# Genetic Variations and Precision Medicine

*by Amal Adel Alzu'bi, PhD; Leming Zhou, PhD, DSc; and Valerie J. M. Watzlaf, MPH, PhD, RHIA, FAHIMA*

## Abstract

The time and costs associated with the sequencing of a human genome have decreased significantly in recent years. Many people have chosen to have their genomes sequenced to receive genomics-based personalized healthcare services. To reach the goal of genomics-based precision medicine, health information management (HIM) professionals need to manage and analyze patients' genomic data. Two important pieces of information from the genome sequence are the risk of genetic diseases and the specific medication or pharmacogenomic results for the individual patient, both of which are linked to a patient's genetic variations. In this review article, we introduce genetic variations, including their data types, relevant databases, and some currently available analysis methods and systems. HIM professionals can choose to use these databases, methods, and systems in the management and analysis of patients' genomic data.

**Keywords**: genetic variation; precision medicine; database

## Introduction

In 2015, the Precision Medicine Initiative (PMI) was launched in order to understand how a person's genetics, environment, and lifestyle can help determine the best approach to prevent or treat disease.[1] In the past, medical practices followed certain protocols and prescribed medications without considering patients' genetic makeup, environment, or lifestyle, even though healthcare providers have a general understanding of the importance of these factors. One major reason was that accurate data about these factors was not easily accessible. Ten years ago, sequencing a human genome was still very expensive, and no reliable and inexpensive methods of collecting environmental and lifestyle data were available. Even electronic health record (EHR) systems were not widely adopted then. Today, all these factors are easily accessible and reliable because of the availability of high-throughput DNA sequencing technologies, wide adoption of mobile health apps and various types of wearable trackers, and the extensive use of EHR systems in almost all hospitals and clinics in the United States. Therefore, it has become feasible to collect a large amount of health-related data from patients and apply the findings in precision medicine.

Precision medicine will make it possible for physicians to choose the best treatment and prevention strategy for each patient according to their personal health information, including their genetic makeup.[2] Some human genetic variations are closely related to certain diseases or individual patient responses to certain medications;[3] therefore, physicians can choose specific treatment options to optimize the patient's outcome. For example, in precision medicine, physicians can choose different medications to help their patients quit smoking by examining the patient's speed of nicotine metabolization.[4]

In the PMI, now renamed All of Us, more than one million Americans will be recruited. Their genomic, EHR, environment, diet, and lifestyle information will be collected and integrated together to

investigate the risk for a range of diseases. Individual responses to commonly used drugs, biomarkers associated with common diseases, and targeted therapies will also be investigated. The recruitment of study participants is ongoing. As stated in the executive summary of the PMI Working Group report, "Precision medicine seeks to redefine our understanding of disease onset and progression, treatment response, and health outcomes through the more precise measurement of molecular, environmental, and behavioral factors that contribute to health and disease. This understanding will lead to more accurate diagnoses, more rational disease prevention strategies, better treatment selection, and the development of novel therapies."[5]

To successfully reach the goals described in the PMI report, extensive collaboration is needed among researchers and professionals from medicine, public health, genetics/genomics, computer science, and health information management (HIM). These professionals need to work together to recruit study participants, develop methods and tools to collect desired data, analyze and interpret the data, and report the results to the public.

In the desired data items from the PMI, personal genomic information is relatively new for HIM researchers and professionals. In this article we introduce one important type of personal genomic data: genetic variations, including their major types, databases, and some currently available variation analysis methods and systems.

## Types of Genetic Variations

A genome is an organism's complete set of DNA sequences. Although people in this world may look different, with different skin color, eye color, hair color, and height, all human genomes are highly similar. More than 99 percent of human genomes from two unrelated individuals are identical.[6–8] The differences among these genomes are *genetic variations*. Most of these genetic variations either do not show noticeable differences or simply make us look different. Some genetic variations can cause diseases.

