PART
I

Critical Concepts

1 Introduction to Population Diversity and Genetic Testing

Michael D. Caldwell, Ingrid Glurich, Kimberly Pillsbury, and James K. Burmester

OVERVIEW OF HOW GENETIC DIVERSITY ARISES

Genetic diversity arises from differences in the genome of humans. Mapping of the human genome has revealed that humans are approximately 99.9 percent identical relative to their DNA sequences, with differences occurring at the rate of one change in every 100 to 300 bases along the sequence of 3 billion bases that comprise the human genome (1, 2). These differences in coding sequences at these sites are termed *single-nucleotide polymorphisms* (SNPs) and are considered significant when they occur with a minimum frequency of 1 percent in a given population. Whereas genetic differences of 0.1 percent among individual humans appear negligible, occurrences of ~10 million SNPs have been cataloged to date. These SNPs may occur in coding or noncoding regions of the genome and may affect gene expression or disease susceptibility. Whereas SNPs are estimated to account for 90 percent of all genetic variability, other genetic differences have been detected and may stem from errors during DNA replication, including copy number variations, insertions, deletions, inversions of bases, or other mutational events caused by environmental factors (3). This genetic variability contributes to genetic diversity among populations as well as its individual members. Genetic diversity has an impact on all manner of human traits from external appearance to disease susceptibility and response to pharmacological agents.

To gain a greater appreciation of the magnitude of the impact that genetic diversity has on human phenotypes, it is best to begin with an exploration of its historical development in modern humans. In general, race and ethnicity, which are largely defined culturally by phenotypic traits or defined by man-made designations of geographic origin, become much less distinct at the genomic level. Mutations can occur within *alleles* that occupy a distinct site within a given gene or genomic region and thereby help to define traits. It is thought that population genetic diversity is largely due to a combination of allelic mutation at specific sites along the genome and the selective pressure from population segregation that occurred during the migration of humans out of Africa over a period of about 200,000 years (4). Each offspring inherits two alleles, one from each parent. Thus, if either parent is not a carrier of the wild-type allele originally encoded in the genome due to local mutations within the allele, the offspring may inherit the mutant allele and be heterozygous for a given trait. The mutation may be dominant (i.e., expressed), recessive (carried silently), or, in some cases, coexpressed. When a new mutation is associated with a beneficial trait, it is thought that *positive selection* occurs, allowing those carrying the beneficial allele to survive, reproduce, and pass on the trait to their offspring at a frequency dictated by its pattern of inheritance. However, the major contribution to genetic diversity occurred because of the geographical isolation that resulted from tribal resettlement following migration that produced colonies of individuals with a reduced genetic pool representative of the founders of each new colony.

Figure 1.1 shows the migration pattern across Europe and Asia. As the oldest "modern humans," Africans have had the most time to accumulate changes in their DNA, making them the most genetically diverse race and the ancestral foundation for the evolution of all other races as they are presently defined. As tribes of humans left Africa and settled in a new location, these migrants distributed a subset of the total gene pool in the locations where they resettled. Figure 1.2 shows the genetic relatedness of Africans with Europeans and Asians. The traceable effect on the evolving gene pool at the site of resettlement exerted by the small number of individuals newly migrated into an area is known as the *founder effect*. The founders reflect their own limited genetic variation relative to the larger population from which they originated. At the same time, any new mutation that arose in the colony was not dispersed back into the ancestral

