# Sample summary page

Created by Susan Fairley, last modified by Nina Habermann on Jan 16, 2016

## Overview

This page provides a summary of the currently available and planned data sets at a sample level.

- Pedigree information

  For the families to which members of the SV trios belong, the pedigree entries for all family members sampled in phase three of the 1000 Genomes Project are included in http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/hgsv_sv_discovery/sv_trio_families.ped. Membership of the trios is indicated in the last column of the file.

**The table at the foot of this page shows the locations of the data sets on the FTP site. Further information can be found by consulting the current.tree.**

## Availability of raw data

This table shows data availability for each sample and each raw data type. Where 'N/A' is present no data is being produced and where 'Available' the data is accessible through the public archives and/or the FTP site. In the remaining empty cells, data has yet to be submitted to the final location. Information on the location of the data can be found in the index files on the FTP site and a summary table is below.

| **Sample name** | HG00512 | HG00513 | HG00514 | HG00731 | HG00732 | HG00733 | NA19238 | NA19239 | NA19240 |
|---|---|---|---|---|---|---|---|---|---|
| **1000 Genomes SRS IDs** | SRS008631 | SRS008632 | SRS008633 | SRS008745 | SRS008746 | SRS008747 | SRS000212 | SRS000213 | SRS000214 |
| **Population** | CHS | CHS | CHS | PUR | PUR | PUR | YRI | YRI | YRI |
| **Relationship** | father | mother | child | father | mother | child | mother | father | child |
| **Gender** | male | female | female | male | female | female | female | male | female |
| **PCR-free high coverage** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **1000 Genomes low coverage** | Available | Available | N/A | Available | Available | Available | Available | Available | Available |
| **1000 Genomes exome** | Available | Available | N/A | Available | Available | Available | Available | Available | Available |
| **Complete genomics** | N/A | N/A | N/A | Available | Available | Available | Available | Available | Available |
| **7kb mate pair** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **3.5kb jumping libraries** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **PacBio genomic** | | | | Available | Available | Available | Available | Available | Available |
| **PacBio RNA-seq** | | | | | | | | | |
| **PacBio methylation** | | | | | | | | | |
| **450k methylation array** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **Strand-seq** | | | | | | | | | |
| **Strand specific RNA-seq** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **Moleculo** | | | | | | | | | |
| **10x** | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive | Available but pending archive |
| **BioNano** | Available | Available | Available | Available | Available | Available | Available | Available | Available |
| **Additional BioNano to 90x** | | | | | | | | | |
| **Hi-C** | | | | | | | | | |
| **aCGH arrays** | Available | Available | Available | Available | Available | Available | Available | Available | Available |

## Data Locations

The following paths represent relative paths with respect to the FTP site ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/

