Quantitative Biology 2015, 3(2): 90–102
DOI 10.1007/s40484-015-0048-8

RESEARCH ARTICLE

Multiple functional linear model for association analysis of RNA-seq with imaging

Junhai Jiang[1], Nan Lin[1], Shicheng Guo[1,2], Jinyun Chen[3] and Momiao Xiong[1,*]

[1] Human Genetics Center, Division of Biostatistics, School of Public Health, The University of Texas, Houston, TX 77030, USA
[2] State Key Laboratory of Genetic Engineering and Ministry of Education, Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China
[3] Department of Epidemiology, Division of OVP, Cancer Prevention and Population Sciences, MD Anderson Cancer Center, The University of Texas, Houston, TX 77030, USA
* Correspondence: momiao.xiong@uth.tmc.edu

Received April 24, 2015; Revised July 18, 2015; Accepted July 20, 2015

Emerging integrative analysis of genomic and anatomical imaging data which has not been well developed, provides invaluable information for the holistic discovery of the genomic structure of disease and has the potential to open a new avenue for discovering novel disease susceptibility genes which cannot be identified if they are analyzed separately. A key issue to the success of imaging and genomic data analysis is how to reduce their dimensions. Most previous methods for imaging information extraction and RNA-seq data reduction do not explore imaging spatial information and often ignore gene expression variation at the genomic positional level. To overcome these limitations, we extend functional principle component analysis from one dimension to two dimensions (2DFPCA) for representing imaging data and develop a multiple functional linear model (MFLM) in which functional principal scores of images are taken as multiple quantitative traits and RNA-seq profile across a gene is taken as a function predictor for assessing the association of gene expression with images. The developed method has been applied to image and RNA-seq data of ovarian cancer and kidney renal clear cell carcinoma (KIRC) studies. We identified 24 and 84 genes whose expressions were associated with imaging variations in ovarian cancer and KIRC studies, respectively. Our results showed that many significantly associated genes with images were not differentially expressed, but revealed their morphological and metabolic functions. The results also demonstrated that the peaks of the estimated regression coefficient function in the MFLM often allowed the discovery of splicing sites and multiple isoforms of gene expressions.

Keywords: imaging; RNA-seq; imaging genomics; functional principal component analysis; functional linear model

INTRODUCTION

There is increasing consensus that imaging measures show closer associations with genomic variants and the penetrance of an individual genomic variant is expected to be higher at the imaging level than at the clinical diagnostic and outcome level. Imaging measures as an endophenotype have a higher power to identify genomic variants that significantly contribute to the development of diseases [1,2]. Integrated genomic and imaging data analysis is a new powerful approach used to uncover the individual variability and mechanism of disease development [3]. Both imaging and genomics generate a huge amount of data that present critical bottlenecks in their analysis. Despite its great success, integrative analysis of unprecedented high dimensional imaging and genomic data faces great conceptual and computational challenges [4].

A key issue to the success of imaging and genomic data analysis is how to reduce dimensions of both imaging and genomic data. Previously investigated methods for imaging information extraction include single region-of-interest (ROI) methods, voxelwise approaches, principal component analysis (PCA), singular value decomposition, self-organizing Map (SOM) and multidimensional scaling (MDS) [5]. However, these multivariate dimen-

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

sion reduction methods do not explore imaging spatial information. They take the set of spectral images as an unordered set of high dimensional pixels [6]. Spatial information is very important for image cluster and classification analysis. To overcome limitations of multivariate dimension reduction and to utilize spatial information of the image signal, we extend the widely used one dimensional functional principal component analysis (FPCA) [7] to high dimensional FPCA to extract imaging signals.

The traditional methods for assessing the relationship between gene expressions measured by microarray and phenotypes are linear regressions [8,9]. However, the rapidly developed next-generation sequencing (NGS) technologies have become the platform of choice for gene expression profiling. RNA-seq for expression profiling offers a comprehensive picture of the transcriptome, with less background noise and a wider dynamic range of expression [10]. Unlike microarrays for measuring gene expression, RNA-seq provides multiple layers of resolutions and transcriptome complexity: the expression at exon, SNP, and positional level, splicing, transcription start sites, polyadenylation sites, post-transcriptional RNA editing across the entire gene, and isoform and allele-specific expression [11]. The current linear regression for modeling association of gene expressions with phenotypes quantifies the expression level of a gene/transcript by a single number that summarizes all the reads mapped to that gene/transcript. A single number measuring gene expression level ignores gene expression variation at the genomic positions. Therefore, linear regression is appropriate for microarray expression data, but may not be good for RNA-seq data.

To overcome these limitations, we propose a multiple functional linear model (MFLM) in which functional principal component scores of images are taken as multiple quantitative traits and RNA-seq profile across a gene is taken as a function predictor for assessing the association of gene expression with imaging signals which can take gene splicing and expression variation at genomic positional levels into account.

RESULTS

To evaluate its performance, the proposed MFLM for integrative imaging and RNA-seq data analysis was applied to images and RNA-seq datasets of ovarian cancer (OV) and kidney renal clear cell carcinoma (KIRC) which were downloaded from the The Cancer Genome Atlas (TCGA) datasets. The ovarian cancer dataset consists of 231 tumor tissue samples with histology images and RNA-seq profiles of 16,598 genes (after quality control). The KIRC dataset consists of 188 (121 tumor and 67 normal tissue samples) with histology images and RNA-seq profiles of 16,775 genes (after quality control). RNA-seq data were created by Illumina HiSeq 2000 PE paired-end RNA sequencing. More detailed information can be downloaded from the TCGA website (http://cancergenome.nih.gov/).

The pathology images are used to study the manifestations of disease. Some tissue samples from the patients are obtained by either surgery, biopsy or autopsy. These tissues are either frozen or placed in formaldehyde for fixation which stabilizes the tissues to prevent decay. Then the fixed samples are sectioned into thin slices and stained with one or more dyes. Finally, the prepared pathology slides are placed under the optical microscope and captured by the charged-couple device (CCD) camera. The pathology images have the ability to identify the pathological change of the patient's tissue at the cellular level such as the shape of the nucleus and the texture of the cell.

FPCA for imaging signal extraction

In our study, we compared our two dimensional FPCA with the traditional PCA by capturing space variation of image signals. To evaluate the performance of the two methods on image compression, we compared the original histology images and reconstructed images by two dimensional FPCA and PCA. The result clearly demonstrated that the reconstructed images by FPCA were much closer to the original images than that by PCA (Figure S1). In addition, 90.3% of the total imaging variation could be explained by the top 30 functional principal components, while only 63.6% of the total imaging variation was explained by the top 30 traditional principal components. Therefore, two dimensional FPCA is a better and more authentic image compression algorithm for image signal capturing, with minimal loss of information and fewer principal components usage than traditional PCA.

Behavior of the MFLM for integrative analysis of RNA-seq and imaging data

In the process of integrative analysis of RNA-seq and image data, histology image data were compressed with our proposed two dimensional FPCA, and the FPC scores were taken as phenotypes. We considered gene expression values at single-base resolution and represented the expression profile of a gene by a functional curve, called a "gene expression function". The pipeline for RNA-seq data processing is given as follows. Bam files were obtained from the TCGA project and raw reads for each gene were extracted from Samtools [12] (revised version to recode the reads into binary data to decrease the storage memory). We used "easyRNASeq" to perform quantile

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

normalization for normalizing the read counts and imputation [13]. We used the Karhunen-Loeve decomposition [7] to decompose the random gene expression function into orthogonal FPCs. The multiple FPC scores for imaging signal extraction were regressed on the FPC scores that were obtained from decomposition of the gene expression functions. In other words, we proposed to use MFLM for integrative analysis of RNA-seq and imaging data (Materials and Methods). Two FPCs that accounted for 81.7% and 88.6% variation of imaging signals for ovarian and KIRC, respectively, were selected as phenotypes. The number of selected FPCs for the RNA-seq which accounted for 95% of the variation of gene expression ranges from 2 to 60. *P*-values for declaring significant association after applying the Bonferroni correction for multiple tests in ovarian cancer and KIRC analysis were 3.012×10^{-6} and 2.98×10^{-6}, respectively. To indirectly examine the validity of MFLM for assessing the association of gene expression with the histology images, we plotted a QQ plot of the test in the MFLM (Figure 1). The QQ plots clearly showed that the false positive rates of the MFLM for detection of the association of gene expression with histology images in both ovarian cancer and KIRC studies were controlled.