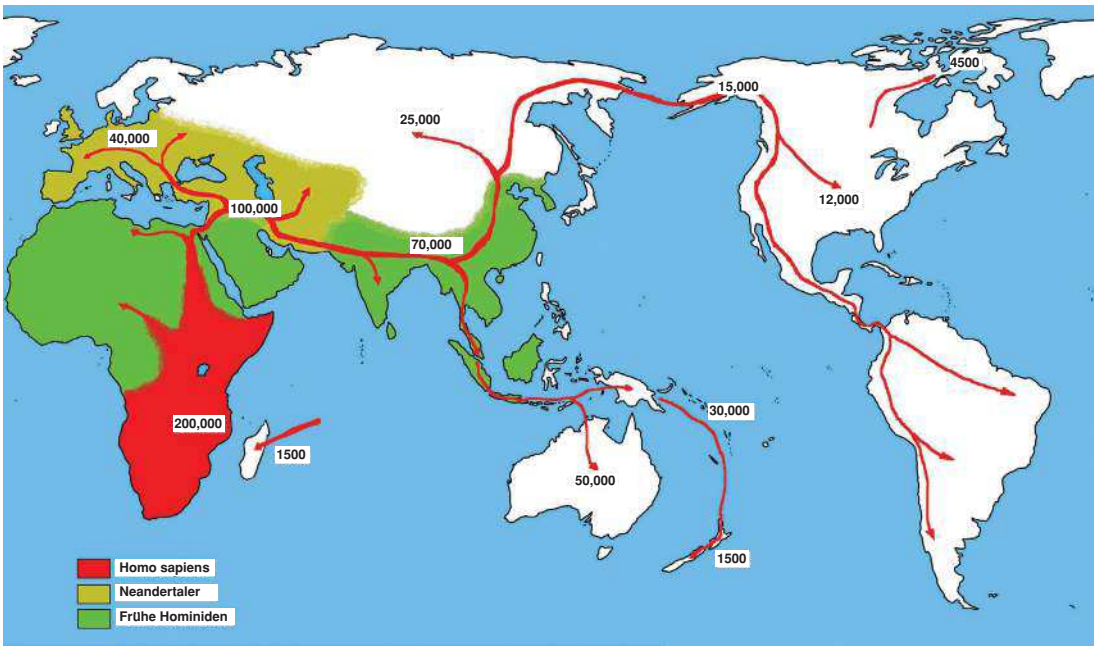

Figure 1.1. Homo sapiens spreading over the world.

foundation population in Africa. This resulted in isolated local offspring that carried the new mutation. When the mutation was associated with disease, the frequency of that disease increased in relative proportion to its allele frequency and penetrance within the population.

Within Africa, genetic diversity correlates with cultural variation and the origin of tribal language (5). Africans are descended from 14 ancestral populations that settled across the continent, establishing cultural and linguistic boundaries. The voluntary migrations out

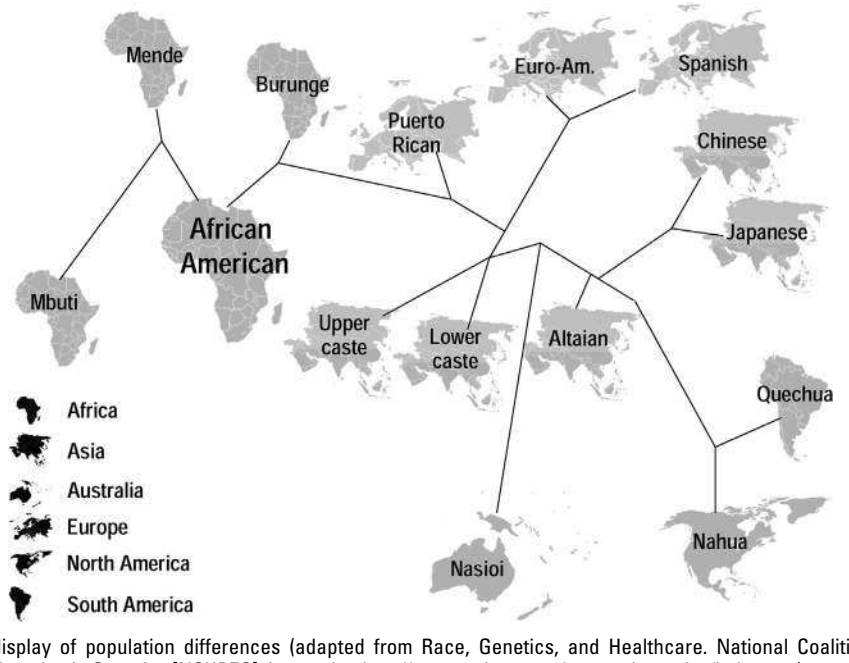

Figure 1.2. A display of population differences (adapted from Race, Genetics, and Healthcare. National Coalition for Health Professional Education in Genetics [NCHPEG]. Located at http://www.nchpeg.org/raceandgenetics/index.asp).

