Validating and genotyping rare SNPs with the Axiom™ Genotyping Solution

Teresa Webster , Jeremy Gollub, Earl Hubbell, Yiping Zhan, Yontao Lu, Tom Asbury, Li Weng, Matthew Purdy, Michael Shapero

Affymetrix, Inc., 3420 Central Expressway, Santa Clara, CA

Affymetrix® Revolutionize life

ABSTRACT #1109

Purpose

Investigators are beginning to explore the role of rare SNPs to better understand the heritability in complex disease and individual differences in disease susceptibility. Due to the relatively low frequency in the population (<5%) and modest gene effect of individual rare variants, a larger numbers of variants and samples need to be assessed to achieve sufficient power and resolution. This requires high-throughput, cost-effective, and accurate methods for validating and genotyping such SNPs. We present components of such a method that uses the Axiom Genotyping Solution, a high-performance microarray technology from Affymetrix.

Method

- Candidate variants from genomic discovery projects are screened by the Axiom Solution against samples from reference populations used by the discovery project with enough diversity to elicit two to three examples of heterozygous genotype.
- Variants are validated based on observation of two to three examples of the minor allele, as well as resolution and reproducibility of genotypes.
- The "Axiom GT1" genotyping algorithm generates genotypes using a Bayesian procedure in which priors for the SNPs are combined with the data. The results are posterior estimates of genotype cluster centers and variances for the missing minor homozygous genotypes, as well as for the heterozygous and major homozygous genotypes. This enables genotyping of rare SNPs in new data sets.

Results

Rare SNPs collected by this process are included in the Axiom Genomic Database of validated SNPs. Genotypes from the reference populations in the screening data are used to estimate the contribution to coverage of rare SNPs in the genome. The current database covers more than 80% of the target set of rare CEU SNPs (MAF < 5%). The target set includes all HapMap SNPs plus validated SNPs from the 1000 Genomes Project.

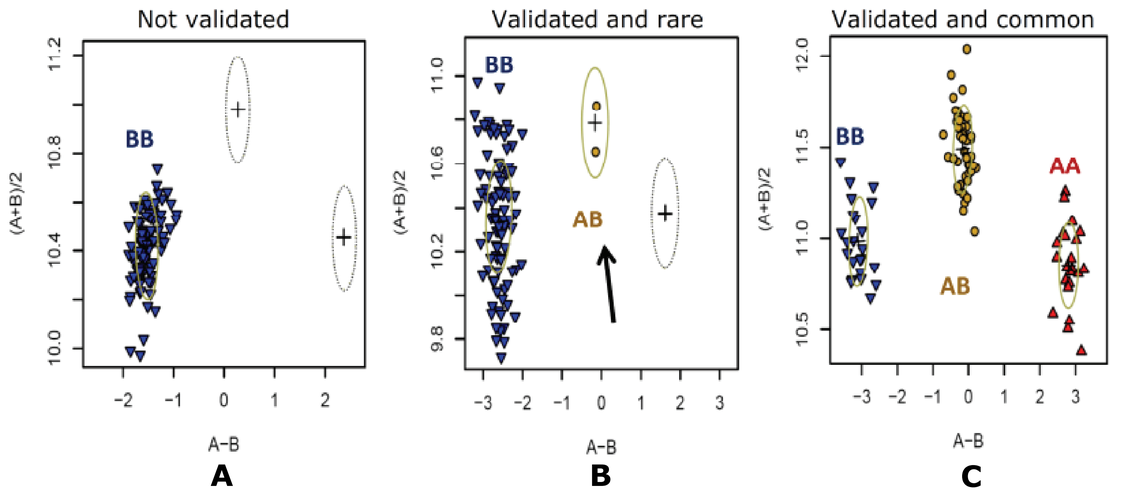
Conclusion

The resulting database is a resource from which multiple SNP panels can be derived, optimizing coverage of rare variants in studies seeking to understand the association of rare SNPs with common diseases.

