Combined Neuroimaging and Gene Expression Analysis of the Genetic Basis of Brain Plasticity Indicates Across Species Homology

Yonatan Dinai,¹ Lior Wolf,^{2*} and Yaniv Assaf^{3,4}

¹Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel

²Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

³Department of Neurobiology, The George S. Wise Faculty of Life-Sciences, Tel Aviv University, Tel Aviv, Israel

⁴Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

Abstract: Brain plasticity and memory formation depend on the expression of a large number of genes. This relationship had been studied using several experimental approaches and researchers have identified genes regulating plasticity through a variety of mechanisms. Despite this effort, a great deal remains unknown regarding the role of different genes in brain plasticity. Previous studies usually focused on specific brain structures and many of the genes influencing plasticity have yet to be identified. In this work, we integrate results of in vivo neuroimaging studies of plasticity with whole-brain gene expression data for the study of neuroplasticity. Brain regions, found in the imaging study to be involved in plasticity, are first spatially mapped to the anatomical framework of the genetic database. Feature ranking methods are then applied to identify genes that are differentially expressed in these regions. We find that many of our highly ranked genes are involved in synaptic transmission and that some of these genes have been previously associated with learning and memory. We show these results to be consistent when applying our method to gene expression data from four human subjects. Finally, by performing similar experiments in mice, we reveal significant cross species correlation in the ranking of genes. In addition to the identification of plasticity related candidate genes, our results also demonstrate the potential of data integration approaches as a tool to link high level phenomena such as learning and memory to underlying molecular mechanisms. *Hum Brain Mapp* 35:5888–5902, 2014. © 2014 Wiley Periodicals, Inc.

Key words: neuronal plasticity; memory; synapses; meta-analysis; microarray analysis; in situ hybridization; magnetic resonance imaging

INTRODUCTION

Neuroplasticity refers to the ability of the brain to adapt and change over time. Recent brain imaging studies suggest that structural plasticity is a significant component of neuroplasticity and can occur within a timescale of hours [Sagi et al., 2012]. As the mechanisms of this phenomenon are still not well understood, neurogenomic approaches might assist in unraveling the cellular and physiological origins of this process.

In recent decades, neurogenomic approaches have been successful in expanding our understanding of the molecular

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Lior Wolf; School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel 69978. E-mail: wolf@cs.tau.ac.il

Received for publication 18 November 2013; Revised 22 May 2014; Accepted 14 July 2014.

DOI: 10.1002/hbm.22592

Published online 22 July 2014 in Wiley Online Library (wileyonlinelibrary.com).

events underlying the complex process of brain plasticity [Alberini, 2009; Davis and Squire, 1984]. Studies using a variety of learning paradigms in different species have revealed specific genes regulating learning and memory processes through a variety of pathways [Mayford and Kandel, 1999; Nedivi et al., 1993; Tischmeyer and Grimm, 1999]. However, many of the genes involved in plasticity have yet to be identified [Alberini, 2009; McClung and Nestler, 2007].

Various approaches have attempted to identify the gene sets that are involved in brain plasticity. Experimental approaches include the use of cDNA microarrays to identify candidate genes that are upregulated after activity [Luo et al., 2001; Lee et al., 2005; Nedivi et al., 1993] and the use of chromatin-immunoprecipitation [Impey et al., 2004; Sun et al., 2005]. These studies often focus on the expression of genes in specific brain structures, typically the hippocampus. In addition, several computational approaches have been developed, focusing on the identification of binding sites for transcription factors known to be important in brain plasticity [Barth and Yassin, 2008; Pfenning et al., 2007].

In this work, we use the progress in the fields of brain imaging and neurogenomics for the study of brain plasticity. Recent developments in structural brain imaging techniques, particularly magnetic resonance imaging (MRI), are allowing the identification of structural changes in gray and white matter that occur with learning [Blumenfeld-Katzir et al., 2011; Zatorre et al., 2012]. Brain plasticity had also been studied using different functional imaging techniques [Kelly and Garavan, 2005; Poldrack, 2000].

Significant progress had also been made in the field of neurogenomics. Several publicly available resources are now providing gene expression maps, or atlases, giving for the first time the opportunity to study where in the brain each gene is expressed. Notably, the Allen Brain Atlas (ABA) provides genome-wide three-dimensional (3D) data, mapped to a unified anatomical framework, for both mice and humans [Hawrylycz et al., 2012; Lein et al., 2007].

We integrate data from these two sources to study the genetic basis of brain plasticity. Using findings from neuroimaging studies of plasticity, we focus on those brain regions which underwent significant change. The spatial gene expression information, available in the ABA resources, is then used to search for the genes which are most specific to these anatomical regions.

Previous studies have demonstrated the usage of spatially mapped gene expression data. They have used the ABA resources to learn the relationship between gene expression patterns and classical anatomical and functional parcellation of the brain [Lein et al., 2007; Yu and Lee, 2011]. The ABA database has also been used to identify genes important in different brain disorders [Park et al., 2012; Piro et al., 2010], and to study the relation between gene expression and brain connectivity [French and Pavlidis, 2011; Wolf et al., 2011]. However, to the best

of our knowledge, there was no previous attempt to use the genetic data provided by ABA for the study of neuroplasticity.

We have used the spatially mapped gene expression data of ABA for the study of brain plasticity. We integrate the ABA genomic data with results from neuroimaging studies of plasticity. This is done by mapping brain areas, found in the imaging study to be involved in plasticity, to the anatomical framework of the ABA atlas. After defining this anatomical region of interest, we use feature selection methods to rank the genes by their specificity to the target region. We repeat this experiment for both human and mice datasets. Functional annotation analysis of our results indicates that many of our candidate genes are known to be related to brain plasticity. In addition, comparing our results from humans and mice reveals a large agreement in the plasticity related candidate genes that we have found.

MATERIALS AND METHODS

Our approach for identification of plasticity related genes is based on integration of data from neuroimaging studies of plasticity and gene expression atlases. In the following sections, we describe in detail the databases used in the study and the methodology used for gene scoring and for the analysis of gene lists.

Neuroimaging Studies of Plasticity

Sagi et al. [2012] have studied the structural changes in brain tissue of individuals following the performance of a spatial learning and memory task. The task was based on a computer car racing game. As a measure of change, they used the mean diffusivity (MD), an index of tissue density derived from diffusion tensor imaging (DTI) scans. MD maps from all subjects were normalized to the Montreal Neurological Institute (MNI) standard coordinate system, and voxel-based statistical analysis was performed. The map of *P*-values derived in the analysis is thresholded at a significance level of 0.05 resulting in a binary map. This binary map, which localizes brain regions where cellular rearrangement was observed, is used in our study.

For further study of brain regions involved in spatial learning and memory, we use a literature based list of stereotactic coordinates of navigation-related structures (Supporting Information Table IV of Orban et al. [2006]). Data is given as a list of MNI coordinates, and is collected from previously published neuroimaging studies that investigated human navigation using virtual reality devices. Brain imaging in these studies was performed using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET).