of Africa occurred from East Africa, whereas the involuntary migrations that occurred with the slave trade were from West Africa. As a result, African Americans have mixed ancestry from the expanse of West Africa. The ratios of genetic origins reflected among African Americans is, on average, 71 percent West African, 8 percent from other parts of Africa, and 13 percent European. The identification of genetic admixture underlying genetic diversity has introduced challenges in integrating genetics into best medical practice. Whereas mapping of the human genome and improved understanding of genetic diversity have given an appreciation of where genetic differences and similarities lie, advancement in our understanding of optimal application of genetics to the management of human health depends on a deeper understanding of how to expand and bridge our application of genetic diversity. Approaches that have been undertaken to advance our understanding include:

- further definition of the human genome through the HapMap Project, and
- association studies that are designed to specifically define clinical manifestations consequential to genetic diversity.

ROLE OF THE HapMap PROJECT

The objective of the International HapMap Project was to establish common patterns of variations in human DNA sequences that could help investigators discover genetic factors influencing vulnerability to disease and variability in drug response (6). The HapMap Project was central to cataloging the identity, position, and frequency of polymorphisms that occur in the human genome across races and ethnicities. This project has developed genome-wide maps for European, African, and Asian populations. The project compiled data on 270 DNA samples from four distinct populations. Samples from two populations represented trios that included two parents and one adult child: these were a Utah (U.S.) population of largely Northern and Western European ancestry ($n = 30$ trios) and a cohort of Yoruba people in Ibadan, Nigeria ($n = 30$ trios). The other two populations consisted of unrelated Japanese from Tokyo, Japan ($n = 45$), and Han Chinese from Beijing, China ($n = 45$). The HapMap Project made it possible to identify 100,736 SNPs that appeared uniquely in each ethnic group (7).

Furthermore, the HapMap Project delineated relationships among SNPs. Ten million SNP sites are estimated within the human genome (6). The HapMap Project capitalized on the observation that SNP alleles in close proximity to each other are often strongly associated and inherited as a "block" (8). When two SNPs are thus associated, they are said to be in linkage disequilibrium (LD). When SNPs exhibit a high degree of LD, their respective alleles are almost always inherited together. This allows detection of the presence of one SNP in an individual's genome to be highly predictive of the presence of the other SNPs in its haplotype block (8, 9). Those SNPs that are used to identify haplotype blocks are called "tag SNPs." The HapMap Project took advantage of the presence of tag SNPs to create the haplotype map of chromosomes (10, 11). Genotyping for tag SNPs as an indicator to detect coinheritance of SNPs in haplotype blocks obviates the need to genotype all 10 million common SNPs within the genome to detect SNPs associated with genetic traits (8, 12, 13).

An example of a haplotype block is shown in Figure 1.3. This pharmacogenetic example highlights the gene VKORC1, which encodes an enzyme involved in the reduction of vitamin K1 into an active cofactor for the synthesis of γ-glutamyl carboxylated clotting factors. The anticoagulant warfarin blocks the activity of this enzyme, thus reducing the levels of proteins essential for clot formation. Variant alleles have been detected within VKORC1 genes, and these define the amount of the enzyme that is made. In turn, the level of activity of the enzyme, along with several other factors defined to date, determines the dose of warfarin required to achieve stable levels of anticoagulation. Haplotype blocks H1 and H2 are present in individuals that require less warfarin to achieve optimal anticoagulation, and haplotype blocks H7, H8, and H9 are present in individuals that require higher doses of warfarin (14). These blocks can be identified by use of the tag SNPs that are highlighted in Figure 1.3, and, as shown, genetic testing of these tag SNPs can be used to modulate the warfarin dose.