| **Data set** | **Description** | **Coverage / Comments** | **Location** |
|---|---|---|---|
| **PCR-free high coverage** | **Alt aware BWA alignments to GRCh38 produced by DCC** | **80x** | data_collections/hgsv_sv_discovery/data/*/*/high_cov_alignment (alt-GRCh38DH) |
| | **Alignments to GRCh37 produced by DCC (not all samples)** | | phase3/data/*/high_coverage_alignment (GRCh37) |
| | **Raw data** | | data_collections/hgsv_sv_discovery/illumina_wgs.sequence.index (analysis group: pcr_free_high) |
| **1000 Genomes low coverage** | **Alt aware BWA alignments to GRCh38 produced by DCC** | **7x** | data_collections/1000_genomes_project/data/*/*/alignment (alt-GRCh38DH) |
| | **Alignments to GRCh37 produced by DCC (not all samples)** | | phase3/data/*/alignment (GRCh37) |
| | **Raw data** | | data_collections/1000_genomes_project/1000genomes.sequence.index (analysis group: low_coverage) |
| **1000 Genomes exome** | **Alt aware BWA alignments to GRCh38 produced by DCC** | **16x** | data_collections/1000_genomes_project/data/*/*/exome_alignment (alt-GRCh38DH) |
| | **Alignments to GRCh37 produced by DCC (not all samples)** | | phase3/data/*/exome_alignment (GRCh37) |
| | **Raw data** | | data_collections/1000_genomes_project/1000genomes.sequence.index (analysis group: exome) |
| **Complete genomics** | **Data from Complete Genomics** | **not applicable** | phase3/data/*/cg_data |
| **7kb mate pair** | **Raw data** | **30-40x** | data_collections/hgsv_sv_discovery/illumina_wgs.sequence.index (analysis group: sv_7kb_mate) |
| **3.5kb jumping libraries** | **Raw data** | | data_collections/hgsv_sv_discovery/illumina_wgs.sequence.index (analysis group: 3.5kb_jump) |
| | **Alignments to versions of GRCh38, see README** | | data_collections/hgsv_sv_discover/working/20151021_3.5kb_jumping (GRCh38DH and GRCh38primary) |
| **PacBio genomic** | **Raw data** | **(parent - parent - child)**<br>**YRI: 18.7 - 16.6 - 36.6**<br>**PUR: 24.3 - 23.9 - 46.9**<br>**CHB: 18.6 - 16.9 - 40.2** | data_collections/hgsv_sv_discovery/smrt.sequence.index (analysis group: sv_smrt) |
| **PacBio RNA-seq** | | **16 SMRT cells of Iso-Seq** | |
| **PacBio methylation** | | **20-40x** | |
| **450k methylation array** | **Raw and analysed data files from Ankit Malhotra** | **485,000 sites**<br>**(Affymetrix Infinium HumanMethylation450 BeadChip array)** | data_collections/hgsv_sv_discovery/working/20151214_450k_methylation/ |
| **Strand-seq** | **(CHS)** | **2.5-3.8x** | |
| **Strand specific RNA-seq** | **Raw data** | | data_collections/hgsv_sv_discovery/illumina_rna.sequence.index (analysis group: strand_specific_rna) |
| | **Alignment to GRCh38 primary, see README** | | data_collections/hgsv_sv_discover/working/20151026_strand_specific_mRNA (GRCh38primary) |
| **Moleculo** | **(YRI)** | **3.8x** | |
| **10x** | **PUR** | **25-33x**<br>**Notes:**<br>**Each dataset contains:**<br>**(i) A BAM file containing barcoded and phased reads**<br>**(ii) A phased VCF file**<br>**(iii) A set of high confidence large-scale (50Kb +) SV calls and a broader set of SV candidates (the two .bedpe files)**<br>**(iv) A file containing the phase calls of the large-scale SVs (the .tsv file)**<br>**(v) A Loupe file for use with our visualization browser**<br>**Documentation for all of the output is here:**<br>**http://software.10xgenomics.com/docs/pipelines/output/overview** | http://s3-us-west-2.amazonaws.com/10x.outbound/10c6adf902689cf3/index.html |
| | **YRI** | | http://s3-us-west-2.amazonaws.com/10x.outbound/70e93e7b402f15b6/index.html |
| | **CHS** | | http://s3-us-west-2.amazonaws.com/10x.outbound/762e270eb720fa1f/index.html |
| **BioNano** | **Raw data** | **(parent - parent - child, aim: 60x)**<br>**YRI: 46.5 - 19.8 - 39.3**<br>**PUR: 33.9 - 31.7 - 30.4**<br>**CHS: 33.3 - 33.9 - 33.4** | data_collections/hgsv_sv_discovery/working/20150709_bionano_optical_maps/ |
| **Additional BioNano to 90x** | | | |
| **Hi-C** | | **NA** | |
| **aCGH arrays** | **Raw data - CEL files** | **750,000 biallelic SNP probes**<br>**(Affymetrix cytoscan HD array)** | data_collections/hgsv_sv_discovery/working/20151103_acgh/ |