An Joint Association Analysis Method for Genomic Sequencing and Neuroimaging Data

January 18, 2019

Xiaoran Tong[1], Olga Vsevolozhskaya[2], Qin Lu[1*],

1 Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, USA **2** Department of Biostatistic, University of Kentucky College of Public Health

Correspondence: Qing Lu
Department of Epidemiology and Biostatistics
College of Human Medicine
Michigan State University
909 Fee Road
East Lansing, MI 48824–1030
qlu@msu.edu

Abstract

Despite heightened interest in the genetic basis of human disease, genome-wide studies have explained relatively little of the heritability of most complex traits and variants identified through these studies have small effect sizes. Although it is hard to provide a single answer to the question of where to find missing heritability, researchers are recently leaning towards using genetic risk factors in combination with clinical biomarkers to gain further knowledge on various phases of complex diseases. For instance, it has been suggested that imaging data that characterizes functional brain abnormalities combined with genetic variability may help researchers identify individuals at risk for Alzheimer's disease. However, a joint analysis of imaging and genetic data presents challenges due to high dimensionality and high redundancy of medical data, leading to a significant decrease in statistical power of any association test. In this work, we incorporate imaging profile to empower the genomic association analysis through the use of the similarity U-statistic. To tackle high dimensionality and high redundancy of imaging data, we use machine learning techniques to replace raw image with abstracted higher order features. We demonstrate flexibility and competitive power of our approach gained by replacing raw neuroimage with high order features through extensive simulation studies. Further, we illustrate potential for discovery of our method through a joint analysis of imaging and genetic data from the Alzheimers Disease Neuroimaging Initiative (ADNI).

Introduction

Despite a decade long effort, genome wide association (GWA) analysis has fell short in providing definitive evidence for casual genetic variants for most common human complex diseases. Although a large number of statistically significant common variants were indeed identified by GWA, only a moderate fraction of heritability has been explained by the totality of these findings (Manolio, 2010; Pandey, 2010). The "rare variant, common disease (RVCD)" hypothesis was aimed to explain the missing heritability which GWA failed to capture. RVCD states that the missing heritability gap could be attributed to rare variants with moderate to large effect sizes that were not covered by GWA (Cirulli and Goldstein, 2010). The Next Generation Sequencing (NGS) project, growing in both numbers and scale over the last decade, offered numerous data sources for the analysis of rare variants. However, the stockpiling data also raised a number of methodological challenges. For one, the variants in the NGS profile are much denser than those in a GWA profile, which poses intense computational and multiple testing burden on the traditional per-variant based screening procedures. Also, as the name suggests, the newly detected rare variants have minor allele frequencies (MAF) close to zero. As a consequence, studies with a small or a moderate sample size may have low statistical power due to the lack of genetic heterogeneity in the sample.

Signal aggregation was proposed as a solution for rare variants' computational issue (Dai et al., 2014; Madsen and Browning, 2009; Purcell et al., 2007; Wei et al., 2014; Wei and Lu, 2015; Wu et al., 2011; Yang et al., 2011). With signal aggregation, instead of screening the whole profile one variant at a time, variants are first grouped according to a certain criteria and then all variants in a group are tested together as a single unit. The aggregation can be achieved by either collapsing all grouped variants into a single variant before a statistical test (Madsen and Browning, 2009), or by testing all variants together with a multivariate approach (Wei et al., 2014; Wei and Lu, 2015; Wu et al., 2011; Yang et al., 2011). Alternatively, the aggregation can be done after per-variant screening by combining group members' statistics. (e.g., p-values) (Dai et al., 2014; Purcell et al., 2007; Zaykin et al., 2002). Grouping and aggregation drastically reduces the number of hypothesis to be tested and improves heterogeneity over any of its member variant. However, the choice of a grouping criteria poses a challenge. The most common choice is to refer to the prior knowledge of biological function, resulting in gene or pathway based grouping (e.g., Vsevolozhskaya et al. (2016)). Alternatively, the grouping can be based on physical distance such as grouping by every few thousand nucleotide base-pairs or by a threshold of linkage disequilibrium (LD) (e.g., Purcell et al. (2007)).

Besides rare variants, an important factor that has been argued to contribute to the unsatisfactory performance of GWA is the fact that complex diseases have intrinsically weak genetic effects due to a large "black box" between the upstream genomic variants and the downstream health outcomes. Therefore, it is desirable to probe this "black box" by incorporating intermediate biological profiles, with the hope that the added information will increase chances of detecting stronger associations, especially when the biomarkers in these new profiles are mediating the genetic casual effect on the disease.

In this paper, we propose a new method that incorporates neuroimaging information into the genomic association analyses and augments statistical power with these "added data." The cortex structure captured by imaging devices is a powerful predictor of a neurological disorder and should be jointly analyzed with a genomic profile. Techniques similar to GAW have been developed for imaging data, given a proper definition of an "image variant" and its value. Taking the structured magnetic resonance imaging (MRI) as an example, it is natural to view a voxel in a pile of slices as a "variant" and the normalized brightness of that voxel as its value. Alternatively, if a three-dimensional (3D) cortex spanned by hundreds of thousands vertices is used, every vertex can be seen as an image variant, while the 3D coordinates of that vertex, the thickness and the curvature of cortex around that vertex, can be seen as its value. These definitions gave rise to a voxel-wise analysis (Ashburner and Friston, 2000; Baron et al., 2001; Chtelat et al., 2005; Smith et al., 2006) or to a vertex-wise analysis (Bernal-Rusiel et al., 2013; Salat et al., 2004), both abbreviated as "VWA" for short. With ideas similar to GWA, VWA applies a per-voxel or per-vertex screening procedure to detect significant loci in the brain. On the one hand, this imaging per-unit analysis is less troublesome than that of a NGS profile because there are no "rare" variants, since imaging data values (e.g., brightness of a voxel or thickness at a vertex) are continuous. On the other hand, imaging profiles are also high dimensional with a large number of voxels or vertices, raising computational and multiple testing issues quiet similar to those encountered with NGS data. Yet again, grouping and aggregation techniques that work with NGS analysis may also help in studies involving imaging data. For example, grouping can be achieved by partitioning a 3D cortex surface into well defined functional anatomical regions (e.g., 68 symmetrical regions, 34 per hemisphere) and considering all vertices in a regions as a single unit for the analysis. Regions so defined will contain from a few hundred to more than ten thousand vertices – a much larger value than a typical number of variants in a gene. However, due to their proximity, vertices exhibit high correlation and redundancy (because they represent a tightly connected brain tissue) and the number of "independent" vertices can be much smaller. One approach to reduce the number of correlated vertices that is recently gaining enormous popularity in computer science is the unsupervised training of deep artificial neural networks (ANN) capable of abstracting high order features from a raw image. The high order features have lower dimension but higher signal-to-noise ratio than the image itself (Glorot and Bengio, 2010; Hinton and Salakhutdinov, 2006; Vincent et al., 2010). Additionally, an unsupervised ANN is capable of cumulatively refining itself with incoming new knowledge. In other words, as long as the future 3D cortex data shares compatible format with the one currently in use, the deep ANN trained today can be re-calibrated to extract more informative features from future data.

In this study, we propose to use the stacked autoencoder (SA), a type of unsupervised deep ANN trained with the maximum likelihood based, gradient guided numerical optimization techniques, to reduce the dimensionality of imaging data (Glorot and Bengio (2010); Vincent et al. (2010)). Then, we adopt a similarity measure based on a U-statistic (Wei et al., 2014; Wei and Lu, 2015) to jointly analyze collapsed genomic and imaging profiles. We show that our newly proposed joint analysis is robust against various types of model misspecification and is faster than the main stream algorithms that currently support multiple aggregated high dimensional components such as SKAT (Wu et al., 2011) or GCTA (Yang et al., 2011). Additionally, we show improvement in statistical power if high order features abstracted by the trained SA are used instead of the raw image profile.

The rest of this paper is organized as follows. In the Methods sections, we detail unsupervised training of stacked autoencoder and the resulting abstraction of high order features from a 3D cortex image, followed by the joint analysis of genomic and imaging profiles (either with the raw cortex data or with the higher order features) via a U-statistics. Further, through simulation studies, we report performance gained by adopting the joint test, the grouping and aggregation strategy on imaging profile, and the replacement of raw imaging with high order features. Finally, we showcase our approach by jointly analyzing case-control imaging and genetic data from the Alzheimers Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005; Nestor et al., 2008).

Methods

Processing Imaging Profile with Stacked Autoencoder

A stacked autoencoder (SA) is an artificial neural network (ANN) mimicking visual processing that abstracts high order features from a raw image. These high order features can typically be more relevant to decision making than the original per-pixel color values of an image. For example, knowing the exact thickness and curvature of every point in the 3D cortex image may be less helpful for a physician in diagnosis of a neurodegenerative disorder than being able to recognize the general location, size and shape of the laceration sites in the same cortex image. Thus, by learning high order features with the SA algorithm, while disregarding trivial and redundant details, we expect to achieve a power boost in a subsequent association analysis, given that these features are appropriate for replacing the unprocessed imaging profile.

