



Published in final edited form as:

*IEEE Symp Comput Intell Bioinforma Comput Biol Proc.* 2009 March 30; 2009: 298–305. doi:10.1109/CIBCB.2009.4925742.

## Data Synthesis and Tool Development for Exploring Imaging Genomic Patterns

Sungeun Kim, Li Shen, Andrew J. Saykin, and John D. West

IU Center for Neuroimaging at the Department of Radiology and the Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN

### Abstract

Recent advances in brain imaging and high throughput genotyping techniques enable new approaches to study the influence of genetic variation on brain structure and function. However, major computational challenges are bottlenecks for comprehensive joint analysis of these high-dimensional image and genomic data. We report our initial progress in developing an imaging genomic browsing system for integrated exploration of neuroimaging and genomic data. We describe a method for synthesizing a set of realistic neuroimaging and genomic data, where the relationships between imaging phenotypes and genotypes are known. This data set is used to demonstrate the functionality of our system, which is designed for effectively exploring the neuroanatomical distribution of statistical results that measure the associations between brain imaging phenotypes and genotypes on a genome-wide scale. The proposed system has substantial potential for enabling discovery of important imaging genomic associations through visual evaluation and can be extended towards several directions.

### I. Introduction

Single nucleotide polymorphisms (SNPs) are DNA sequence variations in which a single base pair is altered. Many SNPs have no effect on cell function, but others predispose individuals to disease or influence their response to a drug. SNPs in single genes have been correlated to neuroimaging measures in brain disorders (e.g., [1], [2]). However, the analytic tools that relate one gene to one scan modality are insufficient to provide insight into the multiple mechanisms and imaging manifestations of brain disorders such as Alzheimer's disease. Thanks to the technological advance in acquiring high throughput multi-modal brain imaging and genetic array data, imaging genomics [3], [4], [5], [6], [7] has recently emerged as a transdisciplinary research field where new strategies are examined to evaluate genetic variation using imaging measures as phenotypes for exploring genetic effects on brain structure and function.

Although genome-wide association studies have been actively performed [8], [9], [10], [11], it remains a highly challenging issue to effectively relate high throughput genotyping data to large scale image data. As pointed out by Glahn et. al. [4], in present imaging genomics studies, researchers usually reduce the image data to a small number of variables (e.g., nine imaging measures used in [5]) or focus on a single SNP or gene (e.g., [3], [6]), to limit the number of statistical tests, control Type I error, and make computation feasible. However, significant reduction in one or both data types greatly limits our capability of identifying

important relationships. Therefore, there is a critical need for new computational strategies that allow for genome-wide and whole brain analyses to facilitate effective pattern discovery from these rich data sources. This is a very challenging problem because both image and genomic data are complex, large-scale and high dimensional.

To address the above pressing issue, we report our work in progress on developing an imaging genomic browsing system, where we apply visualization techniques in the integrated exploration of neuroimaging and genomic data. Visualization has been independently used in both imaging (e.g., to help image-guided diagnosis, surgery and therapy [12]) and genomics (e.g., for identifying biologically meaningful patterns in genomic data [13], [14]). Our goals of having a visual exploratory system in imaging genomics are threefold: (1) to help users better understand the image and genomic data and examine their relationships via visualization and human computer interaction, (2) to navigate and identify potential patterns for further investigation with refined statistical modeling, and (3) to provide a user-friendly visual evaluation scheme for validating results generated by other imaging genomic studies.

In this paper, we present our initial efforts on creating such an imaging genomic browsing system as well as synthesizing a set of realistic image and genomic data for demonstrating the functionality of the system. We currently focus only on structural magnetic resonance imaging (MRI) scans and SNP genotyping arrays. As a future direction, we plan to extend our system to include additional imaging modalities (e.g. functional MRI (fMRI), diffusion tensor imaging (DTI), and positron emission tomography (PET)) as well as more genetic features (e.g., expression arrays, copy number variations). The rest of the paper is organized as follows. Section II describes our method for imaging genomic data synthesis. Section III presents the design of our imaging genomic browser. Section IV demonstrates our experimental results and provides related discussions. Section V concludes this paper.

## II. Data Synthesis

Since imaging genomics is a relatively new field, there is not much public genomic and image data available for conducting studies in this area. Although there are data available from separate genomics and imaging studies on the same diseases, their subject pools are different and so it is hard to merge the data for the purpose of relating imaging measures to genomic features. In order to demonstrate our system, we first describe our method for synthesizing a set of realistic image and genomic data. We make the following two requirements for these synthetic genomic and image data: (1) they should resemble the real data and so look realistic; (2) all the relationships between the genomic and image data are known and can be used to validate the results of any subsequent imaging genomic analysis. By realistic, we mean that several important characteristics of the synthetic data should be similar to those of the real data (e.g., the minor allele frequency and the Hardy-Weinberg Equilibrium test score for genomic data, the range of gray matter density at each voxel location for image data). We describe how the synthetic genomic and image data were created in the following two subsections respectively.