MFLM for integrative analysis of RNA-seq and histology images

Three statistical methods: MFLM with FPC scores as phenotypes, MFLM with image descriptors [14] as phenotypes and multivariate regression model with FPC score as phenotypes and a single gene expression value (level 3 in TCGA datasets) as a regressor were applied to the ovarian cancer and KIRC datasets. For the ovarian cancer dataset, MFLM with FPC scores as phenotypes, MFLM with image descriptors and multivariate regression identified 24, 2 and 0 genes whose expressions were associated with image signals, respectively. Similarly, for the KIRC dataset, MFLM with FPC scores as phenotypes, MFLM with image descriptors and multivariate linear model (MLM) identified 84, 6 and 1 genes whose expressions were associated with image signals, respectively. The results were summarized in Tables 1 and 2.

Several remarkable features from these results were observed. First, the *P*-values calculated from the MFLM with FPC scores as phenotypes were much smaller than that calculated from the MFLM with image descriptors as phenotypes. Two methods assumed the same functional linear model (FLM) for RNA-seq data, but with a different approach to imaging signal reduction. The FPCA can reduce the dimensions of the imaging data more substantially than the traditional image descriptors. Therefore, the degrees of the test statistic in the MFLM with FPC score as phenotypes were much smaller than that in the MFLM with descriptors as phenotypes, which lead to the smaller *P*-values of the tests in the MFLM with the FPC scores as phenotypes. Second, we observed very few significant associations of the gene expression in the MLM. The MLM used the same FPCA for imaging data reduction, but model the gene expression level in a gene as a single value. The results demonstrated that the widely used single value representation of the expression level in the gene overlooked the expression variation across the gene, which led to large *P*-values of the tests. Third, expressions of genes which were associated with the imaging signal may or may not be differentially expressed (Table S1, Figure S2). In other words, significant association of gene expression with imaging signals can provide additional information which differential expres-

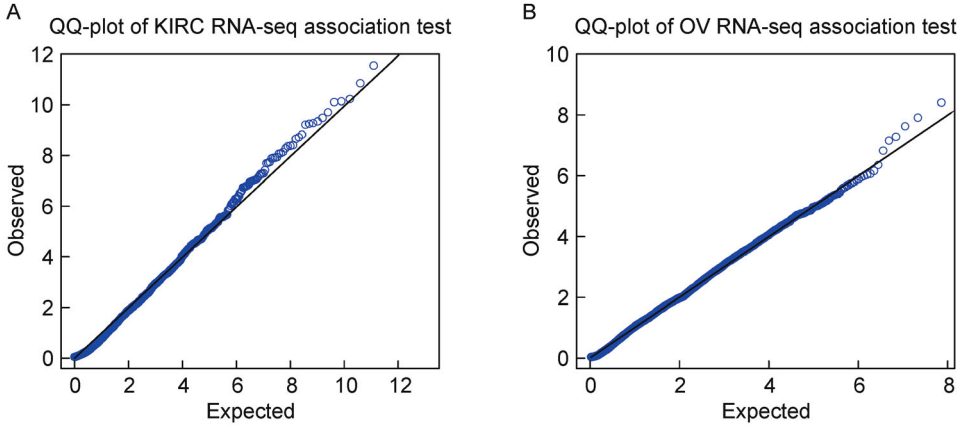

Figure 1. QQ plots for the KIRC dataset and Ovarian Cancer dataset. (A) QQ plot for the KIRC dataset. (B) QQ plot for the Ovarian Cancer dataset.

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

Table 1. ***P*-values of three statistics for testing association of expression with images in ovarian cancer study.**

Gene	*P*-value		
	MFLM_FPC	MFLM_descriptor	MLM
ZNF805	2.31E-10	0.434406	0.932869
LOC653501	3.86E-09	0.024875	0.904242
TMEM170B	1.23E-08	0.006176	0.91104
DRP2	2.38E-08	0.097565	0.589285
OR6V1	5.27E-08	0.002066	0.176132
GPR113	7.09E-08	3.67E-07	0.569642
LOC389765	1.51E-07	0.076075	NA*
ZNF484	4.47E-07	0.006771	0.934051
DNAL1	7.00E-07	0.006636	0.859388
ITGA10	8.72E-07	0.200773	0.6412
NBEAL1	9.43E-07	0.007667	0.711715
IBA57	1.03E-06	0.006602	NA
C16orf52	1.13E-06	0.021002	0.905685
PHKA1	1.31E-06	0.027968	0.715083
PTPRG	1.39E-06	0.949854	0.657859
IFT88	1.64E-06	1.09E-05	0.810783
PARD3B	1.78E-06	0.889289	0.448784
TRAPPC11	1.85E-06	0.367912	NA
LIMD1	2.11E-06	0.468624	0.871478
FAM73A	2.13E-06	0.003969	0.927942
CAPN14	2.45E-06	0.01782	0.479734
CPEB3	2.55E-06	0.026235	0.987898
CDCA2	2.80E-06	0.972605	0.374079
PUS3	3.08E-06	0.78059	0.921797

NA*: Expression (level 3) data were not available.

sions cannot offer. For example, genes *NOTCH1*, *ARHGEF11* and *BRD4* that were associated with imaging signals, but not differentially expressed between tumor and normal tissues were reported to regulate interactions between physically adjacent cells and induce G2/M arrest and trigger apoptosis in renal cell carcinoma [15], associated with kidney injury in the Dahl salt-sensitive rat [16] and kidney disease [17]. Fourth, the MFLM with FPC scores as phenotypes could identify associated genes that showed alternative splicing expression patterns.

To illustrate this, we presented average expression of microtubule associated tumor suppressor 1 (*MTUS1*) in the KIRC study (Figure 2). So far, seven isoforms of *MTUS1* have been discovered. We observed from Figure 2 that a higher expression level in exon 1, exon 2 and exon 15 in the normal samples than that in tumor samples, and alternatively spliced transcript variations encoding different isoforms between tumor and normal samples were substantial. This might indicate that splicing sites affect the tissue structure variation which was measured by imaging signals. *MTUS1* is interacted with microtubules to control cellular architecture and organize microtubule arrays. Express variation of *MTUS1* influences variation in microtubule structure, which in turn causes variation of histology images. Disruption of microtubule-dependent processes is involved in cancer development and metastasis [18]. We should point out that many methods and tools have already been proposed to analyze alternative splicing in RNA-seq between samples, such as DEXseq and DSGseq [19,20]. Fifth, imaging data convey relatively closer association with the disease than traditional phenotypes [21]. The genes significantly associated with imaging will have profound implication in cellular function and disease development.