Several types of genetic variations exist. This article focuses on the following three types of genetic variations:

1. Single nucleotide polymorphisms (SNPs)
2. Short insertions and deletions (INDELs)
3. Copy number variations (CNVs)

## Single Nucleotide Polymorphisms

A DNA polymorphism is any difference in the nucleotide sequence between individuals. SNPs are the most common type of genetic variation. They represent a difference in a single nucleotide.[9] The difference can be one extra nucleotide (insertion) at the specific location, one missing nucleotide (deletion), or a different nucleotide (substitution), compared with the reference genome sequenced in the Human Genome Project. SNPs represent roughly 90 percent of the human DNA polymorphisms.[10] Figure 1 illustrates the concept of SNPs, in which six individuals have either T or C at one specific position in their chromosome 2.

SNPs can occur anywhere in a genome. When SNPs occur within a gene or within the regulatory region (which may regulate the expression of a gene or genetic component) near or inside a gene, they may directly affect the gene function or expression and may increase the risk of developing certain diseases,[11, 12] especially if the gene is critically important for the normal function of cells. SNPs that occur in intergenic regions (or regions between two genes) are less likely to lead to gene function or expression changes or disease.

Research in pharmacogenomics (the study of how genes affect a person's response to medications) indicates that SNPs may be helpful for predicting an individual's response to certain medications, which is critically important for precision medicine.[13] For example, genetic polymorphisms in the *CYP450* genes

(CYP450 enzymes are responsible for metabolizing drugs[14, 15]) affect the metabolism of about 25 percent of all medication treatments.[16, 17] These genetic polymorphisms can influence medication response, causing it to be normal, enhanced, reduced, or neutralized.[18] Additionally, genetic variations in three genes (*CYP2C9*, *VKORC1*, and *CYP4F2*) can result in significant variability in the appropriate dose of warfarin between patients.[19]

## Short Insertions and Deletions

INDELs are insertions or deletions of nucleotides in a DNA sequence. One single nucleotide insertion or deletion is a special case of INDELs. The insertions and deletions can be a long segment (a few hundred or a few thousand nucleotides) of a chromosome.[20] Both insertion and deletion of nucleotides will change the number of nucleotides in a DNA sequence. If either of them occur in a gene, then the protein made by the gene may not function properly. For example, the X-linked recessive disorder previously seen in European royal families[21] results from the insertion of a large segment of DNA leading to inactivation of a protein required for blood clotting. Figure 2 illustrates the concept of INDELs.

In the human genome, approximately 36 percent of short INDELs are in the genetic region.[22] Therefore, some of these INDELs can have an impact on human gene function/expression. Examples of diseases related to short insertions include Huntington disease (CAG repeated more than 35 times in gene *HTT*) and myotonic dystrophy (a segment of DNA repeated too many times in gene *DMPK* or *CNBP*).[23] An example of a disease associated with a short deletion is cystic fibrosis (a deletion of three nucleotides at the 508th position on the *CFTR* gene).[24]

## Copy Number Variations

CNV refers to an intermediate-scale genetic change (in the range from 1,000 nucleotides to several million nucleotides).[25, 26] CNVs include additional copies of a DNA sequence segment (duplications) or losses of a DNA segment (deletions).[27] For example, A-B-C-D is a long DNA sequence in one chromosome; here, A, B, C, and D are DNA segments longer than 1,000 nucleotides. This chromosome in some people can instead have the following sequence: A-B-**B**-C-D (a duplication of B) or A-C-D (a deletion of B). The scale of this type of genetic change may affect multiple genes in certain cases.[28]

There is an increasing concern about the effect of CNVs in the development of complex diseases[29] because a number of CNVs overlap with protein coding regions.[30] CNVs have been detected in genetic regions associated with complex neurological diseases,[31] such as autism,[32] Alzheimer's disease,[33] and schizophrenia.[34]

## Genetic Variation Databases

Many databases have been created to manage the growing number of genomic data sets. These databases help researchers and healthcare professionals obtain genetic information quickly and conveniently.[35] A number of databases contain genetic variations and related information. In this article, we briefly introduce the following two categories of databases:

1. Variation databases that provide information about genetic variations alone.
2. Variations and disease/phenotype databases that provide information about phenotypes and genetic variations. (*Phenotype* refers to the observable characteristics of an organism, such as its morphology, development, physiological properties, and behavior.)