In addition to imaging data from human subjects, we use results from a DTI study in mice, demonstrating that an enriched environment causes significant morphological brain changes [Blumenfeld-Katzir et al., 2010]. In this

work, 16 mice were divided into two groups. The control group (REG) was kept in regular shoebox cages, while the other group (EN) was raised for 2 months in an enriched environment. Mice in the enriched environment were kept in larger cages and larger groups which enabled more complex social interaction. The environment was varied over the period of the experiments and the locations of the toys, tunnels, ladders, and food were frequently changed. Mice of the two groups were scanned in a 7T/30 spectrometer. The obtained DTI data were analyzed to produce parametric maps of apparent diffusion coefficient (ADC). The ADC maps were normalized to a mouse brain template that was segmented into regions using the Paxinos atlas. Following normalization, statistic parametric maps of the voxel-based *t*-test between the EN and REG group for the ADC index were computed. In our experiments, we use a list of atlas brain regions where a significant change was observed (regions which contain voxels with *P*-value smaller than 0.01).

To evaluate our approach for selection of plasticity related genes we have also used a control imaging study which is not directly related to learning and memory. We use data from the DTI study of Shu et al. [2013] who analyzed white matter changes across the whole brain of an amyloid precursor protein/presenilin 1 (APP/PS1) mouse model. In our experiment, we use a list of brain regions where a significant MD change was observed in the APP/PS1 mice compared to wild-type controls.

Gene Expression Atlases

We use the publicly available whole-brain gene expression survey of the Allen Human Brain Atlas (AHBA) [Hawrylycz et al., 2012] (<http://human.brain-map.org/>). The dataset consists of microarray-based gene expression profiles for brain regions sampled throughout the brains of four human donors. Each profile contains information for over 62,000 gene probes. Also provided are MRI scans for each of the subjects. Each tissue sample used for the microarray is annotated with a brain structure name and the corresponding MRI coordinate. For two of the brains, microarray data is given for the entire brain (~900 samples), while for the third and fourth brains samples are provided for the left hemisphere only (~400 samples).

In our experiments with data from mice, we use the Allen Brain Atlas (ABA, <http://mouse.brain-map.org>) which provides cellular resolution expression profiles for ~20,000 genes in the adult mouse brain [Lein et al., 2007]. The primary image data of the atlas were generated by a highly automated and methodical application of in situ hybridization (ISH) procedures using custom gene-specific probes. The high resolution primary data from each experiment were reconstructed in 3D and registered to a reference atlas defined in a uniform 200 micrometer grid (The Allen Reference Atlas). The resulting data consist of spatially aligned $67 \times 41 \times 58$ expression energy volumes

for each gene. Expression energy is defined as the average intensity of pixels in the preprocessed ISH image intersecting that voxel [Ng et al., 2007]. The Allen Reference Atlas is also provided in the common volumetric space as a series of 3D masks for each defined anatomical region.

Spatial Alignment of Neuroimaging Data and Gene Expression Atlases

To study relationships between the human imaging data described above and the gene expression data of AHBA, we first bring the two into a common coordinate space. Since both the binary map, based on the DTI study of Sagi et al. [2012], and the literature based coordinates of Orban et al. [2006] are defined in the MNI coordinate space, we can use the same spatial transformation for the two datasets. We, therefore, perform 3D image registration between the T1 MRI scans of each of the four AHBA postmortem brains and the standard MNI T1 volume. Image registration is done using the SPM8 software package (<http://www.fil.ion.ucl.ac.uk/spm/>) and involves the following steps. First, an affine 3D registration that maximizes the normalized mutual information is computed. In the second stage, the brains are aligned using a nonlinear transformation. Finally, the SPM deformations utility is used to obtain the deformation field computed in the nonlinear stage of the registration. For each of the four AHBA brains, we then use the MRI coordinates of the tissue samples used for microarray profiling, and obtain the corresponding locations in the MNI space.

Spatial alignment is also performed as the first step in analyzing the mice data. To match brain regions that were found in the DTI study of Blumenfeld-Katzir et al. [2010] and in the control study of Shu et al. [2013] to the mice gene expression data of ABA, we use the high resolution coronal atlas images provided by the ABA website. The coronal atlas sections hold detailed 3D annotation in different brain regions including the cortex.

To align the ABA coronal atlas to the 3D voxel data, we first downloaded the full-color images of the atlas from the ABA website. We then used color segmentation and morphological operations to segment the different brain structures in each of the coronal slices. Finally, coronal slices were aligned to their corresponding slices in the provided 3D reference atlas using a two-dimensional similarity transformation (translation, rotation, and scaling). Similarity transformation gives an accurate registration in this case since the same coronal slices were used for the construction of the original ABA 3D reference volume.

Data Representation and Samples Labeling

We follow the same gene selection procedure for both human and mice data. In each case, the data is represented by a measurement matrix X that contains as many rows as

measured genes and as many columns as data points. In our setting, each data point represents a different brain coordinate.

In the experiments involving data from humans, gene expression data is obtained from the AHBA. In this case, the rows of the measurement matrix X correspond to the gene probes analyzed in the microarray experiments of AHBA. The number of columns in X is the number of brain tissue samples analyzed in those experiments. The MRI coordinates for each tissue sample are also provided by AHBA.

In the experiments involving data from mice, data is obtained from the adult mouse ABA. The rows of the measurement matrix X in this case correspond to the genes analyzed in the ISH experiments of ABA. Each column of X corresponds to a different voxel in the ABA reference atlas volume.

In each experiment, binary labels are given per sample (column of X). The labels are given based on whether or not the sample lies within the plasticity related regions of interest defined in the neuroimaging studies. This labeling procedure is detailed below.

In the experiments involving the DTI data of Sagi et al. [2012], labels are given by the following procedure. The MRI coordinate of each tissue sample is mapped to the MNI space using the image registration procedure described above. Those points that lie inside the binary mask of Sagi et al. [2012] or within a distance of 2 voxels (4 mm) from the binary mask are given a positive label (+1). The choice of a 4 mm neighborhood is related to the average size of the tissue samples analyzed in the microarray experiments of AHBA (~2–5 mm, Hawrylycz et al. [2012]). Points with distance greater than four voxels (8 mm) from the binary mask are given a negative label (−1). The remaining points are discarded from the analysis. In the experiments involving the literature based coordinates of Orban et al. [2006], we first define a binary volume in the MNI space by performing a 3D dilation operation on each of the brain coordinates given in the list. We then repeat the same procedure as described above for the data of Sagi et al. [2012], replacing the 3D mask with the computed binary volume.

In the experiments with data from mice, we use the detailed reference volume annotation, described above, to label each voxel and the corresponding column of the measurement matrix X . Voxels that lie in one of the regions specified in the list of Blumenfeld-Katzir et al. [2010], are given a label of +1, while the other voxels are given a label of −1. The same labeling approach was used in the experiment involving brain regions that were found in the control DTI study of Shu et al. [2013].

Gene Ranking

Given the measurement matrix X , and the labels vector Y , numerous feature selection methods could be applied to identify genes that are differentially expressed between

the positive and negative brain areas. Often these methods are used to select genes that are expressed differently in two groups of patients, for example for the identification of disease biomarkers [Jeffery et al., 2006].

We chose to use the method of the area under the receiver operating characteristic (ROC) curve, which was shown to perform well in the task of gene selection [Jeffery et al., 2006; Parodi et al., 2003]. The ROC curve displays the relationship between the proportion of true positive (sensitivity) and false positive (1-specificity) classifications resulting from each possible decision threshold value in a binary classification problem. The area under this curve (AUC) is a measure of classifier accuracy, and when applied to a gene expression profile, it provides an estimate of the probability that a gene is regulated between the two groups. In our setting, we treat the expression values of the k -th gene, $\{X_{kj}\}_{j=1}^N$, as the real-valued outputs of a binary classifier. The importance of each gene is evaluated independently by measuring the gene's ability to predict the class labels given by the vector Y . The evaluation is done by computing the ROC curve for each gene and taking the AUC as an importance score.