In the ovarian cancer study, among the 24 significantly associated genes with histology images, protein tyrosine phosphatase receptor type G (*PTPRG*) that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation, is a functional tumor suppressor gene and involved in ovarian tumorigenesis [22,23]. Cytoplasmic polyadenylation element binding protein 3 (*CPEB3*) that controls cell cycle progression, regulates senescence, establishes cell polarity, promotes tumorigenesis and metastasis [24],

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

Table 2. ***P*-values of three statistics for testing association of expression with images in KIRC study.**

Gene	*P*-value			Gene	*P*-value		
	MFLM (FPC)	MFLM (descriptor)	MLM		MFLM(FPC)	MFLM(descriptor)	MLM
HELZ	6.62E-16	6.08E-01	8.79E-01	*ZNF81*	9.95E-08	2.06E-07	7.25E-01
9-Mar	2.12E-15	1.02E-06	7.58E-01	*GAB2*	1.04E-07	1.34E-02	6.38E-01
MSH5-SAPCD1	8.98E-13	9.79E-03	NA*	*MMP24-AS1*	1.29E-07	3.26E-06	NA
SLC2A12	2.52E-12	9.94E-03	2.76E-08	*LOC647859*	1.43E-07	8.56E-02	1.23E-03
BRWD1	1.26E-11	5.61E-03	9.54E-01	*C2orf68*	1.49E-07	4.83E-03	7.84E-01
RFX7	5.29E-11	1.00E + 00	9.58E-01	*SDR39U1*	1.57E-07	8.83E-04	5.88E-01
C22orf39	6.55E-11	1.29E-03	5.77E-01	*ZRANB3*	1.66E-07	1.03E-03	9.59E-01
NSD1	7.06E-11	1.67E-02	9.74E-01	*PSMC4*	1.71E-07	1.39E-02	8.87E-01
RTF1	1.82E-10	9.49E-01	8.58E-01	*FLJ12825*	1.74E-07	1.39E-04	7.08E-01
MBD5	3.00E-10	1.08E-04	9.33E-01	*ARHGEF11*	2.26E-07	8.24E-03	8.55E-01
ZSCAN16-AS1	4.16E-10	6.08E-02	NA	*LOC100289019*	2.61E-07	8.50E-04	NA
SESN1	4.84E-10	3.42E-01	6.71E-01	*SUFU*	2.79E-07	1.99E-01	5.84E-01
ITGA9	5.12E-10	2.11E-02	9.52E-01	*ZNF555*	3.75E-07	2.16E-02	3.75E-01
PPM1K	5.60E-10	1.48E-01	1.11E-04	*KHNYN*	3.85E-07	1.54E-01	4.62E-01
USP42	1.39E-09	9.79E-01	9.06E-01	*ANKRD11*	4.80E-07	1.00E + 00	8.92E-01
FAM47E-STBD1	1.77E-09	1.11E-02	NA	*BOLA2*	4.82E-07	9.88E-02	8.33E-01
ZNF710	2.05E-09	1.22E-01	9.82E-01	*BOLA2B*	4.82E-07	9.88E-02	NA
TECPR2	3.59E-09	9.53E-04	5.63E-01	*SAPCD1*	4.97E-07	4.24E-01	NA
RASSF8-AS1	3.88E-09	3.08E-03	NA	*SLC9A4*	6.26E-07	2.27E-02	1.87E-01
CCDC93	4.04E-09	1.00E + 00	9.45E-01	*CRYBG3*	6.30E-07	5.18E-03	5.59E-02
NAV2	4.90E-09	4.11E-02	1.17E-01	*SLC15A2*	6.78E-07	1.18E-04	3.63E-05
CYB5B	6.75E-09	5.52E-04	7.11E-01	*BRD4*	7.77E-07	5.46E-01	9.85E-01
ANKRD17	7.55E-09	1.00E + 00	4.47E-01	*ATP6V1C2*	7.80E-07	2.86E-03	1.88E-01
CCDC181	7.98E-09	5.24E-03	NA	*SMAD2*	9.23E-07	9.38E-01	7.52E-01
SPHK2	1.08E-08	3.26E-03	1.83E-02	*ST3GAL6*	1.19E-06	5.01E-01	2.16E-01
KCNN3	1.15E-08	9.47E-01	9.17E-01	*ZMIZ1*	1.26E-06	3.75E-01	9.46E-01
ZFYVE16	1.16E-08	1.98E-02	7.65E-01	*USP34*	1.36E-06	1.74E-03	8.24E-01
CMTM1	1.23E-08	9.99E-01	3.66E-01	*RALGAPA2*	1.38E-06	3.48E-03	1.57E-01
LINC00875	1.69E-08	1.00E + 00	NA	*FRMD4A*	2.03E-06	5.14E-03	7.95E-01
NOTCH1	1.81E-08	1.96E-02	1.80E-01	*PSMA5*	2.12E-06	7.81E-02	9.35E-01
BLZF1	1.87E-08	2.05E-03	8.94E-01	*RIPPLY1*	2.18E-06	1.00E + 00	3.89E-06
CHD2	3.55E-08	1.28E-01	9.47E-01	*ERCC6*	2.20E-06	1.21E-01	7.01E-01
MTUS1	4.71E-08	8.63E-02	2.18E-01	*MINK1*	2.29E-06	1.19E-02	9.14E-01
REV3L	4.96E-08	2.28E-02	9.59E-01	*DIP2C*	2.38E-06	2.51E-03	9.13E-01
LRIG2	5.00E-08	6.78E-03	8.40E-01	*PHLDB2*	2.51E-06	3.45E-03	6.34E-01
DENND1C	5.83E-08	9.41E-01	2.40E-01	*TBC1D24*	2.54E-06	7.79E-01	2.13E-01
TMEM50B	6.77E-08	4.16E-03	8.31E-02	*APBA3*	2.55E-06	7.20E-02	1.29E-01
CELF1	7.92E-08	1.00E + 00	9.13E-01	*TRAK1*	2.60E-06	7.67E-01	1.53E-01
ZSCAN20	8.41E-08	1.23E-06	9.37E-01	*DLC1*	2.60E-06	7.23E-02	8.35E-01
MINA	8.76E-08	1.61E-03	4.43E-03	*NISCH*	2.62E-06	1.00E + 00	2.28E-01
C5AR2	9.01E-08	4.97E-02	NA	*CPEB3*	2.64E-06	9.82E-01	4.91E-03
SSH2	9.68E-08	1.00E-02	3.00E-01	*UFD1L*	2.94E-06	3.03E-03	7.78E-01

NA*: Expression (level 3) data were not available.

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

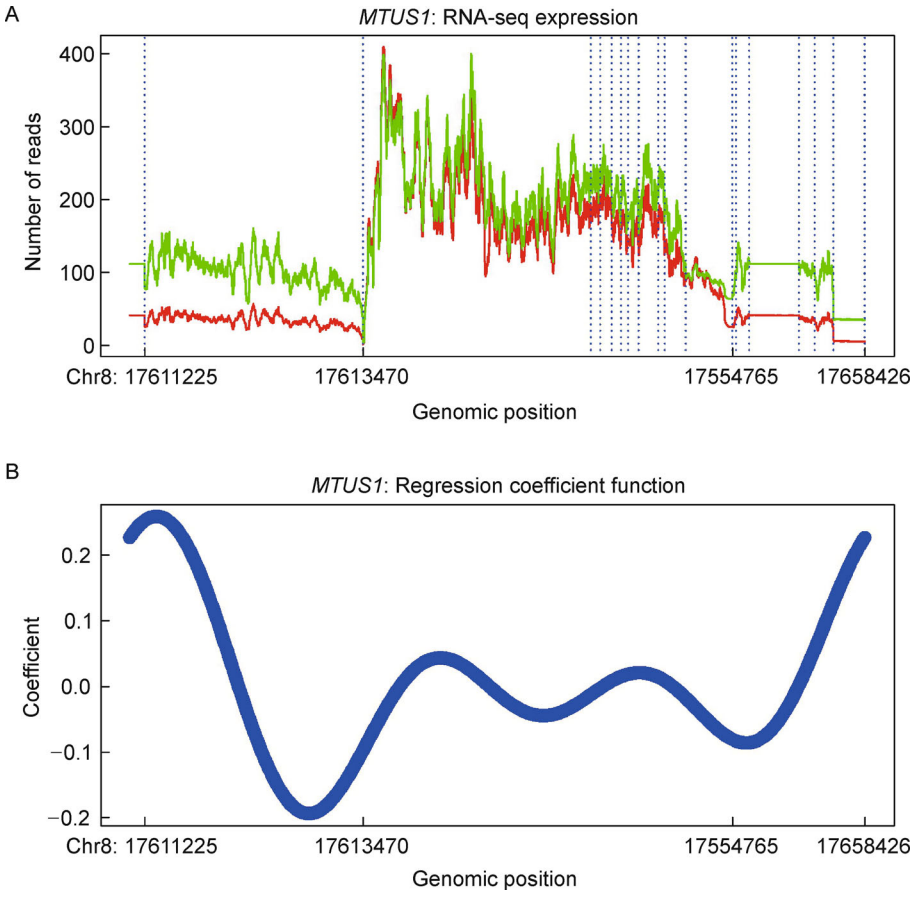

Figure 2. Expression of gene *MTUS1*. (A) RNA-seq curve of Gene *MTUS1*. Number of reads of gene *MTUS1* as a function of the genomic position in the KIRC study, where the green line represents the gene expression profile of normal and the red line represents the gene expression profile of the cancer patient , dashed vertical lines represent exon recombination site where the splicing occurs. Introns were excluded in the plot. (B) Regression coefficient function of gene *MTUS1* in the MFLM.

