GALAXY PIPELINE FOR FASTER WHOLE GENOME GENOTYPE CALLING ON THE GENETITAN PLATFORM

Oleksiy Karpenko, Neil J. Bahroos

Galaxy Community Conference
Chicago
July 26, 2012
OVERVIEW

- What is Affymetrix GeneTitan?
- Axiom® Genome-Wide Population-Optimized Human Arrays
- Affymetrix Power Tools (APT) vs Affymetrix Genotyping Console
- APT in Galaxy at UIC
- Galaxy workflow for Axiom® arrays
- Future work
AFFYMETRIX GENETITAN

- Automated high-throughput solution for monitoring gene expression and genome-wide SNP genotyping
- 16-, 24- and 96-format array plates
- On campus at RRC Core Genomics Facility, College of Medicine West (CMW), Room A-301
One Axiom Genome-Wide PanAFR Plate for 96 samples

~700K SNPs per plate

One Axiom GW (~2.1M) PanAFR Array Set is a set of 3 plates

Also available CEU (European), ASI (Asian), CHB (Chinese Beijing), and custom myDesign plates
POWER TOOLS VS GENOTYPING CONSOLE

Affymetrix Power Tools

Genotyping Console
AFFYMETRIX POWER TOOLS IN GALAXY

- **apt-geno-qc** for quality control
- **apt-probeset-genotype** for genotype calling
- **Write calls in plink format**
- **SNP metrics**

Get Data

Affymetrix Power Tools

- **apt-geno-qc** a single-chip-based genotyping chip quality control tool
- **apt-probeset-genotype** program for determining genotype calls from Affymetrix SNP microarrays
- Write calls in plink format
  - writes PLINK input files from calls, pedigree, annotation and phenotype files
- **SNP metrics** computes Call Rate, FLD, HomFLD, Het Strength Offset, Hom Ratio Offset, Minor Allele count, Number of clusters, AA, AB, BB, and no calls
MINOR CUSTOMIZATIONS

- Registered ".CEL.TAR" file formats for batch uploads
- Modified Galaxy upload tool for better handling custom binary files
- Registered APT ".calls.txt", ".report.txt", "snp-posteriors.txt" file formats
- Registered PLINK formats ".tped", "tmap"
#1. Upload a tarball of .CEL files
#2. Perform QC for each .CEL file
#2. Perform QC for each .CEL file

#3, #4. Filter .CEL files above 0.82 DQC
#5. First run of genotyping
#2. Perform QC for each .CEL file #6, #7.

Filter .CEL files above 97% Call Rate
#8. Second run of genotyping
#9. Calculate different SNP metrics
#10. Upload phenotypes
#11. Output genotypes in PLINK format
FUTURE WORK

- Integrate a few more Affymetrix Power Tools into Galaxy
- Publish the APT tools in Galaxy Tool Shed
- Better support for binary/compressed data files
- Publish the APT genotyping workflow
ACKNOWLEDGEMENTS

- Research Resources Center and Center for Clinical and Translational Science at UIC
- Dr. Damir Herman, Affymetrix
- Morris Chukhman, Tommie Jackson, and the team at the Center for Research Informatics at UIC
- Dr. Rick Kittles, UIC Institute of Human Genetics
- Dr. Zarema Arbieva, UIC Core Genomics Facility