An overview of our gene ranking methodology is presented in Figure 1. The workflow presented is for the analysis of data from humans, but similar steps were applied for the study of the mice data. We performed image registration to relate regions of interest, defined in different neuroimaging studies, to the gene expression measurements of AHBA. Each tissue sample was given a positive or negative label based on whether or not it was contained in the analyzed region of interest as detailed above. Gene ranking based on the area under the ROC curve measure was then applied to identify those genes that were differentially expressed between the positive and negative samples.

Analysis of the Selected Genes

For the AHBA microarray data, expression values are given for gene probes and a gene is often represented by more than one probe. To obtain a list of the top ranked genes, for each gene we take as a score the maximum AUC from the set of probes related to that gene. Similar results are obtained when taking the median score instead.

We perform gene ontology (GO) enrichment analysis for both the human and the mice datasets, using lists of the top 200 ranked genes of each experiment as input to the database for annotation, visualization and integrated discovery (DAVID) bioinformatics tool [Huang et al., 2008]. For background, the complete list of genes in each experiment was used.

In addition, we extracted lists of the top 200 genes associated with different medical disorders from the HuGe database [Yu et al., 2010]. We also obtained lists of presynaptic and postsynaptic related genes from previous studies [Collins, 2006; Liu, 2009]. The significance of the intersection of these lists and lists of the top 200 genes extracted

A. Image Registration and Tissue Samples Labeling

B. Gene Ranking using ROC Analysis

Figure 1.

Methodology. A. The standard MNI T1 reference volume is aligned to the MRI scan of each AHBA donor. Following alignment, the tissue samples analyzed in the microarray experiments of AHBA are separated into groups of positive / negative samples, based on a region of interest found in a plasticity neuroimaging study. **B.** Genes are ranked using the area under the ROC curve as a measure of feature importance.

by the method described above is evaluated using permutation tests. In this experiment, gene ranking and selection is repeated 1,000 times with random assignments of the N

expression profiles to the N brain locations (this can be seen as a shuffling operation on the columns of the expression matrix X , while the labels vector Y is kept fixed).

TABLE I. DAVID functional annotation chart for ABA human subject I for genes selected using the DTI data of Sagi et al. [2012]

Category	Term	Count	P Value	Benjamini
SP_PIR_KEYWORDS	synapse	19	8.40 E -12	1.95 E -09
GOTERM_BP_FAT	GO:0007268~synaptic transmission	21	4.40 E -11	5.00 E -08
GOTERM_CC_FAT	GO:0044456~synapse part	19	4.69 E -11	1.14 E -08
GOTERM_CC_FAT	GO:0045202~synapse	22	5.38 E -11	6.53 E -09
GOTERM_BP_FAT	GO:0019226~transmission of nerve impulse	21	7.73 E -10	4.39 E -07
SP_PIR_KEYWORDS	cell junction	21	6.22 E -09	7.22 E -07
GOTERM_CC_FAT	GO:0030054~cell junction	22	4.88 E -08	3.95 E -06
GOTERM_CC_FAT	GO:0045211~postsynaptic membrane	12	1.18 E -07	7.16 E -06
SP_PIR_KEYWORDS	postsynaptic cell membrane	11	2.09 E -07	1.62 E -05
SP_PIR_KEYWORDS	phosphoprotein	109	5.93 E -07	3.44 E -05
GOTERM_CC_FAT	GO:0008021~synaptic vesicle	9	8.64 E -07	4.20 E -05
GOTERM_MF_FAT	GO:0022836~gated channel activity	16	1.24 E -06	4.26 E -04
GOTERM_BP_FAT	GO:0007267~cell-cell signaling	22	1.24 E -06	4.71 E -04
GOTERM_CC_FAT	GO:0008328~ionotropic glutamate receptor complex	5	9.39 E -06	3.80 E -04
GOTERM_MF_FAT	GO:0005230~extracellular ligand-gated ion channel activity	8	1.31 E -05	0.002243
GOTERM_MF_FAT	GO:0005216~ion channel activity	16	1.83 E -05	0.002093
GOTERM_MF_FAT	GO:0022838~substrate specific channel activity	16	2.55 E -05	0.002187
SP_PIR_KEYWORDS	neurotransmitter receptor	7	2.63 E -05	0.001217
SP_PIR_KEYWORDS	ionic channel	14	2.68 E -05	0.001034
INTERPRO	IPR001508:NMDA receptor	5	3.13 E -05	0.012646
INTERPRO	IPR015683:Glutamate receptor-related	5	3.13 E -05	0.012646
INTERPRO	IPR001320:Ionotropic glutamate receptor	5	3.13 E -05	0.012646
INTERPRO	IPR019594:Glutamate receptor, L-glutamate/glycine-binding	5	3.13 E -05	0.012646
GOTERM_MF_FAT	GO:0004970~ionotropic glutamate receptor activity	5	3.72 E -05	0.002552
GOTERM_MF_FAT	GO:0015267~channel activity	16	3.82 E -05	0.002185
UP_SEQ_FEATURE	splice variant	105	3.92 E -05	0.028467
GOTERM_MF_FAT	GO:0022803~passive transmembrane transporter activity	16	3.92 E -05	0.001927
SP_PIR_KEYWORDS	alternative splicing	105	4.22 E -05	0.001398
GOTERM_CC_FAT	GO:0043005~neuron projection	14	4.24 E -05	0.001469
GOTERM_MF_FAT	GO:0005234~extracellular-glutamate-gated ion channel activity	5	4.67 E -05	0.002004

With each iteration, the size of the intersection of the top 200 selected genes with the target lists of genes is computed. Finally, using the generated distribution and the intersections calculated using the true data a *P*-value is computed for each target genes set.

To check the repeatability of gene ranking for the four AHBA subjects and the two imaging datasets used for labeling, we perform correlation tests. From the gene ranking procedure described above, we obtain for each of the two labelings the AUC scores vectors for the four AHBA subjects, resulting in a set of eight gene scores vectors. We compute the Spearman correlation between each pair of vectors in this set. The *P*-value for the correlation is then calculated by testing the hypothesis of no correlation against the alternative that there is a nonzero correlation.

We perform additional correlation tests to measure the repeatability of genes selected in the experiments with human data and the genes selected in the experiments with mice data. To relate the mice genes to the human genes, we

found the homolog human Entrez ID for each mouse gene symbol using RefGene (<http://refgene.com/>). We obtained a list of 9,743 genes that were mapped correctly, and were represented in both the human and the mice gene expression datasets. We then computed the Spearman correlation between the AUC scores of these genes as calculated with the mice data [Blumenfeld-Katzir et al., 2010], and the scores calculated using the human data.

To calculate a summary score for each gene across the different experiments and species, we perform the following computation. For each gene in the mice-human homolog genes set described above, we calculate the mean of the AUC scores in the different experiments we have performed with the ABA human data (overall 8 scores). We then take the mean of this score with the AUC score computed for the same gene in our experiment with mice data. Finally, we analyze the resulting ranked list of 9,743 genes using the GOrilla tool for GO enrichment analysis and visualization tools [Eden et al., 2009].