and plays a role in ovarian cancer development [25]. In the KIRC study, integrin, alpha 9 (*ITGA9*) that participates in regulation of myotube formation [26], is reported to be involved in renal carcinomas [27], *NOTCH1* that regulates interactions between physically adjacent cells, is reported to trigger apoptosis in renal cell carcinoma [15], and Rho guanine nucleotide exchange factor (*ARHGEF11*) whose expression induces the reorganization of the actin cytoskeleton and the formation of membrane ruffling and filopodia, is associated with kidney injury [16] and key regulators of tumorigenesis [28].

Image associated gene form protein-protein interaction networks

A large proportion of genes whose expression variation was associated with imaging signal variation formed protein-protein interaction networks (Figure 3). In the ovarian cancer study, 30 out of 130 proteins significantly associated genes identified by false discovery rate 0.05 are interacted with each other to form a network. Hub gene *SETDB1,* encoding a histone methyltransferase in the network, is an oncogene and is involved in the development of several cancers [29]. Another hub gene *Glul* that catalyzes the synthesis of glutamine from glutamate and ammonia is involved in cell proliferation, inhibition of apoptosis, and cell signalling, and plays key roles in several cancers [30]. We also observed from the KIRC study that 28 out of the 84 proteins significantly associated genes with imaging signals are interacted to form a network, in which 10 genes are differentially expressed between tumor and normal tissues. A hub gene *REV3L* with 11 degrees in the network is the catalytic subunit of DNA translesion synthesis polymerase ζ. It involves a variety of DNA-damaging, genome stability, cytotoxicity, and resistance to chemotherapeutic agents. Surprisingly, although *REV3L* is not differentially expressed, it is reported to be associated with lung, breast, colon cancers and gliomas [31–33]. The interacted genes *KCNN3, ANKRD17, BRD4, NOTCH1, SMAD2,*

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

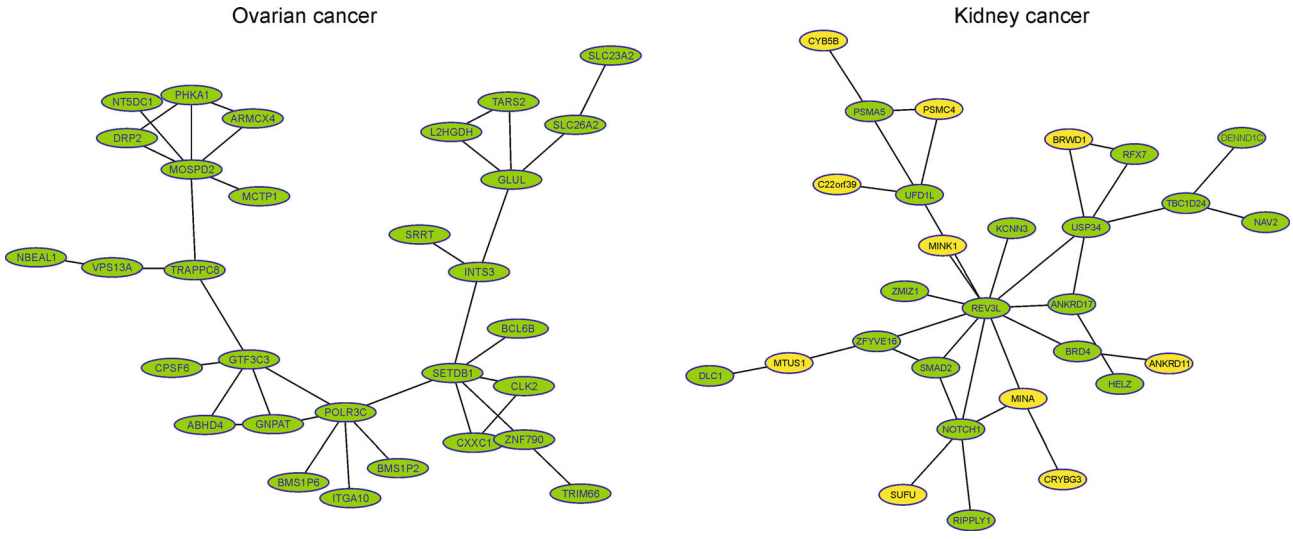

Figure 3. Protein-protein interaction networks in the ovarian cancer study and KIRC study. (A) Protein-protein interaction networks in the ovarian cancer study. Proteins of 30 out of 130 significantly associated genes identified by false discovery rate 0.05 are interacted with each other to form a network in the ovarian cancer study. (B) Protein-protein interaction networks in the KIRC study. Proteins of 28 out of 84 significantly associated genes with images are interacted each other in protein-protein database to form a network in the KIRC study where genes in yellow color were differentially expressed between tumor and normal tissues and dotted vertical lines denote location of splicing sites.

ZMIZ1, UFD1L, and *MINK1* are associated with various cancers [34–40]. Most of these genes are not differentially expressed, but are involved in the formation of cell and tissue structures. Their gene expression variations cause imaging signal variations and are thereby captured by integrative RNA-seq and imaging analysis.

Image associated genes and alternative splicing

We performed FPCA on RNA-seq profiles of each image associated gene and obtained their FPC scores in the ovarian cancer and KIRC studies. Then, we used a hierarchical algorithm to cluster genes based on their FPC scores. The results were shown in Figure 4 and Figure S3. We used DAVID (the Database for Annotation, Visualization and Integrated Discovery) Bioinformatics Resources [41], to extract biological features/meaning of the genes including the role of the gene in metabolization and cell growth regulation as well the gene function location such as membrane, cytoplasm or nuclear. DAVID bioinformatics gene function annotation analysis showed that most image associated genes play important roles in alternative splicing (Figure 2 and Figure S2). We observed that the genes with the similar patterns of alternative splicing sites are grouped together (Figure S4). As Figure S4 showed, the four genes of *TTC23, CPEB3, CAPN14* and *PHKA1* were clustered together since all these genes have large numbers of splicing sites in the end of the genes (3′), and another four genes of *CDCA2, TRAPPC11, PTPRG* and *ITGA10* were clustered together because all these genes have large number of splicing sites in the start of the genes (5′). There is increasing consensus that alternative splicing may affect large and conservative regions of the protein structures and often leads to changes in cell morphologies and phenotypes such as actin cytoskeleton remodeling, regulation of cell-cell junction formation and regulation of cell migrations [42,43]. Variations in alternative splicing of gene expression leads to variations in cell morphologies and phenotypes, thus influencing variations of imaging measures of the cells. This opens a new pathway to cancer development and progression.