## Variation Databases

The National Center for Biotechnology Information at the National Library of Medicine maintains dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), a database of short genetic variations. This database stores all types of genetic variations of less than 50 base pairs in multiple species.[36] Therefore, it includes different types of short genetic variations, such as SNPs, INDELs, and microsatellite markers.

In dbSNP, HIM professionals can find detailed information about each genetic variation, such as the type of the variation, the location of the variation in a chromosome, the relevant genes, and population-specific allele frequencies and genotypes, which are all critically important in precision medicine.[37] For instance, according to one recent journal article,[38] a SNP with ID rs2981579 is associated with breast cancer. A simple query in dbSNP with this SNP ID can tell us that this genetic variation is in the *FGFR2* gene and the specific location of this genetic variation is in chromosome 10, position 121577821; and it occurs in multiple populations, such as European, Asian, and African populations. Therefore, in healthcare practice, HIM professionals can specifically check whether their patients have this genetic variation in this gene (*FGFR2*) and provide the results to physicians.

The dbVar database (http://www.ncbi.nlm.nih.gov/dbvar) archives large-scale genetic variations for multiple species and is based on various genomic studies. A number of large-scale variations are associated with diseases. Data in dbVar are organized according to the study or publication. Genetic variation types in dbVar include INDELs and CNVs.[39] HIM professionals can perform searches in dbVar to determine the type of genetic variation, the location of the variation, the relevant gene, and the clinical relevance of the variation (e.g., pathogenic, likely pathogenic, uncertain significance).

The Database of Genomic Variations (DGV; http://dgv.tcag.ca/dgv/app/home) provides a catalog of human structural variations including INDELs and CNVs. The database is periodically updated according to results reported in peer-reviewed research articles. DGV aims to catalog the highest-quality structural variations in the literature, and it presents the results in a format that is convenient for healthcare professionals and researchers.[40] HIM professionals can search the DGV database using keywords, gene symbols, genetic regions, and the specific location on a chromosome. This database can be helpful for HIM professionals to use when collecting data for specific disease registries, such as cancer, stroke, and diabetes registries.

## Variations and Disease/Phenotype Databases

Online Mendelian Inheritance in Man (OMIM; http://www.ncbi.nlm.nih.gov/omim) is a comprehensive database that catalogs known diseases and their associated genetic variations. Each OMIM record has a summary of one disease and its relevant genes and variations in human genome as reported in the literature.[41] It is particularly useful when HIM professionals want to find all genetic variations associated with one specific disease, or genetic diseases associated with one genetic variation or gene.

The Human Gene Mutation Database (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php) is the gold-standard resource for comprehensive data on published human inherited disease mutations.[42] Each HGMD record includes a reference to the first literature report of a mutation, the associated disease specified in that report, and the gene name, symbol, and chromosomal location. HGMD can be useful to HIM professionals looking for databases to add to a healthcare system data repository. It can also provide useful information on clinical trials, evidence-based best practices, and quality improvement studies.

ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) is a database that provides information about the medical relevance of genetic variations.[43] It archives the relationship between medically significant variations and their phenotypes. ClinVar is strongly related to dbSNP and dbVar because it maintains information about the location of variations in human genome assemblies. Unlike dbSNP and dbVar, Clinvar accepts direct submissions of structured details of phenotypes, and it provides interpretation of the functional and clinical significance of genetic variations. ClinVar can be used by HIM professionals when overseeing the clinical registry function within their healthcare organization.

SNPedia (https://www.snpedia.com/) is a wiki-based database.[44] Researchers convert information presented in large-scale peer-reviewed genomic studies into a machine-readable format and then store the information in SNPedia so that the information is easily accessible to researchers. SNPedia supports personal genome annotation, interpretation, and analysis. This database links the genetic variations to information about diseases or phenotypic traits published in genomic studies. HIM professionals can refer

researchers to the important information provided in this database or use it themselves when compiling annual reports and other business reports that examine linkages of disease.

The use of these databases can be outlined as follows: Short variations can be found in dbSNP. Structural genetic variations can be found in dbVar. Structural variations from healthy people can be searched in DGV. The associations between human SNPs and disease/phenotype can be obtained from SNPedia. OMIM can be used to find the association between human genetic variations and diseases, including an extensive description of relevant genes and phenotypes. HGMD can be used to learn more about the association between human variations and genetic diseases. Table 1 summarizes these genetic variation databases.