In phase I of the HapMap Project, 1 million SNPs were genotyped (7) and another 2.1 million SNPs were genotyped in phase II (15). Park et al. (16) compared the genomic profiles of the four populations in the HapMap Project database and identified the ethnically variant single-nucleotide polymorphisms (ESNPs) using the nearest shrunken centroid method (NSCM). From the top eighty-two ESNPs initially selected to classify the populations, Zhou and Wang (17) established a set of sixty-four SNPs that classified the HapMap ethnic populations more efficiently.

Several studies have emphasized the importance of population diversity within the context of the HapMap database, so that the haplotype maps for different populations would help in the identification of similarities and differences among ethnicities (18). For example, Lin, Hwang, and Tzeng (19) analyzed the practicality of using SNP data from the HapMap database to represent the overall Taiwanese (TWN) population in association studies. Results of the study showed that only the Han Chinese in the Beijing population of the HapMap database characterized the TWN in terms of allele and haplotype frequencies and could be used in association studies to represent the TWN. A study by Ouyang and

	Haplotypes	Frequency		Dose (mg/d)
H1	***C***---C---G---***A***---T---C---***T***---C---***T***---G	12%	2.9	**Low Dose**
H2	***C***---C---G---***A***---G---C---***T***---C---***T***---G	24%	3.0	
H7	***T***---C---G---***G***---T---C---***C***---G---***C***---A	35%	6.0	
H8	***T***---A---G---***G***---T---C---***C***---G---***C***---A	8%	4.8	**High Dose**
H9	***T***---A---C---***G***---T---T---***C***---G---***C***---G	21%	5.5	

Figure 1.3. Polymorphisms are listed in order for positions 381, 861, 2653, 3673, 5808, 6009, 6484, 6853, 7566, and 9041 of reference sequence AY587020 in GenBank (adapted from Rieder et al. *N Eng J Med.* 2005;352:2285–93).

Krontiris (18) found that a highly conserved haplotype structure could be established among ethnic populations having African ancestry and non-African populations. The study also found evidence of common ancestry in all the population groups.

There are plans to extend the HapMap Project to include seven other populations to provide information on less common variants that have the potential to affect individuals (15). These populations are a Luhya cohort in Webuye, Kenya; the Maasai in Kinyawa, Kenya; the Tuscans in Italy; the Gujarati Indians in Houston; the Chinese in Denver; a cohort with Mexican ancestry in Los Angeles; and a cohort with African ancestry in the southwestern region of the United States (15). Whereas it is expected that most common haplotypes are found in all human populations, the frequency of a specific haplotype may differ among populations. Another important future goal of the HapMap Project will be to create molecular phenotypes for the DNA samples and to combine the SNP information with associated structural variations (15).

OVERVIEW OF ASSOCIATION STUDY DESIGN AND CONSIDERATIONS

The availability of the HapMap database has greatly facilitated the study of the relative association of disease traits with specific SNPs or groups of SNPs defined by a haplotype. In these studies, DNA variants from a cohort of individuals affected with the disease of interest are compared with the sequence of a cohort of individuals not affected by the disease. Statistical methods are used to determine whether differences in allele frequency are significant after adjustment for multiple comparisons. Then it is important to test any newly identified genetic association in additional cohorts to replicate the finding and to determine whether the association relates to different racial groups.

The initial approach to association studies was limited to sequencing potential candidate genes for variants that were present at higher frequency in affected individuals in comparison with unaffected controls (20). Some candidate gene studies were based on the detection of a mutation in a gene of an individual affected with the disease. Often these studies were done within families by comparing affected and unaffected family members. Whereas candidate gene analysis is useful in some instances, an inherent limitation to this approach is that families represent a very narrow example of genetic diversity because of their high degree of relatedness. As a result, association may be much more limited or absent in the broader population, which will exhibit higher levels of genetic diversity.

Availability of the HapMap database allowed an alternative approach to association studies in which SNPs are used as genetic markers to discover associations between disease and genomic regions. This indirect approach for detecting sequence variation and disease gene identification is more effective and efficient than the use of the candidate gene approach, because it allows a small set of tag SNPs to identify the common patterns of variation in the genome with a high probability of detection for a disease-gene association (6). Candidate genomic regions thus identified are then more densely mapped by genotyping additional SNPs occurring within the regions that are in close proximity to each other. Detection of high levels of LD among the SNPs helps to narrow the region and allows identification of genes or other genomic regions that underlie the detected association (21).