### A. Genomic data

The NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) [15] is an ongoing five-year public-private partnership to test whether serial MRI, PET, SNP, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's Disease (AD). The ADNI Genetics Working group is presently completing a genotyping project of the ADNI cohort using the Illumina Human610-Quad BeadChips [16]. We expect to demonstrate our imaging

genomic tool using the ADNI data in the near future, since the de-identified ADNI data will be made publicly available to the scientific community. With this anticipation, we create our synthetic genomic data based on a subset of SNPs available on the Human610-Quad BeadChips.

Since this is an initial attempt of building an imaging genomic browsing system, we decided to start with a relatively simple and moderately sized data set. Thus, we can first focus on developing key features of the system. Once the basic system is available, we can then address the scalability issue in future extensions. With this consideration, we decided to synthesize a data set of 300 healthy controls, where each subject contains 3000 genomic features (i.e., SNPs) and a  $182 \times 218 \times 182$  gray matter density map.

To determine 3000 genomic features, we use the minor allele frequency (MAF) file contained in the Illumina Human610-Quad V1 product package. This file contains 592532 SNPs along with their corresponding MAFs as based on the validation panel (51 Utah residents with ancestry from northern and western Europe (CEU), 49 Yoruba in Ibadan, Nigeria (YRI) and 75 Han Chinese in Beijing, China (CHB) + Japanese in Tokyo, Japan (JPT) individuals respectively). In this work, we only use the MAF information associated with 51 CEU individuals. Based on all these SNPs, we randomly select 3000 SNPs with CEU MAF values  $\geq 0.1$  due to the small size of imaging genomic samples. After the selection, genomic information of these sampled SNPs such as chromosome names, positions, RefSNP numbers, and gene symbols are extracted from the Human610-Quad gene annotation file and are included in the synthetic genomic data as genomic annotation information. Shown in Fig. 1 are chromosome names and positions of these chosen SNPs in a Karyo view and this figure is created by using Karyoview on Ensembl web site (<http://www.ensembl.org>).

To assign genotype values to 300 subjects, we use the CEU MAF information associated with each SNP. Given a SNP locus, let  $A$  be the major allele,  $B$  be the minor allele,  $q$  be the MAF, and  $p = 1 - q$ . Based on the CEU MAF information,  $q$  is known. Thus, for this SNP, 300 genotype values ( $AA$ ,  $AB$ , or  $BB$ ) are created in the order of  $p^2$ ,  $2pq$ , and  $q^2$  to satisfy the Hardy-Weinberg Equilibrium [17], i.e., the frequencies of  $AA$ ,  $AB$ , and  $BB$  are close to  $p^2$ ,  $2pq$ , and  $q^2$ , respectively. Then these values are randomly permuted and assigned to the 300 healthy control subjects. Shown in Fig. 2 is the distribution of genotype frequencies of all 3000 SNPs and clearly Hardy-Weinberg Equilibrium holds in our synthetic genomic data. Note that linkage disequilibrium [8] is not modeled in the current data but will be considered in our future extensions.

## B. Image data

As mentioned before, in this work, we focus only on structural MRI data and genomic SNP data. The prototype implementation of our imaging genomic browsing system aims to localize structural changes related to genetic conditions. We employ a widely used voxel-based morphometry (VBM) method [18] to extract a gray matter (GM) density map for each subject and use that to define imaging phenotypes or variables. Below, we first describe how to apply VBM to a set of real MRI data and extract GM maps. Then, we explain how to use these real GM maps to create synthetic GM maps, where we can embed known imaging genomic patterns into these realistic data.

**1) Creating real GM maps**—Our image data synthesis procedure starts from a real data cohort, which consists of 39 healthy controls (HC, age  $70.6 \pm 5.1$ , gender 12M and 27F) from an existing mild cognitive impairment (MCI) study [19]. Note that we only use the controls in this MCI study. Volumetric structural MRI data are acquired on a 1.5 Tesla

General Electric (GE) LX Horizon scanner using a T1-weighted Spoiled Gradient Recalled (SPGR) coronal series with 1.5 mm slice thickness. Further details about image data acquisition are available in [19]. To illustrate the functionality of our imaging genomic system, we employ a VBM approach for examining the relationships between image regions and genotype variables. As a future plan, we will investigate alternative statistical and computational methods for identifying imaging genomic patterns and these new results can be easily incorporated into the existing framework.

VBM analysis [18], [20] involves a voxel-by-voxel statistical analysis of the local concentration of gray or white matter, and can be applied to compare two or more groups of subjects (e.g., genotype groups in our case). Major steps in VBM analysis include: (1) spatially normalize all volumes to the same stereotactic space; (2) segment to obtain gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and other tissue maps; (3) multiply the tissue maps by their relative volume before and after warping as an optional step; (4) smooth normalized segmented images; (5) perform voxel-based statistical parametric tests to contrast groups on GM or other maps.

