allelic differences, polymorphisms, and rare variants—promotes rather than detracts from robustness. This diversity of components may permit the system to adapt in real time. For any given gene, the complexity of the system is built upon gene families and nonsynonymous allelic differences that create different protein homologs that could be further diversified by posttranslational modifications to create a network of similar protein forms. Unlike the classical notion of "one gene, one protein," which implies a solitary gene function, the diversity of protein forms we now know can exist for any gene is consistent with a CAS that uses different forms under distinct conditions of metabolism, invading pathogens, or immune processes to enable the system to operate most effectively.

In a similar vein, the interaction of a network of genes, often referred to as modifiers, could achieve robustness by allowing adaptability to genetic defects. This model is very different from the cell model of a machine or an electronic circuit where each part is specifically engineered to play a particular role under controlled operating conditions (conceptually the same as the one gene, one protein concept). Machines can be highly complicated but lack the diversity that is crucial

to manifest complex adaptive behavior. When unexpected severe circumstances arise, a machine breaks, but a CAS adapts (79). The other CAS characteristic, tipping, refers to what happens when the various indirect parts develop negative synergy. Here the interconnectedness of the system destroys robustness, similar to a disastrous pileup on an icy expressway where drivers realize too late that they cannot control their vehicles and collide one after another. Events like lymphoid tumorigenesis and apoptosis may reflect tipping of the cellular CAS. It will be fascinating to learn how, as the broad genetic landscape is expressed in the cell interactome, the cellular CAS adapts to genetic and environmental perturbation to preserve health or tip into disease. Inbred mouse models may provide valuable insights into human disease by allowing experimental evaluation of different alleles and interacting genes in controlled systems (76).

Genetic changes affect the cellular CAS in ways that belie the traditional paradigm of Mendelian versus complex genetic diseases. Moreover, understanding how these primary determinants of disease have molecular effects on the cell will also give great insights into disease phenotype (76). Sometimes a variety of genes affect the same critical component of the system and adaptability fails. For example, a variety of mutant alleles of several receptor and cytokine genes that all interfere with production or response to interferon-γ cause the clinical entity Mendelian susceptibility to mycobacterial disease (MSMD) (56). However, the same gene can account for two entirely different diseases. Job's syndrome, which features impaired Th17 differentiation, immunodeficiency, abnormally elevated IgE, and connective tissue abnormalities, is caused by germline dominant-interfering STAT3 variants (84). However, STAT3 heterozygous GOF variants have been identified in patients with lymphoproliferation and early-onset autoimmunity (85–87). A similar dichotomy also exists in diseases related to STAT1 (88).

Although we have argued that the adaptability inherent in cells can create a continuum between Mendelian and complex disorders, the ends of the continuum can still be clearly defined. For Mendelian disorders, specific gene identification is achieved by each mutation being necessary and sufficient, within the bounds of penetrance and expressivity considerations, for the disease to occur. Generally the mutations are deleterious to protein function, are rare, and occur with an inheritance pattern consistent with the disease phenotype (22, 31, 51). The mutation reveals the relevant biological pathway and explains the phenotype. By contrast, complex polygenic diseases, which could include multiple sclerosis, type I diabetes, and rheumatoid arthritis, have variants that do not have the same properties as Mendelian disease variants. The associated variants in complex diseases are neither necessary nor sufficient, are not rare, and often are not in the protein-coding regions of the genome (12–14). Typically, they are noncoding and located in regulatory regions. Nevertheless, the identification of genes involved in Mendelian disorders will likely expose critical pathways that are involved in complex diseases (89). Classical studies have shown that the pathways detected in monogenic disorders, such as hypercholesterolemia, later prove to be valuable targets for pharmaceutical intervention for more common related abnormalities (90).

CLINICAL DIAGNOSIS AND INCIDENTAL GENETIC FINDINGS

NGS for identifying gene variants is now an accurate and cost-effective method that complements traditional molecular diagnostic tests in clinical genomics (91, 92). However, once variant implication in disease moves from the research laboratory into clinical medicine, then the DNA sequencing lab must adhere to the Clinical Laboratory Improvement Amendment (CLIA) requirements, and other legal and ethical considerations pertain (93). To report a variant as the cause of a disease, the lab must have a high degree of confidence in the variant assessment (91, 92). As the catalog of pathogenic variants grows and bioinformatics tools improve, it will be increasingly possible to identify disease-causing variants accurately.

However, whole-exome sequencing (WES) or WGS generates large amounts of data that extend beyond a specific set of gene candidates for immunological diseases and provides a wealth of genomic information that is potentially important for the health of the patient. Potential risk alleles, spanning a wide range of diseases, were found in every healthy individual in a reanalysis of data from the 1000 Genomes Project data set (94). For example, an NGS study carried out to discover a gene mutation causing an immune disease might reveal a breast cancer susceptibility allele. Thus, the medical genetics community has begun to focus on how to integrate WES/WGS into medical care by enabling the reporting to physicians of actionable variants (95). However, many issues regarding how to integrate this process into medical care have been raised (96). Databases of clinically important variants, such as ClinVar and Online Mendelian Inheritance in Man (OMIM), will allow access to collected genomic information. This will require ongoing epidemiological analysis of variants at laboratory and hospital levels. It will also create a need for allied genetic counselors who can interpret the genomic information and help physicians provide appropriate diagnostic and prognostic information to patients. The Centers for Disease Control and Prevention has established a working group for developing standards for the use of NGS in clinical diagnosis (97).

