Supplemental Table 1. Glossary of Terms and Concepts in Genetics, Genomic and Genomic Technology

Term	Definition and Discussion
Allele	One of two or more versions of a gene. An individual inherits two alleles for each gene, one from each parent. If the two alleles are the same, the individual is homozygous for that gene. If the alleles are different, the individual is heterozygous. The term *allele* also refers to variation among noncoding DNA sequences.
Base	Also known as nucleobase or nitrogenous base. A nitrogen-containing biological compound found within DNA, RNA, nucleotides, and nucleosides. The primary bases are adenine (A), cytosine (C), guanine ([G] DNA and RNA), thymine ([T] DNA), and uracil ([U] RNA). Base pair (bp) refers to the unit of length of a double-stranded nucleic acid and is composed of two purine (A or G) and pyrimidine (C, T or U) bases on opposite strands of a double-stranded nucleic acid hydrogen bonded to each other.
Biomarker	A biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, a condition or disease. Also called *molecular marker* and

	signature molecule. Biomarkers include, but are not limited to, such phenotypic parameters as specific enzyme or hormone concentration, specific gene phenotype, presence or absence of biological substances. The term *genetic biomarker* implies a detectable and measurable DNA or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other intervention.
Candidate gene association (CGA)	A research strategy used to identify disease-associated genes based on findings of candidate genes in a chromosome region in which a disorder has been mapped.
Capillary sequencer	One of the sequencers used to determine the DNA sequence. DNA is run through an array of 96 gel-filled capillaries—glass tubes about the width of a human hair—rather than through a slab of gel.
Chromatin immunoprecipitation sequencing (ChIP-seq)	A research technique that identifies the physical interactions between DNA and transcription factors as well as the identity of the transcription factors
Cloud computing	A massive distributed computing model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned

	and released with minimal management effort or service-provider interaction (U.S. National Institute of Standards and Technology [NIST], 2011)
Complex disease	A disease caused by the interaction of multiple genes and environmental factors. Also called *multifactorial disease*. Complex diseases can have a variety of different causes and modes of inheritance in different people.
Copy-number variation (CNV)	Variation among individuals in the number of copies of a particular DNA sequence in their genomes.
DNA methylation	Conversion of cytosine in DNA to 5-methyl cytosine (mC), which helps regulate gene expression. DNA methylation, which regulates gene expression, is the main (or the best understood) mechanism of epigenetic memory in the cell. DNA methylation allows particular patterns of gene expression to be stably transmitted to daughter cells. DNA methylation patterns are important, as they influence transcriptionally active regions.
DNA, deoxyribonucleic acid	The chemical inside the nucleus of a cell that carries genetic instructions for making living organisms.

Dominant trait	The trait that is expressed in an individual who is heterozygous for a particular gene.
Endophenotype (intermediate phenotype)	A phenotype that marks one facet of the changes that occur during development of a complex disease.
Enzyme-linked immunosorbent assay (ELISA)	A laboratory technique (immunoassay) that uses antibodies linked to enzymes to detect and measure the amount of a substance in a solution, such as serum. The test relies on use of a solid surface to which the antibodies and other molecules stick. In the final step, an enzyme reaction takes place that causes a color change that can be read using a special machine. ELISAs may be used to help diagnose certain diseases.
Epigenetic	Heritable from mother cell to daughter cell (sometimes from parent to child) but not produced by a change in DNA sequence.
Epigenetics	The study of how age and exposure to environmental factors, such as diet, exercise, drugs, and chemicals, may cause changes in the way genes are switched on and off without changing the actual DNA sequence.
Gene	A functional and physical unit of DNA and heredity passed from parent to offspring. Most genes (protein-coding genes) contain the information for making a specific

	protein.
Gene expression	The process by which a gene gets turned on in a cell to make RNA and proteins. Gene expression may be measured by looking at the RNA, the protein made from the RNA, or the function of the protein in a cell.
Gene silencing	Interruption or suppression of the expression of a gene that may cause disease or be defective in another way, without modification of the gene, at the level of transcription or translation. A gene that would be expressed ("turned on") under normal circumstances is switched off by machinery in the cell. Gene silencing occurs when RNA is unable to make a protein during translation.
Genetic locus	(pl. genetic loci) The specific physical location of a gene or other DNA sequence on a chromosome, like a genetic street address.
Genetics	The study of genes and heredity. Heredity is the passing of genetic information and traits (such as eye color and an increased chance of getting a certain disease) from parents to offspring.
Genomics	The study of the complete genetic material, including genes and their functions, of a

	person as well as interactions of those genes with each other and the environment.
Genotype	The specific genetic contribution to the phenotype as being of a certain type or as corresponding to a certain character. Some phenotypes (traits) are largely determined by the genotype, while others are more influenced by environmental factors. A phenotype can be directly related to a genotype, but not all are, and there is usually not a one-to-one correlation between a genotype and a phenotype. Phenotypes almost always have environmental influences, such as diet and other lifestyle choices.
Heritability	The proportion of the causation of a characteristic that is attributable to genetics.
High throughput	A type of analysis that increases the range, complexity, sensitivity, and accuracy of results by greatly increasing the scale of operations using fast, automated techniques for screening large numbers of substances, including chemicals and genes.
Histones	A family of small basic proteins that form complexes with DNA to create nucleosomes.
Histone (posttranslational) modification	A change that occurs after the translation that produces the histone protein. A histone modification consists of the addition or subtraction of any one of several chemical groups to an individual amino acid of a histone—especially a histone