TABLE II. DAVID functional annotation chart for ABA human subject 2 for genes selected using the DTI data of Sagi et al. [2012]

Category	Term	Count	<i>P</i> Value	Benjamini
GOTERM_CC_FAT	GO:0045202~synapse	20	2.49 E -08	6.47 E -06
SP_PIR_KEYWORDS	synapse	14	3.63 E -07	8.68 E -05
GOTERM_CC_FAT	GO:0044456~synapse part	15	9.54 E -07	1.24 E -04
GOTERM_CC_FAT	GO:0030054~cell junction	19	2.90 E -05	0.002513
SP_PIR_KEYWORDS	cell junction	15	7.60 E -05	0.009038
INTERPRO	IPR000008:C2 calcium-dependent membrane targeting	9	1.33 E -04	0.049226
GOTERM_CC_FAT	GO:0014069~postsynaptic density	7	1.70 E -04	0.010971
INTERPRO	IPR018029:C2 membrane targeting protein	8	1.70 E -04	0.031709
SMART	SM00239:C2	9	3.78 E -04	0.033819
INTERPRO	IPR002290:Serine/threonine protein kinase	11	5.96 E -04	0.072548
GOTERM_CC_FAT	GO:0043005~neuron projection	13	6.16 E -04	0.031526
UP_SEQ_FEATURE	topological domain:Lumenal	14	7.13 E -04	0.355404
GOTERM_CC_FAT	GO:0045211~postsynaptic membrane	8	9.86 E -04	0.041864
KEGG_PATHWAY	hsa04916:Melanogenesis	7	0.001151	0.105739
GOTERM_CC_FAT	GO:0031410~cytoplasmic vesicle	18	0.001151	0.041897
KEGG_PATHWAY	hsa04720:Long-term potentiation	6	0.001284	0.060441
GOTERM_CC_FAT	GO:0016023~cytoplasmic membrane-bounded vesicle	16	0.001712	0.054186
GOTERM_MF_FAT	GO:0005524~ATP binding	29	0.001735	0.468711
GOTERM_CC_FAT	GO:0008021~synaptic vesicle	6	0.001767	0.049823
GOTERM_CC_FAT	GO:0031982~vesicle	18	0.001826	0.046431
SMART	SM00220:S_TKc	11	0.001892	0.082560
GOTERM_BP_FAT	GO:0007268~synaptic transmission	11	0.001923	0.903270
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	29	0.002126	0.321167
INTERPRO	IPR017442:Serine/threonine protein kinase-related	12	0.002140	0.183763
GOTERM_CC_FAT	GO:0044431~Golgi apparatus part	11	0.002197	0.050659
GOTERM_MF_FAT	GO:0004672~protein kinase activity	16	0.002238	0.238110
GOTERM_CC_FAT	GO:0031988~membrane-bounded vesicle	16	0.002343	0.049558
GOTERM_MF_FAT	GO:0004674~protein serine/threonine kinase activity	13	0.002354	0.193094
GOTERM_CC_FAT	GO:0030135~coated vesicle	8	0.002437	0.047628
INTERPRO	IPR000719:Protein kinase, core	14	0.002464	0.170611

RESULTS

Human Data Analysis

For each of the four AHBA subjects, tissue samples were given a binary label based on data from Sagi et al., [2012]. The top 200 genes, selected using the AUC score, were then analyzed with the DAVID functional annotations tool [Huang et al., 2008]. Lists of the top functional annotations for the four subjects are given in Tables I–IV. Functional terms related to synaptic activity appear at the top of the lists for the four subjects. The top functional annotations for Subject 1 (Table I) also include terms related to cell communication and neurotransmitter activity. For Subject 2 (Table II), many of the top functional annotations are again related to synaptic activity and cell communication. A total of six of the top ranked genes are directly related to long-term potentiation, one of the biological processes underlying synaptic plasticity. In the results of Subject 3 (Table III), we noticed that a large percentage of the top genes are related to phosphoproteins. Interestingly, this was also observed in the results of Subject 1 but not in those of Subjects 2 and 4.

Next, we studied the enrichment of genes related to different medical disorders and functional terms within the top genes selected in our human experiments. Results of permutation tests for the four AHBA subjects are shown in Table V. Genes related to autism are enriched in the top 200 genes selected for AHBA Subject 1 (*P*-value 0.017). However, genes related to Alzheimer's disease and dementia are not enriched in our selected genes for the four AHBA subjects. The permutation tests indicate that genes related to postsynaptic activity are enriched in our selected genes (*P*-values 0.002, 0.006, 0.017, and <0.001 for Subjects 1, 2, 3, and 4, respectively). For Subjects 1, 2, and 4, we also find significant intersection with presynaptic genes (*P*-values 0.003, 0.002, and 0.002).

Literature Human Data Analysis

As a validation to our approach and to further study the relationship between gene expression and spatial memory, we used additional neuroimaging data from the literature. We repeated the DAVID functional

TABLE III. DAVID functional annotation chart for ABA human subject 3 for genes selected using the DTI data of Sagi et al. [2012]

Category	Term	Count	<i>P</i> Value	Benjamini
SP_PIR_KEYWORDS	phosphoprotein	110	1.3 E -07	3.02 E -05
GOTERM_CC_FAT	GO:0014069~postsynaptic density	9	8.04 E -07	1.89 E -04
SP_PIR_KEYWORDS	postsynaptic cell membrane	10	1.92 E -06	2.14 E -04
SP_PIR_KEYWORDS	synapse	13	2.04 E -06	1.51 E -04
SP_PIR_KEYWORDS	cell junction	17	3.60 E -06	2.01 E -04
UP_SEQ_FEATURE	compositionally biased region:Ser-rich	17	7.71 E -06	0.005753
GOTERM_CC_FAT	GO:0044456~synapse part	13	1.16 E -05	0.001356
UP_SEQ_FEATURE	compositionally biased region:Poly-Pro	16	2.09 E -05	0.007794
GOTERM_CC_FAT	GO:0045202~synapse	15	2.31 E -05	0.001811
UP_SEQ_FEATURE	short sequence motif:PDZ-binding	7	1.11 E -04	0.027295
GOTERM_CC_FAT	GO:0030054~cell junction	17	1.16 E -04	0.006770
UP_SEQ_FEATURE	splice variant	102	1.33 E -04	0.024475
SP_PIR_KEYWORDS	alternative splicing	102	1.42 E -04	0.006304
INTERPRO	IPR001849:Pleckstrin homology	12	1.82 E -04	0.073779
GOTERM_BP_FAT	GO:0006793~phosphorus metabolic process	23	2.27 E -04	0.197836
GOTERM_BP_FAT	GO:0006796~phosphate metabolic process	23	2.27 E -04	0.197836
KEGG_PATHWAY	hsa04360:Axon guidance	8	2.66 E -04	0.018190
GOTERM_MF_FAT	GO:0004672~protein kinase activity	17	5.86 E -04	0.176841
GOTERM_CC_FAT	GO:0045211~postsynaptic membrane	8	5.89 E -04	0.027324
GOTERM_MF_FAT	GO:0015467~G-protein activated inward rectifier potassium channel activity	3	6.52 E -04	0.102614
SP_PIR_KEYWORDS	kinase	18	6.83 E -04	0.025082
GOTERM_BP_FAT	GO:0019226~transmission of nerve impulse	12	8.16 E -04	0.327331
GOTERM_BP_FAT	GO:0007268~synaptic transmission	11	8.56 E -04	0.242310
GOTERM_BP_FAT	GO:0016310~phosphorylation	19	9.20 E -04	0.200377
GOTERM_MF_FAT	GO:0030695~GTPase regulator activity	13	0.001031	0.107899
GOTERM_BP_FAT	GO:0006468~protein amino acid phosphorylation	17	0.001061	0.186428
GOTERM_MF_FAT	GO:0060589~nucleoside-triphosphatase regulator activity	13	0.001248	0.098495
SMART	SM00233:PH	12	0.001414	0.159746
INTERPRO	IPR011993:Pleckstrin homology-type	11	0.001493	0.269861
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	30	0.001800	0.112761