We use the SPM5 software package [21] to preprocess our image data and created normalized GM maps for our imaging genomic study. The initial GM maps are extracted by segmenting the T1-weighted SPGR volumes after resampling them to  $1\text{ mm}^3$  isotropic voxels. A 12-parameter model is used to spatially normalize the GM maps to the GM prior probability template. The normalized scans are smoothed using an isotropic spatial filter with full width half maximum of  $10\text{mm}$  to help increase signal-to-noise ratio and account for individual differences in gyral anatomy. The smoothed normalized GM maps are used as our imaging phenotypes in the subsequent analyses, where each voxel location corresponds to an imaging variable. Shown in Fig. 3 are sample GM maps in three orthogonal views: the normalized GM maps are shown on top, and the smoothed normalized ones are shown on bottom.

**2) Creating synthetic GM maps**—Based on the 39 real GM maps mentioned above, we create realistic GM maps for 300 synthetic subjects using the following approach. Sample mean  $\mu$  and sample standard deviation  $\sigma$  at each voxel location are computed from 39 real subjects (Fig. 4 shows the mean GM map). The GM values at a voxel location for our synthetic GM maps are then modeled as a normal distribution of  $N(\mu, \sigma)$  using  $\mu$  and  $\sigma$  computed from the real data. We draw 300 GM values for each voxel location based on its corresponding normal distribution and thus generate 300 synthetic volumetric GM maps. These synthetic GM maps represent the imaging phenotypes of our subjects.

The resultant GM maps look discrete due to voxel-by-voxel generation procedure with different sample means and different sample standard deviations although these sample means and sample standard deviations are not completely independent to one another. Therefore, in order to increase the spatial correlations among neighboring voxels and further resemble real image data, the generated volumes are smoothed using an isotropic spatial filter with full width half maximum of  $4\text{mm}$ .

Although the smoothed volumes look similar to real data, there is no association established between these imaging variables (i.e., voxels on the GM maps) and genomic variables created previously. Therefore, we use the following procedure to further alter the GM maps in order to embed some imaging genomic patterns into the data. In the image space, we select four regions of interest (ROIs): thalamus, hippocampus, amygdala, and insula. We associate each ROI with a particular SNP in the genomic space using a particular model. Table I shows these association models together with the related ROIs and SNP indices. We

employ four association models: heterozygotic, additive, dominant, and recessive. Table II shows the relationships between genomic groups (AA, AB, BB) and phenotype groups ( $P_+$ ,  $P_0$ ,  $P_-$ ) for each association model.

Now our task is to modify the GM density value for each voxel within a chosen ROI so that it falls into a particular phenotype group ( $P_+$ ,  $P_0$ , or  $P_-$ ). For  $P_+$  voxels, we increase their values. For  $P_-$  voxels, we decrease their values. For  $P_0$  voxels, we do not change their values. We also define a modulation function  $f(x)$  for each voxel  $x$  in a given ROI to modulate the amount of the value change. In practice, the ROI difference between groups often happens first around its boundary. With this observation, we define  $f(x)$  as the square of the normalized Euclidean distance of the voxel  $x$  from the center of mass of the ROI. Fig. 5 shows a visual representation of this function, from which we see that it can increase or decrease the GM density of the voxel up to 10 percent of the original density. Based on the shape of this function, boundary voxels tend to have more substantial density adjustments than voxels around the center. The details of the modulation procedure is described in Algorithm 1.

### Algorithm 1

Modulation of voxel values for an ROI  $r$ .

- 
- 1 Let  $a(r)$  be the association model assigned for ROI  $r$  in Table I and  $l(r)$  be the corresponding SNP locus
  - 2 Let  $GT(s, l(r))$  be the genotype of subject  $s$  at SNP locus  $l(r)$
  - 3 **for** each subject  $s$  **do**
  - 4 Calculate  $PT(s, r)$ , the phenotype group of its ROI  $r$ , using  $GT(s, l(r))$ ,  $a(r)$  and Table II
  - 5 Let  $m$  be the total number of subjects whose ROI  $r$ 's are in phenotype group  $P_+$ , i.e.,  $m = |\{s | PT(s, r) = P_+\}|$
  - 6 Let  $n$  be the total number of subjects whose ROI  $r$ 's are in phenotype group  $P_-$ , i.e.,  $n = |\{s | PT(s, r) = P_-\}|$
  - 7 Let  $f(x)$  be the modulation function defined in Fig. 5.
  - 8 **for** each voxel  $x$  in the ROI  $r$  of each subject  $s$  **do**
  - 9 Let  $I_{orig}(x)$  be the original voxel value
  - 10 **if**  $PT(s, r) = P_+$  **then**
  - 11 
$$I_{new}(x) = I_{orig}(x) * \left(1 + \frac{n}{m} * f(x)\right)$$
  - 12 **else if**  $PT(s, r) = P_-$  **then**
  - 13 
$$I_{new}(x) = I_{orig}(x) * (1 - f(x))$$
  - 14 Let  $I_{new}(x)$  be the new voxel value after modulation
- 

To relate genotype data with image data, one of four different association models (additive, dominant, recessive, or heterozygotic) is applied to one of four ROIs, as described in Table I. Depending on the association model applied to an ROI and the genotype of the subject at the corresponding SNP locus, the phenotype group of the ROI can be determined (Lines 3-4) and the GM density of each voxel within the ROI can be increased or decreased accordingly

(Lines 5-14). Note that the factor of  $\frac{n}{m}$  in Line 11 is used to maintain the overall mean density as similar to the original mean density as possible. This modulation function simulates a case when GM densities of boundary voxels are more affected by a genomic variation than interior voxels. As a result, voxels, located on the boundary of each ROI, are more affected than interior voxels. Please refer to Section IV for results of sample synthetic images as well as how to use our browsing system to reveal these embedded imaging genomic patterns.