TARGETED THERAPY, PRECISION MEDICINE, AND COMMON IMMUNE DISEASES

An important aspiration in the exploration of the genetic basis of immunological disease is to provide a molecular approach to new therapeutics. This has been termed precision medicine (98). In the pre-NGS era, bone marrow transplantation was the mainstay treatment for genetic diseases of the immune system. We believe that the future of genomics will provide two types of new therapies: (*a*) direct molecular interventions based on the pathway of the mutant protein and (*b*) genome editing combined with lymphocyte or hematopoietic stem cell transplantation (HSCT). Several examples show how gene identification has led directly to new therapeutic concepts for immune diseases (see sidebar "Targeted Therapy: XMEN"). Identification in a group of patients with atopy, immunodeficiency, autoimmunity, and neurocognitive deficits of LOF mutations in *phosphoglucomutase 3* (*PGM3*), which is an essential enzyme for the production of UDP-GlcNAc, is one such example (38, 99, 100). How the defective protein glycosylation selectively alters immunity is unknown, but addition of GlcNAc to cells from the *PGM3*-deficient patients restored intracellular UDP-GlcNAc levels (38). Clinical studies are underway to determine if exogenous nondiabetogenic sugars could be used therapeutically in these patients to ameliorate the immune dysfunction (38). Although HSCT is lifesaving, sugar therapy might be useful as a simple, safer, and inexpensive therapy, especially if the patient is not a good candidate for transplantation (100). Below we recount additional cases of etiological treatments derived from gene identification in XMEN, CHAI/LATAIE, and PASLI diseases.

CHAI AND LATAIE DISEASES

In a cohort of patients with severe immune dysregulation and solid organ lymphocytic infiltration, we identified one subset of patients with an AD mode of inheritance and another subset with an AR mode of inheritance. Molecular investigation revealed heterozygous LOF mutations in *CTLA4* in the patients with AD disease, which we named CTLA-4 haploinsufficiency with autoimmune infiltration (CHAI) disease (9, 46). In patients who looked very similar but lacked *CTLA4* gene mutations, we detected biallelic (homozygous or compound heterozygous) LOF mutations in the *lipopolysaccharide-responsive vesicle trafficking, beach- and anchor-containing* (*LRBA*) gene causing recessive disease, which we termed LRBA deficiency with autoantibodies, regulatory T cell (Treg)

TARGETED THERAPY: XMEN

After the initial discovery of XMEN disease, it was not clear that a direct etiological treatment was possible. The major pathogenesis of XMEN disease appeared to be failure to control a single pathogen, EBV. Magnesium is an obligatory element in all organisms, and its utilization is tightly regulated (101). The most important biological form of magnesium is the divalent cation Mg^{2+}, which is a cofactor for many metabolic reactions, including all ATP-requiring kinase reactions, glycolysis, and nucleic acid enzymology. In addition to causing a defect in TCR signaling, MAGT1 deficiency also substantially decreases cytosolic free basal Mg^{2+} to less than 50% of normal. While studying the biochemical mechanism of reduced NKG2D expression, which contributes to uncontrolled EBV infection, we determined that the decrease in free basal Mg^{2+} was the cause for reduced expression of NKG2D (33). Further experimentation revealed that the reduction of surface NKG2D was due to ubiquitinylation and degradation. Importantly, it was also found that the free basal Mg^{2+} in the cell was in equilibrium with extracellular Mg^{2+}. This led to a series of experiments demonstrating that culturing T or NK cells from XMEN patients in supraphysiological concentrations of Mg^{2+} (>1 mM) restores free basal Mg^{2+} to normal levels and rescues the surface expression and function of NKG2D (33). This then provided the rationale for therapeutic Mg^{2+} supplementation for the XMEN patients. Indeed, chronic Mg^{2+} supplementation resulted in a significant decrease in the fraction of EBV-infected B cells and blood levels of EBV (33, 34). The hope is that Mg^{2+} supplementation, a treatment suggested directly by the gene identity, will be a simple, low-cost way to improve antiviral immunity, suppress EBV, and avoid lymphoma in XMEN patients and perhaps other patients with chronic active EBV.

defects, autoimmune infiltration, and enteropathy (LATAIE) (102–104). Both subsets of patients share the predominant characteristics of autoimmune cytopenias, lymphocytic infiltration of multiple nonlymphoid organs, lymphadenopathy/splenomegaly, and hypogammaglobulinemia. There was variable expressivity in clinical presentation, in which the severity of the mutation appears to play a role. The clinical phenotype ranged from severe, life-threatening interstitial lung disease with enteropathy, autoimmune cytopenias, and CVID to an ALPS-like phenotype (i.e., autoimmune cytopenias with lymphadenopathy and/or splenomegaly) with no other autoimmune manifestations. The fact that a very similar clinical picture resulted from mutations in two different genes with two different modes of inheritance raised the critical question of how the two molecular pathways intersect.

CTLA-4 is a critical immune tolerance checkpoint molecule. It is a cell surface receptor protein on activated T lymphocytes that binds and controls access to the CD80 and CD86 costimulatory molecules that are expressed on the surface of antigen-presenting cells (APCs) with which they are interacting. By constraining costimulation and downregulating T cell responses, it acts as a checkpoint. The importance of this checkpoint control is demonstrated by *Ctla4* homozygous knockout mice that develop fatal, multiorgan lymphocytic infiltration and destruction (105, 106). Patients with CHAI disease develop almost identical manifestations: extensive and destructive lymphocytic infiltrates in multiple organs, mainly the gut, lungs, and/or brain (9, 46). However, discovery of the gene causing CHAI disease was complicated by two factors. First, the disease results from an AD-haploinsufficiency rather than AR mode of inheritance. Second, the penetrance was incomplete. WES was performed with DNA from affected and unaffected family members and analyzed using an AD model. All individuals with CHAI disease had the *CTLA4* mutation, but not all mutation-positive individuals in the family had disease, even though all mutation-positive individuals had reduced CTLA-4 protein. Only after multiple families with mutations were found and the mutations were biochemically demonstrated to impair CTLA-4 protein function or expression was *CTLA4* concluded to be the disease-causing gene.

In addition to conventional T cells, CTLA-4 is a key inhibitory receptor that is constitutively expressed on regulatory T cells (Tregs). In certain contexts, CTLA-4 is crucial for proper Treg suppressive function and limits costimulation by multiple molecular mechanisms, e.g., by negative signaling, by competing with the costimulatory molecule CD28 for the shared ligands CD80 and CD86, and by removing these ligands from APCs through *trans*-endocytosis (107, 108). Consistent with a failure of these functions of CTLA-4, we and others found that patient T cells were hyperproliferative and their Tregs had an abnormal phenotype and were specifically impaired in suppressive function (9, 46). Surprisingly, several patients had low rather than high antibody levels, as would have been expected from the CTLA-4 null mice, thus illustrating the potential limitations of using mouse models for predicting human disease phenotype (106). CHAI patients also had elevated numbers of anergic or exhausted $CD21^{lo}$ B cells, which are putatively autoreactive given that they are often enriched for self-reactive BCRs and are found at increased levels in autoantibody-mediated autoimmune diseases (109, 110). Overall, the clinical presentation of the patients revealed a critical role for CTLA-4 in regulating both B cell and T cell homeostasis.