A stacked autoencoder is comprised of two or more encoders stacked on top of each other. An encoder is an information preserving transformation that produces a more concise output than the input. In other words, an SA is a linear recombination of the input entries followed by an entry-wise non-linear transformation. An ith encoder in a stack of M layers can be written as:

$$\boldsymbol{z}_i^{d_i} = \boldsymbol{s}(\boldsymbol{W}_i^{d_i \times d_{i-1}} \boldsymbol{z}_{i-1}^{d_{i-1}} + \boldsymbol{b}_i^{d_i}), \quad i = 1, \ldots, M, \tag{1}$$

where $\boldsymbol{z}_i^{d_i}$ is the ith encoder's d_i dimensional output, and $\boldsymbol{z}_{i-1}^{d_{i-1}}$ is d_{i-1} dimensional input, which, in turn, is also an output from the encoder down below, (i.e., $(i-1)$th in the stack). A linear recombination of the input is achieved by a $(d_{i-1} \times d_i)$ weight matrix $\boldsymbol{W}_i^{d_i \times d_{i-i}}$ and a $(d_i \times 1)$ offset vector $\boldsymbol{b}_i^{d_i}$. The function $\boldsymbol{s}(\cdot)$ denotes an aforementioned entry-wise non-linear transformation. A popular choice for $\boldsymbol{s}(\cdot)$ includes the "S"-shaped inverse logit function that can mimic neuron activation (Funahashi, 1989; Hansen and Salamon, 1990). Finally, being an encoder demands the output size d_i to be smaller than the input size d_{i-1}, which in turn ensures that dimension reduction and high order feature abstraction actually happens.

Once an encoder is defined, one can assemble a stack of M encoders that accepts size P input vector, $\boldsymbol{x}^P$, and produces size $Q < P$ high order feature vector, $\hat{\boldsymbol{h}}^Q$, by recursively wiring the output of an encoder to the one above it, and ensuring that the input dimensionality at the bottom, d_0, equals to P, and the output dimensionality on the top, d_M, equals to Q. Mathematically, the encoder stack can be written as:

$$\begin{aligned}
\hat{\boldsymbol{h}}^{Q} &= \boldsymbol{z}_{M}^{d_M} \\
\boldsymbol{z}_{M}^{d_M} &= s(\boldsymbol{W}_{M}^{d_M \times d_{M-1}} \boldsymbol{z}_{M-1}^{d_{M-1}} + \boldsymbol{b}_{M}^{d_M}) \\
\boldsymbol{z}_{M-1}^{d_{M-1}} &= s(\boldsymbol{W}_{M-1}^{d_{M-1} \times d_{M-2}} \boldsymbol{z}_{M-2}^{d_{M-2}} + \boldsymbol{b}_{M-1}^{d_{M-1}}) \\
&\vdots \\
\boldsymbol{z}_{i}^{d_i} &= s(\boldsymbol{W}_{i}^{d_i \times d_{i-1}} \boldsymbol{z}_{i-1}^{d_{i-1}} + \boldsymbol{b}_{i}^{d_i}) \\
&\vdots \\
\boldsymbol{z}_{2}^{d_2} &= s(\boldsymbol{W}_{2}^{d_2 \times d_1} \boldsymbol{z}_{1}^{d_1} + \boldsymbol{b}_{2}^{d_2}) \\
\boldsymbol{z}_{1}^{d_1} &= s(\boldsymbol{W}_{1}^{d_1 \times d_0} \boldsymbol{z}_{0}^{d_0} + \boldsymbol{b}_{1}^{d_1}) \\
\boldsymbol{z}_{0}^{d_0} &= \boldsymbol{x}^{P}.
\end{aligned} \tag{2}$$

Here, the P-dimensional input $\boldsymbol{x}^P$ can be viewed as the output of the non-existing 0th encoder. By restricting $P = d_0 > d_1 > d_2 > \cdots > d_{M-1} > d_M = Q$, the encoder stack abstracts Q dimensional high order features from the P dimensional raw profile. To ensure that the output is a concise preservation of the input, the parameters of the stack, that is, the weights and the offsets, $\{\boldsymbol{W}_1, \boldsymbol{b}_1, \boldsymbol{W}_2, \boldsymbol{b}_2, \ldots, \boldsymbol{W}_M, \boldsymbol{b}_M\}$, must be tuned to represent the body of knowledge that generated $\boldsymbol{x}^P$, which in our case is the knowledge of human cortex. To do so, we assemble a stacked encoder counterpart, i.e., a stacked decoder that starts with higher order features and outputs the entire image. The stacked decoders exactly mirror the topology of the initial encoders:

$$\begin{aligned}
\tilde{\boldsymbol{x}}^{P} &= \tilde{\boldsymbol{z}}_{0}^{d_0} \\
\tilde{\boldsymbol{z}}_{0}^{d_0} &= s(\tilde{\boldsymbol{W}}_{1}^{d_0 \times d_1} \tilde{\boldsymbol{z}}_{1}^{d_1} + \tilde{\boldsymbol{b}}_{1}^{d_0}) \\
\tilde{\boldsymbol{z}}_{1}^{d_1} &= s(\tilde{\boldsymbol{W}}_{2}^{d_1 \times d_2} \tilde{\boldsymbol{z}}_{2}^{d_2} + \tilde{\boldsymbol{b}}_{2}^{d_1}) \\
&\vdots \\
\tilde{\boldsymbol{z}}_{i-1}^{d_{i-1}} &= s(\tilde{\boldsymbol{W}}_{i}^{d_{i-1} \times d_i} \tilde{\boldsymbol{z}}_{i}^{d_i} + \tilde{\boldsymbol{b}}_{i}^{d_{i-1}}) \\
&\vdots \\
\tilde{\boldsymbol{z}}_{M-2}^{d_{M-2}} &= s(\tilde{\boldsymbol{W}}_{M-1}^{d_{M-2} \times d_{M-1}} \tilde{\boldsymbol{z}}_{M-1}^{d_{M-1}} + \tilde{\boldsymbol{b}}_{M-1}^{d_{M-2}}) \\
\tilde{\boldsymbol{z}}_{M-1}^{d_{M-1}} &= s(\tilde{\boldsymbol{W}}_{M}^{d_{M-1} \times d_M} \tilde{\boldsymbol{z}}_{M}^{d_M} + \tilde{\boldsymbol{b}}_{M}^{d_{M-1}}).
\end{aligned} \tag{3}$$

The weights of the stacked decoders also mirror the weights of the stacked encoders:

$$\tilde{\boldsymbol{W}}_{M-i}^{d_{M-i-1}\times d_{M-i}} \equiv (\boldsymbol{W}_{M-i}^{d_{M-i}\times d_{M-i-1}})', \quad (i=0,\ldots,M), \tag{4}$$

However, the offsets in the decoders, $\tilde{\boldsymbol{b}}_{M-i}^{d_{M-i-1}}, (i=0,\ldots,M)$, are allowed to be flexible (Vincent et al., 2010). This approach to tune the parameters of the stack is quiet instinctive since decoding is the opposite of encoding. From the bottom to the top, this stack of decoders gradually restores details back to the abstracted state $\tilde{\boldsymbol{z}}_M^{d_M}$ and eventually presents a reconstructed input $\tilde{\boldsymbol{x}}^P$ on its top.

One can combine Eq. (2) and (3) by setting $\tilde{\boldsymbol{z}}_M^{d_M} = \hat{\boldsymbol{h}}^Q$,

$$\begin{aligned}
\tilde{\boldsymbol{x}}^P &= \tilde{\boldsymbol{z}}_0^{d_0} \\
\tilde{\boldsymbol{z}}_0^{d_0} &= s(\tilde{\boldsymbol{W}}_1^{d_0\times d_1}\tilde{\boldsymbol{z}}_1^{d_1} + \tilde{\boldsymbol{b}}_1^{d_0}) \\
&\vdots \\
\tilde{\boldsymbol{z}}_{M-1}^{d_{M-1}} &= s(\tilde{\boldsymbol{W}}_M^{d_{M-1}\times d_M}\tilde{\boldsymbol{z}}_M^{d_M} + \tilde{\boldsymbol{b}}_M^{d_{M-1}}) \\
\tilde{\boldsymbol{z}}_M^{d_M} &= \hat{\boldsymbol{h}}^Q = \boldsymbol{z}_M^{d_M} \\
\boldsymbol{z}_M^{d_M} &= s(\boldsymbol{W}_M^{d_M\times d_{M-1}}\boldsymbol{z}_{M-1}^{d_{M-1}} + \boldsymbol{b}_M^{d_M}) \\
&\vdots \\
\boldsymbol{z}_1^{d_1} &= s(\boldsymbol{W}_1^{d_1\times d_0}\boldsymbol{z}_0^{d_0} + \boldsymbol{b}_1^{d_1}) \\
\boldsymbol{z}_0^{d_0} &= \boldsymbol{x}^P,
\end{aligned} \tag{5}$$

which creates a stacked autoencoder (Vincent et al., 2010). The aforementioned weights and offsets in the encoder stack (the lower half of the SA), alongside with M extra offsets in the decoder stack (the upper half of the SA), constitute the parameters to be calibrated in order to make the encoder stack to be a close represent of the process that generated $\boldsymbol{x}^P$. The calibration is done by minimize the discrepancy between the reconstructed input $\tilde{\boldsymbol{x}}^P$ and the true input $\boldsymbol{x}^P$. The rationale is that if the compact code $\hat{\boldsymbol{h}}$ presented by the encoder stack truly captures the major features of $\boldsymbol{x}^P$, the restored input, $\tilde{\boldsymbol{x}}^P$, should be nearly identical to the original one, except some trivial details.