### III. Imaging genomic browser

We briefly report our efforts on developing an imaging genomic browsing system. Our system is designed for effectively exploring the neuroanatomical distribution of statistical results that measure the associations between brain imaging phenotypes and genotypes on a genome-wide scale. In our initial implementation, one-way analysis of variance (ANOVA) is employed to examine the relationships between imaging phenotypes (i.e., GM densities) and genotypes (i.e., SNP values). ANOVA is a statistical procedure to test the equality of three or more means at one time by using variances. For convenience, we use  $k$ ,  $v$ , and  $p$  to denote the subject index, the voxel index, and the SNP index, respectively. Thus, given subject  $k$ , let  $D_{v,k}$  be its GM density at voxel  $v$ , and let  $G_{p,k}$  be its genotype value for SNP  $p$ . We run ANOVA on each possible combination of  $(v, p)$ 's (i.e., a voxel location and a SNP site) to see if there is an association between the imaging phenotype and genotype.

A user-friendly and dynamic graphical interface (Fig. 6) is developed to enable users to conveniently navigate the data in both the imaging domain and the genomic domain. A brain atlas is typically loaded as an underlay image and shown in the neurological display convention. The statistical map (e.g., F values from ANOVA) can be shown as an overlay image on the brain atlas and the transparency and the color map are adjustable. Two sliders are provided for basic browsing purpose: (1) the image slider is used to browse over the entire image space for a fixed SNP location, and (2) the genomic slider is used to browse along the genomic dimension for a fixed image slice.

Several flexible mechanisms are implemented to help users efficiently navigate and explore potential patterns between imaging and genomic measurements to localize regions of interest for further investigation with refined statistical modeling. For example, users can adjust the threshold to show only significant regions, and the 'prev' and 'next' buttons allow users to ignore all the slices without significant regions and directly jump to the slice with significant regions. To browse in the image domain, users can select any of the three orthogonal views or select a set of ROIs to concentrate on some subregions of the image. A set of ROIs, constructed from the anatomical labels of Tzourio-Mazoyer et al. [22] is provided and users can load user-customized ROIs to further localize statistical maps. To browse in the genomic domain, users can sort the SNPs by their indices, RefSNP numbers, or chromosome positions, or define a set of genes of interest to narrow down the browsing space. Users can also create a movie based on a specified browsing task.

This prototype system is implemented in Matlab (2006a, The MathWorks, Natick, MA), running on a single computer. In order to improve the computational efficiency in ANOVA calculation, all voxels outside gray and white matters are excluded by using a binary brain mask, created from T1-weighted single-subject anatomical template (see Fig. 7). In the next section, we report the performance of the system as well as some initial results of applying the system to synthetic imaging genomic data.

## IV. Results

### A. Data synthesis

We first report data synthesis results. For synthetic genomic data, as mentioned earlier, we selected 3000 SNPs based on the minor allele frequency (MAF) file contained in the Illumina Human610-Quad V1 product package, and synthesized 300 subjects with these SNPs. Fig. 1 shows the Karyoview of these SNPs and Fig. 2 shows the genotype frequencies for each selected SNP in our 300 synthetic subjects, satisfying Hardy-Weinberg Equilibrium.

For synthetic image data, we synthesized gray matter (GM) density maps for our 300 synthetic subjects based on a real structural MRI data set containing 39 healthy controls. Shown in Fig. 8 and Fig. 9 are sample GM maps in three orthogonal views for five real subjects and for five synthetic subjects respectively. Overall distribution of GM densities between these two groups of subjects look similar to each other, but Fig. 9 exhibits less variation in the details of brain structure than Fig. 8. All panels in Fig. 9 look similar to Fig. 4 and this is due to the method of image synthesis.

## B. Performance of the imaging genomic browser

To demonstrate the efficiency of our imaging genomic browser, we ran the browser on our synthetic image and genomic data (i.e., 300 subjects, 3000 SNPs,  $182 \times 218 \times 182$  voxels) and measured the computation time of one-way ANOVA. A notebook PC with Intel Core2Duo CPU of 2.53 GHz clock speed and 4 GB RAM was used, running Windows XP Professional 64-bit edition and Matlab R2007B without the multithreading function. Fig. 10 shows an averaged result of the test. The measurement of the computation time starts before image data are loaded into memory and ends after a computed f-map is rendered on top of an underlay and this measurement is repeated 50 times. The imaging genomic browser uses a binary brain mask (Fig. 7) to accelerate the computation. In Fig. 10, the top panel plots the number of voxels inside the brain mask on a slice against the slice number, and the bottom panel shows a linear increase in the computation time as the number of voxels within each slice increases.