## Genetic Variation Analysis

After a genome is sequenced, one critically important step is to analyze the sequence data. Specific to genetic variations, there are four types of analysis:

1. Variation identification,
2. Variation annotation,
3. Variation visualization, and
4. Variation and disease association.

These tasks are typically done by genetics/genomics researchers instead of HIM professionals; therefore, we will only briefly describe them. The more clinically relevant variation analysis task is genetic variation filtering and identification of associated diseases via existing tools, which is described at the end of this section.

### Variation Identification

Variation identification is the process of identifying genetic variations from DNA sequence data.[45] Usually variation identification can be done by comparing a newly sequenced human genome with the reference human genome and identifying candidate sites at which one or more samples differ from the reference sequence.[46] Several tools for identifying variations, such as CRISP, GATK, and SAMtools, are available.[47, 48]

### Variation Annotation

Variation annotation refers to the classification and prediction of the functional impact of variations, followed by filtering and prioritization of the ones that cause diseases.[49] Several tools are available for variation annotations; these tools include ANNOVAR,[50] VEP,[51] and SVA.[52]

### Variation Visualization

Variation visualization refers to the validation and visual representation of genetic variations.[53] IGV[54] and UCSC Genome Browser[55] (https://genome.ucsc.edu) are commonly used variation visualization tools at various levels of detail, from base pair level to chromosome level.

### Variation and Disease Association

The study of genetic diseases has a long history.[56] However, only after the recent improvements in DNA sequencing,[57] researchers can now study genetic variations in the whole genome and their association with diseases. For instance, in recent years, genome-wide association studies (GWAS) have been widely used for the discovery of genetic risk factors associated with human diseases.

## Genetic Variation Analysis Tools for Clinical Purposes

In this section, we present several tools that help HIM professionals manage and analyze genetic variations conveniently. After all, once a patient's genomic data is integrated into the patient's EHR, HIM

professionals will be the key personnel to use existing software tools to analyze and manage the data and make the analysis results available to physicians for precision medicine practice.

An integrated system of patient genomic information management and analysis for healthcare professionals was created to enable genetic variation analysis and management in the healthcare environment.[58] In this Java-based system, HIM professionals can upload the patient's genetic variation data in VCF format (VCF–Variant Call Format, a widely used text file format for describing genetic variations) and obtain reports about the risks of genetic disease and the patient's response to certain medications.

VCF-Explorer is a software tool that can be used to perform analysis for large genetic variation files in the VCF format.[59] The software program can run on various computational platforms, from laptops to high-performance computers. HIM professionals can use the graphical user interface (GUI) to filter variations based on specific samples or variant-level annotations.

VCF.Filter is a user-friendly and standalone tool that allows users to search for genetic variations that cause rare diseases.[60] The software provides a GUI that helps users design filtering rules and apply them to any VCF file.

The myVCF desktop application allows users to explore, query, visualize, and export genetic variation data.[61] The application provides a browser interface that is compatible with Windows, MacOS, and UNIX systems. It can manage multiple VCF files, and it allows users to browse the information contained in the VCF file. The application provides a flexible search engine that allows users to perform queries about region/variant coordinates, gene symbols, or dbSNP IDs.

VariantDB is a web-based annotation and filtering platform that automatically annotates variations with allele frequencies, functional impact, pathogenicity predictions, and pathway information.[62]

After patients have their genome sequenced, HIM professionals can use variation identification programs to determine genetic variations in patients. They can use the software tools mentioned in this article to filter genetic variations associated with certain diseases, and they can visualize these genetic variations, perform searches in variation-disease databases to learn details about these genetic variations, and explain the results to physicians. Physicians can then design a personalized treatment plan for each patient. In this entire process, HIM professionals do not need to know the detailed implementations of these software programs and databases, but they do need to have basic knowledge of them and know how to search in these databases and use these tools. The obtained results (variations and reports from variation analyses) can be stored in the EHR so that they can be used to support the physician's treatment decisions, clinical coding, and patient outcomes. In other words, working on genomic data is a natural extension of the HIM professional's typical job functions, such as managing data, performing data analysis using software programs, and protecting patients' data security and privacy.[63]

## Conclusion

This article provides a brief introduction to genetic variations and some existing databases and data analysis software programs. Because HIM professionals have the skills for managing and processing large-scale clinical data, they may need only a minimum amount of training in genomics (such as basic concepts, frequently used databases, and data analysis programs) to be able to understand genomic data and extract desired information for clinical practice. In this way, HIM professionals can play a key role in the implementation of precision medicine nationwide.