Tuning the encoder stack is equivalent to solving the following optimization problem:

$$\boldsymbol{\Theta}^* = \min_{\boldsymbol{\Theta}} \sum_{k=1}^{N} d(\tilde{\boldsymbol{x}}_k^P, \boldsymbol{x}_k^P), \quad \boldsymbol{\Theta} = \cup_{i=1}^{M} \{\boldsymbol{W}_i, \boldsymbol{b}_i, \tilde{\boldsymbol{b}}_i\}, \tag{6}$$

where k indices N training samples. The objective function, d, measures the disagreement between the reconstructed and the original input. A popular form of d is a cross-entropy:

$$d(\tilde{\boldsymbol{x}}_k^P, \boldsymbol{x}_k^P) = \sum_{j=1}^{P} [x_{j,k} \log \tilde{x}_{j,k} + (1 - x_{j,k}) \log (1 - \tilde{x}_{j,k})], \tag{7}$$

where j indices P entries of the input.

Optimization of a large number of parameters (cardinality of $|\boldsymbol{\Theta}| = \sum_{i=1}^{M} d_i d_{i-1} + d_i + d_{i-1}$) is achieved by stochastic gradient descent (SGD) (Bottou, 2010; Zhang, 2004), which is also called the back propagation (BP) algorithm by neural network literature concerning computations of high dimensional gradients (Cun et al., 1990; Fahlman, 1988; Vogl et al., 1988). For practical implementation of SGD and BP algorithms, a Python (www.python.org, 2017) library Theano (Theano Development Team, 2016) can be used.

Optimization of deep neural networks, (i.e., an SA with many layers of encoders and decoders) is challenging due to the increased prevalence of local minimum site in the error terrain and relatively slow convergence rates. To address these issues, we follow the "deep learning" trend, which is recently becoming increasingly popular and by which a layer-wise greedy pre-training procedure is applied prior to fine-tuning the entire network (Bengio et al., 2007; Vincent et al., 2010). To do so, the output of the ith encoder, $\boldsymbol{z}_i$, is disconnected from the encoder above it and rewired to its decoder's counterpart, immediately forming a single layered autoencoder, which is then calibrated by minimizing the intermediate reconstruction loss $d(\boldsymbol{z}_{i-1}, \tilde{\boldsymbol{z}}_{i-1})$. That is,

$$\begin{aligned}
\tilde{\boldsymbol{z}}_{i-1} &= s(\tilde{\boldsymbol{W}}_i \tilde{\boldsymbol{z}}_i + \tilde{\boldsymbol{b}}_i) \\
\tilde{\boldsymbol{z}}_i &= \hat{\boldsymbol{h}}_i = \boldsymbol{z}_i \qquad\qquad (i = 0 \ldots M) \\
\boldsymbol{z}_i &= s(\boldsymbol{W}_i \boldsymbol{z}_{i-1} + \boldsymbol{b}_i). \\
\boldsymbol{\theta}_i^* &= \min_{\boldsymbol{\theta}_i} d(\boldsymbol{z}_{i-1}, \tilde{\boldsymbol{z}}_{i-1}), \qquad \boldsymbol{\theta}_i = \{\boldsymbol{W}_i, \boldsymbol{b}_i, \tilde{\boldsymbol{b}}_i\},
\end{aligned} \tag{8}$$

This optimization problem is much easier than the one in Eq. (5) due to a smaller number of parameters ($|\boldsymbol{\theta}_i| = d_i d_{i-1} + d_i + d_{i-1}$). Then, after all M single-layer autoencoders are pre-trained, the encoders and decoders are wired back to Eq. (5) and fine-tuned together, resulting in faster convergence and less likely chances of "going down the wrong pit" in the terrain of $d(\tilde{\boldsymbol{x}}_k^P, \boldsymbol{x}_k^P)$.

For the current project, we let each encoder halve its input (i.e., $d_i = d_{i-1}/2$). Then, for each cortex region we form an $M = 4$ layered SA, achieving a 16 fold dimension reduction. Further, we use higher order features extracted by the SA in place of the raw imaging profile in the subsequent statistical analysis.

Joint Test with the U-statistic

We use a similarity measure based on the U-statistic (Wei et al., 2014; Wei and Lu, 2015) to jointly test for an association between a phenotype of interest (either quantitative or qualitative) and genomic and imaging profiles combined. To derive the statistic, three kernel functions measuring pairwise similarity are chosen for each profile. The measurement, $f(\cdot)$, can be flexible to suit specific characteristics of a profile (e.g. bounded or not, continuous or discrete), as long as f is symmetric and has a finite second moment (i.e., $f(x_i, x_j) \equiv f(x_j, x_i)$ and $E(f^2(x_i, x_j)) < \infty$).

For genomic profile coded by minor allele count (0, 1 or 2), we use the Identical By State (IBS) kernel:

$$f_G(\boldsymbol{g}_{i\cdot}, \boldsymbol{g}_{j\cdot}) = \frac{\sum_{m=1}^{|\boldsymbol{G}|} w_m (2 - |g_{im} - g_{jm}|)}{2 \sum_{m=1}^{|\boldsymbol{G}|} w_m},$$

where (i, j) indices a pair of observations for subjects i and j, m indices a genomic variant (i.e., a SNP) in the testing unit (e.g., a gene $\boldsymbol{G}$), and w_m is the weight assigned to the mth variant according to *a prior* hypothesis. For example, weights can be assigned based on the allele frequency (AF) as $w_m = \sqrt{AF(g_{.m})[1 - AF(g_{.m})]}$ to emphasize the effect of rare variants.

The imaging profile can either be a raw 3D cortical surface vertices or the high order features abstracted from them by an SA. For both, we can use the Gaussian kernel, which is well suited for continuous values,

$$f_V(\boldsymbol{v}_{i\cdot}, \boldsymbol{v}_{j\cdot}) = \exp\left[-\frac{1}{|\boldsymbol{V}|} \sum_{m=1}^{|\boldsymbol{V}|} (v_{im} - v_{jm})^2\right].$$

Here, m indices a variant (e.g., a vertex) in a cortex region $\boldsymbol{V}$ (e.g., superior-temporal, entorhinal, ect.).

Then, the entry-wise product of genomic and imaging kernels forms the joint predictor kernel

$\boldsymbol{K}^J$ as follows:

$$K_{ij}^J = f_G(\boldsymbol{g}_{i.}, \boldsymbol{g}_{j.}) f_V(\boldsymbol{v}_{i.}, \boldsymbol{v}_{j.}).$$

Further, the entries in the predictor kernel $\boldsymbol{K}^J$ are centered by subtracting two marginal means (for both subjects i and j) and by adding the overall mean (Wei and Lu, 2015), so that:

$$\tilde{K}_{ij}^J = K_{ij}^J - \frac{1}{N}\sum_{k=1}^{N} K_{ik}^J - \frac{1}{N}\sum_{l=1}^{N} K_{lj}^J + \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N} K_{lk}^J,$$

where N is the number of subjects.

