Chapter 5
Direct-to-Consumer Genetic Testing

Richard A. Stein
Princeton University, USA

ABSTRACT

The 1953 discovery of the DNA double-helical structure by James Watson, Francis Crick, Maurice Wilkins, and Rosalind Franklin, represented one of the most significant advances in the biomedical world (Watson and Crick 1953; Maddox 2003). Almost half a century after this landmark event, in February 2001, the initial draft sequences of the human genome were published (Lander et al., 2001; Venter et al., 2001) and, in April 2003, the International Human Genome Sequencing Consortium reported the completion of the Human Genome Project, a massive international collaborative endeavor that started in 1990 and is thought to represent the most ambitious undertaking in the history of biology (Collins et al., 2003; Thangadurai, 2004; National Human Genome Research Institute). The Human Genome Project provided a plethora of genetic and genomic information that significantly changed our perspectives on biomedical and social sciences. The sequencing of the first human genome was a 13-year, 2.7-billion-dollar effort that relied on the automated Sanger (dideoxy or chain termination) method, which was developed in 1977, around the same time as the Maxam-Gilbert (chemical) sequencing, and subsequently became the most frequently used approach for several decades (Sanger et al., 1975; Maxam & Gilbert, 1977; Sanger et al., 1977). The new generations of DNA sequencing technologies, known as next-generation (second generation) and next-next-generation (third generation) sequencing, which started to be commercialized in 2005, enabled the cost-effective sequencing of large chromosomal regions during progressively shorter time frames, and opened the possibility for new applications, such as the sequencing of single-cell genomes (Service, 2006; Blow, 2008; Morozova and Marra, 2008; Metzker, 2010).

DOI: 10.4018/978-1-61692-883-4.ch005

INTRODUCTION

While $10 allowed the sequencing of a single base in 1985, approximately 10,000 bases were generated, at the same cost, two decades later (Shendure et al. 2004; Pettersson et al. 2009). Next-generation sequencing methodologies, in addition to their lower costs, enabled large chromosomal regions to be sequenced in shorter time frames – for example, over 100 million base pairs can now be read, in 200-400 base-pair fragments, in as little as 4 hours (Imelfort et al., 2009).

As a result of these advances, it took only two months in 2008, and less than 1 million dollars, to sequence a diploid human genome, that of James Watson (Wheeler et al., 2008). All these developments promise to bring the $1,000 genome goal, an important milestone of personalized medicine, closer to reality.

The Human Genome Project revealed that 20,000-25,000 genes are encoded in the genome, less than the approximately 50,000-100,000 that were predicted a few years earlier (Fields et al., 1994; International Human Genome Sequencing Consortium, 2004). As a result of gene-disease and gene-phenotype connections that are continually unveiled, genetic testing has experienced an unprecedented development.

INTER-INDIVIDUAL GENETIC VARIATIONS

The Human Genome Project unveiled similarities and differences at the DNA level within and between populations. Most importantly, it revealed that two unrelated individuals are approximately 99.9% identical at the DNA level (Shastry, 2002; Feuk et al., 2006). A very interesting part of the genome revolves around the remaining 0.1%, which constitutes inter-individual differences. Several million such differences, which occur approximately once every 300-1000 base pairs, exist between two random genomes, and they became known as **S**ingle **N**ucleotide **P**olymorphisms (SNPs) (Brookes, 1999; Manolio et al., 2008). SNPs were linked to many medical conditions including diabetes, cancer, and asthma, and some of them increase, while others decrease the risk to develop a specific disease (Kaklamani et al., 2008; Ionita-Laza et al., 2009; Kilpivaara et al., 2009). To characterize genomic variations and provide a public database of common genome-wide sequence variants, the International HapMap Project was initiated in October 2002 (The International HapMap Consortium, 2003). Its second-generation map, published in 2007, reported over 3.1 million SNPs, and provided a valuable resource to study gene-disease interactions (The International HapMap Consortium, 2007).

While SNPs were previously thought to represent the only source of inter-individual genetic variation, it was recently revealed that various genes or chromosomal regions exhibit inter-individual variations in their copy numbers, and this became known as **C**opy **N**umber **V**ariation (CNV). For example, a cluster of several antimicrobial β-defensin genes was shown to be present in 2 to 12 copies per diploid genome in the chromosome of different individuals, and this variation is thought to shape the inter-individual differences in susceptibility to infectious diseases (Hollox et al., 2003). CNVs range in size from 1 kb to 3 Mb (Feuk, 2006; Freeman et al., 2006; Kehrer-Sawatzki, 2007) and are thought to account for more genomic variation, as far as the number of nucleotides is concerned, than all SNPs combined. Recent estimates predicted that two human genomes might differ from each other by as much as 20 million bases due to CNVs alone, which would thus account for approximately 5 times more variation than all SNPs together (Kehrer-Sawatzki, 2007). CNVs were described in almost 3000 genes (Kehrer-Sawatzki, 2007), and while some of them do not appear to be associated with disease, others were linked to conditions that include cancer, autism, schizophrenia, epilepsy, and amyotrophic lateral sclerosis (Gonzalez et al.,