These findings left unresolved the pathogenesis of AR LATAIE disease. *LRBA* mutations were first discovered to be disease causative in 2012 by means of genetic linkage analysis of several consanguineous families with hypogammaglobulinemia and autoimmunity, followed by sequencing of positional candidate genes (104). Shortly thereafter, other patients with *LRBA* deficiency were identified by NGS (102, 103). All mutations caused loss of LRBA protein; however, the molecular function of LRBA or its role in the immune system was initially unknown. A study of the biochemical basis of LATAIE disease not only revealed both the function of LRBA and why LATAIE patients resembled CHAI patients but also revealed targeted therapies (111). LATAIE patients with life-threatening interstitial lung disease, which was refractory to conventional immunosuppressants, were treated with abatacept (a CTLA4-immunoglobulin fusion protein) for several years before their genetic etiology was known. With this treatment, the patients' autoimmune conditions and lung function dramatically improved. A clue to the connection between this remarkable treatment effect and *LRBA* mutations was the tantalizing homology of LRBA to vesicle-trafficking molecules (112–114). This link led to the hypothesis that LRBA plays a key role in controlling protein trafficking of CTLA-4, which is known to be tightly regulated through recycling vesicles (111, 115). Consistent with this hypothesis, loss of LRBA, both in LATAIE patients and in other experimental settings, was found to cause a profound posttranslational loss of CTLA-4 protein.

CTLA-4 is stored in intracellular vesicles and cycles, after T cell stimulation, to the cell surface, where it must reside to perform its inhibitory function (115, 116). LRBA was shown to bind to CTLA-4, and in its absence, CTLA-4 transited to lysosomes and was rapidly degraded. Inhibition of lysosomes with the drug chloroquine could rescue the loss of CTLA-4 in LRBA-deficient cells. Thus, LRBA appears to protect CTLA-4 from lysosomal degradation and therefore help maintain intracellular pools of CTLA-4 for rapid mobilization to the cell surface for inhibitory function. The biochemical link between LRBA and CTLA-4 also explained why abatacept was such an effective treatment for the LRBA-deficient patients. By traveling through the blood to sites of T cell activation and providing a therapeutic cap on costimulation by blocking CD80 and CD86, abatacept could make up for the loss of endogenous CTLA-4 on the surface of conventional and regulatory T cells. Altogether, the discovery of CHAI and LATAIE diseases, both autoimmune diseases resulting from CTLA-4 deficiency, emphasizes how critical this molecule is for immune tolerance and homeostasis, given that reduction of CTLA-4 or loss of a single allele can throw the immune system out of balance. It also provides a vivid illustration of how research on haploinsufficiency genetic disorders can lead to a greater understanding of immunologic processes.

Uncovering the genetic and molecular etiology of CHAI and LATAIE diseases yielded novel insights into immune regulation as well as a rational basis for precision medicine treatments for

the diseases. Lymphoproliferation and autoimmunity in CHAI and LATAIE diseases both stem from loss of CTLA-4. Thus, treatment with abatacept as a CTLA-4 replacement therapy, clinically demonstrated to be effective in LATAIE disease, could also be effective for CHAI disease and perhaps other autoimmune diseases involving insufficient CTLA-4. Additionally, LRBA was found to regulate the turnover and degradation of CTLA-4 in lysosomes. Since the lysosomal inhibitor chloroquine could prevent lysosomal loss of CTLA-4 in LRBA-deficient cells, it or its pharmacologically safer alternative hydroxychloroquine could also be effective in treating CHAI or LATAIE disease by boosting the levels of CTLA-4 protein. Interestingly, hydroxychloroquine is already used as a therapy for rheumatoid arthritis and lupus (117–119), so it is worth investigating whether its efficacy may be partly attributed to an augmentation of CTLA-4. If it were proven to be effective, hydroxychloroquine would be a very inexpensive alternative to abatacept. In summary, our knowledge of not only the genes but also their biochemical and molecular roles in the cell suggests molecular theories that might be used to develop dramatically effective precision medicines for previously undefined diseases.

PASLI DISEASE

Multiple patients have presented with an AD immunodeficiency syndrome with recurrent sinopulmonary infections, immunoglobulin synthesis defects, predisposition to EBV and/or cytomegalovirus viremia, and lymphoproliferative disease (27, 28, 120–122). NGS of the patients' genomes defined a subgroup who carry mutations in the *PIK3CD* gene encoding the leukocyte-enriched p110δ subunit of phosphatidylinositol 4′,5′-bisphosphate 3′-kinase (PI3K) (27, 28, 120–122). In unrelated families, one of three germline, heterozygous mutations resulting in amino acid substitutions N334K, E525K, or E1021K were identified. Intriguingly, N334K and E525K align precisely with the hyperactivating substitutions N345K and E545K in p110α, resulting from somatic mutations in tumor cells (27, 28). Molecularly, these two changes disrupt inhibitory contacts between the regulatory p85α protein and the catalytic p110 proteins of the PI3K complex. Each of these mutations would be expected to augment enzymatic activity. This would explain the dominance of the heterozygous mutant allele. The recurrence of such mutations strongly suggests amino acid vulnerabilities in enzyme activity regulation. We named the specific constellation of disease features associated with this genetic etiology p110δ-activating mutations causing senescence, lymphadenopathy, and immunodeficiency (PASLI) (the disease has also been called activated PI3K-δ syndrome, or APDS) (27, 28). Since the initial descriptions, approximately 100–150 people with *PIK3CD* mutations causing PASLI disease have been identified, and the initial list of three mutation sites has been expanded to 5 DNA sites where mutations cause specific amino acid substitutions that increase enzyme activity. The prevalence of heterozygous GOF mutations in *PIK3CD* suggests that p110δ-hyperactivating substitutions are a significant contributor to disease among immunodeficient lymphoproliferative individuals.

