

Data and Code for "Stable clonal contribution of lineage-restricted stem cells to human hematopoiesis"

Author: Tetsuichi Yoshizato

Note:

To run the scripts properly, you will need to modify the file paths in the code to match your local directory structure.

Sample information

Data

- `data/pb_info.txt`
Peripheral blood count data
 - `data/pb_ref_values.txt`
Peripheral blood reference values
 - `data/sample_info.txt`
Sample metadata
 - `data/sample_info_wide.txt`
Sample metadata (wide format)
-

CH driver genes

Data

- `data/CH_35genes.rds`
List of 35 clonal hematopoiesis (CH) driver genes
 - `data/cols_gene.rds`
Color codes used for CH genes
 - `data/gene_refseq_id.txt`
RefSeq ID list for ECTS genes
-

ECTS

Mutations detected in bulk bone marrow mononuclear cells (BMMNCs) by error-corrected targeted DNA capture sequencing (ECTS)

Data

- `data/ECTS/BMMNC_ECTS_mutation.txt`
Mutation list from ECTS analysis of BMMNC samples

Code

- `scripts/ECTS_analysis.R`
R script for analyzing the ECTS mutation list and generating figures
-

ddPCR

Data

- `data/ddPCR/ddPCR_1st_visit_result.txt`
ddPCR results for mutations detected by ECTS in first-visit samples
- `data/ddPCR/ddPCR_serial_result.txt`
ddPCR results for mutations detected by ECTS in serial samples
- `data/ddPCR/ddPCR_WGS_result.txt`
ddPCR results for mutations detected by single-colony whole-genome sequencing (WGS)
- `data/ddPCR/ddPCR_transplanted_mouse_result.txt`
ddPCR results for mutations analyzed in NSG mouse transplantation experiments

Code

- `scripts/ddPCR_model_cellnum.stan`
Stan code for estimating mutant cell fractions from ddPCR results
 - `scripts/ddPCR_stan_model.R`
R script for running the Stan model
 - `scripts/ddPCR_processing_stan_res.R`
R script for processing the Stan model output
 - `scripts/ddPCR_post_stan_analysis_for_1st.R`
R script for analyzing preprocessed ddPCR data from first-visit samples
 - `scripts/ddPCR_post_stan_analysis_for_serial.R`
R script for analyzing preprocessed ddPCR data from serial samples
 - `scripts/ddPCR_post_stan_analysis_for_WGS.R`
R script for analyzing preprocessed ddPCR data for mutations detected by single-colony WGS
 - `scripts/ddPCR_post_stan_analysis_for_transplanted_mouse.R`
R script for analyzing preprocessed ddPCR data from transplanted mouse samples
-

Flowcytometry profile for HSPCs and mature cells

Number of cells in hematopoietic stem/progenitor cells (HSPCs) and mature cells measured by flow cytometry.

Two types of panels were used for HSPC evaluation: one including ProB cells and one without.

For samples tested with both panels, only the values from the panel that included ProB cells are reported.

For samples evaluated with both HSPC panels, the panel including ProB cells reports only the ProB values; all other lineage data are taken from the panel without ProB cells.

See Extended Data Fig. 3 for the gating strategy.

Data

- `data/FACS_profile/HSPC_cell_num.txt`
Number of HSPCs in each sample
- `data/FACS_profile/mature_cell_num.txt`
Number of mature cells in each sample

Whole-exome sequencing (WES) in bulk BMMNCs

Data

- `data/WES/BMMNC_WES_mutation.txt`
Mutation list detected by WES in bulk BMMNCs
- `data/WES/BMMNC_WES_cnv.txt`
Copy number variation (CNV) list detected by WES in bulk BMMNCs

Single-colony whole-genome sequencing (WGS)

Data

- `data/scWGS/HD**_scWGS_mutation.txt`
Mutation list detected by single-colony WGS for each donor
- `data/scWGS/HD**_scWGS_snp_sensi.txt`
Sensitivity of single nucleotide polymorphisms (SNPs) per colony for each donor
- `data/scWGS/cnv_all.txt`
CNV list across all samples
- `data/scWGS/sv_all.txt`
List of structural variants (SVs) across all samples
- `data/scWGS/for_mpboot/HD**.phy.contree`
MPBoot results (phylogenetic trees) for each donor
- `data/scWGS/cosmicV2_SBSblood_freq.txt`
COSMIC V2 mutational patterns

- `data/scWGS/trinuc_freq.txt`
Trinucleotide mutational patterns
- `data/scWGS/genotype_info_for_trinuc_mut.txt`
Genotype information for trinucleotide plot

Code

- `scripts/phylo_tree.R`
R script for visualizing phylogenetic trees from MPBoot output
Requires downloading functions from:
https://github.com/emily-mitchell/normal_haematopoiesis/tree/main
 - `scripts/phylo_tree2.R`
R script for statistical analysis integrating phylogenetic tree and ddPCR data
Requires downloading functions from:
https://github.com/emily-mitchell/normal_haematopoiesis/tree/main
Requires outputs from `scripts/phylo_tree.R` and `scripts/ddPCR_processing_stan_res.R`
 - `scripts/scWGS_VAF_hist.R`
R script for generating VAF distribution plots per colony
 - `scripts/scWGS_mutational_signature.R`
R script for generating mutational signature heatmaps
-

Single-colony genotyping

Data

- `data/sc_genotype/genotyped_colonies.txt`
List of genotyped colonies
 - `data/sc_genotype/HD*_sc_genotype.txt`
Genotyping results per colony for each donor
 - `data/sc_genotype/probes_sc_genotype.txt`
Probes used for genotyping
-

Engrafted human cells in transplanted Mice

Data

- `data/mouse_engraft/human_cell_in_transplanted_mouse.txt`
Number of human cells detected by flow cytometry
See Extended Data Fig. 9 for gating strategy