The newest approach to association studies is the *genome-wide association study* (GWAS) or *whole genome association study*. This approach involves a more global examination of genetic variation across the entire genome for the purpose of detecting genetic associations and linking them to observable traits. These studies use a case-control design in which both cases and controls are genotyped. Data are analyzed using bioinformatics approaches to identify regions exhibiting genetic variability between case and control cohorts. If a higher frequency of an SNP is detected in people with the disease, a level of association is defined. This approach allows

identification of genomic regions associated with disease in a non–hypothesis-driven manner (22). A list of genetic associations identified by GWAS is summarized at http://www.genome.gov/gwastudies/. In many GWAS studies, the variant that is identified as associated with the disease is unlikely to be the variant that actually causes the disease. In many cases, the variant is in LD with the causal variant, and additional functional studies are required to understand the physiological role of the variants.

Several genotyping platforms are available to conduct GWAS and other association studies. The Affymetrix genotyping array is designed to use randomly selected, evenly spaced SNPs across the whole genome. The Illumina platform uses fewer SNPs, but uses HapMap tag SNPs. Because the use of tag SNPs is more efficient than using the same number of randomly selected SNPs, the difference in power can be as large as 20 percent, with coverage in Africans being most difficult. The use of tag SNPs for association studies results in a 5 percent to 10 percent reduction of power compared with using all HapMap SNPs (23).

LINKAGE DISEQUILIBRIUM STRUCTURE IN POPULATIONS OF DIVERSE GEOGRAPHICAL ORIGIN

The history of the population being studied affects LD patterns within the population. As the ancestral population from which all others evolved, it follows that the most diverse haplotype structure is found in Africa. As distance from Africa increases, diversity in haplotype structure decreases, reflecting the origin of humans in Africa and migration and segregation patterns around the world. This is consistent with each group of migrants drawing a sample, but not a complete set, of haplotypes from the population at the original site they left. Linkage disequilibrium patterns describe population history, human migration, natural selection, and localization of preferred recombination sites.

Recombination hot spots are loci prone to exhibit variability in sequence and are similar across different racial groups. However, environmental pressures have heavily influenced resulting genetic expression. "Bottlenecks" are one example of how environment can have an impact on genetic expression. Genetic bottlenecks are brought about by cataclysmic events that markedly reduce the number of humans, consequentially causing the extinction of many genetic lineages within the population, and thereby decreasing genetic diversity. One example of a bottleneck occurred about 70,000 years ago in Africa (24). It is proposed that the supereruption of the volcano Toba spewed volcanic ash into the atmosphere, resulting in drought and famine. As a result, it is estimated that only 5,000 females survived in Africa. When a bottleneck reduces a population that later expands, the result can be *genetic drift* in which changes in allele frequency occur independently of selection pressure, sometimes at the cost of elimination of beneficial adaptations. Bottlenecks and genetic drift are associated with decreased LD.

Population stratification due to *population admixture*, in which a cohort reflects a mixture of several subpopulations, may also have an impact on LD. In this scenario, a subpopulation with altered frequencies for expression of a given allele will cause the perception of an increased LD. For example, population admixture can occur in the presence of founder effect where a rare allele in the originating generation has penetrated one of the subpopulations in the admixture resulting in allele frequencies that are incongruent with natural selection. This may result in the misleading appearance of a higher frequency of a genetic disease within the combined population. To control for type 1 error due to the presence of population admixture in association studies, data are subjected to Hardy-Weinberg equilibrium testing (25).

There is low portability of HapMap tag SNPs in Africans, because the LD length in these individuals is smaller than the length in whites and Asians. In contrast to European and Asian data, where tag SNPs are valuable for identifying a haplotype block, these same tag SNPs do not efficiently identify haplotype blocks in Africans. By contrast, tag SNPs transfer well among white populations living in distant regions of the world, but they do not transfer nearly as well between more distant populations such as African Americans and Africans (26–28). It is also difficult to estimate how well tag SNPs will cover rare SNPs.

