

Sequencing data for "Stable clonal contribution of lineage-restricted stem cells to human hematopoiesis"

Author: Tetsuichi Yoshizato

DNA Sequencing Data Overview

This repository contains three types of DNA sequencing data:

- **Error-corrected DNA capture sequencing (ECTS)**
- **Bulk whole-exome sequencing (WES)**
- **Single-colony whole-genome sequencing (WGS)**

All sequencing was performed on an Illumina NovaSeq 6000 at the National Genomics Infrastructure in Stockholm, using paired-end sequencing mode.

All BAM files were converted to CRAM format using samtools version 1.5 with the following command:

```
samtools view -T {reference_GRCh37_fasta_file} -C -o {output_CRAM_file}
{input_BAM_file}
```

For detailed information on each file, please refer to CRAM_file_description.xlsx.

Error-Corrected DNA Capture Sequencing (ECTS)

- Performed on bulk bone marrow mononuclear cells (BMMNCs) from all 93 healthy donors
- Capture probe information is available in the Supplementary Table of the paper
- Sequencing reads were aligned to the GRCh37 reference genome
- Reads sharing the same unique molecular identifier (UMI) were grouped, and consensus reads were generated using fgbio FilterConsensusReads
- Deposited CRAM files were generated from BAM files containing consensus reads
- MD5 checksums and file sizes for each ECTS CRAM file are listed in [ECTS_cram_to_upload_md5_file_size.txt](#)

Bulk Whole-Exome Sequencing (WES)

- Performed on BMMNC DNA from 20 out of 93 healthy donors aged over 71 years
 - Libraries prepared using the IDT Lotus DNA Library Prep Kit and xGen Exome Hyb Panel v2
 - Paired buccal swab DNA was used to filter out single-nucleotide polymorphisms (SNPs) and sequencing errors
 - Sequencing reads were aligned to the GRCh37 reference genome
 - MD5 checksums and file sizes for each WES CRAM file are listed in [WES_cram_to_upload_md5_file_size.txt](#)
-

Single-Colony Whole-Genome Sequencing (WGS)

- Performed on non-amplified DNA extracted from 333 genotyped single colonies from 10 out of 93 healthy donors
- Paired buccal swab DNA from the same donors was used to filter out SNPs and sequencing errors
- Sequencing reads were aligned to the GRCh37 reference genome
- MD5 checksums and file sizes for each WGS CRAM file are listed in

`scWGS_cram_to_upload_md5_file_size.txt`