National Institute on Drug Abuse Center for Genetic Studies
Poly Substance Dependence Genetics Initiative, Study 11 Genotypes Dataset 1 (DS1)
30MAY2013

Documentation from Study 11 (23MAY2013)

- data from different generations of Affymetrix chips (from the 10K to 6.0) were used to generate these data; location analogs were created when SNPs were not present on later generation chips; paper that describes the process in detail is in preparation.

- approximately half the genotypes were obtained on one of 3 slightly different Affy 10K chips, with the rest obtained from newer chips as the older ones were no longer available; when a SNP was retired from a given chip, markers that flanked its location that were available from the later generation of chips were used; the idea was to use all of the data from the later chips.

- GRR (Abecasis) was used to find Mendelian errors and to identify potential relationship problems; mean IBD estimates of <.375 were used to differentiate half-sibs from full siblings; sporadic Mendelian errors within families were dealt with by zeroing out the SNP for the family; SNPs with large numbers of Mendelian errors across families were discarded.

- These SNP calls are currently being validated against genotypes generated on Affy 6.0 chips and called through Birdsuite and Beagle.

Processing by NIDA (30MAY2013)

- deposited text files were read into SAS; subject directives were applied to zero out genotypes, leaving the empty record as a placeholder in the pedigree; resulting data written to text
- a column of zeros was added to the fourth column of the map files to represent missing base-pair location to conform to plink format; cM provided.
- each text file read into plink binary files, then concatenated into one binary file with following plink command, where nida_study11_concat_plink.txt contained list of chr2-chr22 plink files:

```
--bfile      nida_study11_chr1
--merge-list nida_study11_concat_plink.txt
--make-bed
--out        nida_poly_study11_ds1
```