Lastly, for a phenotype profile generated by an unknown distribution, we first normalized it by the rank normal quantile transformation:

$$q_i = \frac{\Phi^{-1}[\text{rank}(y_i - 0.5)]}{N},$$

where y_i is the phenotypic measure for the ith subject. Next, we project the transformed phenotype $\boldsymbol{q}$ onto the space spaned by P covariate $\boldsymbol{X} = [\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_P]$ through a linear regression. The residual terms of this regression are then used to build a similarity measure for phenotypes via the cross-product kernel (Wei et al., 2014):

$$\hat{H}_{ij} = \tilde{f}_Y(q_i, q_j) = \frac{(q_i - \hat{q}_i)(q_j - \hat{q}_j)}{(\boldsymbol{q} - \hat{\boldsymbol{q}})'(\boldsymbol{q} - \hat{\boldsymbol{q}})/(N - P - 1)}.$$

where $\hat{\boldsymbol{q}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{q}$ is the predicted mean phenotype vector, given covariates $\boldsymbol{X}$. After this projection, phenotype similarities are already centered and one can use $\hat{\boldsymbol{H}}.$ as the response kernel. To make the test statistics more robust against residual confounding, Wei et al. (2015) also suggested to project the similarity kernel among predictors $\tilde{K}^J$ onto the covariates space and use the resulting residuals as a new measure orthogonal to $\boldsymbol{X}$,

$$\hat{\boldsymbol{K}}^J = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\tilde{\boldsymbol{K}}^J(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}').$$

Then, the joint similarity U-statistic, U_J, can be defined as the mean of entry-wise products between the response kernel $\boldsymbol{H}$ and the predictor kernel $\boldsymbol{K}^J$, excluding self-pairs $i = j$, as:

$$U_J = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \hat{H}_{ij} \hat{K}^J_{ij}$$

Under the null hypothesis of no association between any of the profiles, U_J should be close to 0 since all centered kernel measurements should have zero means. Conversely, significant deviations of U_J from 0 will imply an association. Formally, under H_0, U_J follows a weighted mixture of the χ^2_1 distributions (Wei et al., 2014),

$$U_J \sim \frac{1}{N} \sum_{m=1}^{\infty} \alpha_m \sum_{l=1}^{N} \hat{\lambda}_l (\chi^2_{1,ml} - 1),$$

where α_m is an eigenvalue of response similarity kernel function $f_Y(\cdot, \cdot)$, approximated by $\boldsymbol{H}$, and $\hat{\lambda}_l$ is an eigenvalue of predictor kernel $\hat{\boldsymbol{K}}^J$. When a cross-product kernel is used for f_Y, $\alpha_1 = 1$, and $\alpha_m = 0, (m > 1)$, and the mixture can be simplified as (Wei et al., 2014),

$$U_J \sim \frac{1}{N} \sum_{l=1}^{\infty} \hat{\lambda}_l \chi^2_{1,l}.$$

A p-value based on this chi-squared mixture can be approximated by the Davis method (Davies, 1980).

Alternatively, two simpler tests can be performed by dropping either the imaging or the genomic kernels. That is,

$$K^G_{ij} = f_G(\boldsymbol{g}_{i.}, \boldsymbol{g}_{j.}), \quad \text{or} \quad K^V_{ij} = f_V(\boldsymbol{v}_{i.}, \boldsymbol{v}_{j.}),$$

which are subsequently center and projected onto $\boldsymbol{X}$ just as in the case of $\boldsymbol{K}^J$. The simplified test based on the:

$$U_G = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \hat{H}_{ij} \hat{K}^G_{ij}$$

will test for an association between a phenotype and the genomic profile. And,

$$U_V = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \hat{H}_{ij} \hat{K}^V_{ij}$$

will test for an association between a phenotype and the imaging profile. These two tests are more specific and correspond to more parsimonious models. However, they also run the risk of model misspecification, a sensitivity to which will be explored in a simulation study.

The imaging similarity measure described just above aggregates all signals of all vertices in one testing unit $\boldsymbol{V}$. For comparison purposes, in our simulation studies, we also implemented a per-vertex analysis, that is, the vertex-wise analysis (VWA). Briefly speaking, for the per-vertex analysis, we first smoothed the imaging profile with a Gaussian filter to reduce noise and grind away trivial details in a 3D cortex. Next, for each one of the $|\boldsymbol{V}|$ vertices, we performed a test based on the U_J or U_V statistic, and the resulting $|\boldsymbol{V}|$ p-values were corrected against multiple testing according to the false discovery rate (FDR) criteria (Benjamini and Hochberg, 1995). If at least one FDR corrected p-value was below the 0.05 threshold, the entire testing unit of $|\boldsymbol{V}|$ vertices was declared statistically significant.

Result

Simulation Study Setup

To mimic the real sequencing and imaging data structure, we based our simulations on observations from 806 participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005; Nestor et al., 2008) with information over both NGS and MRI profiles. The 3D cortex was re-built from MRI data using Freesurfer software (Fischl, 2012). In each simulation iteration, we picked a gene, $\boldsymbol{G}$, and a cortex region of 512 vertices, $\boldsymbol{V}$, (roughly a circle of 28mm in diameter), and assigned an effect size to 5% of the genomic and imaging variants. Non-zero effect sizes were drawn from a standard normal distribution. Then, we generated two phenotypes solely based on a genomic or an imaging effect, $\boldsymbol{Y}_G$ and $\boldsymbol{Y}_V$, respectively; added them up to get an additive effect $\boldsymbol{Y}_A$; obtained their entry-wise product for an interaction effect $\boldsymbol{Y}_I$. More formally,

$$\begin{aligned}
&\boldsymbol{Y}_G = \boldsymbol{\beta}_G \boldsymbol{X}_G + \boldsymbol{\epsilon}_G, \quad \beta_{\cdot G} \sim N(0,1) \times \text{Bernoulli}(0.05),\\
&\boldsymbol{Y}_V = \boldsymbol{\beta}_V \boldsymbol{X}_V + \boldsymbol{\epsilon}_V, \quad \beta_{\cdot V} \sim N(0,1) \times \text{Bernoulli}(0.05),\\
&\boldsymbol{Y}_A = \boldsymbol{Y}_G + \boldsymbol{Y}_V, \qquad \boldsymbol{Y}_I = \boldsymbol{Y}_A + \boldsymbol{Y}_G \circ \boldsymbol{Y}_V.
\end{aligned} \tag{9}$$

We were interested in evaluating the statistical power of a joint test under two scenarios: (a) when a phenotype was purely genomic or imaging based, and (b) when a phenotype was based on a combination of imaging and genomic effects (either additive or interaction). We evaluated the power of joint statistic, U_J, versus genomic-based statistic, U_G, and imaging-based statistic, U_V, for eight sample sizes ($n = 100, 200, \ldots, 800$) and four effects (genetic only, imaging only, additive, and interaction). Statistical power of the three tests is summarized in Figure 1. Based on the top row of Figure 1, it is evident that the two simpler statistics (U_G and U_V) perform the best if the phenotypes were truly genomic or imaging based only. However, when the source of variation is misspecified (i.e., imaging for a genetic-based statistic or genetic for an image-based statistic), the power of the two simpler test is no different from the size of the test. In contrast, the joint statistic, U_J, performs fairly well in all cases, with power close to the correct parsimonious model in the case of a single source of variation (either genetic or imaging), and outperforms both U_G and U_V when phenotype involves a combination of genomic and imaging effects (the bottom row of Fig. 1).

Next, we evaluated the performance of imaging-based tests with a cortical region-based analysis versus a per-vertex wise analysis (VMA) with the false discovery rate (FDR) correction. Note, here we skipped the gene-based test, since the statistic U_G is not sensitive to imaging profile. The results are shown in Figure 2. The aggregated test (solid lines) overpowers VWA (dashed lines) by a large margin under all scenarios. Interestingly, when the kernels are completely misspecified, the type I error rate of VWA is below 0.05 (Figure 2, top left panel), which means that tests of each vertex (512 of them) are correlated, thus causing the FDR correction to be conservative.

We further investigated the performance of the imaging-based tests with the use of high order features versus raw imaging profiles. The results are summarized in Figure 3. Based on this figure, we can conclude that replacing raw imaging profile by higher order features results in a test at least as powerful, as the one based on raw data. Additionally, the top left panel of Fig. 3 shows that using high order features does not deviate the type I error rate away from the nominal 0.05 level.

Finally, we generated dichotomous phenotypes by applying the inverse logit function to the four continuous phenotypes and drawing the case/control status from the resulting probabilities. In every scenario, the power performance was very similar to the ones with continuous phenotypes. These additional results are presented in the Appendix to this manuscript.

Real Data Analysis

To illustrate our proposed methodology in real-life scenario, we analyzed data of 327 individuals from ADNI data. Among them, 47 were definitively diagnosed with Alzheimer's Disease (AD), and 280 were healthy controls. For the genomic testing units, we used $40,039$ gene regions. For the imaging testing units, we used high order features of 68 cortical regions abstracted from the raw imaging using 68 SAs trained with all 806 subjects in ADNI sample. To adjust for possible confounders, we included known AD risk factors as covariates, namely, age, sex, race, ethnicity, years of education, marital status, smoking history, and APOE $\epsilon4$ haplotype count (Helmer et al., 1999; Qiu et al., 2009). In total, there were $40,039 \times 68 = 2,722,652$ gene/brain region combinations to test with a joint U-statistic, U_J. For comparison purposes, we also performed two simplified tests, U_G and U_V. Triplets of negative log transformed p-values (P_J, P_G, P_V) are presented in Figure 4, ordered by ascending P_J.