Shortly after discovery of heterozygous *PIK3CD* mutations, a second set of patients with a similar clinical phenotype, i.e., PASLI disease, were shown to harbor heterozygous splice site mutations in the ubiquitously expressed *PIK3R1* gene (39, 40). *PIK3R1* encodes the binding partner for p110δ called p85α that ensures stability, localization, and regulation of the p110δ protein. Notably, all the patients (approximately 30–40 people have been identified) shared a splice site mutation that results in skipping of exon 11 and production of an in-frame transcript encoding a protein lacking important residues in the inter-SH2 domain of p85α. Loss of these residues alters binding of the p85α regulatory subunit to the p110δ catalytic subunit such that the former still guides stability and localization but cannot regulate p110δ enzyme activity, resulting in PI3K hyperactivation. Given the similar molecular effect, i.e., hyperactivation of PI3K due to unrestrained

p110δ enzyme activity, it is apparent why the clinical and cellular phenotypes within the immune system would be similar in patients with p110δ and p85α alterations. However, unlike p110δ expression, p85α expression is not restricted to the immune system. Despite ubiquitous p85α expression, these patients with *PIK3R1* mutations do not show dramatic nonimmune phenotypes due to p110α or p110β hyperactivation. This is an area of active investigation and may shed light on unique features of the association between p85α and p110δ (as opposed to p110α or p110β). The condition of patients with *PIK3CD* or *PIK3R1* mutations can be referred to as PASLI-CD or PASLI-R1, respectively.

PI3Ks are a family of crucial signaling enzymes that transduce signals from tyrosine kinase receptors and G protein–coupled receptors by phosphorylating the hydroxyl group of the 3′ position of the inositol ring of phosphatidylinositol. This leads to generation of phosphatidylinositol 3-phosphate [PI(3)P], phosphatidylinositol (3,4)-bisphosphate [$PI(3,4)P_2$], and phosphatidylinositol (3,4,5)-trisphosphate [$PI(3,4,5)P_3$]. The last of these products, typically formed in the inner leaflet of the plasma membrane, serves as a docking site for phosphoinositide-binding protein domains such as the pleckstrin homology (PH) domain. There are two types of class I PI3K molecules that are activated by different stimuli, and these have the same heterodimeric structure comprising one 110-kD enzymatic subunit and a regulatory subunit that can vary in size. The class IA PI3K molecules, which include p110δ and p85α, are activated by receptor tyrosine kinases to phosphorylate phosphatidylinositol (4,5)-bisphosphate [$PI(4,5)P_2$] at the 3 position, generating the $PI(3,4,5)P_3$ (123). The generation of $PI(3,4,5)P_3$ at the membrane initiates signaling by attracting the PH domain–containing proteins, including phosphoinositide-dependent kinase-1 (PDK1) and protein kinase B (PKB), also known as AKT. These stimulate further pathways downstream, including the mammalian target of rapamycin (mTOR) kinase, which promotes cell growth, proliferation, and survival. Hence, activating mutations in class IA PI3K genes cause lymphoproliferation through the well-known proliferative mTOR pathway, but why this is associated with immunodeficiency loomed large as a key to understanding the pathogenesis and devising a new way to treat disease.

Investigation at the cellular level revealed that PI3K and mTOR hyperactivation in PASLI disease created imbalances in the subsets of T and B lymphocytes. On the T cell side, PASLI patients show a marked reduction in naive T cells and a corresponding increase in effector-type T cell subsets that have become replicatively and functionally senescent with short telomeres and expression of the senescence marker CD57 on $CD8^+$ T cells (28). Increased glucose uptake in the diseased T cells betrayed metabolic changes driven by PI3K/mTOR signaling that contributed to precocious effector cell maturation and led to most T cells converting into terminally differentiated cells that failed to function adequately (28, 40). On the B cell side, PASLI patients show defective B cell development with a preponderance of $CD10^+$ transitional B cells and reduced frequency of $CD27^+$ memory B cells. Functionally, patient B cells fail to properly secrete class-switched immunoglobulins, particularly those specific to polysaccharide antigens. Histological analysis indicates PASLI patients have prominent germinal centers that lack a mantle zone (28). Thus, putting the cell surface signaling in overdrive caused the lymphocytes to differentiate abnormally.

The knowledge gained by the genetic and biochemical characterization of these patients immediately provided new therapeutic concepts. Fortunately, the mTOR kinase downstream of PI3K-AKT has been successfully targeted in vivo by the immunosuppressant rapamycin [approved by the US Food and Drug Administration (FDA)]. The defects in PASLI disease provided a clear rationale for using PI3K/mTOR inhibitors to treat this disorder. Indeed, data from an initial evaluation of the efficacy of rapamycin therapy in one patient suggest that blocking this intrinsic drive toward effector differentiation at least partially restores the balance of naive, memory, and effector CD8 T cells and markedly improves lymphadenopathy and splenomegaly (28). A larger cohort of

patients treated with rapamycin is being studied (G. Uzel, personal communication). Thus, a counterintuitive therapeutic—treating an immunodeficiency with an immunosuppressant—becomes a rational targeted approach once the biochemical mechanism is clear.

Although rapamycin therapy holds great promise for an immediately useful treatment for people with PASLI disease, mTOR is only one pathway that is stimulated by activated AKT. The others, including NF-κB, GSK3, FOXO, and MDM2, will not be directly affected by rapamycin, so the effects of hyperactive PI3K would not be expected to be fully reversed. Consequently, the ideal therapy would be to directly and specifically target p110δ with an inhibitor to restrain activity to the normal range in PASLI patients. Specific targeting of PI3K isoforms has been the subject of intense research in the cancer field owing to the high frequency with which this pathway is upregulated by direct mutation, gene amplification, or upstream receptor activation (124). Indeed, a p110δ-specific inhibitor called idelalisib (formerly CAL101 or GS1101) has shown promise as a therapy for chronic lymphocytic leukemia (CLL) and is now FDA approved for use in combination with rituximab for relapsed CLL (125, 126). We anticipate that directly inhibiting the source of augmented signaling in PASLI patients will maximize clinical benefit by enabling production of new B and T lymphocytes that escape the PI3K-driven development and differentiation defects.