Amal Adel Alzu'bi, PhD, is an assistant professor in the Department of Computer Information Systems at Jordan University of Science and Technology in Irbid, Jordan.

Leming Zhou, PhD, DSc, is an associate professor in the Department of Health Information Management at the University of Pittsburgh in Pittsburgh, PA.

Valerie J. M. Watzlaf, MPH, PhD, RHIA, FAHIMA, is an associate professor in the Department of Health Information Management at the University of Pittsburgh in Pittsburgh, PA.

# Notes

1.  Collins, F. S., and H. Varmus. "A New Initiative on Precision Medicine." *New England Journal of Medicine* 372, no. 9 (2015): 793–95.
2.  Wynn, R., K. Adams, R. Kowalski, et al. "The Patient in Precision Medicine: A Systematic Review Examining Evaluations of Patient-Facing Materials." *Journal of Healthcare Engineering* 2018 (2018): 9541621.
3.  Juengst, E., and M. McGowan. "Why Does the Shift from 'Personalized Medicine' to 'Precision Health' and 'Wellness Genomics' Matter?" *AMA Journal of Ethics* 20, no. 9 (2018): E881–E890.
4.  Lerman, C., R. A. Schnoll, and L. W. Hawk. "Use of the Nicotine Metabolite Ratio as a Genetically Informed Biomarker of Response to Nicotine Patch or Varenicline for Smoking Cessation: A Randomised, Double-Blind Placebo-controlled Trial." *The Lancet Respiratory Medicine* 3, no. 2 (2015): 131–38.
5.  Precision Medicine Initiative Working Group. *The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine*. 2015, p. 1. Available at https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf.
6.  Shafer, A. "Understanding Genetics." *The Tech*, Stanford University, 2006.
7.  Lander, E. S., L. M. Linton, B. Birren, et al. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409, no. 6822 (2001): 860–921.
8.  Venter, J. C., M. D. Adams, E. W. Myers, et al. "The Sequence of the Human Genome." *Science* 291, no. 5507 (2001): 1304–51.
9.  Altshuler, D., V. J. Pollara, C. R. Cowles, et al. "An SNP Map of the Human Genome Generated by Reduced Representation Shotgun Sequencing." *Nature* 407, no. 6803 (2000): 513–16.
10. Varela, M. A., and W. Amos. "Heterogeneous Distribution of SNPs in the Human Genome: Microsatellites as Predictors of Nucleotide Diversity and Divergence." *Genomics* 95, no. 3 (2010): 151–59.
11. Xiong, Q., N. Ancona, E. R. Hauser, et al. "Integrating Genetic and Gene Expression Evidence into Genome-wide Association Analysis of Gene Sets." *Genome Research* 22, no. 2 (2012): 386–97.
12. Talseth-Palmer, B. A., and R. J. Scott. "Genetic Variation and Its Role in Malignancy." *International Journal of Biomedical Science* 7, no. 3 (2011): 158–71.
13. Carlson, B. "SNPs—A Shortcut to Personalized Medicine." *Genetic Engineering & Biotechnology News* 28, no. 13 (2008).
14. Leon-Cachon, R. B., J. A. Ascacio-Martinez, and H. A. Barrera-Saldana. "Individual Response to Drug Therapy: Bases and Study Approaches." *Revista de Investigación Clínica* 64, no. 4 (2012): 364–76.
15. Ventola, C. L. "Role of Pharmacogenomic Biomarkers in Predicting and Improving Drug Response: Part 1: The Clinical Significance of Pharmacogenetic Variants." *Pharmacy and Therapeutics* 38, no. 9 (2013): 545–60.
16. Ibid.
17. Ma, J. D., K. C. Lee, and G. M. Kuo. "Clinical Application of Pharmacogenomics." *Journal of Pharmacy Practice* 25, no. 4 (2012): 417–27.