The genomic based U_G statistic (crosses in Figure 4) never reached statistical significance after FDR correction for $40,039$ gene-based tests. The signal from imaging test U_V (diamonds in Figure 4) in general is stronger than from the genomic U_G, reflecting the fact that the upstream genomic effect is rather weak compared with the downstream cortex structure, which is a strong indicator of a neurological disorder. The magnitude of p-values from the joint test U_J (solid dots in Figure 4) lies between the p-value magnitude of the two simpler tests, leaning closer to the imaging based p-values. Notice how U_J "borrows" information from the imaging profile to enhance the signal of genomic based U_G. That is, when both U_G and U_V in the triplet are moderately significant, the joint statistic U_J is likely to be more significant than either U_G and U_V alone, reaching the 0.05 threshold even after FDR correction for $2,722,652$ tests (top left section of Figure 4). These result suggest the existence of strong interaction, such as the genetic effect on cortex mediated by AD, or the genetic effect on AD mediated by cortex.

Table 1 listed our top 20 most significant findings. Based on results presented in Table 1, it is evident that all top 20 signals involve the same cortex region: left-superior-temporal, where the excessive lose of neurons, copper transport and anti-oxidant protein ceruloplasmin, and the shrinkage of tissue are significantly associated with the onset of AD and its progression to DAT (Dementia of the Alzheimer Type) (Connor et al., 1993; Gmez-Isla et al., 1997; Mountjoy et al., 1983). The gene involved in the most significant joint U statistics, *IGLV1-44*, is preferentially associated with amyloidosis mediated by heart (Perfetti et al., 2012), while the amyloidosis in the cortex is a well known precursor of AD (Ghiso and Frangione, 2002). Gene *CDH4* (row 6, Table 1) was previously found to be significantly associated with the overall brain volume and the risk of AD (Seshadri et al. (2007), rs1970546, p = 3.7×10^{-8}). *FAM72C* (row 8, Table 1) encodes protein p17, which is found upregulated in amyloidosis mediated neuron damage (Nehar et al., 2009a) in AD mouse model (Nehar et al., 2009b). The RNA gene *RP11-638L3.1* (9th entry in Table 1, alias of the hypothetical protein LOC643542) is affiliated with major depressive and attention deficit hyperactivity disorders (Shi et al., 2011). The carboxipeptidase-encoding gene *CPXM1* (entry 10, Table 1) was found transiently upregulated during post-stroke healing and scaring in mice (Buga et al., 2012). Cerebellar-degeneration-related antigen-2 (*CDR2*, 15th entry, table 1) is normally transcripted in cerebellar Purkinje neurons and brainstem neurons, and abnormally transcripted in PCD (paraneoplastic cerebellar degeneration) tumer cells (Corradi et al., 1997). The rapamycin-sensitive gene *MIS18BP1* is identified as one of the screening and therapeutic targets for AD (Nagy, 2015). The zinic finger encoding gene *ZDHHC15* (last entry in Table 1) is implicated in neurological diseases including AD, schizophrenia, and X-linked mental retardation (Mansouri et al., 2005; Young et al., 2012).

In addtion, we collected the most siginificant tests for each cortex region, showing the top 20 in Table 2. Other than left-superior-temporal (and the symmetric right-superior-temporal), the imaging statistics (U_V) alone detected a handful of cortex regions to be significantly associated with AD, which is unsurprising given the abundant evidence linking them to AD. For example, 40% volume loss of entorhinal among AD cases (Juottonen et al., 1998); the appearence and disappearence of negative correlation between motor cortex excitability and the thickness of cuneus (Niskanen et al., 2011) when the disease progress from MCI (minor cognitive impairment) to AD; the improvement in distinguishing MCI from DAT (Dementia of the Alzheimer Type) by incorporating fusiform volume into the diagnostic model (Convit et al., 1997); and the significant reduction in mid-temporial neuron density in DAT cases. The joint test (U_J) points out gene *FAM72C* again together with a number of cortex regions, reflecting the fact that *FAM72C* is a predictor of cortical neuron loss (Nehar et al., 2009a). Another frequently detected gene is *ZNF749*, expressed in the cerebral tissue, whose enzyme product annotates IDE (insulin-degrading enzyme) via zinic iron binding that in turn digest amyloid-β peptide – the very molecule causes amyloidosis (Farris et al., 2003; Qiu and

Folstein, 2006), thus the mutation of *ZNF749* may affect the efficiency of IDE, partially explaining the link between type 2 diabetes, hyperinsulinaemia, and the elevated risk of AD.

A handful of pseudogenes was also selected by the joint test, with the corresponding complete genes relating to the risk of neurological disorders one way or another. The second most significant U_J statistics involves pseudogene *NBEAP2* (the first entry in Table 1). The complete gene *NBEA* (neurobeachin) encodes a member of A-kinase anchor protein involved in neuronal post-Golgi membrane traffic, which is a candidate gene associated with neurodevelopment disorder (e.g., autism) (Castermans et al., 2003; Creemers et al., 2014; Volders et al., 2011). *RPL21* (ribosomal protein L21, of pseudogene *RPL21P89*) is found upregulated in sleep deprived (Cirelli et al., 2006), and aged mouse (Bouter et al., 2014). *RPL41* (ribosomal protein L21, of pseudogene *RPL41P2*) is associated with ATF4 degradation (Wang et al., 2011), which is activated in the brain of AD patients (Lewerenz and Maher, 2009). The missense mutation of *HNRNPA1* (heterogeneous nuclear ribonucleoprotein, of pseudogene *HNRNPA1P19*) is found causing ALS (amyotrophic lateral sclerosis), a progressive neurodegenerative disease. *HSPD1P13* is jointly significant with multiple cortex regions (table 2); its corresponding complete gene *HSPD1* (Heat Shock Protein Family D Member 1) is one of the key transcripts of mitochondrial unfolded protein response (mtUPR) process, activated when damaged or unfolded protein (i.e., amyloid-β and prion) abnormally accumulate within mitochondrial, which is found upregulated in AD cases, and is suggested to be involved in selective neuron vulnerability (S Beck et al., 2016). In all, these pseudogenes hint the significance of their corresponding complete genes, along with the evolution history in terms of prevention and response against neurological disorders.

Discussion

In this 'proof-of-concept' paper, we demonstrated a new method to integrate two types of high dimensional profiles (i.e., genetic and imaging data in this study) in order to enhances detection power of genomic risk factors by augmenting it with additional information over imaging profile but maintained a robust performance even if imaging bio-markers do not affect health outcomes. To reduce the dimensionality of imaging data, we built stacked autoencoders (SAs) – a type of artificial neural network through machine learning, – to extract high order features from cortex regions. The simulation study showed that these high order features substantially reduced the dimensionality of data without losing any, or even rising statistical power in the association analysis over the one from a test based on the raw imaging profile. Also, the unsupervised training makes full use of the training materials. That is, the materials excluded from the association analysis due to uncertain diagnosis also build SAs. In the future, imaging data from alternative sources, not limited to those collected for AD diagnosis, can serve as training materials to optimized the SAs with respected to the prediction of AD, or other neuralogical disorders.

We also showed that grouping and signal aggregation of imaging profile also improves the power of association tests. Based on our simulation results, it was clear that per-vertex wise analysis with the FDR correction is inferior to the cortical region-based analysis due to high physical proximity and thus correlation among vertices attributes (e.g., gray matter thickness). However, if no clear grouping criteria can be specified for imaging profiles (e.g., due to an overlap in biological pathways, which leads to the same variant appearing in more than one group) then one can still rely on a per-variant screening procedure. Signal aggregation is, after all, a compromise between accuracy, power, and efficiency, only per-vertex analysis can precisely locate significant variants, while sparing the effort of deciding on a grouping scheme. Alternatively, voxels in the MRI can be easily divided into equally sized volumes satisfying any resolution criteria, which balances accuracy and power without having to consult any prior biological knowledge.

This study provideds an initial expoloration of SAs' capabilities and showed artificial neural network based feature extraction is a promising tool to pre-processing high-dimensional data. In addition, we used the joint U-test to combine extracted imaging features with genomic data to show improvement in association analysis. With more efficient hardware, the SAs can grow a lager number of layers, produce more compact output2, and preferably lower the reconstruction losses. A deeper SA will hopefully extract more meaningful features for the subsequent statistical analysis.

In this study, we took a rather conservative approach to conduct simulation. For real conditions with a clear physiology though, the benefit of high order features over raw imaging data can be larger. For example, we know that a deep artificial neural network is imitating visual processing that focuses on major cortical features such as the dozens of gyri and sulci recognizable by the naked eye, but the simulated effects were randomly assigned to 5% of the variants scattered all over the cortex, not bounded by large, visible features.