annotation analysis for the top 200 genes selected using the literature based coordinates of Orban et al. [2006] for each of the four AHBA subjects. For Subjects 1, 2, and 4 (Supporting Information Tables I, II, and IV), we find that the top genes are again related to synaptic and neurotransmitter activity. However, in the genes selected for Subject 3 (Supporting Information Table III) the top functional annotations are not directly related to synaptic activity. It is important to note that unlike Subjects 1 and 2, the gene expression data for AHBA Subjects 3 and 4 is provided only for the left hemisphere. Hence, only a partial set of coordinates is used to define our region of interest for the gene selection procedure.

For this dataset, we also examined the enrichment of different medical disorders and functional groups related genes in the top selected genes. Results of the permutations tests are provided in Supporting Information Table V. We find that presynaptic related genes are enriched in the lists of Subjects 1, 2, and 4 (*P*-value <0.001, 0.007, and 0.016, respectively).

Repeatability Test of Human Data

Despite the apparent differences in the enrichment test results, the selected genes are well correlated across different experiments. The Spearman correlation coefficients between the various ranked lists that were obtained and the corresponding *P*-values are provided in Supporting Information Table VI. We find a high correlation between different subjects, indicating the stability of the results for different gene expression experiments. A high correlation is also found for rankings obtained with the two neuroimaging data sources we have used. The repeatability obtained across the different human subjects is also portrayed in Figure 2. Each plate in the figure shows the distribution of gene scores, obtained using the DTI data of Sagi et al. [2012] for a pair of AHBA subjects.

Mice Data Analysis

To further validate the utility of our method in selecting plasticity related genes, we have also used mice data. We

TABLE IV. DAVID functional annotation chart for ABA human subject 4 for genes selected using the DTI data of Sagi et al. [2012]

Category	Term	Count	P Value	Benjamini
GOTERM_CC_FAT	GO:0045211~postsynaptic membrane	14	2.59 E -09	6.20 E -07
GOTERM_CC_FAT	GO:0014069~postsynaptic density	11	4.70 E -09	5.62 E -07
SP_PIR_KEYWORDS	postsynaptic cell membrane	12	2.20 E -08	4.57 E -06
GOTERM_CC_FAT	GO:0030054~cell junction	23	4.09 E -08	3.26 E -06
SP_PIR_KEYWORDS	synapse	15	5.78 E -08	6.01 E -06
GOTERM_CC_FAT	GO:0044456~synapse part	16	7.29 E -08	4.36 E -06
GOTERM_CC_FAT	GO:0005886~plasma membrane	68	9.91 E -07	4.74 E -05
GOTERM_MF_FAT	GO:0005216~ion channel activity	18	1.09 E -06	3.47 E -04
SP_PIR_KEYWORDS	cell junction	18	1.11 E -06	7.67 E -05
GOTERM_BP_FAT	GO:0006811~ion transport	25	1.20 E -06	0.001239
SP_PIR_KEYWORDS	ionic channel	16	1.34 E -06	6.97 E -05
SP_PIR_KEYWORDS	membrane	97	1.34 E -06	5.59 E -05
GOTERM_BP_FAT	GO:0030030~cell projection organization	17	1.54 E -06	7.95 E -04
GOTERM_CC_FAT	GO:0045202~synapse	17	1.54 E -06	6.15 E -05
GOTERM_BP_FAT	GO:0030001~metal ion transport	19	1.61 E -06	5.55 E -04
GOTERM_MF_FAT	GO:0022838~substrate specific channel activity	18	1.61 E -06	2.56 E -04
GOTERM_MF_FAT	GO:0015267~channel activity	18	2.59 E -06	2.75 E -04
GOTERM_MF_FAT	GO:0022803~passive transmembrane transporter activity	18	2.68 E -06	2.13 E -04
GOTERM_BP_FAT	GO:0031175~neuron projection development	14	2.82 E -06	7.31 E -04
GOTERM_CC_FAT	GO:0044459~plasma membrane part	47	2.92 E -06	9.98 E -05
GOTERM_BP_FAT	GO:0030182~neuron differentiation	18	3.36 E -06	6.97 E -04
GOTERM_BP_FAT	GO:0006812~cation transport	20	4.41 E -06	7.62 E -04
GOTERM_MF_FAT	GO:0022836~gated channel activity	15	7.52 E -06	4.78 E -04
SP_PIR_KEYWORDS	ion transport	20	1.05 E -05	3.66 E -04
GOTERM_BP_FAT	GO:0032989~cellular component morphogenesis	16	1.65 E -05	0.002435
GOTERM_BP_FAT	GO:0006816~calcium ion transport	10	1.80 E -05	0.002328
GOTERM_BP_FAT	GO:0000902~cell morphogenesis	15	2.16 E -05	0.002480
UP_SEQ_FEATURE	region of interest:Calmodulin-binding	7	2.36 E -05	0.017026
GOTERM_CC_FAT	GO:0030666~endocytic vesicle membrane	6	2.73 E -05	8.14 E -04
SP_PIR_KEYWORDS	neurotransmitter receptor	7	2.87 E -05	8.51 E -04

TABLE V. Intersection of the top 200 selected genes using the DTI data of Sagi et al. [2012] with the top 200 genes related to different medical disorders and functional groups

	Subject 1 P-Value	Subject 2 P-Value	Subject 3 P-Value	Subject 4 P-Value
Alzheimer's Disease	0.667	0.670	1.000	0.904
Autistic Disorder	0.017	0.730	0.190	0.698
Dementia	0.448	0.690	1.000	0.908
Schizophrenia	0.054	0.307	0.757	0.556
Coronary Disease	0.819	0.819	0.853	1.000
Diabetes	0.324	1.000	0.337	0.860
Glioma	1.000	1.000	0.602	1.000
Hypertension	0.157	0.864	0.595	1.000
Obesity	0.267	0.224	0.914	1.000
Postsynaptic	0.002	0.006	0.017	<0.001
Presynaptic	0.003	0.002	0.208	0.002

integrated the results of the DTI study of Blumenfeld-Katzir et al. [2010] with gene expression data from the adult mouse brain provided by ABA. Using a similar approach to that described for the human datasets, genes were ranked based on the AUC score. The top 200 genes were then analyzed with the DAVID functional annotation tools. The most significant functional terms found in the analysis are listed in Table VI. Similar to the human results, most of the top functional annotations are related to synaptic activity, cell communication and neurotransmitter activity. In addition, eight of the top 200 genes are directly related to learning and memory (P -value < 0.0001).