In the experiment of using our synthetic image data, the mean number of voxels is around  $1.5 \times 10^4$  (Standard Deviation (SD):  $1.1 \times 10^4$ ) in an axial slice, around  $1.3 \times 10^4$  (SD:  $0.9 \times 10^4$ ) in a coronal slice, and around  $1.5 \times 10^4$  (SD:  $1.1 \times 10^4$ ) in a sagittal slice. Therefore, a computation time per one slice is expected to be between 1 and 1.5 sec (for 300 subjects) in a computational environment similar to the testing one. This performance is acceptable to explore moderately sized data sets such as our synthetic image and genomic data. However, it is not fast enough for a very large-scale and complex data, for example, the ADNI cohort containing 800 subjects each with 600K SNPs. Therefore, one future direction is to improve the computational performance to enable users to efficiently explore very large-scale image and genomic data.

## C. Application of the imaging genomic browser

In the synthetic image data, four regions of interest are associated with four SNPs (see Table I) and the imaging genomic browser easily captures these relationships between regions of interest and SNPs from f-maps of one-way ANOVA. Shown in Fig. 11 and Fig. 12 are f-maps rendered on several axial slices and on sagittal slices, respectively. These figures clearly reveal relationships between four regions of interest and SNPs are quickly identified by users. All f-maps are color-coded using the colormap drawn by each panel of the figure.

For further review of these results, movie clips exhibiting association maps of several SNPs with the synthetic image data are created from the imaging genomic browser and linked at <http://imaging.indyrad.iupui.edu/testwiki/index.php/IGB>. We plan to document and package our software, create more synthetic genomic and image data using our method or existing ones such as GenomeSimla [23], and release them on the same website. The released package will include utility functions for importing real image and genomic data.

## V. Conclusions

In this paper, initial efforts toward developing an imaging genomic browsing system are presented. This includes a data synthesis method to create a set of realistic image and genomic data, the demonstration of fundamental features of this initial system, and its performance measurement. Encouraging initial experimental results suggest that the proposed system has substantial potential for enabling discovery of important imaging genomic associations through visual evaluation. This can be expected to yield an enhanced understanding of the topography of polygenetic influences on brain structure and function in normal individuals and patients with neurological and psychiatric disorders.

This browsing system currently focuses only on structural magnetic resonance imaging (MRI) scans and SNP genotyping arrays and performs one-way ANOVA. The system can be extended towards many future directions: (1) include additional imaging modalities (e.g., functional MRI, diffusion tensor imaging, and positron emission tomography), (2) include additional genomic features (e.g., expression arrays, copy number variations), and (3) incorporate more statistical methods that are appropriate for finding potential relationships between phenotype and genotype data. In addition, we plan to improve the computational performance of this browsing system by employing general-purpose distributed computing systems. This improvement of computational efficiency can help users explore very large-scale genomic and image data to identify potential patterns between them. We also expect to demonstrate this system using real imaging genomic data such as ADNI [15] when available.

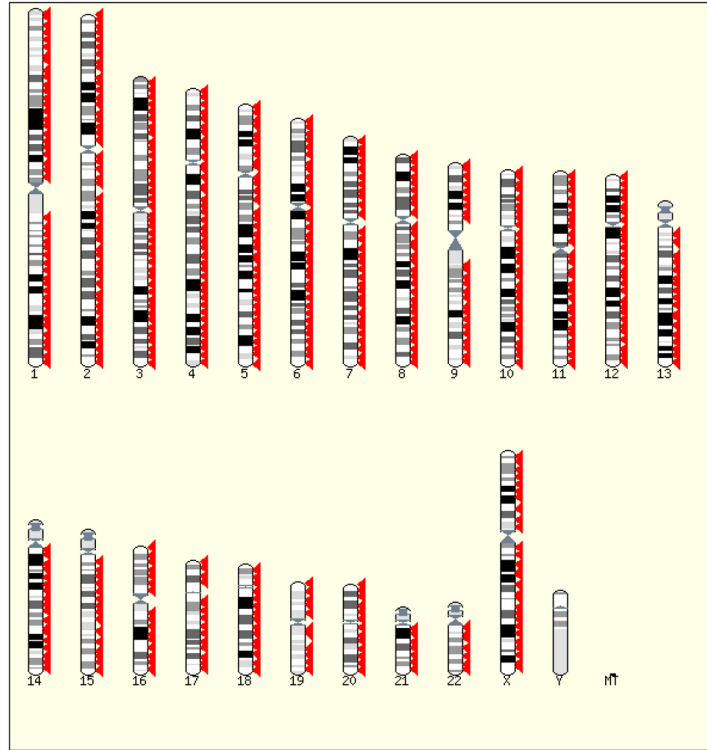
## Acknowledgments

This work was supported in part by NIBIB R03 EB008674, NIA R01 AG19771, NCI R01 CA101318 and U54 EB005149 from the NIH, Foundation for the NIH, and grant #87884 from the Indiana Economic Development Corporation (IEDC).

