# Standard Operating Procedure:

# SNP Genotyping Requirements for Rare Disease Whole Genome Sequencing Samples by Genomic Laboratory Hubs

## Contents

# Introduction

This Standard Operating Policy (SOP) defines the process of Single Nucleotide Polymorphism (SNP) genotyping of rare disease DNA samples by each of the seven Genomic Laboratory Hubs (GLHs) prior to submission to the Whole Genome Sequencing (WGS) Provider within the Genomic Medicine Service (GMS).

The multistep process involved in delivering WGS requires effective quality control in terms of sample checking and tracking to be in place to minimise sample mix up. As the process also involves external providers the potential for error is increased and additional checks are required to be in place. SNP genotyping enables a sample identity verification step to be performed once the WGS data has been returned to the GLH by Genomics England.

- The rare disease samples commissioned for NHS clinical diagnostic purposes for WGS are supplied by NHS England through the Genomic Laboratory Hubs (GLHs), to the WGS Provider in an agreed format for the WGS provider to handle and sequence.
- Prior to submission for WGS, the GLHs will perform SNP genotyping on all rare disease samples. SNP genotyping is performed to enable the identity of the submitted sample to be verified against the returned sequence.
- During the implementation of WGS in the GMS a single dedicated GLH will plate all DNA samples for WGS on behalf of all GLHs into a 96-well format acceptable for the WGS Provider to handle and will dispatch the samples to the WGS Provider. All GLHs will develop a process to plate their own WGS samples and within 18 months of WGS being delivered in the GMS all seven GLHs will dispatch their own WGS samples directly to the WGS Provider.
- The process of determining the SNP genotyping will be performed using a number of specified technologies currently in use across the GLH network. These assays shall be accredited to ISO 15189.

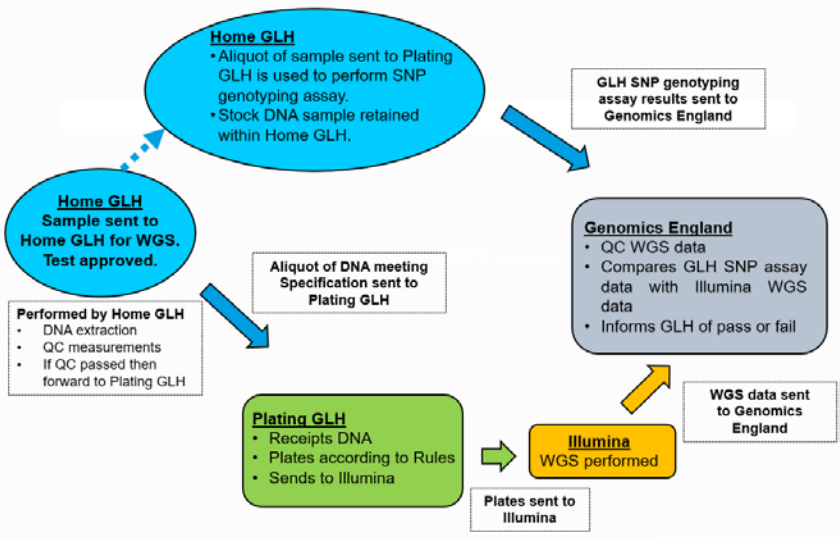
# 1. SNP Genotyping Pathway

## 1.1 Summary

The GLHs shall be responsible for the provision of SNP genotyping of DNA samples to ensure the correct assignment of the WGS data returned to the GLHs from the WGS Provider and Genomics England to the sample submitted for WGS.

The DNA extraction GLH shall perform SNP genotyping on rare disease samples submitted for WGS and shall supply this data to Genomics England for comparison against the WGS data once available.

*Figure 1: High level summary of the SNP genotyping pathway for the GMS*




## 2. SNP Genotyping Requirements

The SNP genotyping assay in use across each of the GLHs shall be validated in house and be accredited to ISO 15189.

In summary, the detailed minimum requirements of assay are:

- 24 SNPs (Pengelly *et al*., Genome Medicine 2013, 5:89)
- High data quality ≥99% call rates
- High accuracy ≥99.75%
- Designed for high throughput, fast workflow

### 2.1 Methodologies/Assays

Several existing platforms and technologies are currently in use across the GLH network therefore each GLHs can continue to perform any currently validated assays providing they comply with minimum specification.

### 2.2 Selection and Number of SNPs

There must be a minimum number of 24 SNPs in any assay which comply with the following criteria:

1. Represent biallelic substitutions, excluding substitutions of complementary bases
2. Be technically amenable to accurate whole exome sequencing (WES) and orthogonal genotyping i.e. must not be present in large scale genomic repeats
3. Must not alter the primary sequence of the encoded protein or have an associated Online Mendelian Inheritance in Man (OMIN) record
4. Must be located at least 10bp from the exon boundaries
5. Must not be situated in regions with a high sequence similarity to non-target regions
6. Must be outside of linkage disequilibrium with other selected SNPs

Appendix 1 lists all current SNPs in use across the GLH network

### 2.3 Cost

The assay of choice needs to be cost effective therefore high-throughput, low cost assays are preferable.