A Control Imaging Study in Mice

To evaluate the effect of the selected neuroimaging study on the gene groups identified by our approach, we have also used results of a DTI study which did not focus on changes related to a specific learning paradigm. We use a list of brain regions found in the work of Shu et al. [2013] who analyzed white matter changes across the whole brain of an APP/PS1 mouse model. We integrated

TABLE VI. DAVID functional annotation chart for the top 200 genes selected using the ABA adult mouse gene expression atlas, and the DTI results of Blumenfeld-Katzir et al. [2010]

Category	Term	Count	P Value	Benjamini
SP_PIR_KEYWORDS	postsynaptic cell membrane	11	5.16 E -08	1.02 E-05
GOTERM_CC_FAT	GO:0005886~plasma membrane	57	5.61 E -08	1.08 E-05
GOTERM_MF_FAT	GO:0022836~gated channel activity	15	2.33 E -07	7.30 E-05
KEGG_PATHWAY	mmu04020:Calcium signaling pathway	14	2.58 E -07	2.01 E-05
SP_PIR_KEYWORDS	synapse	13	4.67 E -07	4.60 E-05
SP_PIR_KEYWORDS	cell membrane	37	8.02E -07	5.27 E-05
GOTERM_BP_FAT	GO:0030001~metal ion transport	18	1.44 E -06	0.001523
SP_PIR_KEYWORDS	phosphoprotein	86	2.08 E -06	1.02 E-04
GOTERM_MF_FAT	GO:0046873~metal ion transmembrane transporter activity	14	2.12 E -06	3.31 E-04
SP_PIR_KEYWORDS	membrane	78	2.43 E-06	9.56 E -05
GOTERM_MF_FAT	GO:0005261~cation channel activity	13	2.43 E -06	2.53 E -04
GOTERM_BP_FAT	GO:0006812~cation transport	19	2.70 E -06	0.001425
GOTERM_MF_FAT	GO:0005216~ion channel activity	15	3.09 E -06	2.42 E -04
GOTERM_CC_FAT	GO:0045202~synapse	15	3.23 E -06	3.12 E -04
GOTERM_CC_FAT	GO:0045211~postsynaptic membrane	10	4.30 E -06	2.77 E -04
GOTERM_MF_FAT	GO:0022838~substrate specific channel activity	15	4.43 E -06	2.77 E -04
GOTERM_MF_FAT	GO:0005244~voltage-gated ion channel activity	11	4.94 E -06	2.58 E -04
GOTERM_MF_FAT	GO:0022832~voltage-gated channel activity	11	4.94 E -06	2.58 E -04
GOTERM_MF_FAT	GO:0015267~channel activity	15	5.20 E -06	2.32 E -04
GOTERM_MF_FAT	GO:0022803~passive transmembrane transporter activity	15	5.20 E -06	2.32 E -04
GOTERM_BP_FAT	GO:0007267~cell-cell signaling	14	5.28 E-06	0.001857
GOTERM_BP_FAT	GO:0006811~ion transport	22	5.44 E -06	0.001435
GOTERM_CC_FAT	GO:0043005~neuron projection	13	5.60 E -06	2.70 E -04
GOTERM_CC_FAT	GO:0044456~synapse part	12	8.13 E -06	3.14 E -04
GOTERM_CC_FAT	GO:0014069~postsynaptic density	7	1.05 E -05	3.39 E -04
GOTERM_MF_FAT	GO:0022843~voltage-gated cation channel activity	9	2.03 E -05	7.93 E -04
GOTERM_BP_FAT	GO:0006816~calcium ion transport	9	2.43 E -05	0.005119
GOTERM_BP_FAT	GO:0007611~learning or memory	8	2.60 E -05	0.004566
GOTERM_MF_FAT	GO:0005249~voltage-gated potassium channel activity	8	3.08 E -05	0.001070
GOTERM_BP_FAT	GO:0060079~regulation of excitatory postsynaptic membrane potential	5	3.76 E -05	0.005667

this data with the ABA atlas of the adult mouse brain and repeated the process of gene ranking. The top 200 genes for this experiment were also analyzed using DAVID functional annotation tools and the top terms appear in Table VII. We find that the top genes are dominated by ion channel activity related genes, while terms related to synaptic activity are less represented compared to the results obtained with the data of Blumenfeld-Katzir et al. [2010].

Across Species Analysis

To study the similarity between the genes selected for mice and humans we constructed a mice-human homolog genes set, obtaining a set of 9,743 genes. We then computed the Spearman correlation between the gene scores obtained for mice (using the DTI study of Blumenfeld-Katzir et al. [2010]) and the corresponding gene scores in

the experiments with human data (Supporting Information Table VII). Significant correlation (P -value < 0.0001) is found between the gene scores vector from the mice experiment and the eight vectors of gene scores from the human data (4 subjects $\times$ 2 imaging studies). Finally, we calculated a summary score for each gene in the homolog data set by averaging the AUC scores found for the gene in our experiments. The resulting ranked list of genes was analyzed using the GOrilla tool for GO enrichment analysis tools [Eden et al., 2009] (Figure 3).

DISCUSSION

We have presented a computational approach for identifying plasticity related genes. Our method is based on integration of spatially mapped gene expression and new brain imaging techniques for identifying structural and functional changes that occur with learning.

Figure 2.

Relationship of gene scores for each pair of subjects in the AHBA. Each point in the scatter plots corresponds to a single gene. The AUC scores were computed based on the brain regions identified in the DTI study of Sagi et al. [2012].

We found that many of the known plasticity related genes are identified using the described approach. Some of the top ranked genes are directly related to learning and memory while many of the other genes are related to biological processes involved in plasticity, such as synaptic transmission and dendrite development. In our experiments with the whole brain microarray data from human subjects, we found high intersubject repeatability in the ranking of genes. Functional annotation analysis revealed that the top ranked genes from different subjects share a large functional overlap. In our experiments with data from mice, we again identified many genes known to be involved in learning and memory. Finally, we found significant cross-species repeatability in the ranking of genes for humans and mice.

Underscoring the strength of the observations is the fact that the analysis was not performed on the same samples. The genetic database and the in vivo MRI results were obtained from different subjects. The fact that high correlation, in the relevant biological domain and across species, is found indicates that the identified genes are related to neuroplasticity beyond the intersubject variability across the population, revealing a deep common background to this process.

In a control experiment, we used the voxel based analysis results of a study that found MD changes in mice carrying APP/PS1 compared with wild-type controls [Shu et al. 2013]. We chose this work since some of the regions

reported in the plasticity imaging study of Blumenfeld-Katzir et al. [2010] overlap those reported in Shu et al. [2013], but there are a few regions that do not overlap. While in the experiment involving plasticity imaging data we mainly found genes affecting synaptic activity (see Table VI), the MRI-gene correlation of Shu et al. [2013] highlights mainly ion channel related genes. This observation supports the validity of our observations since it shows that the MRI-gene correlations are sensitive to the affected regions and specific to the task or condition that caused the MD changes.

