Appendix 1: Glossary of Terms

Term	Formal Definition	Source	Reference
Allele	Mutually exclusive alternative forms of the same gene occupying at the same locus on homologous chromosomes, differing in DNA sequence and governing the same biochemical and developmental process.	NCI Thesaurus	C16277
Chromosomal Phase	The property shared by DNA features found on the same chromosome.	This Document	
Consensus Sequence	A sequence pattern derived from the alignment of multiple sequences that represents the nucleotide or amino acid most likely to occur at each position in a sequence.	NCI Thesaurus	C40982
Genome Assembly	An annotated assembly of genome sequences created by the assimilation of data pieces from numerous sources.	NCI Thesaurus	C73517
Genotype	At its broadest level, genotype includes the entire genetic constitution of an individual. It is often applied more narrowly to the set of alleles present at one or more specific loci.	NCI Thesaurus	C16631
Genotyping	The process of assessing genetic variation present in an individual.	NCI Thesaurus	C45447
Genotyping Assay	An assay that generates data about a genotype from a specimen of genomic DNA. A variety of techniques and instruments can be used to produce information about sequence variation at particular genomic positions.	OBI	OBI_0000435
Genotyping Performance	The relative accuracy of genotyping results generated by different genotyping assays.	This Document	
HLA Nomenclature	Naming system for HLA alleles as defined by the WHO Nomenclature Committee for Factors of the HLA System.	This Document	
KIR Nomenclature	Naming system for KIR alleles as defined by the KIR Nomenclature Committee, a sub-committee of the WHO	This Document	

	Nomenclature Committee for Factors of the HLA System.		
Metadata	Data about data; information that describes another set of data.	NCI Thesaurus	C52095
Nomenclature	A naming system for describing genetic variants at a genetic locus in a defined, systematic manner.	This Document	
Novel Polymorphisms	Variants that are not described in the current version of a reference allele sequence database.	This Document	
Null Allele	A mutation that results in either no gene product or the absence of function at the phenotypic level.	NCI Thesaurus	C93103
Polymorphism	The regular and simultaneous occurrence of alternative nucleotide sequences at equivalent locations of the genetic material among individuals of a single interbreeding population that are not maintained by recurrent mutation. Polymorphisms may range in size from single nucleotide to large nucleotide sequence variation visible at the chromosomal level.	NCI Thesaurus	C17004
Primary Data	The fundamental representation of collected data (what is considered fundamental is defined by community consensus and may be ambiguous or mutable).	NCI Thesaurus	C84339
Primer Target Location	Chromosomal position of the DNA sequence complementary to that of the primer.	This Document	
Reference Allele Sequence Database	A database that relates known allele sequences to allele names assigned by a nomenclature authority.	This Document	
Reference Sequence (Refseq)	The Reference Sequence (refseq) collection aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms. Refseq standards serve as the basis for medical, functional, and diversity studies; they provide a stable reference for gene	NCI Thesaurus	C45335

	identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analyses. Refseqs are used as a reagent for the functional annotation of some genome sequencing projects, including those of human and mouse.		
Unreferenced Sequence	Regions of the consensus sequence for which no homologous sequence is available in the pertinent reference allele sequence database.	This Document	

Appendix 2: Explanation of Acronyms

Acronyms	Meaning
BED	Browser Extensible Data; a format used to define the lines displayed in a genome browser track.
EMBL	European Molecular Biology Laboratory
FASTA	Text based format for representing nucleotide or peptide sequences (not an acronym)
FASTQ	Text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores
GL String	Genotype List String
GTR	NCBI Genetic Testing Registry
HLA	Human Leukocyte Antigen
IMGT	ImMunoGeneTics
IPD	Immuno Polymorphism Database
KIR	Killer-cell Immunoglobulin-like Receptors
NCI	National Cancer Institute
NGS	Next Generation Sequencing; aka High Throughput Sequencing
OBI	Ontology for Biomedical Investigations
SBT	Sequencing Based Typing
SFF	Standard Flowgram Format; 454, contains information about the flowgram, called sequence, qualities, recommended quality and adaptor clipping
SSO	Sequence Specific Oligonucleotide hybridization
SSP	Sequence Specific Primer amplification
VCF	Variant Call Format
WHO	World Health Organization