## 3. Sample matching

### 3.1 Power of Discrimination

The ability of a set of SNPs to discriminate whether two independent samples are from the same individual can be estimated as the probability that two samples from different individuals have the same genotype at a biallelic SNP given their population ancestry (allele frequency of that SNP) and their relatedness. Here we assume that the two samples are of the same ancestry. Because the SNPs are assumed to be in linkage equilibrium, the probabilities of various sites are independent and can be, therefore, simply multiplied to obtain the probability that all SNPs under examination have the same genotype.

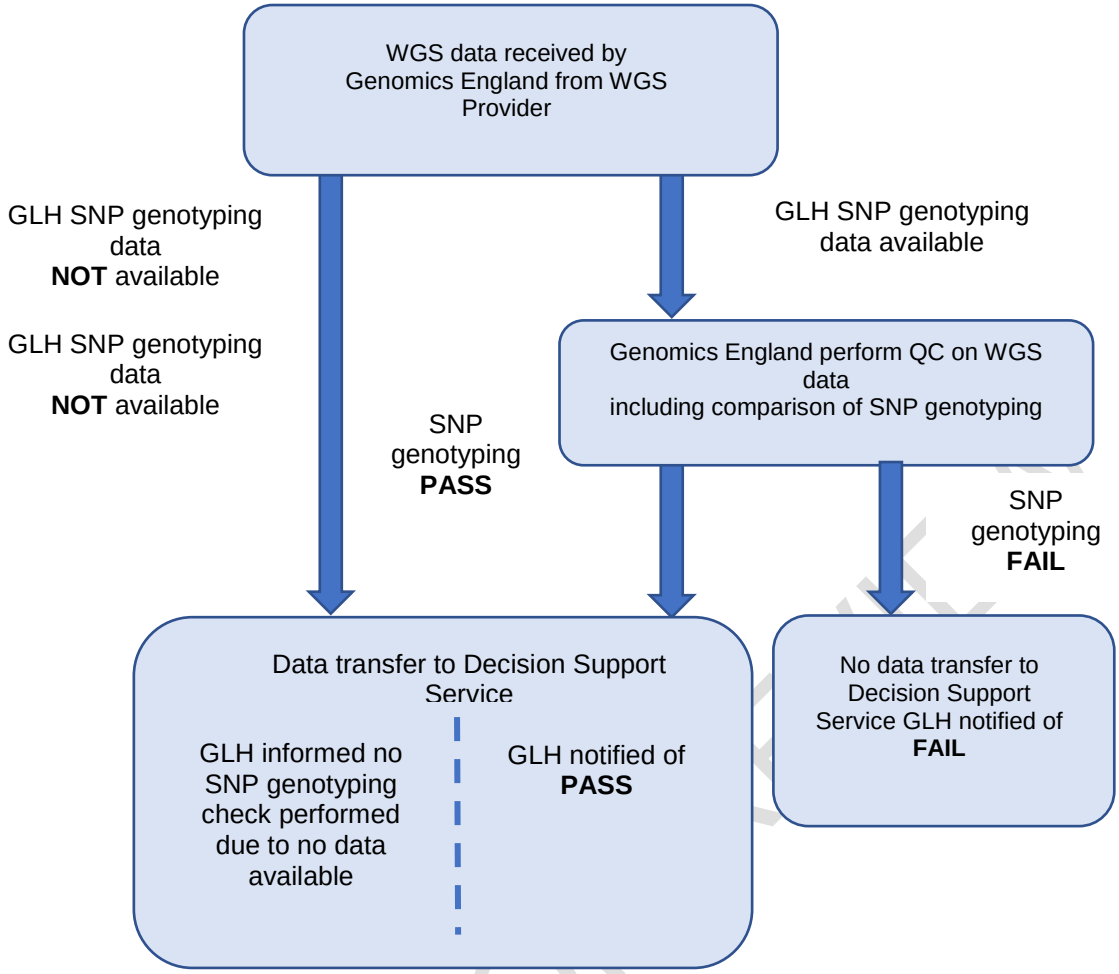
The single site probability is:

Pr(IBS=2)=z0*(p^4 + 4*p^{2}*q^{2} + q^4) + z1*(p^3 + p^{2}*q + p*q^{2} +q^3) + z2*(p^2 + 2*p*q + q^2),
where p is the allele frequency, q=1-p, and the z are the number of alleles shared IBD.

The formula can be easily derived by considering the states where the genotype is identical in both individuals ( { AA,AA ; Aa,Aa ; aa,aa }) in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950838/table/TB2/

## 4. SNP Genotyping Analysis

*Figure 2 – Summary of the SNP genotyping informatics pathway*




## 4.1 Accuracy

High level of accuracy and reproducibility (>99% call rates with >99.7% accuracy) must be achieved.