Our results on human gene expression data are limited by the availability of a small number of subjects in the atlas. Although our results already show significant intersubject repeatability, the reliability of the proposed method should be tested when the atlas is extended to include data from more human subjects. It is important to note that for only two of the four subjects (Subjects 1 and 2) in the ABA human dataset, gene expression data is available for both hemispheres. As we integrate this data with whole brain asymmetric data from neuroimaging studies, we expect a higher reliability and specificity for the genes found for Subjects 1 and 2. Additional data from two human subjects was added to the ABA resources recently. However, since the microarray data from these subjects was normalized using an algorithm different from the one applied to the microarray data used here, we have not included this data in our experiments.

TABLE VII. DAVID functional annotation chart for the top 200 genes selected using the ABA adult mouse gene expression atlas, and the DTI results of Shu et al. [2013]

Category	Term	Count	P Value	Benjamini
GOTERM_MF_FAT	GO:0022832~voltage-gated channel activity	14	3.60 E -08	9.89 E -06
GOTERM_MF_FAT	GO:0005244~voltage-gated ion channel activity	14	3.60 E -08	9.89 E -06
UP_SEQ_FEATURE	region of interest:Segment H5 (pore-forming)	7	1.77 E -07	9.24 E -05
GOTERM_MF_FAT	GO:0022836~gated channel activity	16	2.23 E -07	3.07 E -05
GOTERM_BP_FAT	GO:0030001~metal ion transport	20	3.17 E -07	2.65 E -04
SP_PIR_KEYWORDS	synapse	14	4.40E -07	8.13 E -05
GOTERM_MF_FAT	GO:0005216~ion channel activity	17	5.84 E -07	5.36 E -05
GOTERM_MF_FAT	GO:0005249~voltage-gated potassium channel activity	10	5.89 E -07	4.05 E -05
GOTERM_MF_FAT	GO:0022843~voltage-gated cation channel activity	11	6.99 E -07	3.84 E -05
GOTERM_BP_FAT	GO:0006812~cation transport	21	7.82 E -07	3.27 E -04
GOTERM_MF_FAT	GO:0022838~substrate specific channel activity	17	8.83 E -07	4.05 E -05
GOTERM_MF_FAT	GO:0022803~passive transmembrane transporter activity	17	1.04 E -06	4.08 E -05
GOTERM_MF_FAT	GO:0015267~channel activity	17	1.04 E -06	4.08 E -05
UP_SEQ_FEATURE	short sequence motif:Selectivity filter	8	1.13 E -06	2.95 E -04
SP_PIR_KEYWORDS	voltage-gated channel	11	1.42 E -06	1.31 E -04
GOTERM_MF_FAT	GO:0005261~cation channel activity	14	1.47E -06	5.04 E -05
GOTERM_BP_FAT	GO:0006811~ion transport	24	2.64 E -06	7.35 E -04
GOTERM_MF_FAT	GO:0005267~potassium channel activity	10	4.97 E -06	1.52 E -04
GOTERM_CC_FAT	GO:0005886~plasma membrane	54	7.31 E -06	0.001373
GOTERM_MF_FAT	GO:0046873~metal ion transmembrane transporter activity	14	9.78 E -06	2.69 E -04
SP_PIR_KEYWORDS	membrane	84	1.05 E -05	6.46 E -04
SP_PIR_KEYWORDS	ionic channel	14	1.36 E -05	6.31 E -04
SP_PIR_KEYWORDS	ion transport	18	4.68 E -05	0.001731
GOTERM_BP_FAT	GO:0006813~potassium ion transport	10	5.15 E -05	0.010698
SP_PIR_KEYWORDS	cell membrane	35	5.23E -05	0.001611
SP_PIR_KEYWORDS	potassium channel	7	7.00 E -05	0.001848
GOTERM_CC_FAT	GO:0034705~potassium channel complex	7	8.00 E -05	0.007493
GOTERM_CC_FAT	GO:0008076~voltage-gated potassium channel complex	7	8.00 E -05	0.007493
SP_PIR_KEYWORDS	potassium transport	8	1.04 E -04	0.002394
GOTERM_BP_FAT	GO:0015672~monovalent inorganic cation transport	13	1.15 E -04	0.019002

In our experiments with genes related to different medical disorders, we have not seen a significant overlap of our top ranked genes with genes enriched in memory related disorders such as dementia and Alzheimer's disease. It is possible that to use our approach to find genes related to specific disorders, the regions of interest used for gene ranking should be defined in a more targeted way, for example using imaging studies specific to the medical disorder of interest. Comparing the results of such targeted searches and the results obtained for healthy subjects could possibly lead to subsets of genes that distinguish plasticity in the normal brain and in different disorders.

Previous studies of the genes affecting neuroplasticity often involved experimental techniques to identify genes that are upregulated after activity [Luo et al., 2001; Lee et al., 2005; Nedivi et al., 1993]. A targeted search for genes that are rapidly activated after specific types of learning is

not feasible with the data integration approach we present, which is based on gene expression atlases. However, by focusing on brain regions which are susceptible to undergoing task-related changes, our method allows us to seek genes that participate in the cellular processes that underlie neuroplasticity.

We have found significant agreement in the ranking of genes for mice and humans. This high correlation was achieved despite a difference in the stimuli used for the imaging studies in mice and humans. For instance, while the human DTI study [Sagi et al., 2012] uses a spatial navigation protocol of only hours, the mice study uses a 2 month environmental enrichment protocol [Blumenfeld-Katzir et al., 2010]. Although these two stimuli are conceptually different, what is common to both is that they rely on the same basic biological process—neuroplasticity—the ability of the neural system to remodel itself mainly at the synaptic level, to new experiences. Thus, it is likely that

P-value > 10 ⁻³	10 ⁻⁵ < P-value < 10 ⁻³	10 ⁻⁷ < P-value < 10 ⁻⁵	10 ⁻⁹ < P-value < 10 ⁻⁷	P-value < 10 ⁻⁹
----------------------------	---	---	---	----------------------------

Figure 3.

Biological process GO categories significantly enriched in the mouse-human homologue dataset constructed. The top functional categories were found using the GOrilla tools for visualization of enriched GO terms in ranked gene lists [Eden et al., 2009].

different studies on neuroplasticity will highlight similar gene groups. An open question that should be studied in future work is whether we can define, using the proposed approach, subsets of genes that are more likely to be related to plasticity that occurs in different cognitive domains. Answering this question necessitates comparing different species with a very similar task compared with other tasks.

We believe that future research can benefit from the proposed approach as a tool for identifying genes related to plasticity or other neurological phenomena. Additional spatially mapped gene expression data from humans, which will become available in the next few years, will increase the reliability of the candidate genes found in such studies. Moreover, the integration of gene expression datasets and results from imaging studies can be facilitated with researchers in the neuroimaging field sharing detailed information regarding their spatial findings and using standard coordinate systems to report their results.

CONCLUSIONS

This work demonstrates the possibility to infer biological insights by the integration of knowledge from genomics and neuroimaging. Exploring the direct relation between imaging findings across a large cohort of subjects and specific gene expression can be done only on nonhuman; and even then it requires an enormous work on limited brain areas. The results of this article show that combining imaging observations with a genomic database enable to extract a link between genes and tissue changes that is biologically relevant. Although such approach might highlight only a percentile of the involved genes, it certainly provides a first indication on the wide, solid genetic basis of neuroplasticity (that was challenged in our case). Indeed, the results of this work, obtained in two different species and using different imaging protocols, suggest that the highlighted genes correspond to a very basic and common process of neuroplasticity. The analysis pipeline presented here can be applied to any imaging observation (e.g., revealing the genomic origins of the imaging findings for the comparison between Alzheimer's disease patients and healthy controls) underscoring the robustness of this approach. It is anticipated that once hundreds of such comparisons will be performed a better view of the link between brain physiology and gene expression could be yielded across the entire spectra of cognitive domain and process that the brain undergo.