Image associated genes and ingenuity pathway analysis

We used the Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, (http://www.ingenuity.com)) that is a web-based functional analysis tool for comprehensive omic data to study the function of genes significantly associated with image. In other words, the list of the identified genes which were significantly associated with imaging signals were input into the IPA. The IPA transforms a list of genes (with or without accompanying expression information) into a set of relevant networks based on extensive records maintained in the Ingenuity Pathways Knowledge Base (IPKB) [1,2]. This knowledge base has been abstracted into a large network, called the Global Molecular Network, composed of thousands of

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

genes and gene products that interact with each other. In the ovarian cancer study, unloading 24 significantly associated genes with imaging into the IPA software, the identified network with the highest score 27 (P-value $< 10^{-27}$) is the cancer network. We observed that 11 out of the 24 genes were included in the cancer network (Figure 5). Figure 5 showed that these genes are mostly regulated by miR-128, miR-92, miR-34 and miR-27a. These microRNAs play an important role in tumorigenesis and cancer development, especially in ovarian cancer [44–48]. Figure 5 also showed that the gene *CPEB3* was a translation regulator and played an important role in the development of ovarian cancer [25]. In the KIRC study, IPA analysis of 84 genes that were significantly associated with imaging identified the network of cellular function and maintenance, hematological system development and function, and inflammatory response with the highest score 54 (Figure S5). We observed that 23 out of the 84 genes were included in the network. Genes *SMAD2, NOTCH1, ERCC6* and *NSD1* are transcription regulators that mediate multiple signaling pathways. It was reported that *SMAD2* might serve as novel prognostic markers in clear cell renal cell carcinoma patients [49], *NOTCH1* played an important role in oncogenesis of the KIRC [15]. Translation regulator *CELF1* regulate pre-mRNA alternative splicing resulting in multiple transcript variants encoding different isoforms. *CELF1* suppressed the proliferation of cancer [50]. Figure S5 also showed that most genes in this network interact with the Akt, p38MARK, PI3K, and NFκB complex. The PI3K/Akt, NF-κB, and MAPK pathways have been reported to be involved in nephrogenesis, and these pathways are activated in human renal cell carcinoma [51–54].

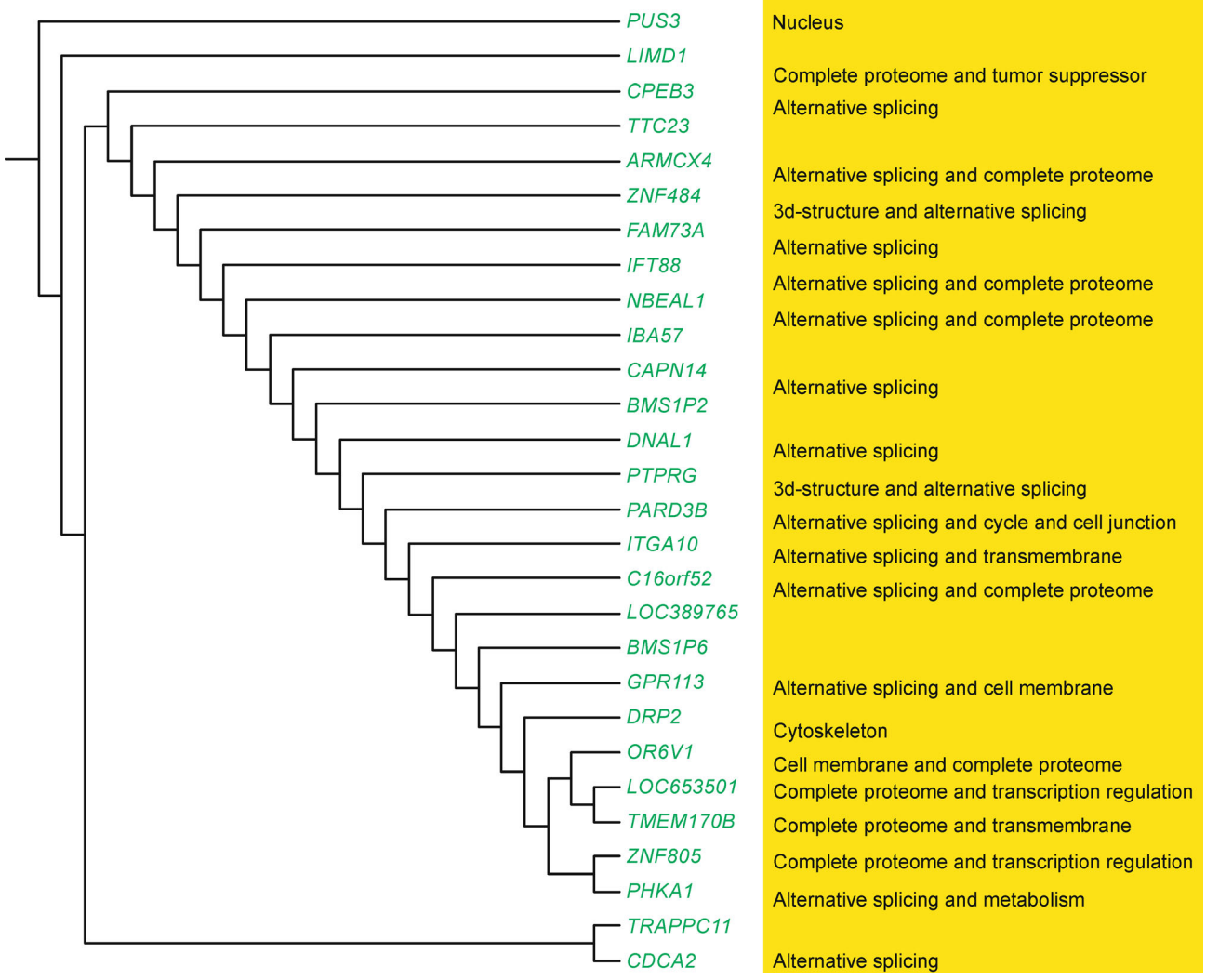

Figure 4. Clusters of image associated genes in the ovarian cancer study by *k*-means clustering algorithms. Dendrogram of the differential expressed or image-significant associated genes based on FPCA scores. The result showed that the majority of these genes were related with alternative splicing.