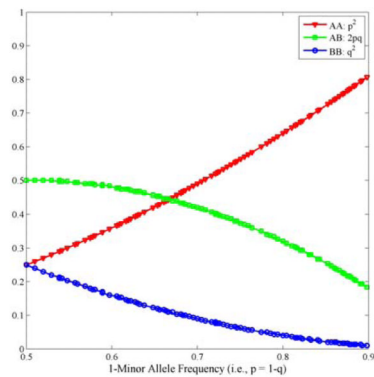
## References

1. Lind J, Larsson A, Persson J, Ingvar M, Nilsson LG, Bckman L, Adolfsson R, Cruts M, Sleegers K, Van Broeckhoven C, Nyberg L. Reduced hippocampal volume in non-demented carriers of the apolipoprotein E epsilon4: relation to chronological age and recognition memory. *Neurosci Lett*. 2006; 396:23–7. [PubMed: 16406347]
2. Wishart HA, Saykin AJ, Rabin LA, Santulli RB, Flashman LA, Guerin S, Mamourian AC, Belloni D, Rhodes CH, McAllister TW. Increased prefrontal activation during working memory in cognitively intact APOE E4 carriers. *American Journal of Psychiatry*. Sep.2006 163:1603–1610. [PubMed: 16946187]
3. Ahmad RH, Emily MD, Daniel RW. Imaging genetics: Perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biological Psychiatry*. 2006; 59:888–897. [PubMed: 16442081]
4. Glahn DC, Thompson PM, Blangero J. Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping*. 2007; 28:488–501. [PubMed: 17440953]
5. Seshadri S, DeStefano A, Au R, Massaro J, Beiser A, Kelly-Hayes M, Kase C, D'Agostino R, DeCarli C, Atwood L, Wolf P. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the framingham study. *BMC Medical Genetics*. 2007; 8:S15. [PubMed: 17903297]
6. Shen, L.; Saykin, AJ.; Chung, MK.; Huang, H. Morphometric analysis of hippocampal shape in mild cognitive impairment: An imaging genetics study; *IEEE 7th International Symposium on Bioinformatics & Bioengineering*; Oct. 2007; p. 211-217.

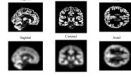
7. Viding E, Williamson DE, Hariri AR. Developmental imaging genetics: Challenges and promises for translational research. *Development and Psychopathology*. 2006; 18:877–892. [PubMed: 17152405]
8. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006; 7:781–91. [PubMed: 16983374]
9. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005; 6:95–108. [PubMed: 15716906]
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]
11. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. *Nat Protoc*. 2007; 2:2492–501. [PubMed: 17947991]
12. NAMIC. 3D slicer web page. Available at <http://www.slicer.org>, accessed on March 31, 2008
13. Reif, DM.; Dudek, SM.; Shaffer, CM.; Wang, J.; Moore, JH. Exploratory visual analysis of pharmacogenomic results; Pacific Symposium on Biocomputing. 2005. p. 296-307.
14. Reif DM, Israel MA, Moore JH. Exploratory visual analysis of statistical results from microarray experiments comparing high and low grade glioma. *Cancer Informatics*. 2007; 2:19–24. [PubMed: 19390666]
15. ADNI. The alzheimer's disease neuroimaging initiative (ADNI). Available at <http://www.adni-info.org/>, accessed on November 10, 2008
16. Illumina. Whole-genome genotyping & cnv analysis: human610-quad beadchip. Available at <http://www.illumina.com/pages.ilmn?ID=248>, accessed on November 10, 2008
17. Stern C. The Hardy-Weinberg law. *Science*. 1943; 97:137–138. [PubMed: 17788516]
18. Ashburner J, Friston K. Voxel-based morphometry-the methods. *NeuroImage*. 2000; 11:805–821. [PubMed: 10860804]
19. Saykin AJ, Wishart HA, Rabin LA, et al. Older adults with cognitive complaints show brain atrophy similar to that of amnesic MCI. *Neurology*. 2006; 67:834–842. [PubMed: 16966547]
20. Ashburner J, Friston K. Why voxel-based morphometry should be used. *NeuroImage*. 2001; 14:1238–1243. [PubMed: 11707080]
21. Wellcome Department of Imaging Neuroscience. The statistical parametric mapping software package. London, UK: Available at <http://www.fil.ion.ucl.ac.uk/spm/>, accessed on March 31, 2008
22. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labelling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*. 2002; 15:273–289. [PubMed: 11771995]
23. Ritchie, MD. genomeSIMLA. Available at <http://chgr.mc.vanderbilt.edu/genomeSIMLA>, accessed on December 31, 2008