In the future, it may be desirable to build SAs for genomic regions as well, but a major improvement in computation is required due to the large number of genes, even if 98% of them have fewer variants than an average cortex region.

Finally, an immediate expansion of the proposed methodology is possible by incorporating other "omics" data with proper U kernels, which may aid future association analysis. Some common candidates include transcriptomics (which is close to the upstream genomics in the casual pathway) and the symbiotic microbiome that complements human genome.

Legends

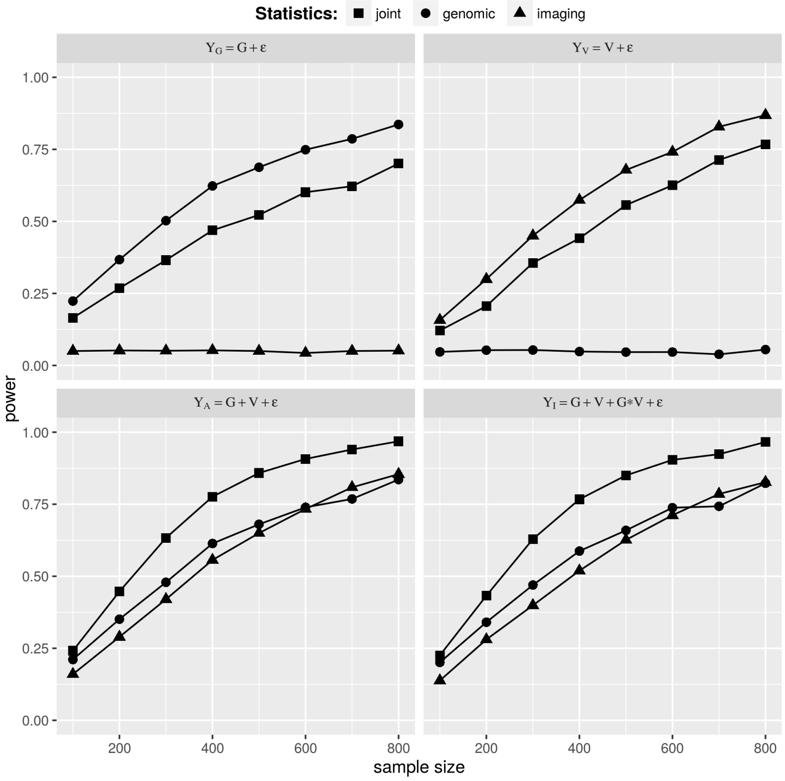

Figure 1: Statistical power of the joint U-test versus genetic U-test and imaging U-test.

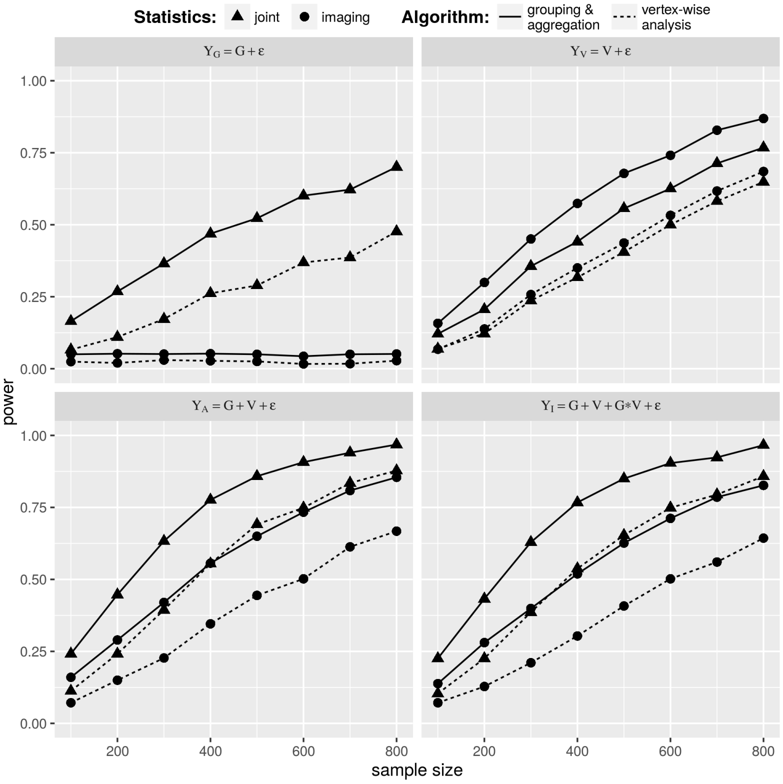

Figure 2: Statistical power of the joint U-test and the imaging U-test with grouping versus vertex-wise analysis.

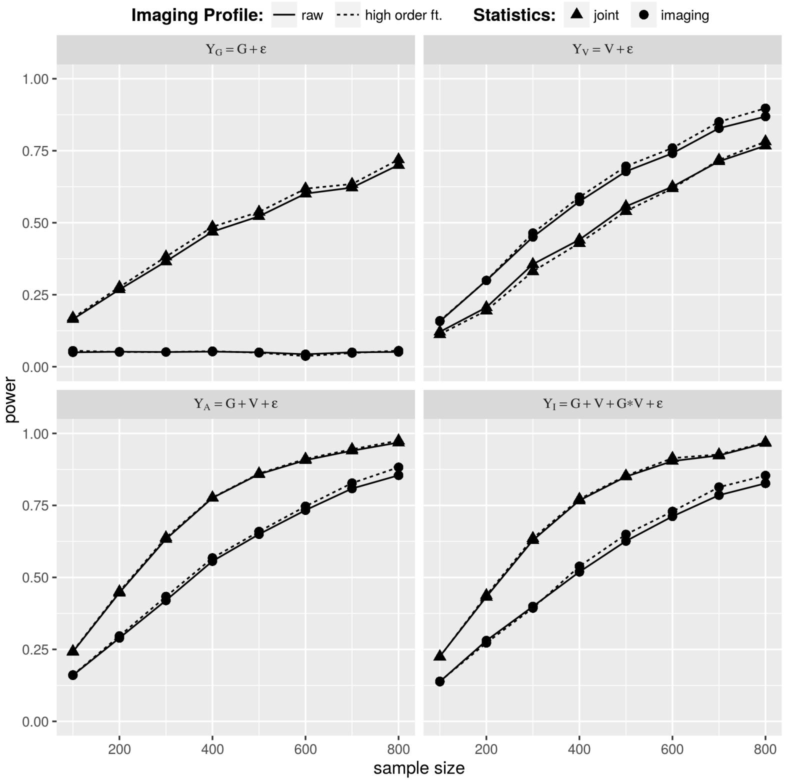

Figure 3: Statistical power of the joint U-test and the imaging U-test with high order features versus original vertices.

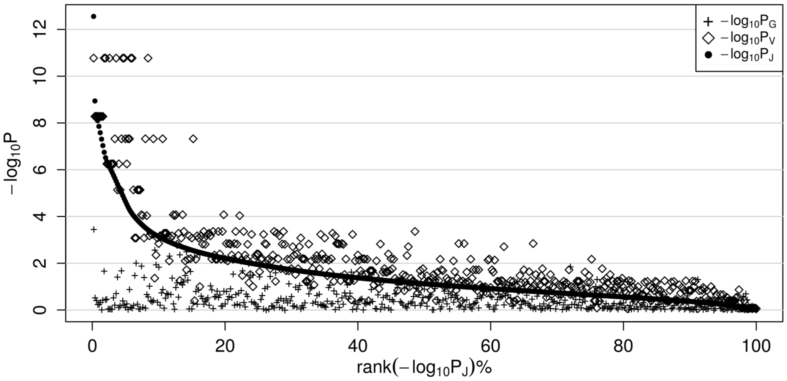

Figure 4: Triplets of p-values (P_J, P_G, P_V) for a joint test, gene-based only and imaging-based only from real data analysis.

