**Human Population Genetics Data at DeepThought – Dr. Schrodi Lab**

Shicheng Guo, Steven Schrodi @ **03/20/2020**

**1, Personal Medical Research Project (PMRP, 18,604 samples)**

* 1. PLINK FORMAT (PHASE1 and PHASE2 MERGED DATA)

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R2.bed

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R2.bim

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R2.fam

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R3.pgen

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R3.psam

/mnt/sas0/AD/sguo234/pmrp/MCRI20000\_R3.pvar

* 1. VCF FORMAT (non-imputation and non-phasing)

/mnt/sas0/AD/sguo234/pmrp/MCRI20000.vcf.gz

/mnt/sas0/AD/sguo234/pmrp/MCRI20000.vcf.gz.tbi

* 1. IMPUTED and PHASED PMRP DATA

/mnt/sas0/AD/sguo234/pmrp/MCRI20000.impute.phase.vcf.gz

/mnt/sas0/AD/sguo234/pmrp/MCRI20000.impute.phase.vcf.gz.tbi

**2. Exome-Sequencing data to PMI (Personal Medicine Imitate)**

2.1 INDIVIDUAL VCF FILES

/mnt/sas0/AD/sguo234/exome/ 226764-a.filtered.vcf.gz

……………………………………………………………….

……………………………………………………………….

/mnt/sas0/AD/sguo234/exome/ 282139-a.filtered.vcf.gz

/mnt/sas0/AD/sguo234/exome/ 296044-a.filtered.vcf.gz

2.2 MERGED VCF

/mnt/sas0/AD/sguo234/exome/Final.vcf.gz

/mnt/sas0/AD/sguo234/exome/final.rs.vcf.gz

/mnt/sas0/AD/sguo234/exome/exome.rename.vcf.gz.gz

**3. Cytokine RNA-SEQ Dataset**

3.1 Saved in DataLocker since DeepThought don’t have enough space to store it.

**4. dbSNP153 in hg19**

Deepthought don’t have enough space, hence, dbSNP153 in hg38 was not generated.

/mnt/sas0/AD/sguo234/db/dbSNP153/dbSNP153.hg19.fam

/mnt/sas0/AD/sguo234/db/dbSNP153/dbSNP153.hg19.fam

/mnt/sas0/AD/sguo234/db/dbSNP153/dbSNP153.hg19.fam

/mnt/sas0/AD/sguo234/db/dbSNP153/ dbSNP153.norm.hg19.vcf.gz

/mnt/sas0/AD/sguo234/db/dbSNP153/ dbSNP153.norm.hg19.vcf.gz.tbi

**5. 1K Northeast Asians Genome Project (**1,781 Northeast Asians Genome)

/mnt/sas0/AD/sguo234/db/nard/NARD\_MAF.hg38.vcf.gz

/mnt/sas0/AD/sguo234/db/nard/NARD\_MAF.hg38.zip

**6. ANNOVAR annotation to dbSNP153 for Loss-of-function prediction**

**7. 1000 Genome phase 3 data in VCF and PLINK formats**

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.fam

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.bim

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.bed

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.pvar.zst

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.pgen.zst

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3.psam.zst

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3\_1000G.vcf.gz

/mnt/sas0/AD/sguo234/db/1000Genome/all\_phase3\_1000G.vcf.gz.tbi

**8. Software installed in DeepThought Server:**

Rvtests: version 20150104: /mnt/sas0/AD/sguo234/tools/rvtests/executable

Eagle-2.4.1: /mnt/sas0/AD/sguo234/tools/Eagle\_v2.4.1

samtools-1.10: /mnt/sas0/AD/sguo234/tools/

TrimGalore-0.6.5: /mnt/sas0/AD/sguo234/tools/

htslib-1.10.2: /mnt/sas0/AD/sguo234/tools/

gcta\_1.93.0beta: /mnt/sas0/AD/sguo234/tools/

FastQC-0.11.9: /mnt/sas0/AD/sguo234/tools/

bcftools-1.10.2: /mnt/sas0/AD/sguo234/tools/

R version 3.6.0: /usr/bin/R

Perl v5.16.3: /usr/bin/perl

ZSTD: /mnt/sas0/AD/sguo234/tools/zstd