CONCLUSIONS

With the dramatic drop in the cost of NGS, our data banks are bulging to the breaking point with DNA sequences. It is clear that bioinformatics, although enormously powerful, is not adequate on its own to accurately assess the medical significance of most new gene variants. This requires biochemical and molecular validation, which is slow, expensive, and labor intensive (**Figure 2**). However, our hope is that we have made the case in this review that investing the time and resources for the appropriate biochemical investigation of gene variants yields important fundamental insights into immunology as well as new, targeted therapies for genetic diseases. We believe that technological advances provide an optimistic outlook for understanding the genetics of immune disorders. (The distinguishing characteristics of XMEN, PASLI, and CHAI/LATAIE diseases are listed in **Table 1**.) NGS and other molecular technologies have greatly expedited the identification of disease-causing variants. As of February 2015, OMIM contained data on 2,937 genes associated with 4,163 Mendelian phenotypes, but the genes underlying approximately 50% of all known Mendelian phenotypes are still undiscovered, and an untold number of additional Mendelian phenotypes have yet to be defined. We have found that an integrative approach to genomics with biochemical, molecular, and cellular examination of the phenotype defined in the clinic is most effective to understand pathways in the healthy and diseased immune system. We have provided examples of how causative gene identification in human disorders is already leading to unexpected new therapeutic interventions with less cost and morbidity than traditional HSCT. Another exciting frontier is the development of new technologies in genome engineering that could be used to directly correct the DNA lesion in patient cells. This will be more readily deployable in immune diseases and other hematopoietic disorders than for solid organ diseases. If the major effect of the gene defect is in cells of hematopoietic origin, then severe defects can potentially be corrected with genome editing of hematopoietic stem cells that could be used for autologous HSCT. If the defect is somewhat less severe, then alternative strategies such as correction of somatic cells (e.g., T cells) followed by reinfusion of the altered cells could be employed.

The ultimate success of clinical genomics will require further integration of knowledge at the molecular and clinical levels longitudinally. Also, well-defined genetic entities can help in defining the prognosis of immunological diseases, predicting the response to medical interventions, and providing appropriate genetic counseling. Necessary organizational efforts on the clinical

Table 1 Distinguishing characteristics of CVID, ALPS, CAEBV, XMEN, PASLI, and CHAI/LATAIE diseases

	CVID	ALPS	CAEBV	XMEN	PASLI	CHAI/LATAIE
Sinopulmonary infections	Yes	No	No	Variable	Yes	Yes
Immunoglobulin defects	Yes	Yes	Variable	Yes	High IgM	Yes, frequent
Lymphoproliferative disease	Variable	Yes	Yes	Variable	Yes	Yes
Viremia	Variable	No	EBV	EBV	Often CMV and/or EBV	
EBV + B cell lymphoma risk	Variable	No	Yes	Yes	Yes	
Lymphocyte tissue infiltration	Variable	No	Yes	No	Yes	Yes
Expanded cell population	Variable	Double-negative αβ T cells	$CD3^+CD8^+$	$CD3^+CD8^+$(low CD4:CD8) $CD8^+NKG2D^-$	$CD3^+CD8^+$ $CD3^+CCR7^-$ $CD8^+CD57^+$ $CD20^+CD10^+$	More memory T cells and $CD21^{lo}$ B cells
Disease gene product	Variable	FAS, FASL, caspase-10	Variable	MAGT1	Phosphoinositide 3-kinase p110δ or p85α	CTLA4 or LRBA

Abbreviations: ALPS, autoimmune lymphoproliferative syndrome; CAEBV, chronic active EBV; CHAI, CTLA-4 haploinsufficiency with autoimmune infiltration; CMV, cytomegalovirus; CVID, common variable immunodeficiency; EBV, Epstein-Barr virus; HGG, hypogammaglobulinemia; LATAIE, LRBA deficiency with autoantibodies, Treg defects, autoimmune infiltration, and enteropathy; PASLI, p110δ-activating mutation causing senescent T cells, lymphadenopathy, and immunodeficiency; XMEN, X-linked immunodeficiency with magnesium defect, EBV infection, and neoplasia.

side include understanding the sociology of the public's interaction with the medical delivery system, developing electronic medical records that can be integrated with genomics and medical research, and incentivizing medical professionals to record phenotypic features according to the pathophysiology of the disease. At some point in the future, medical genetics may cease to be an independent specialty by fully integrating with medical practice.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We apologize to colleagues whose work we could not cite or could only superficially discuss due to space limitation. We thank our clinical colleagues Gulbu Uzel, Michael Jordan, and Helen Matthews. We also thank Helen Su and Morgan Similuk for critically reading the manuscript and Ryan Kissinger for figure preparation. We thank Bill Paul for his leadership and enthusiastic support of this research program. This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