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

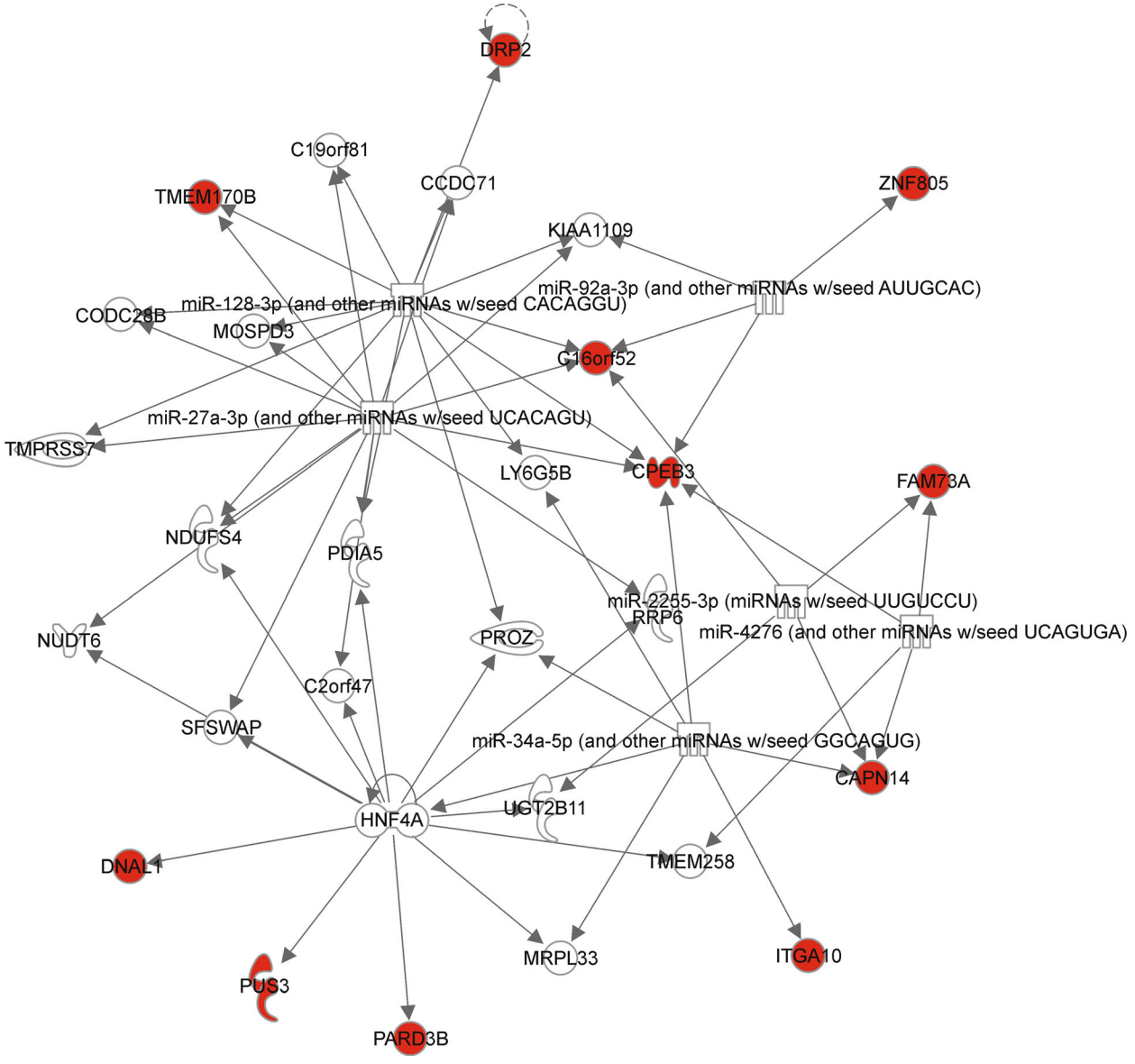

Figure 5. The top protein-protein physical/functional interaction network generated by Ingenuity Pathway Analysis for ovarian cancer. Genes in red node were identified as significantly associated genes with imaging signals in our study.

DISCUSSION

The current major focus of RNA-seq data analysis is to identify differentially expressed genes [55] and the major paradigm of RNA-seq data analysis is to test differences in gene expression level that is measured by a single value summarizing statistic. However, there is increasingly recognition that the differential expression feature of genes may not be a unique source to cause disease. Changes in cell morphologies and motility can also influence development and progression of diseases. In this paper we have presented a MFLM with FPC scores of imaging measures as phenotypes for the integrative analysis of imaging and RNA-seq data and offered a new alternative paradigm for RNA-seq data analysis. We have also shifted the paradigm of RNA-seq data analysis from the single value representation of gene expression to the random function representation of RNA-seq profiles which takes gene expression variation at the genomic positional level into account. Our study has made several remarkable findings.

The first finding is that imaging and RNA-seq analysis

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

can detect cancer susceptibility genes that are not differentially expressed. Changes in cell morphologies, motility and phenotypes play important roles in the development and progression of the cancer. Genes causing these changes may not be differentially expressed between tumor and normal tissue samples and hence cannot be detected by gene differential expression analysis. The integrative analysis of imaging and RNA-seq data opens a new avenue for identifying cancer causing genes.

The second finding is that the function feature of image associated genes is alternative splicing. Surprisingly, we found that the peaks of regression coefficient functions in the MFLM of imaging and RNA-seq data analysis were located in the splicing sites. Alternative splicing often changes the protein structures, cell morphologies and phenotypes [42,43]. These changes generate variation of histology images of tumor tissues, which in turn provide information for discovery of image associated genes.

The third finding is that the widely used single value representation of the expression level in the gene overlooks the expression variation at the genomic positional level across the gene and hence has great limitations to identify image associated genes.

As demonstrated in the real data analysis, the MFLM showed great promise as a tool for integrative analysis of imaging and RNA-seq data. However, to date, very few integrative analyses of imaging and RNA-seq data have been performed. The results presented in this paper are among the first such studies and hence are considered preliminary. The number of selected orthogonal basis functions in the expansion of RNA-seq function will influence the performance of the integrative analysis of imaging and RNA-seq data. Genome-wide imaging and RNA-seq data analysis still pose great challenges. The main purpose of this paper is to stimulate discussion on the optimal strategies for genome-wide imaging and RNA-seq data analysis.

MATERIALS AND METHODS

Two dimensional functional principal component analysis

One dimensional FPCA has been well developed (7). Now we extend one dimensional FPCA to two dimensional FPCA. Consider a two dimensional region. Let s and t denote coordinates in the s axis and t axis, respectively. Let $x(s,t)$ be a centered image signal located at s and t of the region. The signal $x(s,t)$ is a function of locations s and t.

Consider a linear combination of functional values:

$$f = \int_S \int_T \beta(s,t)x(s,t)\mathrm{d}s\mathrm{d}t,$$

where $\beta(s,t)$ is a weight function. To capture the variations in the random functions, we chose weight function $\beta(s,t)$ to maximize the variance of f, which, after imposing a constraint to make the solution unique, leads to the following optimization problem:

$$\begin{aligned} \max \quad & \int_S \int_T \int_S \int_T \beta(s_1,t_1)R(s_1,t_1,s_2,t_2)\beta(s_2,t_2)\mathrm{d}s_1\mathrm{d}t_1\mathrm{d}s_2t_2, \\ \text{s.t.} \quad & \int_S \int_T \beta^2(s,t)\mathrm{d}s\mathrm{d}t = 1. \end{aligned} \tag{1}$$

where $R(s_1,t_1,s_2,t_2) = \mathrm{cov}\,(x(s_1,t_1),x(s_2,t_2))$ is the covariance function of the image signal function $x(s,t)$. By variation calculus [56], we obtain the eigenequation as a solution to the optimization problem (1):

$$\int_S \int_T R(s_1,t_1,s_2,t_2)\beta_j(s_2,t_2)\mathrm{d}s_2\mathrm{d}t_2 = \lambda\beta_j(s_1,t_1), \quad j = 1,2,\cdots,J, \tag{2}$$

for an appropriate eigenvalue λ, where $\beta_j(s,t)$ is an eigenfunction. The random functions $x_i(s,t)$ can be expanded in terms of eigenfunctions as

$$x_i(t,s) = \sum_{j=1}^{K} \xi_{ij}\beta_j(s,t), i = 1,2,\cdots,N, \tag{3}$$

where $\xi_{ij} = \int_S \int_T x_i(t,s)\beta_j(s,t)\mathrm{d}s\mathrm{d}t$, $i = 1,2,\cdots,N$, $j = 1,2,\cdots,J$ are FPC scores (See Supplementary Materials 1).

Multivariate functional linear model for integrative analysis of imaging and RNA-seq data

We take K FPC scores as K quantitative traits. Assume that n individuals are sampled. Let $y_{ik}, k = 1,2,\cdots,K$, be K trait values of the i-th individual. Consider a genomic region $[a, b]$. Let $x_i(t)$ be a RNA-seq profile, the number of reads as a function of the genomic position t, of the i-th individual defined in the regions $[a, b]$. The multivariate functional linear model (MFLM) for integrative analysis of imaging and RNA-seq data can be defined as

$$y_{ik} = \alpha_{0k} + \int_T \alpha_k(t)x_i(t)\mathrm{d}t + \varepsilon_{ik}, \tag{4}$$

where α_{0k} is an overall mean, $\alpha_k(t)$ is a regression coefficient function for the k-th trait, $k = 1,2,\cdots,K$, ε_{ik} are independent and identically distributed normal variables with mean of zero and covariance matrix Σ.

We assume that both trait values and RNA-seq profiles are centered. The RNA-seq profiles $x_i(t)$ are expanded in terms of the orthonormal basis function as:

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

$$x_i(t) = \sum_{j=1}^{J} \xi_{ij}\phi_j(t), \quad (5)$$

where $\phi_j(t)$ are sequences of the orthonormal basis functions. Substituting equation (5) into equation (4), we obtain

$$y_{ik} = \sum_{j=1}^{J} \xi_{ij}\alpha_{kj} + \varepsilon_{ik}, i = 1, 2, \cdots, n, \ k = 1, 2, \cdots, K, \quad (6)$$

where $\alpha_{kj} = \int_T \alpha_k(t)\phi_j(t)\mathrm{d}t$. The parameters α_{kj} are referred to as genetic additive effect scores for the k-th trait.

