Individuals table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| family\_id | int | Unique family identifier |  |
| individual\_id | int | Unique individual identifier | Primary |
| father\_id | int | Individual identifier of the father |  |
| mother\_id | int | Individual identifier of the mother |  |
| sex | tinyint | Male=1, Female=2 |  |
| sex\_method | varchar | Method of estimating the sex |  |
| ethnicity | varchar | Inferred ancestral population based on 1000 Genomes classification |  |
| ethnicity\_superpopulation | varchar | Inferred ancestral super-population based on 1000 Genomes classification |  |
| ethnicity\_method | varchar | Method of inferring ancestry ex: PCA or reported |  |

Samples table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| sample\_id | int | Unique sample identifier (One individual can have multiple samples) | Primary |
| individual\_id | int | Unique individual identifier | Foreign from table Individuals |
| data\_type | varchar | Type of data ex: WGS/SNP genotyping etc. |  |
| external\_lab\_id | varchar | Unique original sample identifier |  |
| external\_source | varchar | Source of sample collection |  |
| phasing\_method | varchar | Method used for phasing ex: trio, duo or related |  |
| impute\_method | varchar | Method used for imputation ex: MIS, TOPMed |  |
| impute\_panel | varchar | Reference panel used for imputing ex: 1000 Genomes |  |
| import\_date | date | The date the sample was imported to the database |  |

PathogenicMutations table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| mutation\_id | int | Unique identifier for a specific mutation | Primary |
| disease | varchar | Unique name of the disease (use OMIM name by default) |  |
| disease\_id | varchar | Unique OMIM abbreviation (This should be the same as the first part of DCV parameter) |  |
| omim\_id | int | Unique OMIM identifier for the disease variant/mutation |  |
| gene | varchar | Gene containing the mutation |  |
| genomic\_region | varchar | Exonic, intronic etc. |  |
| inheritance\_model | varchar | Autosomal dominant (AD), recessive (AR) or X-linked dominant or recessive (XLD or XLR) |  |
| chromosome | varchar | Chromosome ex: chr1 |  |
| start\_position\_hg19 | int | Start base pair position of the variant in hg19 |  |
| end\_position\_hg19 | int | End base pair position of the variant in hg19 |  |
| start\_position\_hg38 | int | Start base pair position of the variant in hg38 |  |
| end\_position\_hg38 | int | End base pair position of the variant in hg38 |  |
| start\_position\_cM | double | Start position of the variant in cM |  |
| end\_position\_cM | double | End position of the variant in cM |  |

IndividualsWithKnownMutations table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| individual\_id | int | Unique individual identifier | Foreign from Individuals |
| mutation\_id | int | Unique identifier for a specific mutation | Foreign from table PathogenicMutations |
| genotype | tinyint | Number of copies of the identified disease variant ex: 1 or 2 |  |
| validated | boolean | Is the disease variant validated? |  |
| validation\_method | varchar | Validation method ex: RP-PCR, bioinformatic etc. |  |
| Validation\_note | varchar | Extra note on validation, if any |  |

GeneticMarkers table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| marker\_id | bigint | Unique identifier of genetic markers | Primary |
| rs\_id | varchar | Reference SNP identifier |  |
| chromosome | varchar | Chromosome ex: chr1 |  |
| position\_hg19 | int | Base pair position in hg19 |  |
| position\_hg38 | int | Base pair position in hg38 |  |
| position\_cM | double | position in cM |  |
| reference\_allele | varchar | Reference allele |  |
| alternate\_allele | varchar | Alternate allele |  |
| marker\_type | varchar | SNP (FoundHaplo currently uses only SNPs) |  |
| maf\_gnomad\_ALL | float | General allele frequency |  |
| maf\_gnomad\_AFR | float | Allele frequency in African population |  |
| maf\_gnomad\_NFE | float | Allele frequency in Non-finnish European population |  |
| maf\_gnomad\_FIN | float | Allele frequency in Finnish population |  |
| maf\_gnomad\_AMR | float | Allele frequency in American population |  |
| maf\_gnomad\_EAS | float | Allele frequency in East Asian population |  |
| maf\_gnomad\_SAS | float | Allele frequency in South Asian population |  |

Genotypes table

|  |  |  |  |
| --- | --- | --- | --- |
| **Field name** | **Data type** | **Description** | **Key** |
| marker\_id | bigint | Unique identifier of genetic markers | Primary |
| sample\_id | int | Unique sample identifier (One individual can have multiple samples) | Foreign from table Samples |
| genotype | tinyint | “0” denotes the reference allele and “1” denotes the alternate allele |  |
| imputed | boolean | Is the marker imputed? |  |
| imputation\_quality | float | Imputation quality score |  |