LITERATURE CITED

1. McKusick VA. 1963. Genetics in medicine and medicine in genetics. *Am. J. Med.* 34:594–99
2. Bruton OC. 1952. Agammaglobulinemia. *Pediatrics* 9:722–28
3. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, et al. 1993. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 72:279–90
4. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, et al. 1993. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* 361:226–33
5. Conley ME, Casanova JL. 2014. Discovery of single-gene inborn errors of immunity by next generation sequencing. *Curr. Opin. Immunol.* 30:17–23
6. Picard C, Fischer A. 2014. Contribution of high-throughput DNA sequencing to the study of primary immunodeficiencies. *Eur. J. Immunol.* 44:2854–61
7. Doudna JA, Charpentier E. 2014. Genome editing: the new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096
8. Chun HJ, Zheng L, Ahmad M, Wang J, Speirs CK, et al. 2002. Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature* 419:395–99
9. Kuehn HS, Ouyang W, Lo B, Deenick EK, Niemela JE, et al. 2014. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science* 345:1623–27
10. Pennisi E. 2000. Human genome. Finally, the book of life and instructions for navigating it. *Science* 288:2304–7
11. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225–29
12. Cho JH, Feldman M. 2015. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nat. Med.* 21:730–38
13. Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90:7–24
14. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. 2011. Clan genomics and the complex architecture of human disease. *Cell* 147:32–43
15. Cirulli ET, Goldstein DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 11:415–25
16. Bittles AH, Black ML. 2010. Consanguinity, human evolution, and complex diseases. *PNAS* 107(Suppl. 1):1779–86
17. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton Univ. Press
18. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell* 155:27–38
19. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* 14:125–38
20. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, et al. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151:1431–42
21. Ponting CP. 2012. Loaded dice for human genome mutation. *Cell* 151:1399–400
22. Veltman JA, Brunner HG. 2012. De novo mutations in human genetic disease. *Nat. Rev. Genet.* 13:565–75
23. Holzelova E, Vonarbourg C, Stolzenberg MC, Arkwright PD, Selz F, et al. 2004. Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *N. Engl. J. Med.* 351:1409–18
24. Jing H, Zhang Q, Zhang Y, Hill BJ, Dove CG, et al. 2014. Somatic reversion in dedicator of cytokinesis 8 immunodeficiency modulates disease phenotype. *J. Allergy Clin. Immunol.* 133:1667–75
25. Al-Herz W, Bousfiha A, Casanova JL, Chatila T, Conley ME, et al. 2014. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies expert committee for primary immunodeficiency. *Front. Immunol.* 5:162
26. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature* 508:469–76
27. Angulo I, Vadas O, Garcon F, Banham-Hall E, Plagnol V, et al. 2013. Phosphoinositide 3-kinase delta gene mutation predisposes to respiratory infection and airway damage. *Science* 342:866–71

28. Lucas CL, Kuehn HS, Zhao F, Niemela JE, Deenick EK, et al. 2014. Dominant-activating germline mutations in the gene encoding the PI(3)K catalytic subunit p110δ result in T cell senescence and human immunodeficiency. *Nat. Immunol.* 15:88–97
29. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, et al. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335:823–28
30. Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, et al.; 1000 Genomes Proj. Consort. 2012. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.* 91:1022–32
31. Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, et al. 2011. Finding disease variants in Mendelian disorders by using sequence data: methods and applications. *Am. J. Hum. Genet.* 89:701–12
32. Li FY, Chaigne-Delalande B, Kanellopoulou C, Davis JC, Matthews HF, et al. 2011. Second messenger role for Mg^{2+} revealed by human T-cell immunodeficiency. *Nature* 475:471–76
33. Chaigne-Delalande B, Li FY, O'Connor GM, Lukacs MJ, Jiang P, et al. 2013. Mg^{2+} regulates cytotoxic functions of NK and CD8 T cells in chronic EBV infection through NKG2D. *Science* 341:186–91
34. Li FY, Chaigne-Delalande B, Su H, Uzel G, Matthews H, Lenardo MJ. 2014. XMEN disease: a new primary immunodeficiency affecting Mg^{2+} regulation of immunity against Epstein-Barr virus. *Blood* 123:2148–52
35. Dobbs K, Dominguez Conde C, Zhang SY, Parolini S, Audry M, et al. 2015. Inherited DOCK2 deficiency in patients with early-onset invasive infections. *N. Engl. J. Med.* 372:2409–22
36. Okada S, Markle JG, Deenick EK, Mele F, Averbuch D, et al. 2015. Impairment of immunity to *Candida* and *Mycobacterium* in humans with bi-allelic RORC mutations. *Science* 349:606–13
37. Zhang Q, Davis JC, Lamborn IT, Freeman AF, Jing H, et al. 2009. Combined immunodeficiency associated with DOCK8 mutations. *N. Engl. J. Med.* 361:2046–55
38. Zhang Y, Yu X, Ichikawa M, Lyons JJ, Datta S, et al. 2014. Autosomal recessive phosphoglucomutase 3 (PGM3) mutations link glycosylation defects to atopy, immune deficiency, autoimmunity, and neurocognitive impairment. *J. Allergy Clin. Immunol.* 133:1400–9.e5
39. Deau MC, Heurtier L, Frange P, Suarez F, Bole-Feysot C, et al. 2014. A human immunodeficiency caused by mutations in the PIK3R1 gene. *J. Clin. Investig.* 124:3923–28
40. Lucas CL, Zhang Y, Venida A, Wang Y, Hughes J, et al. 2014. Heterozygous splice mutation in PIK3R1 causes human immunodeficiency with lymphoproliferation due to dominant activation of PI3K. *J. Exp. Med.* 211:2537–47
41. Price S, Shaw PA, Seitz A, Joshi G, Davis J, et al. 2014. Natural history of autoimmune lymphoproliferative syndrome associated with FAS gene mutations. *Blood* 123:1989–99
42. Siegel RM, Frederiksen JK, Zacharias DA, Chan FK, Johnson M, et al. 2000. Fas preassociation required for apoptosis signaling and dominant inhibition by pathogenic mutations. *Science* 288:2354–57
43. Oftedal BE, Hellesen A, Erichsen MM, Bratland E, Vardi A, et al. 2015. Dominant mutations in the autoimmune regulator AIRE are associated with common organ-specific autoimmune diseases. *Immunity* 42:1185–96
44. Boisson B, Quartier P, Casanova JL. 2015. Immunological loss-of-function due to genetic gain-of-function in humans: autosomal dominance of the third kind. *Curr. Opin. Immunol.* 32:90–105
45. Rieux-Laucat F, Casanova JL. 2014. Immunology. Autoimmunity by haploinsufficiency. *Science* 345: 1560–61
46. Schubert D, Bode C, Kenefeck R, Hou TZ, Wing JB, et al. 2014. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat. Med.* 20:1410–16
47. Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, et al. 2001. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat. Genet.* 27:20–21
48. Noguchi M, Yi H, Rosenblatt HM, Filipovich AH, Adelstein S, et al. 1993. Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* 73:147–57
49. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, et al. 2015. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am. J. Hum. Genet.* 97:67–74
50. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, et al. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* 367:1921–29

51. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, et al. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12:745–55
52. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* 132:1077–130
53. Shawky R. 2014. Reduced penetrance in human inherited disease. *Egypt. J. Med. Hum. Genet.* 15:103–11
54. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, et al. 2015. Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* 162:391–402
55. Zlotogora J. 2003. Penetrance and expressivity in the molecular age. *Genet. Med.* 5:347–52
56. Bustamante J, Boisson-Dupuis S, Abel L, Casanova JL. 2014. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin. Immunol.* 26:454–70
57. Ciancanelli MJ, Huang SX, Luthra P, Garner H, Itan Y, et al. 2015. Infectious disease: life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science* 348:448–53
58. Gilissen C, Hoischen A, Brunner HG, Veltman JA. 2012. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* 20:490–97
59. Knecht C, Krawczak M. 2014. Molecular genetic epidemiology of human diseases: from patterns to predictions. *Hum. Genet.* 133:425–30
60. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. 2010. Massively parallel sequencing and rare disease. *Hum. Mol. Genet.* 19:R119–24
61. Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, et al. 2013. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121:3830–37
62. Liadaki K, Sun J, Hammarstrom L, Pan-Hammarstrom Q. 2013. New facets of antibody deficiencies. *Curr. Opin. Immunol.* 25:629–38
63. Lougaris V, Tampella G, Baronio M, Vitali M, Plebani A. 2014. The genetic heterogeneity of common variable immunodeficiency (CVID): an update. *J. Vaccines Vaccin* 5:223
64. Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. 2013. Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy. *Genomics* 102:223–28
65. Sengupta M, Sarkar D, Ganguly K, Sengupta D, Bhaskar S, Ray K. 2015. In silico analyses of missense mutations in coagulation factor VIII: identification of severity determinants of haemophilia A. *Haemophilia* 21:662–69
66. van der Velde KJ, Kuiper J, Thompson BA, Plazzer JP, van Valkenhoef G, et al. 2015. Evaluation of CADD scores in curated mismatch repair gene variants yields a model for clinical validation and prioritization. *Hum. Mutat.* 36:712–19
67. Wu J, Li Y, Jiang R. 2014. Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies. *PLOS Genet.* 10:e1004237
68. Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. 2014. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. *J. Exp. Med.* 211:2137–49
69. Zhang Y, Su HC, Lenardo MJ. 2015. Genomics is rapidly advancing precision medicine for immunological disorders. *Nat. Immunol.* 16:1001–4
70. Mohr SE, Smith JA, Shamu CE, Neumuller RA, Perrimon N. 2014. RNAi screening comes of age: improved techniques and complementary approaches. *Nat. Rev. Mol. Cell Biol.* 15:591–600
71. Oliveira JB, Bidere N, Niemela JE, Zheng L, Sakai K, et al. 2007. NRAS mutation causes a human autoimmune lymphoproliferative syndrome. *PNAS* 104:8953–58
72. Plotkin JB. 2010. Transcriptional regulation is only half the story. *Mol. Syst. Biol.* 6:406
73. Chakravarti A, Clark AG, Mootha VK. 2013. Distilling pathophysiology from complex disease genetics. *Cell* 155:21–26
74. Candotti F, O'Shea JJ, Villa A. 1998. Severe combined immune deficiencies due to defects of the common gamma chain-JAK3 signaling pathway. *Springer Semin. Immunopathol.* 19:401–15
75. Miller JN, Pearce DA. 2014. Nonsense-mediated decay in genetic disease: friend or foe? *Mutat. Res. Rev. Mutat. Res.* 762:52–64

76. Kousi M, Katsanis N. 2015. Genetic modifiers and oligogenic inheritance. *Cold Spring Harb. Perspect. Med.* 5:a017145
77. Girard SL, Dion PA, Bourassa CV, Geoffroy S, Lachance-Touchette P, et al. 2015. Mutation burden of rare variants in schizophrenia candidate genes. *PLOS ONE* 10:e0128988
78. Parikshak NN, Gandal MJ, Geschwind DH. 2015. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* 16:441–58
79. Miller JH, Page SE. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton Univ. Press
80. Holland JH. 2014. *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. Cambridge, MA: MIT Press
81. Gursky VV, Surkova SY, Samsonova MG. 2012. Mechanisms of developmental robustness. *Biosystems* 109:329–35
82. Naujokat C, Fuchs D, Berges C. 2007. Adaptive modification and flexibility of the proteasome system in response to proteasome inhibition. *Biochim. Biophys. Acta* 1773:1389–97
83. Rossi A, Kontarakis Z, Gerri C, Nolte H, Holper S, et al. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524:230–33
84. Milner JD, Brenchley JM, Laurence A, Freeman AF, Hill BJ, et al. 2008. Impaired T_H17 cell differentiation in subjects with autosomal dominant hyper-IgE syndrome. *Nature* 452:773–76
85. Flanagan SE, Haapaniemi E, Russell MA, Caswell R, Lango Allen H, et al. 2014. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat. Genet.* 46:812–14
86. Haapaniemi EM, Kaustio M, Rajala HL, van Adrichem AJ, Kainulainen L, et al. 2015. Autoimmunity, hypogammaglobulinemia, lymphoproliferation, and mycobacterial disease in patients with activating mutations in STAT3. *Blood* 125:639–48
87. Milner JD, Vogel TP, Forbes L, Ma CA, Stray-Pedersen A, et al. 2015. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. *Blood* 125:591–99
88. Boisson-Dupuis S, Kong XF, Okada S, Cypowyj S, Puel A, et al. 2012. Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. *Curr. Opin. Immunol.* 24:364–78
89. Emison ES, Garcia-Barcelo M, Grice EA, Lantieri F, Amiel J, et al. 2010. Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *Am. J. Hum. Genet.* 87:60–74
90. Brown MS, Goldstein JL. 1992. Koch's postulates for cholesterol. *Cell* 71:187–88
91. Katsanis SH, Katsanis N. 2013. Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* 14:415–26
92. Rizzo JM, Buck MJ. 2012. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prev. Res.* 5:887–900
93. Javitt GH, Carner KS. 2014. Regulation of next generation sequencing. *J. Law Med. Ethics* 42(Suppl. 1):9–21
94. Tabor HK, Auer PL, Jamal SM, Chong JX, Yu JH, et al. 2014. Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. *Am. J. Hum. Genet.* 95:183–93
95. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, et al. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* 15:565–74
96. Burke W, Antommaria AH, Bennett R, Botkin J, Clayton EW, et al. 2013. Recommendations for returning genomic incidental findings? We need to talk! *Genet. Med.* 15:854–59
97. Gargis AS, Kalman L, Bick DP, da Silva C, Dimmock DP, et al. 2015. Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat. Biotechnol.* 33:689–93
98. Collins FS, Varmus H. 2015. A new initiative on precision medicine. *N. Engl. J. Med.* 372:793–95
99. Sassi A, Lazaroski S, Wu G, Haslam SM, Fliegauf M, et al. 2014. Hypomorphic homozygous mutations in phosphoglucomutase 3 (PGM3) impair immunity and increase serum IgE levels. *J. Allergy Clin. Immunol.* 133:1410–19.e13