POPULATION STRATIFICATION ISSUES IN STUDY DESIGN

In association studies using SNPs, a subpopulation of people with an index condition are compared with otherwise-matched individuals derived from the same population. However, the same limitations that apply to familial candidate gene analysis may apply when this association is tested in populations that exhibit higher genetic diversity in comparison with the index population. An understanding of population structure is critical to avoiding false associations and allows assessment of the extent to which population stratification may affect the results of an association study (29). Nonrandom distribution of individuals in an association study results in *population stratification*. In the presence of population admixture, an association may be inferred if the trait distribution and allele prevalence differ between the subpopulations, even though no biological influence on the trait is present and the locus is not linked to any gene that influences the trait. In this case, population stratification

can lead to the false association of a DNA variant with a disease.

Another form of population stratification leading to false association in a case-control design occurs when the population being studied has subpopulations, and the case and control groups have different proportions of each subpopulation represented in each study arm (30). Allele and genotype frequencies between the case and control groups should be the same. Several current methods for assessing population stratification involve SNP frequency analyses (31, 32). Whereas association studies are generally performed using large sample sizes to increase the strength of the association and the significance of the findings, it is important to determine whether population stratification is present to avoid false-positive results. For example, a study by Yamaguchi-Kabata et al. (29) showed that, within the same Japanese population, genetic differentiation created subpopulations according to geographic region. Different stratifications of the Japanese subpopulations in the case and control groups caused more false-positive results with increasing sample size (29).

The choice of the optimal statistical method for a population-based association study should be a function of study population sampling. Four population-based statistical methods were evaluated in different population stratification levels for their ability to reduce the influence of the stratification in an association study (33). Data from the HapMap Project were used to imitate a stratified population, and the four statistical methods were applied, including the traditional case-control test (TCCT), structured association (SA), genomic control (GC), and principal components analysis (PCA) (26, 33–35). PCA had a low rate of false-positive results with a high level of accuracy, making it the best choice to correct for population stratifications in association studies (33). SA and PCA had comparable results as long as adequate ancestral informative markers were included in the SA analysis (33). The ability of GC to correct for population stratification depended on the level of stratification; the GC analysis was effective only in studies with low levels of population stratification (33).

Most GWAS attempt to correct for population stratification during analysis of the genetic data. Several methods have been developed to correct for admixture in association studies (34–37). Other issues in study design that have an impact on the analysis of association include small sample size, small effect of alleles on disease causation, parent of origin effects, interaction between environment and genes, and copy number changes throughout the genome at genes of interest. Because cases are more likely than controls to be related in diseases with a genetic basis, the assumption of independence of observations is violated. This may lead to overestimation of the size of the association (38).

POPULATION STRATIFICATION ISSUES RESULTING FROM EXTRAPOLATING RESULTS FROM ONE POPULATION TO ANOTHER

The effect of population structure on common SNP variation is considerable and emphasizes the need for understanding both ancestry and stratification in association studies (39). Some variants linked to common disease risk in Americans of European descent have significantly different frequencies from those in other American ancestral populations, making it difficult to study the effects of the variant on disease risk (28, 40). This also limits the ability to extrapolate findings for a single population to the entire human genome.

Admixture among races is common and increasing, making it difficult to apply personalized medicine to patient treatment based solely on the frequency of polymorphisms of medical relevance in a racial population. Personalized treatment of disease and response to drug therapy rely on inherited genetic individuality. This individuality necessitates genetic testing of each patient for the polymorphisms prevalent in the racial background. However, because of admixture, it becomes necessary to expand the testing panel to cover disease-causing polymorphisms frequent in other racial groups to ensure that patients of mixed background receive complete genetic testing appropriate to them or to concomitantly use panels that define ethnic diversity. The application of genetic testing to pharmacogenetics requires that all potentially important alleles are tested on the basis of the race of the patient. Fortunately, multiplex genetic testing platforms are available, making it possible to test all of the known relevant functional alleles in genes that modify drug response.