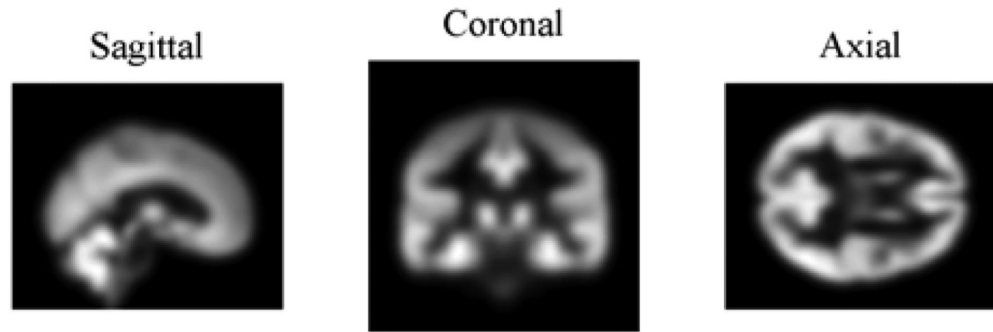
**Fig. 1.** Karyoview of selected SNPs, created by using Karyoview on <http://www.ensembl.org>.



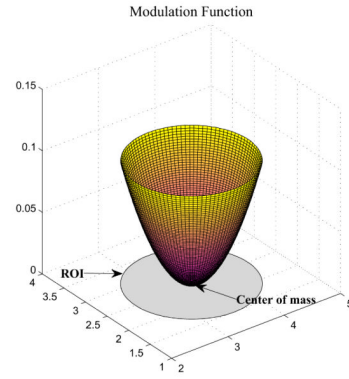
**Fig. 2.** Hardy-Weinberg Equilibrium holds for synthetic SNP data, where genotype frequencies are shown for each selected SNP on the plot.



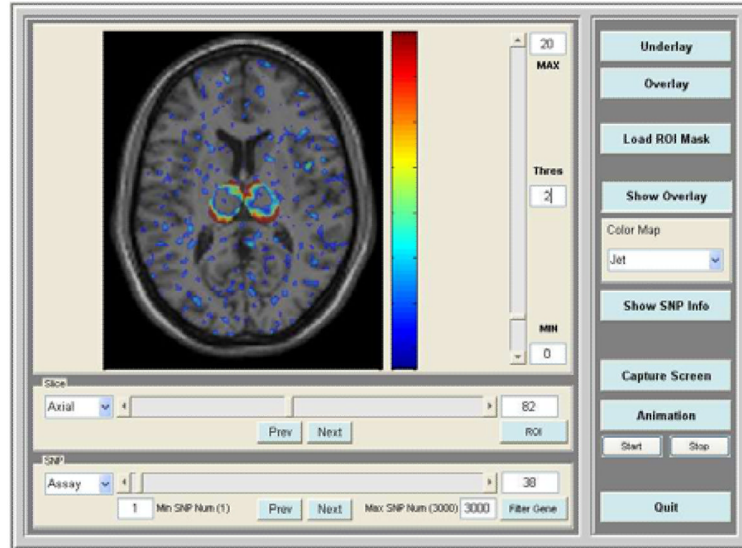
**Fig. 3.** Sample gray matter (GM) maps in three orthogonal views: the normalized GM maps are shown on top, and the smoothed normalized ones are shown on bottom.



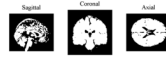
**Fig. 4.** Average gray matter (GM) maps over 39 real HC subjects in three orthogonal views.



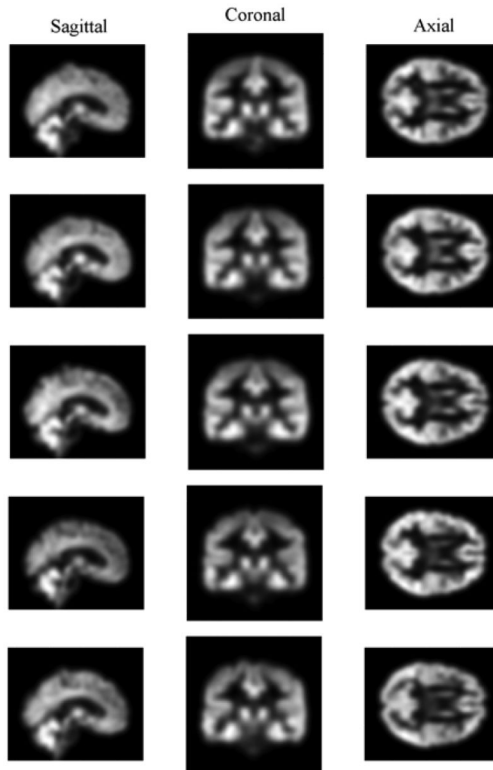
**Fig. 5.** Modulation function for gray matter density.



