Introduction
A major issue for the successful design of disease association studies using SNPs as surrogate markers is the selection of a set effective variants not only to detect association but to simplify the association test. A number of strategies have been proposed for the selection of SNPs based on empirical patterns of LD. Integrating all these criteria can be challenging: the algorithms required are complex to deploy and all the necessary annotations are deposited in heterogeneous databases. To simplify this process, we have developed SNPbrowser™ Software, a tool to assist in the knowledge-based selection of markers for association studies.

Material and Methods
Genotyping data
We genotyped DNA from 45 African-Americans, 45 Caucasians, 45 Chinese, and 45 Japanese, all unrelated individuals. Over 250,000 TaqMan® SNP Genotyping Assays were used to genotype these samples, resulting in up to 160,000 SNPs with reliable genotype calls and with a minimum call-rate of 90%, at least one of these populations (De La Vega et al., 2002). Only SNPs having MAF values > 10% and samples, resulting in up to 160,000 SNPs with reliable genotype calls and with a minimum call rate of 90%, at least one of these populations (De La Vega et al., 2002). Only SNPs having MAF values > 10% and result in a genotype call rate of at least 90% and a marker call rate of > 90% were considered for analysis.

Results
Selection of SNP Markers for Genetic Association Studies of Complex Disease Based on Empirical Knowledge of the Genome-wide Patterns of Linkage Disequilibrium

We developed SNPbrowser™ Software, a visualization suite to facilitate the selection of a set effective variants not only to detect association but to simplify the association test. A number of strategies have been proposed for the selection of SNPs based on empirical patterns of LD. Integrating all these criteria can be challenging: the algorithms required are complex to deploy and all the necessary annotations are deposited in heterogeneous databases. To simplify this process, we have developed SNPbrowser™ Software, a tool to assist in the knowledge-based selection of markers for association studies.

To implement these computationally demanding methods, optimally minimum sets of tag SNPs selected from each chromosome and population (http://cedar.genetics.soton.ac.uk/public_html/helpld.html). For more information on the theory and implementation, see De La Vega, et al., 2002.

A number of filters are available to select subsets of SNPs of interest, including filtering by minor allele frequency as determined in the HapMap project or during genotyping. Power calculations for genetic case/control studies using gene-centric data: De La Vega, et al., 2002.

Conclusions
Finally, once the SNPbrowser™ Software accepts the ideal set of SNPs for an association study, simply clicking on the results bar adds the SNP assays to the researcher’s order list, and one additional click adds the Applied Biosystems’ TaqMan® SNP Genotyping Assays to the selected list. The selected SNP list can also be exported in the same format as the Applied Biosystems’ submission of the TaqMan® Geenotyping System for the ordering of multiplex oligonucleotide ligase-dramphor probes.

Acknowledgements
We are very grateful to Andrew Collins (U. Southampton), for providing the LDMAP software to estimate LD maps and fine mapping studies through the selection of the most informative SNPs for meaningful and powerful study results. The new easy-to-use TaqMan® SNP Genotyping Assays simplifies the selection of SNP marker sets for a ‘tagging’ SNP sets of SNP density enrichment.

To supplement the validated assays, we introduced two new assay collections, the TaqMan® Coding SNP Genotyping Assays and the TaqMan® Pre-Designed SNP Genotyping Assays. A SNP Density Selection Wizard allows the selection of valid SNPs based on empirical patterns of LD. Additional, the SNP density selection can be performed with the physical map (e.g. on the LD map) and/or the physical map (e.g. on the LD map). Additionally, the SNP Wizard allows prioritization of the selection of high-sensitivity supplemental SNPs through user-defined prioritization of criteria, including minor allele frequency as determined in the HapMap project or during genotyping.

We are very grateful to Andrew Collins (U. Southampton), for providing the LDMAP software to estimate LD maps and fine mapping studies through the selection of the most informative SNPs for meaningful and powerful study results. The new easy-to-use TaqMan® SNP Genotyping Assays simplifies the selection of SNP marker sets for a ‘tagging’ SNP sets of SNP density enrichment.

Finally, once the SNPbrowser™ Software accepts the ideal set of SNPs for an association study, simply clicking on the results bar adds the SNP assays to the researcher’s order list, and one additional click adds the Applied Biosystems’ TaqMan® SNP Genotyping Assays to the selected list. The selected SNP list can also be exported in the same format as the Applied Biosystems’ submission of the TaqMan® Geenotyping System for the ordering of multiplex oligonucleotide ligase-dramphor probes.
**Selection of SNP Markers for Genetic Association Studies of Complex Disease Based on Empirical Knowledge of the Genome-wide Patterns of Linkage Disequilibrium**

Francisco M. De La Vega, Hadar Isaac, Charles R. Scafe, Bjarni V. Halldórsson, and Eugene Spier

Applied Biosystems, 850 Lincoln Centre Dr., Foster City, CA 94404, USA

---

**Introduction**

A major issue for the successful design of disease association studies using SNPs as surrogate markers is the selection of a cost-effective subset of the available variants that would provide high statistical power to detect the association. A number of strategies have been proposed for the selection of SNPs based on empirical patterns of LD. Integrating all these criteria and methods can be challenging: the algorithms required are complex to deploy and all the necessary annotations are deposited in heterogeneous databases. To simplify this process, we have developed SNPbrowser™ Software, a tool to assist in the knowledge-based selection of markers for association studies.