100. Stray-Pedersen A, Backe PH, Sorte HS, Morkrid L, Chokshi NY, et al. 2014. PGM3 mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. *Am. J. Hum. Genet.* 95:96–107
101. Maguire ME, Cowan JA. 2002. Magnesium chemistry and biochemistry. *Biometals* 15:203–10
102. Alangari A, Alsultan A, Adly N, Massaad MJ, Kiani IS, et al. 2012. LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency. *J. Allergy Clin. Immunol.* 130:481–88.e2
103. Burns SO, Zenner HL, Plagnol V, Curtis J, Mok K, et al. 2012. LRBA gene deletion in a patient presenting with autoimmunity without hypogammaglobulinemia. *J. Allergy Clin. Immunol.* 130:1428–32
104. Lopez-Herrera G, Tampella G, Pan-Hammarstrom Q, Herholz P, Trujillo-Vargas CM, et al. 2012. Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am. J. Hum. Genet.* 90:986–1001
105. Tivol EA, Borriello F, Schweitzer AN, Lynch WP, Bluestone JA, Sharpe AH. 1995. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* 3:541–47
106. Waterhouse P, Penninger JM, Timms E, Wakeham A, Shahinian A, et al. 1995. Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* 270:985–88
107. Walker LS, Sansom DM. 2011. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. *Nat. Rev. Immunol.* 11:852–63
108. Wing K, Yamaguchi T, Sakaguchi S. 2011. Cell-autonomous and -non-autonomous roles of CTLA-4 in immune regulation. *Trends Immunol.* 32:428–33
109. Warnatz K, Denz A, Drager R, Braun M, Groth C, et al. 2002. Severe deficiency of switched memory B cells ($CD27^+IgM^-IgD^-$) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* 99:1544–51
110. Isnardi I, Ng YS, Menard L, Meyers G, Saadoun D, et al. 2010. Complement receptor 2/$CD21^-$ human naive B cells contain mostly autoreactive unresponsive clones. *Blood* 115:5026–36
111. Lo B, Zhang K, Lu W, Zheng L, Zhang Q, et al. 2015. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science* 349:436–40
112. Wang JW, Howson J, Haller E, Kerr WG. 2001. Identification of a novel lipopolysaccharide-inducible gene with key features of both a kinase anchor proteins and chs1/beige proteins. *J. Immunol.* 166:4586–95
113. Gebauer D, Li J, Jogl G, Shen Y, Myszka DG, Tong L. 2004. Crystal structure of the PH-BEACH domains of human LRBA/BGL. *Biochemistry* 43:14873–80
114. de Souza N, Vallier LG, Fares H, Greenwald I. 2007. SEL-2, the *C. elegans* neurobeachin/LRBA homolog, is a negative regulator of *lin-12/Notch* activity and affects endosomal traffic in polarized epithelial cells. *Development* 134:691–702
115. Qureshi OS, Kaur S, Hou TZ, Jeffery LE, Poulter NS, et al. 2012. Constitutive clathrin-mediated endocytosis of CTLA-4 persists during T cell activation. *J. Biol. Chem.* 287:9429–40
116. Mead KI, Zheng Y, Manzotti CN, Perry LC, Liu MK, et al. 2005. Exocytosis of CTLA-4 is dependent on phospholipase D and ADP ribosylation factor-1 and stimulated during activation of regulatory T cells. *J. Immunol.* 174:4803–11
117. HERA Study Group. 1995. A randomized trial of hydroxychloroquine in early rheumatoid arthritis: the HERA Study. *Am. J. Med.* 98:156–68
118. Clegg DO, Dietz F, Duffy J, Willkens RF, Hurd E, et al. 1997. Safety and efficacy of hydroxychloroquine as maintenance therapy for rheumatoid arthritis after combination therapy with methotrexate and hydroxychloroquine. *J. Rheumatol.* 24:1896–902
119. Costedoat-Chalumeau N, Dunogue B, Morel N, Le Guern V, Guettrot-Imbert G. 2014. Hydroxychloroquine: a multifaceted treatment in lupus. *Presse Med.* 43:e167–80
120. Crank MC, Grossman JK, Moir S, Pittaluga S, Buckner CM, et al. 2014. Mutations in PIK3CD can cause hyper IgM syndrome (HIGM) associated with increased cancer susceptibility. *J. Clin. Immunol.* 34:272–76
121. Hartman HN, Niemela J, Hintermeyer MK, Garofalo M, Stoddard J, et al. 2015. Gain of function mutations of PIK3CD as a cause of primary sclerosing cholangitis. *J. Clin. Immunol.* 35:11–14

122. Kracker S, Curtis J, Ibrahim MA, Sediva A, Salisbury J, et al. 2014. Occurrence of B-cell lymphomas in patients with activated phosphoinositide 3-kinase delta syndrome. *J. Allergy Clin. Immunol.* 134:233–36
123. Cantley LC. 2002. The phosphoinositide 3-kinase pathway. *Science* 296:1655–57
124. Yuan TL, Cantley LC. 2008. PI3K pathway alterations in cancer: variations on a theme. *Oncogene* 27:5497–510
125. Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, et al. 2014. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* 370:997–1007
126. Shah A, Mangaonkar A. 2015. Idelalisib: a novel PI3Kδ inhibitor for chronic lymphocytic leukemia. *Ann. Pharmacother.* 49:1162–70