Brief description of the ICA matrices

Brief experimental details:
The cells in this project include the delivery of:
Cord blood (CB) 10X measurements 8 individuals (donor)
Bone Marrow (BM) 10X measurements 8 individuals (donor)
For each donor we prepared 8 independent 10X channels

- Bone marrow donor 6, channel 3 was failed thus not sequenced
- We identified 274,182 cells from the bone marrow data set
- We identified 253,910 cells from the cord blood data set

Raw counts:
Raw counts in 10X hdf5 format separated by tissue type.
Generated by cell ranger with GRCh38, standard 10X reference
Selected top 6000 cell barcodes (ranked by total UMI per cell barcode).

- For each channel we approximated loaded 7000 cell and expect 4000 cells given the 10X capture rate (10X documentation/brochure)
- Since our samples contain a variety of cell types and some cell types are known for small cell sizes (e.g. T cells), we did not use CellRanger's automated functionality for determining the number of cells per channel but forced cellranger to report the top 6000 "cells" and then filtered low quality ones with our own criteria
- **Note: people should filter out low quality cells before they analyze the raw count matrix**

Cellranger commands:

1) Make fastq files:

{CellRanger} mkfastq --id={ID} --run={Run_ID} --csv={bcl_spreadsheet} --jobmode=local --localcores={number of threads} --localmem={total memory}

2) Generate raw gene count matrices:

{CellRanger} count --id={ID} --transcriptome={cellranger_ref_GRCh38} --fastqs={DATA_PATH}/fastqs/fastq_path --sample={Sample_ID} --force-cells 6000 --nosecondary --jobmode=local --localcores={number of threads} --localmem={total memory}

Files:
Ica_cord_blood_h5.h5 (produced by merging cellranger-produced hdf5 files)
Ica_bone_marrow_h5.h5 (produced by merging cellranger-produced hdf5 files)