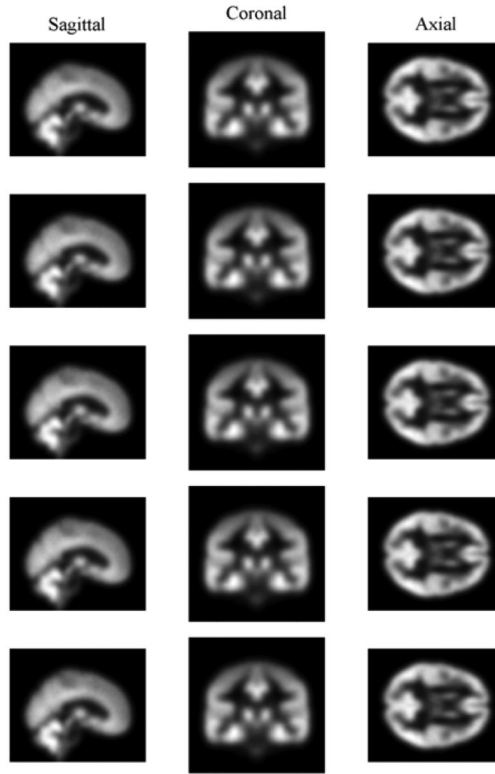
**Fig. 6.**  
Graphical user interface of the initial imaging genomic browser.



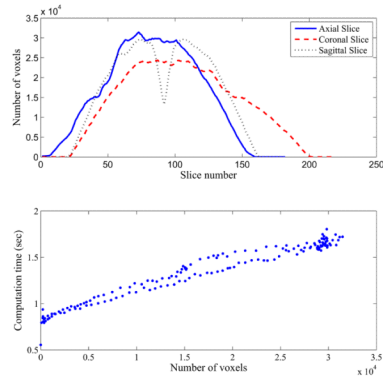
**Fig. 7.**  
Binary mask created from T1-weighted single-subject anatomy in three orthogonal views.



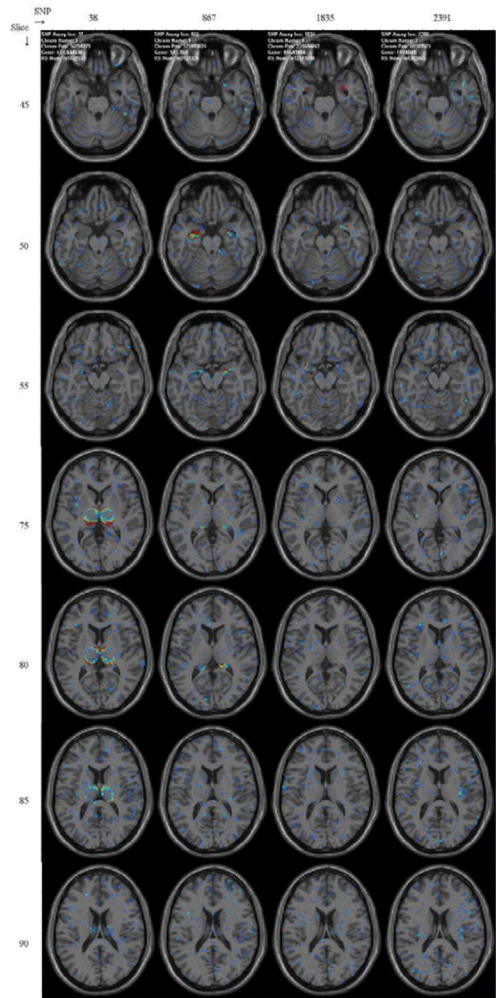
**Fig. 8.**  
Real GM maps for five real subjects.



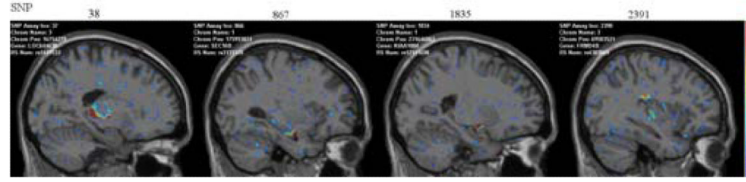
**Fig. 9.**  
Simulated GM maps for five synthetic subjects.



**Fig. 10.** Measurement of computation time of one-way ANOVA. On the top panel, the number of voxels inside the brain mask (Fig. 7) on a slice is plotted against the slice number. On the bottom panel, the computation time is plotted against the number of voxels involved in the computation.



**Fig. 11.** Axial view. F-maps of one-way ANOVA, thresholded at the level of 3, depicting significant associations of four regions of interest with four SNPs. Values of F-map between 0 and 20 are color-coded using the colormap ranging from blue to red next to each panel.



**Fig. 12.** Sagittal view. F-maps of one-way ANOVA, thresholded at the level of 3, depicting significant associations of four regions of interest with four SNPs. Values of F-map between 0 and 20 are color-coded using the colormap ranging from blue to red next to each panel.

**TABLE I**

Synthetic associations between ROIs (imaging phenotypes) and SNPs (genotypes)

Association model	ROI	SNP index
Heterozygotic model	Thalamus	38
Additive model	Hippocampus	867
Dominant model	Amygdala	1835
Recessive model	Insula	2391

**TABLE II**

Relationships between genotype groups (AA, AB, BB) and phenotype groups ( $P_+$ ,  $P_0$ ,  $P_-$ ) for each association model.

	AA	AB	BB
Heterozygotic model	$P_+$	$P_-$	$P_+$
Additive model	$P_+$	$P_0$	$P_-$
Dominant model	$P_+$	$P_-$	$P_-$
Recessive model	$P_+$	$P_+$	$P_-$