References

- Axiom™ Genotyping Solution: www.affymetrix.com/axiom
- International HapMap Project: http://hapmap.ncbi.nlm.nih.gov/

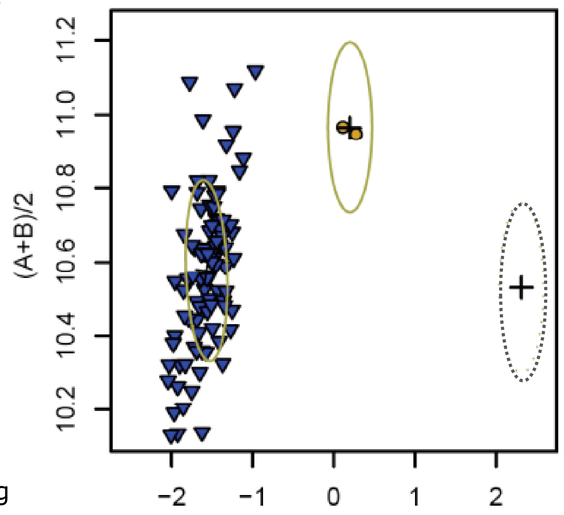
Verify that a candidate SNP is real and characterize minor allele frequency



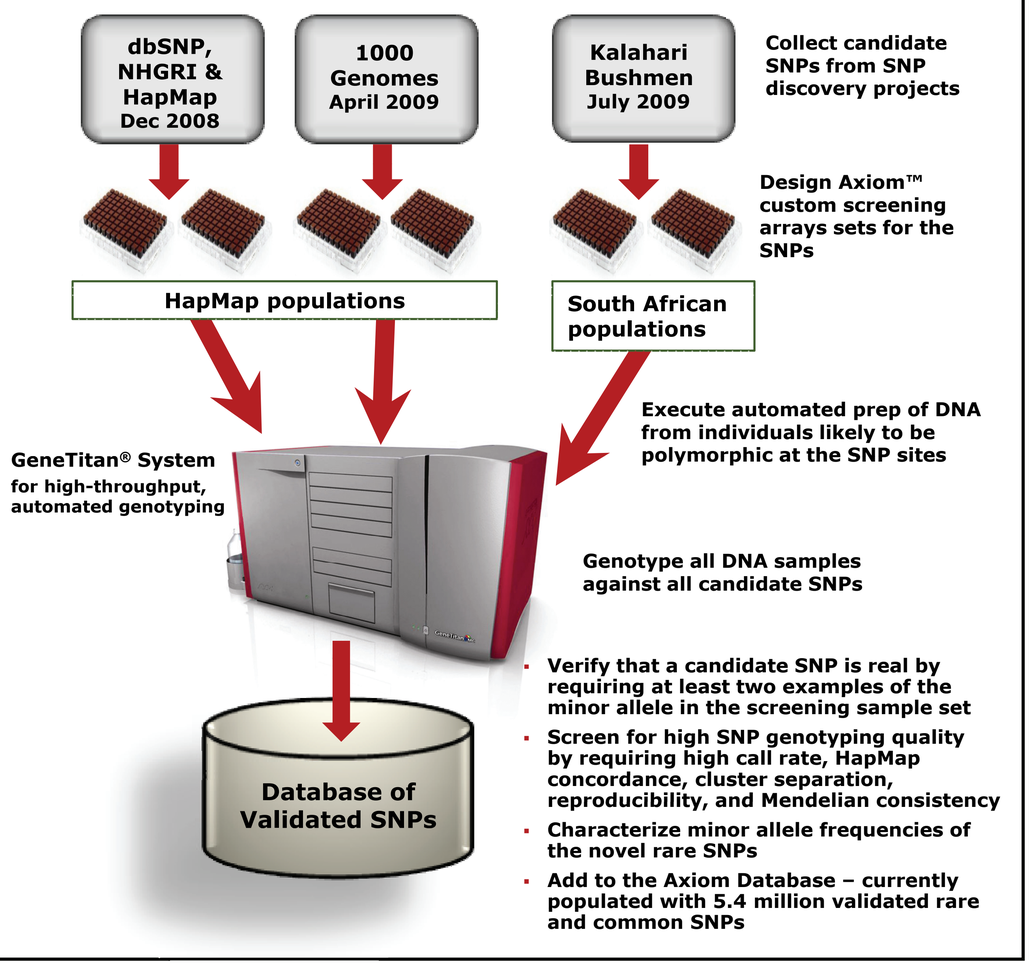
YRI genotype clusters for three candidate SNPs from the 2009 1000 Genomes Project release. SNPs should be polymorphic for 90 YRI individuals based on 1000 Genomes Project genotypes. Cluster separation is in the $\log_2$ (allele signal) difference space (X axis). Each point corresponds to a genotype in the YRI genome, colored according to the Axiom-called genotype (red, gold, blue, and gray correspond to AA, AB, BB, and no call). **A)** The SNP has no examples of the minor allele, so the SNP site is not validated. **B)** The SNP has two examples of the minor allele (the AB cluster identified by the arrow), so it is both valid and rare with YRI MAF = 1% (2 minor alleles/180 chromosomes). **C)** The SNP has many examples of the minor allele, so it is valid and common (MAF > 5%).