ACKNOWLEDGMENT

The authors would like to thank Ido Tavor, Yaniv Sagi, and Shir Hofstetter for data collection and analysis of the humans DTI data and Tamar Blumenfeld Katzir for data collection and analysis of the mice DTI data.

REFERENCES

- Alberini CM. (2009): Transcription factors in long-term memory and synaptic plasticity. *Physiol Rev* 89:121–145.
- Barth AL, Yassin L. (2008): Activity-dependent gene transcription in neurons: Defining the plasticity transcriptome. In: Dudek SM, editor. *Transcriptional Regulation by Neuronal Activity*. Springer US. pp 295–312.
- Blumenfeld-Katzir T, Michaelson DM, Assaf Y. (2010): Interaction between ApoE genotype and environmental stimulation studied with diffusion tensor imaging. In: *Proceedings of 16th Annual Meeting of the Organization for Human Brain Mapping (OHBM)*. Barcelona, Spain, June 6–10, 2010.
- Blumenfeld-Katzir T, Pasternak O, Dagan M, Assaf Y. (2011): Diffusion MRI of structural brain plasticity induced by a learning and memory task. *PLoS One* 6:e20678.
- Collins MO, Husi H, Yu L, Brandon JM, Anderson CNG, Blackstock WP, Choudhary JS, Grant SG. (2006): Molecular characterization and comparison of the components and multi-protein complexes in the postsynaptic proteome. *J Neurochem* 97:16–23.
- Davis HP, Squire LR. (1984): Protein synthesis and memory: A review. *Psychol Bull* 96:518–559.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. (2009): GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.
- French L, Pavlidis P. (2011): Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput Biol* 7:e1001049.
- Hawrylycz MJ, Lein S, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, Daly BD, Dang C, Datta S, Dee N, Dolbeare TA, Faber V, Feng D, Fowler DR, Goldy J, Gregor BW, Haradon Z, Haynor DR, Hohmann JG, Horvath S, Howard RE, Jeromin A, Jochim JM, Kinnunen M, Lau C, Lazars ET, Lee C, Lemon TA, Li L, Li Y, Morris JA, Overly CC, Parker PD, Parry SE, Reding M, Royall JJ, Schalkin J, Sequeira PA, Slaughterbeck CR, Smith SC, Sodt AJ, Sunkin SM, Swanson BE, Vawter MP, Williams D, Wohnoutka P, Zielke HR, Geschwind DH, Hof PR, Smith SM, Koch C, Grant SG, Jones AR. (2012): An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–399.
- Huang DW, Sherman BT, Lempicki RA. (2008): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH. (2004): Defining the CREB regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell* 119:1041–1054.
- Jeffery IB, Higgins DG, Culhane AC. (2006): Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics* 7:359.
- Kelly AC, Garavan H. (2005): Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex* 15:1089–1102.
- Lee KH, Ryu CJ, Hong HJ, Kim J, Lee EH. (2005): CDNA microarray analysis of nerve growth factor-regulated gene expression profile in rat PC12 cells. *Neurochem Res* 30:533–540.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen

- L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramie AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivasay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR. (2007): Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176.
- Liu R, Hannenhalli S, Bucan M. (2009): Motifs and cis-regulatory modules mediating the expression of genes co-expressed in presynaptic neurons. *Genome Biol* 10(R72).
- Luo Y, Long JM, Spangler EL, Longo DL, Ingram DK, Weng NP. (2001): Identification of maze learning-associated genes in rat hippocampus by cDNA microarray. *J Mol Neurosci* 17:397–404.
- Mayford M, Kandel ER. (1999): Genetic approaches to memory storage. *Trends Genet* 15:463–470.
- McClung CA, Nestler EJ. (2007): Neuroplasticity mediated by altered gene expression. *Neuropsychopharmacology* 33:3–17.
- Nedivi E, Hevroni D, Naot D, Israeli D, Citri Y. (1993): Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* 363:718–722.
- Ng L, Pathak SD, Kuan C, Lau C, Dong H, Sodt A, Dang C, Avants B, Yushkevich P, Gee JC, Haynor D, Lein E, Jones A, Hawrylycz M. (2007): Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE/ACM Trans Comput Biol Bioinform* 4:382–393.
- Orban P, Rauchs G, Balteau E, Degueldre C, Luxen A, Maquet P, Peigneux P. (2006): Sleep after spatial learning promotes covert reorganization of brain activity. *Proc Natl Acad Sci USA* 103: 7124–7129.
- Park B, Lee W, Han K. (2012): Modeling the interactions of Alzheimer-related genes from the whole brain microarray data and diffusion tensor images of human brain. *BMC Bioinformatics* 13(Suppl 7):S10.
- Parodi S, Muselli M, Fontana V, Bonassi S. (2003): ROC curves are a suitable and flexible tool for the analysis of gene expression profiles. *Cytogenet Genome Res* 101:90–91.
- Pfennig AR, Schwartz R, Barth AL. (2007): A comparative genomics approach to identifying the plasticity transcriptome. *BMC Neurosci* 8:20.
- Piro RM, Molineris I, Ala U, Provero P, Di Cunto F. (2010): Candidate gene prioritization based on spatially mapped gene expression: An application to XLMR. *Bioinformatics* 26:i618–i624.
- Poldrack RA. (2000): Imaging brain plasticity: Conceptual and methodological issues—A theoretical review. *NeuroImage* 12: 1–13.
- Sagi Y, Tavor I, Hofstetter S, Tzur-Moryosef S, Blumenfeld-Katzir T, Assaf Y. (2012): Learning in the fast lane: New insights into neuroplasticity. *Neuron* 73:1195–1203.
- Shu X, Qin YY, Zhang S, Jiang JJ, Zhang Y, Zhao LY, Shan D, Zhu WZ. (2013). Voxel-based diffusion tensor imaging of an APP/PS1 mouse model of Alzheimer's disease. *Mol Neurobiol* 48: 78–83.
- Sun YM, Greenway DJ, Johnson R, Street M, Belyaev ND, Deuchars J. (2005): Distinct profiles of REST interactions with its target genes at different stages of neuronal development. *Mol Biol Cell* 16:5630–5638.
- Tischmeyer W, Grimm R. (1999): Activation of immediate early genes and memory formation. *Cell Mol Life Sci* 55:564–574.
- Wolf L, Goldberg C, Manor N, Sharan R, Ruppin E. (2011): Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput Biol* 7:e1002040.
- Yu H, Lee D. (2011): Functional analysis of human whole brain regions based on gene expression. In: *Proceedings of the ACM fifth international workshop on Data and Text Mining in Biomedical Informatics*. pp 71–74, DTMBio 2011, Glasgow, UK.
- Yu W, Clyne M, Khoury MJ, Gwinn M. (2010): Phenopedia and Genopedia: Disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* 26:145–146.
- Zatorre RJ, Fields RD, Johansen-Berg H. (2012): Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nat Neurosci* 15:528–536.