Tables

Table 1: Top 20 most significant joint test, overall

GENE	CORTEX	$\lvert V \rvert$	$\lvert G \rvert$	P_G	P_V	P_J
IGLV1-44	l.superiortemporal	7271	174	3.51×10^{-04}	$1.68 \times 10^{-11*}_{+}$	$2.77 \times 10^{-13*}_{+}$
NBEAP2	l.superiortemporal	7271	238	1.19×10^{-04}	$1.68 \times 10^{-11*}_{+}$	$4.74 \times 10^{-13*}_{+}$
RPL21P89	l.superiortemporal	7271	90	6.36×10^{-04}	$1.68 \times 10^{-11*}_{+}$	$5.14 \times 10^{-13*}_{+}$
LOC102724504	l.superiortemporal	7271	59	1.41×10^{-03}	$1.68 \times 10^{-11*}_{+}$	$5.56 \times 10^{-13*}_{+}$
CNTNAP3P8	l.superiortemporal	7271	40	1.08×10^{-03}	$1.68 \times 10^{-11*}_{+}$	$6.17 \times 10^{-13*}_{+}$
CDH4	l.superiortemporal	7271	9464	4.64×10^{-03}	$1.68 \times 10^{-11*}_{+}$	$6.96 \times 10^{-13*}_{+}$
HNRNPA1P19	l.superiortemporal	7271	17	8.88×10^{-04}	$1.68 \times 10^{-11*}_{+}$	$7.80 \times 10^{-13*}_{+}$
FAM72C	l.superiortemporal	7271	174	9.28×10^{-06}	$1.68 \times 10^{-11*}_{+}$	$7.82 \times 10^{-13*}_{+}$
RP11-638L3.1	l.superiortemporal	7271	4067	1.41×10^{-1}	$1.68 \times 10^{-11*}_{+}$	$9.49 \times 10^{-13*}_{+}$
CPXM1	l.superiortemporal	7271	208	8.78×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$1.08 \times 10^{-12*}_{+}$
LOC101929612	l.superiortemporal	7271	256	1.44×10^{-2}	$1.68 \times 10^{-11*}_{+}$	$1.15 \times 10^{-12*}_{+}$
LOC100996517	l.superiortemporal	7271	34	6.77×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$1.20 \times 10^{-12*}_{+}$
IGLV5-45	l.superiortemporal	7271	179	3.44×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$1.23 \times 10^{-12*}_{+}$
MIS18BP1	l.superiortemporal	7271	553	4.95×10^{-3}	$1.68 \times 10^{-11*}_{+}$	$1.35 \times 10^{-12*}_{+}$
CDR2	l.superiortemporal	7271	260	1.82×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$1.39 \times 10^{-12*}_{+}$
RPL41P2	l.superiortemporal	7271	87	6.04×10^{-3}	$1.68 \times 10^{-11*}_{+}$	$1.59 \times 10^{-12*}_{+}$
LOC101927737	l.superiortemporal	7271	157	7.20×10^{-3}	$1.68 \times 10^{-11*}_{+}$	$1.60 \times 10^{-12*}_{+}$
IGLV1-47	l.superiortemporal	7271	138	1.44×10^{-2}	$1.68 \times 10^{-11*}_{+}$	$1.60 \times 10^{-12*}_{+}$
IGLV7-46	l.superiortemporal	7271	130	9.15×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$1.69 \times 10^{-12*}_{+}$
ZDHHC15	l.superiortemporal	7271	80	1.56×10^{-3}	$1.68 \times 10^{-11*}_{+}$	$1.73 \times 10^{-12*}_{+}$

`*:  below 0.05 after Bonferroni correction`
`+:  below 0.01 after FDR correction`

Table 2: top 20 most significant joint test, per cortex region

GENE	CORTEX	$\lvert V \rvert$	$\lvert G \rvert$	P_G	P_V	P_J
IGLV1-44	l.superiortemporal	7271	174	3.51×10^{-4}	$1.68 \times 10^{-11*}_{+}$	$2.77 \times 10^{-13*}_{+}$
ZNF749	l.entorhinal	1102	321	2.67×10^{-5}	$5.28 \times 10^{-9*}_{+}$	$2.63 \times 10^{-11*}_{+}$
FAM72C	r.superiortemporal	6868	174	9.28×10^{-6}	$4.75 \times 10^{-8*}_{+}$	$2.14 \times 10^{-10*}_{+}$
ZNF749	r.entorhinal	902	321	2.67×10^{-5}	$5.62 \times 10^{-7*}_{+}$	$1.08 \times 10^{-9*}_{+}$
FAM72C	l.cuneus	1630	174	9.28×10^{-6}	$7.27 \times 10^{-6*}_{+}$	$4.22 \times 10^{-9*}_{+}$
ZNF749	l.fusiform	4714	321	2.67×10^{-5}	$8.43 \times 10^{-5*}_{+}$	$4.54 \times 10^{-8}_{+}$
FAM72C	l.middletemporal	4452	174	9.28×10^{-6}	$4.38 \times 10^{-4*}_{+}$	$5.14 \times 10^{-8}_{+}$
FAM72C	r.cuneus	1638	174	9.28×10^{-6}	$8.31 \times 10^{-4}_{+}$	$6.33 \times 10^{-8}_{+}$
ZNF749	l.temporalpole	839	321	2.67×10^{-5}	$9.20 \times 10^{-5*}_{+}$	$7.05 \times 10^{-8}_{+}$
FAM72C	r.precuneus	7975	174	9.28×10^{-6}	$1.44 \times 10^{-3}_{+}$	$7.43 \times 10^{-8}_{+}$
HSPD1P13	l.pericalcarine	1912	86	1.69×10^{-5}	$6.74 \times 10^{-4*}_{+}$	$1.12 \times 10^{-7}_{+}$
FAM72C	r.fusiform	4661	174	9.28×10^{-6}	$5.83 \times 10^{-4*}_{+}$	$1.18 \times 10^{-7}_{+}$
HSPD1P13	r.pericalcarine	1823	86	1.69×10^{-5}	$5.22 \times 10^{-4*}_{+}$	$1.50 \times 10^{-7}_{+}$
FAM72C	r.precentral	10705	174	9.28×10^{-6}	6.73×10^{-3}	$2.15 \times 10^{-7}_{+}$
HSPD1P13	r.paracentral	3831	86	1.69×10^{-5}	1.91×10^{-2}	$2.41 \times 10^{-7}_{+}$
ZNF749	r.temporalpole	817	321	2.67×10^{-5}	$1.55 \times 10^{-3}_{+}$	$3.30 \times 10^{-7}_{+}$
FAM72C	l.precentral	10740	174	9.28×10^{-6}	6.62×10^{-3}	$3.94 \times 10^{-7}_{+}$
FAM72C	l.superiorfrontal	12179	174	9.28×10^{-6}	$1.96 \times 10^{-3}_{+}$	$4.39 \times 10^{-7}_{+}$
FAM72C	l.postcentral	9519	174	9.28×10^{-6}	1.71×10^{-2}	$5.69 \times 10^{-7}_{+}$
ZNF749	l.insula	5229	321	2.67×10^{-5}	$1.52 \times 10^{-3}_{+}$	$6.69 \times 10^{-7}_{+}$

`*:  below 0.05 after Bonferroni correction`
`+:  below 0.01 after FDR correction`

References

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometrythe methods. *NeuroImage*, 11(6):805 – 821.

Baron, J., Chtelat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., and Eustache, F. (2001). In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer's disease. *NeuroImage*, 14(2):298 – 309.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al. (2007). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., and Sabuncu, M. R. (2013). Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *NeuroImage*, 66:249 – 260.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer.

Bouter, Y., Kacprowski, T., Weissmann, R., Dietrich, K., Borgers, H., Brau, A., Sperling, C., Wirths, O., Albrecht, M., Jensen, L. R., Kuss, A. W., and Bayer, T. A. (2014). Deciphering the molecular profile of plaques, memory decline and neuron loss in two mouse models for alzheimers disease by deep sequencing. *Frontiers in Aging Neuroscience*, 6:75.

Buga, A.-M., Scholz, C. J., Kumar, S., Herndon, J. G., Alexandru, D., Cojocaru, G. R., Dandekar, T., and Popa-Wagner, A. (2012). Identification of new therapeutic targets by genome-wide analysis of gene expression in the ipsilateral cortex of aged rats after stroke. *PLOS ONE*, 7(12):1–14.

Castermans, D., Wilquet, V., Parthoens, E., Huysmans, C., Steyaert, J., Swinnen, L., Fryns, J.-P., Van de Ven, W., and Devriendt, K. (2003). The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *Journal of Medical Genetics*, 40(5):352–356.

Chtelat, G., Landeau, B., Eustache, F., Mzenge, F., Viader, F., de la Sayette, V., Desgranges, B., and Baron, J.-C. (2005). Using voxel-based morphometry to map the structural changes associated with rapid conversion in mci: A longitudinal {MRI} study. *NeuroImage*, 27(4):934 – 946.

Cirelli, C., Faraguna, U., and Tononi, G. (2006). Changes in brain gene expression after long-term sleep deprivation. *Journal of Neurochemistry*, 98(5):1632–1645.

Cirulli, E. T. and Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, 11(6):415–425.

Connor, J. R., Tucker, P., Johnson, M., and Snyder, B. (1993). Ceruloplasmin levels in the human superior temporal gyrus in aging and alzheimer's disease. *Neuroscience Letters*, 159(12):88 – 90.

Convit, A., Leon, M. D., Tarshish, C., Santi, S. D., Tsui, W., Rusinek, H., and George, A. (1997). Specific hippocampal volume reductions in individuals at risk for alzheimers disease. *Neurobiology of Aging*, 18(2):131 – 138.