## 4.2 Comparison against sequencing-based genotypes

SNP genotypes will be compared against genotypes ascertained from the whole genome sequencing data. The comparison determines the probability of a given number of observed genotypes to be concordant between the two samples, assuming that the two individuals are unrelated or that they are full siblings. The probability is calculated for all major populations available in GNOMAD. Only SNPs that passed on both the SNP assay and the WGS genotyping are used. Therefore, the total number of SNPs used in the comparison may be lower than those present in the assay.

A “FAIL” is called when the probability of genotype matching assuming that the other sample comes from a full sibling, given the number of SNPs that pass in both assays, is greater than 10^-4, for the least favourable population (maximum probability across all populations). That means more than 1 in 10,000 chances of a sample swap. Such a level of evidence is roughly equivalent to saying that we will tolerate 1 sample swap across 1,000,000 tests assuming a swap rate of 1% and that all swaps occur with

siblings. This threshold will be re-evaluated at later stages as more empirical evidence of its consequences becomes available.

In the case of a “FAIL” for a sample showing in the interpretation portal, the sample will have passed internal checks for ‘sex, Mendelian inconsistencies and family relatedness’.
For those samples that fail checks of sex, Mendelian inconsistencies and family relatedness, the Sample Matching Service will run, and results will be applied to the query resolution ticket raised with the GLH.

# 5. Detailed Technical specification for developers

## 5.1 The detailed technical specification

The detailed technical specification for the GLHs to develop and submit the SNP data to NGIS can be found on NHS Futures
(https://future.nhs.uk/connect.ti/home/grouphome)

# 6. Appendix 1

### SNPs in use across the GLH Network – genomic position according to GRCh38

| SNP/RS ID | Chromosome | Position (Hg19) |
|---|---|---|
| rs1410592 | 1 | 179551371 |
| rs2229546 | 1 | 67395837 |
| rs1805087 | 1 | 236885200 |
| rs2229546 | 1 | 67395837 |
| rs1410592 | 1 | 179551371 |
| rs497692 | 2 | 168932506 |
| rs10203363 | 2 | 227032260 |
| rs497692 | 2 | 168932506 |
| rs10203363 | 2 | 227032260 |
| rs10498027 | 2 | 214955289 |
| rs2229267 | 2 | 169235885 |
| rs2819561 | 3 | 4362083 |
| rs4688963 | 4 | 5748177 |
| rs4688963 | 4 | 5748177 |
| rs3088052 | 5 | 139121126 |
| rs309557 | 5 | 83538811 |
| rs309557 | 5 | 83538811 |
| rs2942 | 6 | 146434004 |

| | | |
|---|---|---|
| rs2942 | 6 | 146434004 |
| rs2256135 | 6 | 152143704 |
| rs2747662 | 6 | 152145539 |
| rs17548783 | 7 | 48410560 |
| rs17548783 | 7 | 48410560 |
| rs17548783 | 7 | 48410560 |
| rs712700 | 7 | 127610853 |
| rs4735258 | 8 | 93923709 |
| rs4735258 | 8 | 93923709 |
| rs4735258 | 8 | 93923709 |
| rs639225 | 9 | 27202872 |
| rs1381532 | 9 | 97428498 |
| rs1381532 | 9 | 97428498 |
| rs1381532 | 9 | 97428498 |
| rs10883099 | 10 | 98459557 |
| rs2275271 | 10 | 103054405 |
| rs10883099 | 10 | 98459557 |
| rs6163 | 10 | 102837167 |
| rs10883099 | 10 | 98459557 |
| rs4617548 | 11 | 16111867 |
| rs4617548 | 11 | 16111867 |
| rs7300444 | 12 | 884764 |
| rs7300444 | 12 | 884764 |
| rs9532292 | 13 | 38859469 |
| rs9532292 | 13 | 38859469 |
| rs9532292 | 13 | 38859469 |
| rs2287016 | 14 | 75579515 |
| rs2297995 | 14 | 50302999 |
| rs2297995 | 14 | 50302999 |
| rs4577050 | 15 | 34236747 |
| rs4577050 | 15 | 34236747 |
| rs2070203 | 16 | 70269677 |
| rs2070203 | 16 | 70269677 |
| rs2285475 | 17 | 10639154 |
| rs1037256 | 17 | 73201609 |
| rs9962023 | 18 | 23833905 |
| rs11080572 | 18 | 12351343 |
| rs2298628 | 18 | 49929553 |
| rs2228611 | 19 | 10156401 |
| rs2228611 | 19 | 10156401 |
| rs2241714 | 19 | 41363487 |
| rs2228611 | 19 | 10156401 |
| rs10373 | 20 | 6119441 |
| rs2076584 | 20 | 19990061 |
| rs10373 | 20 | 6119441 |
| rs753381 | 20 | 41168825 |

| | | |
|---|---|---|
| rs4148973 | 21 | 42903480 |
| rs4148973 | 21 | 42903480 |
| rs4148973 | 21 | 42903480 |
| rs4675 | 22 | 20787012 |