	belonging to a nucleosome.
Histone acetylation	Covalent linkage of acetyl groups to histone proteins associated with DNA, particularly in active chromatin. Histone acetylation may be involved in regulation of gene expression or modification of chromatin structure.
Hybridization	The process of combining two complementary single-stranded DNA or RNA molecules and allowing them to form a single double-stranded molecule through base pairing. Hybridization is a part of many important laboratory techniques such as polymerase chain reaction and Southern blotting.
Inbred strains	Animals that are nearly identical to each other in genotype due to long inbreeding. Inbred strains of animals are frequently used for experiments where, for reproducibility of conclusions, all the test animals should be as similar as possible. In human genetics, the coefficient of inbreeding is the proportion of a person's genes that are identical by descent.
Indel	Abbreviation for *insertion and deletion*, which refers to types of mutation involving

	either addition or loss of genetic material. An insertion-or-deletion mutation can be small, involving a single extra or missing DNA base pair, or large, involving a piece of a chromosome.
Mass spectrometry	An analytical technique wherein ions are separated according to their ratio of charge to mass. The atomic weight of the particle can be deduced from the mass spectrum produced.
Mendelian disease	A genetic disease caused in whole or in part by Mendelian inheritance, where genes and traits are passed from parents to children, following Mendel's first law of independent segregation of the alleles at the same locus conveyed by each parent. Medelian inheritance includes autosomal dominant, autosomal recessive, and sex-linked genes.
Messenger RNA (mRNA)	A processed gene transcript that carries protein-coding information to the ribosomes.
MicroRNA (miRNA)	Short (21−22 nt) RNA molecules encoded within normal genomes that have a role in regulation of gene expression and maybe also of chromatin structure.
Multiple testing correction	Or *multiple comparisons.* Statistical adjustment used when considering a set of statistical inferences simultaneously for preventing errors in inference (Benjamini,

	2010). The expected errors when testing many statistical inferences simultaneously include confidence intervals that fail to include their corresponding population parameters or hypothesis tests that incorrectly reject the null hypothesis. One such adjustment is the *false discovery rate* (FDR), which is a statistical method used for multiple testing corrections to control the expected proportion of false positives (Type I error) among rejected tests or hypotheses. FDR is less stringent than other multiple testing corrections and, thereby, will allow a percentage of false positives.
Next-generation sequencer	An automated machine that performs massively parallel sequencing to facilitate the rapid determination of the DNA sequence of large numbers of strands or segments of DNA.
Noncoding RNA (ncRNA)	RNA that does not contain genetic code for a protein.
-ome (suffix)	A totality of some sort.
Epigenome	Description of all the epigenetic modifications across the whole genome. Unlike the genome (DNA sequence), which is consistent across cells, different cells within an organism have different epigenomes, which may change with time in

	response to environmental cues.
Genome	The complete set of genetic material of an organelle, cell or organism. The human genome—the sum total of hereditary information in a person—consists of 3 x 10^9 bp of DNA divided into 25 molecules, the mitochondrial DNA molecule plus the 24 different chromosomal DNA molecules.
Proteome	The totality of proteins in a cell or organism.
Transcriptome	The total set of different RNA transcripts in a cell or tissue.
Phenotype	An observable characteristics of a cell or organism, such as height, eye color, or blood type. Also called *trait.* Phenotype may also define a whole compendium of traits together.
Protein	A large complex molecule made up of one or more polypeptide chains.
Real-time reverse transcription polymerase chain reaction (RT qPCR)	A laboratory method to quantify mRNA, employing the following steps: 1) the reverse transcriptase (RT)-dependent conversion of RNA into cDNA (complementary DNA), 2) the amplification of the cDNA using the PCR amplification and 3) the detection and quantification of amplification products in real time (Gibson, Heid, & Williams, 1996)

Recessive	Description of a characteristic that is manifested only in the homozygote.
Reference genome	A digital nucleic acid sequence database, assembled by scientists as a representative example of a species' set of genes.
RNA, ribonucleic acid	A chemical similar to a single strand of DNA that delivers DNA's genetic message to the cytoplasm of a cell where proteins are made.
RNA-seq	A full-length cDNA sequencing application used to get information about a sample's comprehensive and quantitative RNA content.
Sequence alignment	A way of arranging two or more amino acid or base sequences from an organism or organisms in such a way as to line up individual residues in each sequence with residues in the other sequence(s) for aligning areas of the sequences sharing common properties.
Single nucleotide polymorphism (SNP)	A position in the genome where two, or occasionally three, alternative nucleotides are common in the population. May be pathogenic or neutral.
Trait	A specific characteristic or phenotype of an organism.

Note. Compiled from Lea, 2009; National Cancer Institute, 2012, 2013; National Human Genome Research Institute, 2012; Strachan & Read, 2010.

References

1. Benjamini, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. Biometrical Journal, 52,708–721. doi:10.1002/bimj.200900299

2. Gibson, U. E., Heid, C. A., & Williams, P. M. (1996). A novel method for real time quantitative RT-PCR. Genome Research, 6,995–1001. doi:10.1101/gr.6.10.995

3. National Cancer Institute. (2012, October 31). NCI term browser: NCI thesaurus (Version 12.10e). Retrieved from http://ncit.nci.nih.gov/

4. National Cancer Institute. (2013). NCI dictionary of cancer terms widget. Retrieved from http://www.cancer.gov/global/widgets/dictionary

5. National Human Genome Research Institute. (2012). Talking glossary of genetic terms. Retrieved from http://www.genome.gov/glossary/

6. U.S. National Institute of Standards and Technology. (2011). The NIST definition of cloud computing: Recommendations of the National Institute of Standards and Technology. Retrieved from http://dx.doi.org/10.6028/NIST.SP.800-145