18. Leon-Cachon, R. B., J. A. Ascacio-Martinez, and H. A. Barrera-Saldana. "Individual Response to Drug Therapy: Bases and Study Approaches." *Revista de Investigación Clínica* 64, no. 4 (2012): 364–76.

19. Salari, K., H. Watkins, and E. A. Ashley. "Personalized Medicine: Hope or Hype?" *European Heart Journal* 33, no. 13 (2012): 1564–70.

20. Clancy, S. "DNA Deletion and Duplication and the Associated Genetic Disorders." *Nature Education* 1, no. 1 (2008): 23.

21. Rogaev, E. I., A. P. Grigorenko, G. Faskhutdinova, et al. "Genotype Analysis Identifies the Cause of the 'Royal Disease.'" *Science* 326, no. 5954 (2009): 817.

22. Mills, R. E., C. T. Luttig, C. E. Larkins, et al. "An Initial Map of Insertion and Deletion (INDEL) Variation in the Human Genome." *Genome Research* 16, no. 9 (2006): 1182–90.

23. Bergeron, B. *Case Studies in Genes and Disease: A Primer for Clinicians*. Philadelphia, PA: American College of Physicians, 2004.

24. Massie, J., and M. Delatycki. "Cystic Fibrosis Carrier Screening." *Paediatric Respiratory Reviews* 14, no. 4 (2013): 270–75.

25. Stankiewicz, P. and J. R. Lupski. "Structural Variation in the Human Genome and Its Role in Disease." *Annual Review of Medicine* 61 (2010): 437–55.

26. Pollex, R. L., and R. A. Hegele. "Copy Number Variation in the Human Genome and Its Implications for Cardiovascular Disease." *Circulation* 115, no. 24 (2007): 3130–38.

27. Eichler, E. E. "Copy Number Variation and Human Disease." *Nature Education* 1, no. 3 (2008): 1.

28. Check, E. "Human Genome: Patchwork People." *Nature* 437, no. 7062 (2005): 1084–86.

29. Beckmann, J. S., X. Estivill, and S. E. Antonarakis. "Copy Number Variants and Genetic Traits: Closer to the Resolution of Phenotypic to Genotypic Variability." *Nature Reviews Genetics* 8, no. 8 (2007): 639–46.

30. Sebat, J., B. Lakshmi, J. Troge, et al. "Large-Scale Copy Number Polymorphism in the Human Genome." *Science* 305, no. 5683 (2004): 525–28.

31. Girirajan, S., C. D. Campbell, and E. E. Eichler. "Human Copy Number Variation and Complex Genetic Disease." *Annual Review of Genetics* 45 (2011): 203–26.

32. Gai, X., H. M. Xie, J. C. Perin, et al. "Rare Structural Variation of Synapse and Neurotransmission Genes in Autism." *Molecular Psychiatry* 17, no. 4 (2012): 402–11.

33. Zheng, X., F. Y. Demirci, M. M. Barmada, et al. "Genome-wide Copy-Number Variation Study of Psychosis in Alzheimer's Disease." *Translational Psychiatry* 5 (2015): e574.

34. Singh, S. M., C. A. Castellani, and R. L. O'Reilly. "Copy Number Variation Showers in Schizophrenia: An Emerging Hypothesis." *Molecular Psychiatry* 14, no. 4 (2009): 356–58.

35. Lathe, W., III, J. Williams, M. Mangan, and D. Karolchik. "Genomic Data Resources: Challenges and Promises." *Nature Education* 1, no. 3 (2008): 2.

36. Wheeler, D. L., T. Barrett, D. A. Benson, et al. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 35 (2007): D5–D12.

37. NCBI Resource Coordinators. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 41 (2013): D8–D20.

38. Michailidou, K., P. Hall, A. Gonzalez-Neira, et al. "Large-Scale Genotyping Identifies 41 New Loci Associated with Breast Cancer Risk." *Nature Genetics* 45, no. 4 (2013): 353–61, 361e1–361e2.