The Axiom GT1 algorithm enables genotyping of rare SNPs in new samples

Genotyping rare SNPs. Once a SNP is validated, the Axiom Solution's ability to genotype it even in the absence of the minor allele is very robust, because the genotype model is always trained on actual observed instances of the minor allele. The Axiom GT1 algorithm builds SNP-specific models using a Bayesian procedure in which priors (green ovals) for the SNP genotype clusters are combined with the observed data (blue and gold points) to obtain posterior estimates of genotype cluster centers and variance. The posterior cluster center (+) and variance (dashed oval) are produced for the unpopulated genotype cluster (AA) as well. This estimation is influenced by the prior information, and covariances between all three genotype centers. The SNP-specific models produced by training on the screening data are used as priors to guide the Axiom GT1 genotyping of new sample sets.

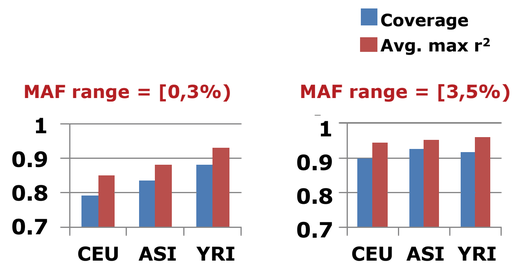


Screen candidates to validate and characterize rare SNPs



- Verify that a candidate SNP is real by requiring at least two examples of the minor allele in the screening sample set
- Screen for high SNP genotyping quality by requiring high call rate, HapMap concordance, cluster separation, reproducibility, and Mendelian consistency
- Characterize minor allele frequencies of the novel rare SNPs
- Add to the Axiom Database – currently populated with 5.4 million validated rare and common SNPs

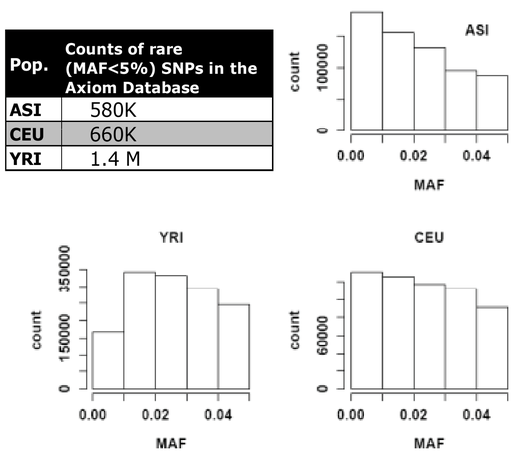
Coverage of rare SNPs in the genome by the Axiom Genomic Database



Coverage of rare SNPs in the genome-wide target set. The genome-wide target set SNPs consist of (1) all HapMap SNPs and (2) SNPs in the 1000 Genomes Project 2009 release that were validated by Axiom screens. The rare SNP target sets are the subsets in the given minor allele frequency (MAF) range. *Coverage* = percentage of target set SNPs tagged by a SNP in the Axiom Database with $r^2 > 0.8$. *Max* r^2 = maximum of r^2 values between a target set SNP and the SNPs in the Axiom Database. *Avg. max* r^2 = average over the *max* r^2 values.

Counts of validated rare SNPs available for association studies using Axiom Genotyping Arrays

Pop.	Counts of rare (MAF<5%) SNPs in the Axiom Database
ASI	580K
CEU	660K
YRI	1.4 M



Counts and distributions of rare SNPs for three populations: (1) ASI (Asian) is a combination of HapMap CHB and JPT, (2) HapMap CEU (Caucasian), (3) HapMap YRI (Yoruba).