Characterization of common SNPs among ancestral populations helps test the influence of common variants on disease risk (41). The degree to which common variants may account for disease risk across populations depends in part on whether alleles are common, or at least shared, among different populations. Even if SNPs are common among differing populations, the frequency of variants may differ between groups (27). Common alleles that influence the risk of common disease vary greatly in frequency across populations (27). Population-specific natural selection may have perpetuated variations in frequency of the genetic variants that play a role in the common disease risk for different populations (28). A study comparing the differentiation in frequencies of disease-associated SNPs and random SNPs in the genomes of Europeans and Africans found that ethnicity was not a good predictor of either disease-associated or random variants (28). Thus, ethnicity cannot easily be tied to risk of genetically based common diseases, because frequencies of

the risk alleles do not vary significantly among ancestral populations.

COMMON VARIANTS VERSUS RARE VARIANTS

The "common disease, common variant" hypothesis states that the genetic impact on common diseases is due to a limited number of DNA variants with a frequency greater than 1 percent within the disease-causing gene. At the outset of the HapMap Project, there were few data supporting the identification of genes for common disease. As a result, the common disease, common variant hypothesis remained controversial. The alternate hypothesis, the "rare variant hypothesis," states that inherited predisposition to common diseases is the result of additive affects of multiple low-frequency dominant alleles, each independently contributing variants of multiple genes, each of which confers a moderate but detectable increase in risk (42). Taken to the extreme, the rare variant hypothesis would predict that each disease-causing mutation might only be found once in the population. In many cases, rare variants are most pronounced in specific ethnic groups because of a founder effect. Both the common disease/common variant and common disease/rare variant causes of disease are correct, depending on the gene and disease examined. In the case where the risk-associated allele is very rare, by definition, multiple genes must be involved to manifest a common disease. By contrast, the common disease/common variant hypothesis holds that genetic variants that cause disease should be present at least at 1 percent to 5 percent across a population.

Rare variants are important in inherited disease, especially where the mutation results in a highly penetrant manifestation of disease. In these cases, the mutation causes disease in almost all individuals that inherit the mutation. Approximately 5 percent of colorectal cancers are inherited in a familial Mendelian manner. These cancers are caused by highly penetrant deleterious mutations in genes such as HNPCC, mismatch repair genes (MLH1 and MSH2), Wnt signaling genes (APC, AXIN1, and CTNNB1), and others. Similarly, approximately 5 percent of breast cancer results from rare inherited mutations in the BrCa1, BrCa2, or FANCF genes.

In many cases, multiple genes have been identified for the common diseases, with each new variant contributing a small amount (1.0–1.5) to the odds ratio (OR) attributed to development of disease. Because of the low contribution to OR, it has been important to do very large studies and then replicate the observation in several additional cohorts. In addition, because of the small ORs, studies have had to account for confounding effects of population structure, population admixture, and the testing of a very large number of SNPs.

POSITIVE SELECTION AND DISEASE INCIDENCE

Genetic diversity among populations is thought to support survival of the population through positive selection. Positive selection predicts an increase in protective gene alleles that support fitness and is thereby considered associated with the emergence of new phenotypes. Population segregation also clearly adds to this selection. Selection is also thought to be driven by environmental factors present in a given geographical location that increases the risk of a population for certain diseases. The increase in frequency of a protective gene allele is thought to counterbalance the environmental factor resulting in increased fitness of individuals with the protective allele, whereas those with disease susceptibility exhibit lower survival. Because environmental factors are not evenly distributed across all geographic portions of the world, the protective allele would then increase in frequency in localized regions but not in other distant populations. Genes affected by positive selection could cover a diverse set of phenotypic characteristics including host-pathogen interactions, reproduction, dietary influences, and facial and body appearances and attributes.

An example of positive selection that occurred in Africa at the level of the β-globin gene was the emergence of red blood cell sickling based on protection against malarial infection of the red blood cells by *Plasmodium falciparum* and *Plasmodium vivax* organisms. Individuals that carried one mutant allele had greater resistance to malaria than homozygous wild-type individuals and thus were more protected from malarial disease. As a result, the allele frequency has risen under selective pressure in malaria-infested regions of Africa. However, this trait can also be deleterious, because individuals that are homozygous for the mutant β-globin allele develop sickle-cell anemia.