---

**Material and Methods**

**Genotype data**

We genotyped DNAs from 45 African-Americans, 46 Caucasians, 45 Chinese, and 45 Japanese, all unrelated individuals. Over 250,000 TaqMan® SNP Genotyping Assays were used to genotype these samples, resulting in up to 100,000 SNPs with reliable genotype calls and with a minimum call rate of 80%, in at least one of these populations (De La Vega et al., 2002). Only SNPs having MAP values > 10% and that have passed Hardy Weinberg Equilibrium test at a p-value < 0.05 were considered for analysis.

---

**Results**

**SNPbrowser™ Software Visualization and Query Tools**

The visualization panel consists of a chromosome map viewer representing the location in the physical map of up to 100,000 validated, and optionally up to 2 million Pre-designed, TaqMan® SNPs Genotyping Assays, and their relationship to human genes and their exons. The display differentiates between coding SNPs, which can be prioritized in some studies, and Unlabeled and Pre-designed assays. SNPbrowser™ Software shows the location of the SNPs on the physical (0) and linkage disequilibrium map, while horizontally lines indicate the spans of blocks of high LD determined by two algorithms. Applied Biosystems’ metric LD map was experimentally generated from over 20 million genotypes determined in four major populations and provides information on how to best position SNPs across the genes or regions of interest in a study. A metric LD map, expressed in LD units (LDU), places SNPs on a coordinate system where distances between SNPs are directly related to the degree of LD between them. For example, SNPs in perfect LD (completely correlated) have zero distance between them, whereas SNPs with no significant correlation are separated by over three LDUs in this map. Analogous to the genetic map expressed in centi-Morgans used for selecting markers for linkage studies in families, the LD map can be used to efficiently select markers for population-based disease association studies with high statistical power. Details of SNP allele frequency on the four populations, and other annotations, can be easily visualized. A number of filters are available to select subsets of SNPs of interest, including filtering by minor allele frequency in one or more the populations. SNPs and genes can be easily located by searching by a number of keywords, including SNP gene, and transcript identifiers, as well as assembly coordinates. A powerful batch search feature provided in the software is very useful when large candidate gene lists are being considered for SNP selection—a click on the search result window immediately opens the genome viewer and zooms the viewer on the region of interest. Also, previously visited locations can be remembered through bookmarks for easy access in subsequent sessions.

---

**Optimizing Marker Coverage**

To supplement the validated assays, we introduced two new assay collections, the TaqMan® Coding SNP Genotyping Assays and the TaqMan® Expressed SNP Genotyping Assays. A SNP Density Selection function in the SNP Wizard allows supplementing validated assays with additional SNPs when their density is not sufficient for the coverage requirements or to select SNPs in a specific DNA pattern. The SNP density selection can be performed with the physical map (bp) or the LD map (LDUs). Additionally, the SNP Wizard allows prioritization of the selection of high confidence supplemental SNPs through user-defined prioritization of criteria, including: minor allele frequency as determined in the Hammap project or during the Applied Genomics Initiative resequencing efforts, and independent evidence of discovery of a SNP from multiple data sources (so-called double-hit SNPs).

---

**Conclusions**

We developed SNPbrowser™ Software, a visualization suite to facilitate the selection of assays for genetic association studies utilizing prior empirical knowledge of the profile of linkage disequilibrium across genomic regions. It includes a library of almost two million human SNP assays, facilitating the knowledge-driven design of association and fine mapping studies through the selection of the most informative SNPs for meaningful and powerful study results. The new easy-to-use SNP Wizard simplifies the selection of SNP marker sets with either a “tagging” SNP or SNP density selection workflow. Finally, the software provides an easy interface for ordering ready-to-use TaqMan® SNP Genotyping Assays or submitting a list of SNPs for the design of custom SNPs™ Genotyping System multiplex CLA assays.

---

**Acknowledgements**

We are very grateful to Andrew Collins (U. Southampton), for providing the LDMAP software to estimate LDUs, and to Dave Gordon (Rockefeller U.), who provided custom software to calculate power. We acknowledge the valuable feedback and support provided by Katherine Rogers, Anthony D'Ors, Rachel Reardon, Fentim Williams, Michael Rhodes, Timitne Harbsorne, Tresor Woolage, Kelly McDonald, Raymond Feilner and the SQA team.

For Research Use Only—Not for use in diagnostic procedures.

Applied Biosystems is a registered trademark. Applied, Applied Biosystems, and TaqMan are trademarks of Applied Corporation or its subsidiaries in the US and/or certain other countries.

TaqMan is a registered trademark of Roche Molecular Systems, Inc.

© 2004 Applied Biosystems. All rights reserved.