Equation (6) can be rewritten in a matrix form:

$$Y = \xi\alpha + \varepsilon.$$

The standard least square estimators of α and the variance covariance matrix Σ are given by

$$\hat{\alpha} = (\xi^T\xi)^{-1}\xi^T\left(Y - \overline{Y}\right),$$

$$\hat{\Sigma} = \frac{1}{n}(Y - \xi\hat{\alpha})^T(Y - \xi\hat{\alpha}).$$

Denote the matrix $(\xi^T\xi)^{-1}\xi^T$ by A. Then, the estimator of the parameter α is given by

$$\hat{\alpha} = A\left(Y - \overline{Y}\right).$$

The variance-covariance matrix of the estimator of the parameter α is given by

$$\begin{aligned}\Lambda &= \mathrm{var}(vec(\hat{\alpha})) = (I_k \otimes A)(\Sigma \otimes I_n)(I_k \otimes A^T) \\ &= \Sigma \otimes (AA^T).\end{aligned} \quad (7)$$

An essential problem in the QTL analysis or in the integrative analysis of imaging and RNA-seq data is to test the association of a gene with imaging phenotype. Formally, we investigate the problem of testing the following hypothesis:

$$\alpha_k(t) = 0, \forall t \in [a, b], k = 1, 2, \cdots, K,$$

which is equivalent to testing the hypothesis:

$$H_0 : \alpha = 0.$$

Define the test statistic for testing the association of a gene with K quantitative traits as

$$T = vec(\hat{\alpha})^T\Lambda^{-1}vec(\hat{\alpha}). \quad (8)$$

Let

$$r = \mathrm{rank}(\Lambda).$$

Then, under the null hypothesis $H_0 : \alpha = 0$, T is asymptotically distributed as a central $\chi^2_{(KJ)}$ or $\chi^2_{(r)}$ distribution if J components are taken in the expansion equation (5) (Supplementary Materials 2).

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-015-0048-8.

ACKNOWLEDGEMENTS

The project described was supported by grants 1R01AR057120–01, 1R01HL106034–01 and 1R01 MH101054 from the National Institutes of Health, NHLBI and NIMH, respectively. The authors wish to acknowledge the contributions of the research institutions, study investigators, field staff and study participants in creating the TCGA datasets for biomedical research.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Junhai Jiang, Nan Lin, Shicheng Guo, Jinyun Chen and Momiao Xiong declare that they have no conflict of interest.

Junhai Jiang, Nan Lin, Shicheng Guo, Jinyun Chen and Momiao Xiong conformed to the Helsinki Declaration of 1875, as revised in 2000(5) concerning Human and Animal Rights, and that they followed out policy concerning Informed Consent as shown on Springer.com.

REFERENCES

1. Hibar, D. P., Kohannim, O., Stein, J. L., Chiang, M. C. and Thompson, P. M. (2011) Multilocus genetic analysis of brain images. Front. Genet., 2, 73
2. Liu, J. and Calhoun, V. D. (2014) A review of multivariate analyses in imaging genetics. Front. Neuroinform., 8, 29
3. Stingo, F. C., Guindani, M., Vannucci, M. and Calhoun, V. D. (2013) An integrative Bayesian modeling approach to imaging genetics. J. Am. Stat. Assoc., 108, 876
4. Chi, E. C., Allen, G. I., Zhou, H., Kohannim, O., Lange, K. and Thompson, P. M. (2013) Imagine genetics via sparse canonical correlation analysis. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging., 740–743.
5. Burges, C. J. C. (2010) Dimension reduction: A guided tour. Found. Trends Mach. Learn., 2, 275–365
6. Gupta, M. R. and Jacobson, N. N. P. (2006) Wavelet principal component analysis and its application to hyperspectral images. Image Processing, IEEE Int. Conf., 1585–1588.
7. Ramsay, J. O. and Silverman, B. W. (2005) Functional Data Analysis. 2nd edition. Heidelberg: Springer, 147–172.
8. Ray, M. and Zhang, W. (2009) Integrating gene expression and phenotypic information to analyze Alzheimer's disease. J. Alzheimers Dis., 16, 73–84
9. Wu, T., Sun, W., Yuan, S., Chen, C. H. and Li, K. C. (2008) A method for analyzing censored survival phenotype with gene expression data. BMC Bioinformatics, 9, 417
10. Sun, Z. and Zhu, Y. (2012) Systematic comparison of RNA-Seq normalization methods using measurement error models. Bioinformatics, 28, 2584–2591
11. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