The identification of genes underlying common diseases holds substantial potential impact on human health. However, genetic diversity will challenge whether genetic observations made in one population will be applicable to another population. Over time, the allelic frequencies at specific sites may have undergone changes across distinct populations, and alleles that are important in the development of complex diseases in one population may not be relevant in another population because of the genetic divergence of the populations. It is therefore imperative that the impact of genetics and the environment be compared across populations to accurately assess disease susceptibility, response to therapy, and environmental triggers for disease.

GENETIC DIVERSITY AND PERSONALIZED MEDICINE

A thorough understanding of genetic diversity brings with it the promise of personalized medicine. Broadly defined, *personalized medicine* is health management informed by knowledge of the underlying genetics of each individual. For example, medical management of patients by the use of a personalized medicine approach would assess genetically encoded capacity for drug metabolism and disease susceptibility. Personalized medicine represents a substantial shift in the current population-based paradigm for medical practice (evidence-based medicine) that extrapolates algorithms based on the average observed global experience of the broader population and applies them to its individual members. For example, scientists have classically conducted drug trials using randomized case-control experimental designs within a given population and administer only those drugs that achieve a defined standard of safety and efficacy to the population at large. However, it is well known that there is considerable interindividual variation in patient response to pharmacological agents attributable to genetically based capacity for drug metabolism (43). In many cases, the polymorphisms that determine drug *A*bsorption, *D*istribution, *M*etabolism, and *E*limination (ADME) vary markedly in frequency across different individuals and racial heritages.

An understanding of population diversity is central to developing personalized medicine. For example, to develop relevant genetic tests that predict response to medications, it is imperative to understand the genetic variation prevalent among diverse races that define drug responsiveness (44). Furthermore, it is important to recognize how population diversity confounds the definition of the genetic impact on disease susceptibility and how this must be taken into account when designing studies that test disease association with genetic factors.

As our understanding of genetic diversity expands, it grows increasingly clearer that the "one size fits all" approach to medicine is no longer relevant in most circumstances. An example of the growing acceptance of the concept of personalized medicine can be found in the approval of the drug Bidil in 2005 by the Food and Drug Administration specifically for African Americans with heart failure. This was possible only after strong drug efficacy was noted in this subpopulation, but not in the broader population (44). A deeper understanding of genetic diversity among populations has concomitantly blurred the artificial distinctions of race and ethnicity that have risen historically based on phenotypic characteristics and social organization. Ideally, to realize personalized medicine, as genetic diversity among populations and their individual members becomes more defined and the optimization of study design increases, a clearer understanding of disease processes relevant to specific subpopulations will emerge. These developments will allow for identification of relevant risk factors, for preventive medicine strategies, and for development of effective postdiagnostic pharmacological treatment and medical care tailored in a population- and patient-specific manner.

REFERENCES

1. Guttmacher AE & Cullins FS. Genomic medicine – a primer. *N Engl J Med.* 2002;**347**:1512–20.
2. Altshuler D, Daly MJ, & Lander ES. Genetic mapping in human disease. *Science.* 2008;**322**:881–8.
3. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tüzün E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, & Eichler EE. Mapping and sequencing of structural variation from eight human genomes. *Nature.* 2008;**453**:56–64.
4. Relethford JH. Genetic evidence and the modern human origins debate. *Heredity.* 2008;**100**:555–63.
5. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, & Williams SM. The genetic structure and history of Africans and African Americans. *Science.* 2009;**324**:1035–44.
6. International HapMap Consortium. The International HapMap Project. *Nature.* 2003;**426**:789–94.
7. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;**437**:1299–1320.
8. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, & Altshuler D. The structure of haplotype blocks in the human genome. *Science.* 2002;**296**:2225–9.
9. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, & Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science.* 2004;**304**:581–4.
10. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, & Lander ES. High-resolution haplotype structure in the human genome. *Nature Genet.* 2001;**29**:229–32.
11. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, & Todd JA. Haplotype tagging for the identification of common disease genes. *Nature Genet.* 2001;**29**:233–7.