39. Lappalainen, I., J. Lopez, L. Skipper, et al. "DbVar and DGVa: Public Archives for Genomic Structural Variation." *Nucleic Acids Research* 41 (2013): D936–D941.

40. MacDonald, J. R., R. Ziman, R. K. Yuen, et al. "The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome." *Nucleic Acids Research* 42 (2014): D986–D992.

41. Hamosh, A., A. F. Scott, J. S. Amberger, et al. "Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders." *Nucleic Acids Research* 33 (2005): D514–D517.

42. Stenson, P. D., M. Mort, E. V. Ball, et al. "The Human Gene Mutation Database: Building a Comprehensive Mutation Repository for Clinical and Molecular Genetics, Diagnostic Testing and Personalized Genomic Medicine." *Human Genetics* 133, no. 1 (2014): 1–9.

43. Landrum, M. J., J. M. Lee, M. Benson, et al. "ClinVar: Improving Access to Variant Interpretations and Supporting Evidence." *Nucleic Acids Research* 46 (2018): D1062–D1067.

44. Cariaso, M., and G. Lennon. "SNPedia: A Wiki Supporting Personal Genome Annotation, Interpretation and Analysis." *Nucleic Acids Research* 40 (2012): D1308–D1312.

45. Stitziel, N. O., A. Kiezun, and S. Sunyaev. "Computational and Statistical Approaches to Analyzing Variants Identified by Exome Sequencing." *Genome Biology* 12, no. 9 (2011): 227.

46. 1000 Genomes Project Consortium. "A Map of Human Genome Variation from Population-Scale Sequencing." *Nature* 467, no. 7319 (2010): 1061–73.

47. Quintánsa, B., A. Ordóñez-Ugaldea, P. Cacheiroa, et al. "Medical Genomics: The Intricate Path from Genetic Variant Identification to Clinical Interpretation." *Applied & Translational Genomics* 3, no. 3 (2014): 60–67.

48. Altmann, A., P. Weber, C. Quast, et al. "vipR: Variant Identification in Pooled DNA Using R." *Bioinformatics* 27, no. 13 (2011): i77–i84.

49. McCarthy, D. J., P. Humburg, A. Kanapin, et al. "Choice of Transcripts and Software Has a Large Effect on Variant Annotation." *Genome Medicine* 6, no. 3 (2014): 26.

50. Pabinger, S., A. Dander, M. Fischer, et al. "A Survey of Tools for Variant Analysis of Next-Generation Genome Sequencing Data." *Briefings in Bioinformatics* 15, no. 2 (2014): 256–78.

51. McLaren, W., B. Pritchard, D. Rios, et al. "Deriving the Consequences of Genomic Variants with the Ensembl API and SNP Effect Predictor." *Bioinformatics* 26, no. 16 (2010): 2069–70.

52. Ge, D., E. K. Ruzzo, K. V. Shianna, et al. "SVA: Software for Annotating and Visualizing Sequenced Human Genomes." *Bioinformatics* 27, no. 14 (2011): 1998–2000.

53. Nielsen, C. B., M. Cantor, I. Dubchak, et al. "Visualizing Genomes: Techniques and Challenges." *Nature Methods* 7, no. 3S (2010): S5–S15.

54. Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. "Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration." *Briefings in Bioinformatics* 14, no. 2 (2013): 178–92.

55. Raney, B. J., T. R. Dreszer, G. P. Barber, et al. "Track Data Hubs Enable Visualization of User-defined Genome-wide Annotations on the UCSC Genome Browser." *Bioinformatics* 30, no. 7 (2014): 1003–5.

56. Sturtevant, A. "The Linear Arrangement of Six Sex-linked Factors in Drosophila, as Shown by Their Mode of Association." *Journal of Experimental Biology* 14 (1913): 43–59.

57. Bras, J., R. Guerreiro, and J. Hardy. "Use of Next-Generation Sequencing and Other Whole-Genome Strategies to Dissect Neurological Disease." *Nature Reviews Neuroscience* 13, no. 7 (2012): 453–64.