of exons from RNA-seq data. Genome Res., 22, 2008–2017
12. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009) The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079
13. Delhomme, N., Padioleau, I., Furlong, E. E. and Steinmetz, L. M. (2012) easyRNASeq: a bioconductor package for processing RNA-Seq data. Bioinformatics, 28, 2532–2533
14. Lowe, D. G. (1999) Object recognition from local scale-invariant features. The Proceedings of the Seventh IEEE International Conference on Computer Vision, 2, 1150–1157.
15. Wu, K., Zhang, L., Lin, Y., Yang, K. and Cheng, Y. (2014) Inhibition of γ-secretase induces G2/M arrest and triggers apoptosis in renal cell carcinoma. Oncol Lett, 8, 55–61
16. Williams, J. M., Johnson, A. C., Stelloh, C., Dreisbach, A. W., Franceschini, N., Regner, K. R., Townsend, R. R., Roman, R. J. and Garrett, M. R. (2012) Genetic variants in *Arhgef11* are associated with kidney injury in the Dahl salt-sensitive rat. Hypertension, 60, 1157–1168
17. Zhang, G., Liu, R., Zhong, Y., Plotnikov, A. N., Zhang, W., Zeng, L., Rusinova, E., Gerona-Nevarro, G., Moshkina, N., Joshua, J., et al. (2012) Down-regulation of NF-κB transcriptional activity in HIV-associated kidney disease by BRD4 inhibition. J. Biol. Chem., 287, 28840–28851
18. Hernandez, P. and Tirnauer, J. S. (2010) Tumor suppressor interactions with microtubules: keeping cell polarity and cell division on track. Dis. Model. Mech., 3, 304–315
19. Liu, R., Loraine, A. E. and Dickerson, J. A. (2014) Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. BMC Bioinformatics, 15, 364
20. Wang, W., Qin, Z., Feng, Z., Wang, X. and Zhang, X. (2013) Identifying differentially spliced genes from two groups of RNA-seq samples. Gene, 518, 164–170
21. Rasetti, R. and Weinberger, D. R. (2011) Intermediate phenotypes in psychiatric disorders. Curr. Opin. Genet. Dev., 21, 340–348
22. Della Peruta, M., Martinelli, G., Moratti, E., Pintani, D., Vezzalini, M., Mafficini, A., Grafone, T., Iacobucci, I., Soverini, S., Murineddu, M., et al. (2010) Protein tyrosine phosphatase receptor type γ is a functional tumor suppressor gene specifically downregulated in chronic myeloid leukemia. Cancer Res., 70, 8896–8906
23. van Niekerk, C. C. and Poels, L. G. (1999) Reduced expression of protein tyrosine phosphatase gamma in lung and ovarian tumors. Cancer Lett., 137, 61–73
24. D'Ambrogio, A., Nagaoka, K. and Richter, J. D. (2013) Translational control of cell growth and malignancy by the CPEBs. Nat. Rev. Cancer, 13, 283–290
25. Hansen, C. N., Ketabi, Z., Rosenstierne, M. W., Palle, C., Boesen, H. C. and Norrild, B. (2009) Expression of CPEB, GAPDH and U6snRNA in cervical and ovarian tissue during cancer development. APMIS, 117, 53–59
26. Ooishi, R., Shirai, M., Funaba, M. and Murakami, M. (2012) Microphthalmia-associated transcription factor is required for mature myotube formation. Biochim. Biophys. Acta, 1820, 76–83
27. Senchenko, V. N., Liu, J., Loginov, W., Bazov, I., Angeloni, D., Seryogin, Y., Ermilova, V., Kazubskaya, T., Garkavtseva, R., Zabarovska, V. I., et al. (2004) Discovery of frequent homozygous deletions in chromosome 3p21.3 LUCA and AP20 regions in renal, lung and breast carcinomas. Oncogene, 23, 5719–5728
28. Gu, J., Wu, X., Dong, Q., Romeo, M. J., Lin, X., Gutkind, J. S. and Berman, D. M. (2006) A nonsynonymous single-nucleotide polymorphism in the PDZ-Rho guanine nucleotide exchange factor (Ser1416Gly) modulates the risk of lung cancer in Mexican Americans. Cancer, 106, 2707–2715
29. Rodriguez-Paredes, M., Martinez de Paz, A., Simó-Riudalbas, L., Sayols, S., Moutinho, C., Moran, S., Villanueva, A., Vázquez-Cedeira, M., Lazo, P. A., Carneiro, F., et al. (2014) Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. Oncogene, 33, 2807–2813
30. Zhou, C., Chen, H., Han, L., Wang, A. and Chen, L. A. (2014) Identification of featured biomarkers in different types of lung cancer with DNA microarray. Mol. Biol. Rep., 41, 6357–6363
31. Knobel, P. A., Kotov, I. N., Felley-Bosco, E., Stahel, R. A. and Marti, T. M. (2011) Inhibition of REV3 expression induces persistent DNA damage and growth arrest in cancer cells. Neoplasia, 13, 961–970
32. Doles, J., Oliver, T. G., Cameron, E. R., Hsu, G., Jacks, T., Walker, G. C. and Hemann, M. T. (2010) Suppression of Rev3, the catalytic subunit of Polζ, sensitizes drug-resistant lung tumors to chemotherapy. Proc. Natl. Acad. Sci. USA, 107, 20786–20791
33. Varadi, V., Bevier, M., Grzybowska, E., Johansson, R., Enquist, K., Henriksson, R., Butkiewicz, D., Pamula-Pilat, J., Tecza, K., Hemminki, K., et al. (2011) Genetic variation in genes encoding for polymerase ζ subunits associates with breast cancer risk, tumour characteristics and survival. Breast Cancer Res. Treat., 129, 235–245
34. Chantôme, A., Potier-Cartereau, M., Clarysse, L., Fromont, G., Marionneau-Lambot, S., Guéguinou, M., Pagès, J. C., Collin, C., Oullier, T., Girault, A., et al. (2013) Pivotal role of the lipid Raft SK3-Orai1 complex in human cancer cell migration and bone metastases. Cancer Res., 73, 4852–4861
35. Ioana, M., Angelescu, C., Burada, F., Mixich, F., Riza, A., Dumitrescu, T., Alexandru, D., Ciurea, T., Cruce, M. and Saftoiu, A. (2010) *MMR* gene expression pattern in sporadic colorectal cancer. J Gastrointestin Liver Dis, 19, 155–159
36. Yan, Y., Yang, F. Q., Zhang, H. M., Li, J., Li, W., Wang, G. C., Che, J. P., Zheng, J. H. and Liu, M. (2014) Bromodomain 4 protein is a predictor of survival for urothelial carcinoma of bladder. Int. J. Clin. Exp. Pathol., 7, 4231–4238
37. Bokhari, A. A., Lee, L. R., Raboteau, D., Hamilton, C. A., Maxwell, G. L., Rodriguez, G. C. and Syed, V. (2014) Progesterone inhibits endometrial cancer invasiveness by inhibiting the TGFβ pathway. Cancer Prev. Res., 7, 1045–1055
38. Zhang, B., Jia, W. H., Matsuda, K., Kweon, S. S., Matsuo, K., Xiang, Y. B., Shin, A., Jee, S. H., Kim, D. H., Cai, Q., et al., (2014) Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat. Genet., 46, 533–542
39. Chen, X., Ran, Z. H., Tong, J. L., Nie, F., Zhu, M. M., Xu, X. T. and Xiao, S. D. (2011) RNA interference (RNAi) of Ufd1 protein can sensitize a hydroxycamptothecin-resistant colon cancer cell line SW1116/HCPT to hydroxycamptothecin. J. Dig. Dis., 12, 110–116
40. Hwang, J. and Pallas, D. C. (2014) STRIPAK complexes: structure, biological function, and involvement in human diseases. Int. J. Biochem. Cell Biol., 47, 118–148
41. Landau, W. M. and Liu, P. (2013) Dispersion estimation and its effect on test performance in RNA-seq data analysis: a simulation-based comparison of methods. PLoS One, 8, e81415
42. Birzele, F., Csaba, G. and Zimmer, R. (2008) Alternative splicing and protein structure evolution. Nucleic Acids Res., 36, 550–558

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

43. Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., Burge, C. B. and Gertler, F. B. (2011) An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS Genet., 7, e1002218
44. Siegel, R., Naishadham, D. and Jemal, A. (2012) Cancer statistics, 2012. CA: Cancer J. Clin., 62, 10–29
45. Li, M., Fu, W., Wo, L., Shu, X., Liu, F. and Li, C. (2013) miR-128 and its target genes in tumorigenesis and metastasis. Exp. Cell Res., 319, 3059–3064
46. Xu, L., Xiang, J., Shen, J., Zou, X., Zhai, S., Yin, Y., Li, P., Wang, X. and Sun, Q. (2013) Oncogenic MicroRNA-27a is a target for genistein in ovarian cancer cells. Anticancer. Agents Med. Chem., 13, 1126–1132
47. Ohyagi-Hara, C., Sawada, K., Kamiura, S., Tomita, Y., Isobe, A., Hashimoto, K., Kinose, Y., Mabuchi, S., Hisamatsu, T., Takahashi, T., et al. (2013) miR-92a inhibits peritoneal dissemination of ovarian cancer cells by inhibiting integrin α5 expression. Am. J. Pathol., 182, 1876–1889
48. Corney, D. C., Hwang, C. I., Matoso, A., Vogt, M., Flesken-Nikitin, A., Godwin, A. K., Kamat, A. A., Sood, A. K., Ellenson, L. H., Hermeking, H., et al. (2010) Frequent downregulation of miR-34 family in human ovarian cancers. Clin. Cancer Res., 16, 1119–1128
49. Park, J. H., Lee, C., Suh, J. H., Chae, J. Y. and Moon, K. C. (2013) Nuclear expression of Smad proteins and its prognostic significance in clear cell renal cell carcinoma. Hum. Pathol., 44, 2047–2054
50. Wu, L. N., Xue, Y. J., Zhang, L. J., Ma, X. M. and Chen, J. F. (2013) Si-RNA mediated knockdown of *CELF1* gene suppressed the proliferation of human lung cancer cells. Cancer Cell Int., 13, 115
51. Sourbier, C., Lindner, V., Lang, H., Agouni, A., Schordan, E., Danilin, S., Rothhut, S., Jacqmin, D., Helwig, J. J. and Massfelder, T. (2006) The phosphoinositide 3-kinase/Akt pathway: a new target in human renal cell carcinoma therapy. Cancer Res., 66, 5130–5142
52. Sourbier, C., Danilin, S., Lindner, V., Steger, J., Rothhut, S., Meyer, N., Jacqmin, D., Helwig, J. J., Lang, H. and Massfelder, T. (2007) Targeting the nuclear factor-κB rescue pathway has promising future in human renal cell carcinoma therapy. Cancer Res., 67, 11668–11676
53. Huang, D., Ding, Y., Luo, W. M., Bender, S., Qian, C. N., Kort, E., Zhang, Z. F., VandenBeldt, K., Duesbery, N. S., Resau, J. H., et al. (2008) Inhibition of MAPK kinase signaling pathways suppressed renal cell carcinoma growth and angiogenesis *in vivo*. Cancer Res., 68, 81–88
54. Dormoy, V., Danilin, S., Lindner, V., Thomas, L., Rothhut, S., Coquard, C., Helwig, J.J., Jacqmin, D., Lang, H., Massfelder, T. (2009) The sonic hedgehog signaling pathway is reactivated in human renal cell carcinoma and plays orchestral role in tumor growth. Mol. Cancer, 8, 123.2803450.
55. Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res., 35, W169–175
56. Sagan, H. (1969) Introduction to the Calculus of Variations. New York: Dover Publications, Inc.

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2015