Corradi, J. P., Yang, C., Darnell, J. C., Dalmau, J., and Darnell, R. B. (1997). A post-transcriptional regulatory mechanism restricts expression of the paraneoplastic cerebellar degeneration antigen cdr2 to immune privileged tissues. *Journal of Neuroscience*, 17(4):1406–1415.

Creemers, J. W. M., Nuytens, K., and Tuand, K. (2014). *Neurobeachin Gene in Autism*, pages 825–844. Springer New York, New York, NY.

Cun, L., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems*, pages 396–404. Morgan Kaufmann.

Dai, H. D., Leeder, J. S., and Cui, Y. (2014). A modified generalized fisher method for combining probabilities from dependent tests. *Frontiers in Genetics*, 5(32).

Davies, R. B. (1980). Algorithm as 155: The distribution of a linear combination of χ 2 random variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3):323–333.

Fahlman, S. E. (1988). An empirical study of learning speed in back-propagation networks. Technical report, Air Force Wright Aeronautical Laboratories.

Farris, W., Mansourian, S., Chang, Y., Lindsley, L., Eckman, E. A., Frosch, M. P., Eckman, C. B., Tanzi, R. E., Selkoe, D. J., and Guénette, S. (2003). Insulin-degrading enzyme regulates the levels of insulin, amyloid β-protein, and the β-amyloid precursor protein intracellular domain in vivo. *Proceedings of the National Academy of Sciences*, 100(7):4162–4167.

Fischl, B. (2012). Freesurfer. *NeuroImage*, 62(2):774 – 781. 20 {YEARS} {OF} fMRI20 {YEARS} {OF} fMRI.

Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183 – 192.

Ghiso, J. and Frangione, B. (2002). Amyloidosis and alzheimers disease. *Advanced Drug Delivery Reviews*, 54(12):1539 – 1551. Current treatments and therapeutic targets in Alzheimer's disease.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *International conference on artificial intelligence and statistics*, pages 249–256.

Gmez-Isla, T., Hollister, R., West, H., Mui, S., Growdon, J. H., Petersen, R. C., Parisi, J. E., and Hyman, B. T. (1997). Neuronal loss correlates with but exceeds neurofibrillary tangles in alzheimer's disease. *Annals of Neurology*, 41(1):17–24.

Hansen, L. K. and Salamon, P. (1990). Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12:993–1001.

Helmer, C., Damon, D., Letenneur, L., Fabrigoule, C., Barberger-Gateau, P., Lafont, S., Fuhrer, R., Antonucci, T., Commenges, D., Orgogozo, J., and Dartigues, J. (1999). Marital status and risk of alzheimers disease: A french population-based cohort study. *Neurology*, 53(9):1953.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

Juottonen, K., Laakso, M., Insausti, R., Lehtovirta, M., Pitkänen, A., Partanen, K., and Soininen, H. (1998). Volumes of the entorhinal and perirhinal cortices in alzheimers disease. *Neurobiology of aging*, 19(1):15–22.

Lewerenz, J. and Maher, P. (2009). Basal levels of eif2α phosphorylation determine cellular antioxidant status by regulating atf4 and xct expression. *Journal of Biological Chemistry*, 284(2):1106–1115.

Madsen, B. E. and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*, 5(2):e1000384.

Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine*, 363(2):166–176. PMID: 20647212.

Mansouri, M. R., Marklund, L., Gustavsson, P., Davey, E., Carlsson, B., Larsson, C., White, I., Gustavson, K.-H., and Dahl, N. (2005). Loss of zdhhc15 expression in a woman with a balanced translocation t (x; 15)(q13. 3; cen) and severe mental retardation. *European journal of human genetics*, 13(8):970–977.

Mountjoy, C., Roth, M., Evans, N., and Evans, H. (1983). Cortical neuronal counts in normal elderly controls and demented patients. *Neurobiology of Aging*, 4(1):1 – 11.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L. (2005). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869 – 877. Alzheimer's Disease: 100 Years of Progress.

Nagy, Z. (2015). Therapeutic targets for alzheimer's disease. US Patent App. 14/413,659.

Nehar, S., Mishra, M., and Heese, K. (2009a). Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *{FEBS} Letters*, 583(19):3247 – 3253.

Nehar, S., Mishra, M., and Heese, K. (2009b). Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *{FEBS} Letters*, 583(19):3247 – 3253.

Nestor, S. M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J. L., Fogarty, J., and Bartha, R. (2008). Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database. *Brain*, 131(9):2443–2454.

Niskanen, E., Knnen, M., Mtt, S., Hallikainen, M., Kivipelto, M., Casarotto, S., Massimini, M., Vanninen, R., Mervaala, E., Karhu, J., and Soininen, H. (2011). New insights into alzheimer's disease progression: A combined tms and structural mri study. *PLOS ONE*, 6(10):1–9.

Pandey, J. P. (2010). Comment on genomewide association studies and assessment of risk of disease. *New England Journal of Medicine*, 363(21):2076–2077. PMID: 21083406.

Perfetti, V., Palladini, G., Casarini, S., Navazza, V., Rognoni, P., Obici, L., Invernizzi, R., Perlini, S., Klersy, C., and Merlini, G. (2012). The repertoire of light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, iglv1-44. *Blood*, 119(1):144–150.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575.

Qiu, C., Kivipelto, M., and von Strauss, E. (2009). Epidemiology of alzheimers disease: occurrence, determinants, and strategies toward intervention. *Dialogues Clin Neurosci*, 11(2):111–128.

Qiu, W. Q. and Folstein, M. F. (2006). Insulin, insulin-degrading enzyme and amyloid- peptide in alzheimer's disease: review and hypothesis. *Neurobiology of Aging*, 27(2):190 – 198.

S Beck, J., J Mufson, E., and E Counts, S. (2016). Evidence for mitochondrial upr gene activation in familial and sporadic alzheimers disease. *Current Alzheimer Research*, 13(6):610–614.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., Morris, J. C., Dale, A. M., and Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14(7):721–730.

Seshadri, S., DeStefano, A. L., Au, R., Massaro, J. M., Beiser, A. S., Kelly-Hayes, M., Kase, C. S., D'Agostino, R. B., DeCarli, C., Atwood, L. D., and Wolf, P. A. (2007). 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*, 8(1):S15.

Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., Lawson, W. B., DePaulo, J. R., Gejman, P. V., Sanders, A. R., Johnson, J. K., Adams, P., Chaudhury, S., Jancic, D., Evgrafov, O., Zvinyatskovskiy, A., Ertman, N., Gladis, M., Neimanas, K., Goodell, M., Hale, N., Ney, N., Verma, R., Mirel, D., Holmans, P., and Levinson, D. F. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular Psychiatry*, 16(2):193–201.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., et al. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505.

Theano Development Team (2016). Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408.

Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., and Alkon, D. L. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, 59(4):257–263.

Volders, K., Nuytens, K., and WM Creemers, J. (2011). The autism candidate gene neurobeachin encodes a scaffolding protein implicated in membrane trafficking and signaling. *Current molecular medicine*, 11(3):204–217.

Vsevolozhskaya, O. A., Zaykin, D. V., Barondess, D. A., Tong, X., Jadhav, S., and Lu, Q. (2016). Uncovering local trends in genetic effects of multiple phenotypes via functional linear models. *Genetic epidemiology*, 40(3):210–221.

Wang, A., Xu, S., Zhang, X., He, J., Yan, D., Yang, Z., and Xiao, S. (2011). Ribosomal protein rpl41 induces rapid degradation of atf4, a transcription factor critical for tumour cell survival in stress. *The Journal of pathology*, 225(2):285–292.

Wei, C., Elston, R. C., and Lu, Q. (2015). A weighted U statistic for association analysis considering genetic heterogeneity. *ArXiv e-prints*.

Wei, C., Li, M., He, Z., Vsevolozhskaya, O., Schaid, D. J., and Lu, Q. (2014). A weighted u-statistic for genetic association analyses of sequencing data. *Genetic Epidemiology*, 38(8):699–708.

Wei, C. and Lu, Q. (2015). A generalized similarity u test for multivariate analysis of sequencing data. *arXiv preprint arXiv:1505.01179*.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89(1):82 – 93.

www.python.org (2017). Python program language.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82.

Young, F. B., Butland, S. L., Sanders, S. S., Sutton, L. M., and Hayden, M. R. (2012). Putting proteins in their place: Palmitoylation in huntington disease and other neuropsychiatric diseases. *Progress in Neurobiology*, 97(2):220 – 238. The Neurotoxicity of Mutant Proteins20 Years after the discovery of the first mutant gene involved in neurodegeneration.

Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H., and Weir, B. S. (2002). Truncated product method for combining p-values. *Genetic epidemiology*, 22(2):170–185.

Zhang, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML '04, pages 116–, New York, NY, USA. ACM.