58. Alzu'bi, A., and L. Zhou. "An Integrated Patient Genomic Information Management and Analysis System for Healthcare Professionals." In *2017 IEEE International Conference on Healthcare Informatics*. Los Alamitos, CA: IEEE Computer Society, 2017, 107–13.

59. Akgün, M., and H. Demirci. "VCF-Explorer: Filtering and Analysing Whole Genome VCF Files." *Bioinformatics* 33, no. 21 (2017): 3468–70.

60. Müller, H., R. Jimenez-Heredia, A. Krolo, et al. "VCF.Filter: Interactive Prioritization of Disease-linked Genetic Variants from Sequencing Data." *Nucleic Acids Research* 45, no. W1 (2017): W567–W572.

61. Pietrelli, A., and L. Valenti. "myVCF: A Desktop Application for High-Throughput Mutations Data Management." *Bioinformatics* 33, no. 22 (2017): 3676–78.

62. Vandeweyer, G., L. Laer, B. Loeys, et al. "VariantDB: A Flexible Annotation and Filtering Portal for Next Generation Sequencing Data." *Genome Medicine* 6, no. 10 (2014): 74.

63. Alzu'bi, A., L. Zhou, and V. Watzlaf. "Personal Genomic Information Management and Personalized Medicine: Challenges, Current Solutions, and Roles of HIM Professionals." *Perspectives in Health Information Management* (Spring 2014).

**Table 1**

Summary of Genetic Variation Databases

| Name | Purpose | Species |
|------|---------|---------|
| dbSNP | Stores all types of genetic variations less than 50 base pairs | All species |
| dbVar | Archives large-scale genetic variations | All species |
| DGV | Provides a catalog of structural variations, including insertion, deletion, and copy number variations | Human |
| OMIM | Catalogs known diseases with their genetic component | Human |
| HGMD | Includes a reference to the first literature report of a mutation, the associated disease state, and the gene name, symbol, and chromosomal location | Human |
| ClinVar | Archives the medically significant variations and phenotypes | Human |
| SNPedia | Converts the information in large-scale peer-reviewed genomic studies into machine-readable format | Human |

**Figure 1**

Illustration of Single Nucleotide Polymorphisms (SNPs)
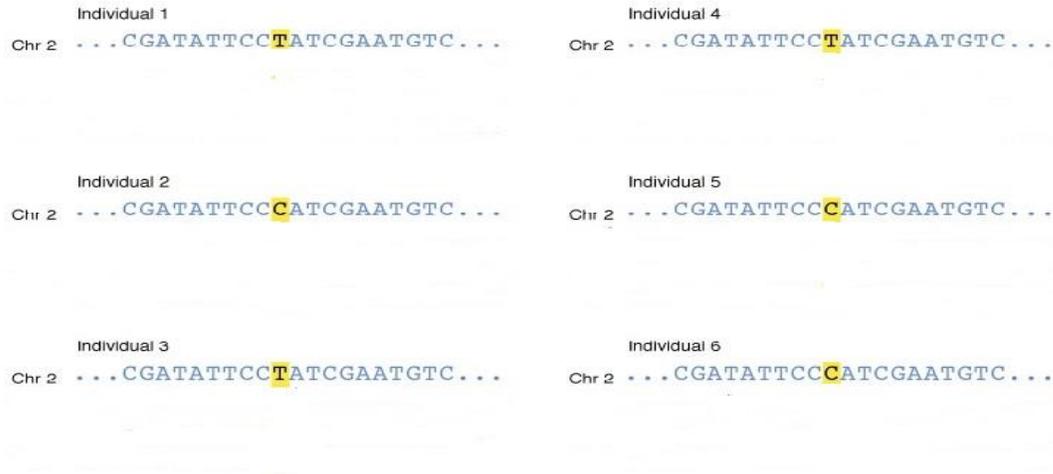


*Source:* National Human Genome Research Institute. Available at
http://www.genome.gov/glossary/index.cfm?p=viewimage&id=185.

**Figure 2**

Illustration of Single-Base Insertions and Deletions (INDELs)

Sequence: A C T G - A A A

Insertion: A C T G G A A A

**Insertion of G**

Sequence: A C T G C T C A A A

Deletion : A C T G C - C A A A

**Deletion of T**