BMC Genomics

DNA methylation profiles capturing breast cancer heterogeneity

Xiao Chen[1†], Jianying Zhang[2†] and Xiaofeng Dai[3*]

Abstract

Background: As one of the most described epigenetic marks in human cancers, DNA methylation plays essential roles in gene expression regulation and has been implicated in the prognosis and therapeutics of many cancers. We are motivated in this study to explore DNA methylation profiles capturing breast cancer heterogeneity to improve breast cancer prognosis at the epigenetic level.

Results: Through comparisons on differentially methylated CpG sites among breast cancer subtypes followed by a sequential validation and functional studies using computational approaches, we propose 313 CpG, corresponding to 191 genes, whose methylation pattern identifies the triple negative breast cancer subtype, and report cell migration as represented by extracellular matrix organization and cell proliferation as mediated via MAPK and Wnt signalings are the primary factors driving breast cancer subtyping.

Conclusions: Our study offers novel CpGs and gene methylation patterns with translational potential on triple negative breast cancer prognosis, as well as fresh insights from the epigenetic level on breast cancer heterogeneity.

Background

Breast cancers are highly heterogeneous, which can at least be classified into luminal, HER2 positive (HER2p) and triple negative (TN) types of tumors given their intrinsic differences in the transcriptional expression pattern and clinical outcome association [1, 2]. TN breast cancers are malignant, lack effective targeted therapy, and is not homogeneous that complicates its diagnosis and therapeutics. DNA methylation plays essential roles in numerous cellular processes such as embryonic development, genomic imprinting, cell differentiation and senescence, deregulation of which contributes to several human diseases including cancers [2]. DNA methylation markers are more chemically and biologically stable than RNA and most proteins [3], thus have emerged as an important class of diagnostic or prognostic markers [4–6], with some of which already being applied in clinics [3]. In this regard, several computational approaches have been established to model methylation patterns including, e.g., Bayesian network that has been applied to analyze chromatin interactions [7]. Here, we are

motivated to identify the primary DNA methylation profiles that captures the heterogeneity of breast cancers and can be used to distinguish TN tumors from the rest breast cancer subtypes, with the aim of identifying epigenetic marks or targets facilitating the diagnosis and therapeutics of TNBCs and gaining insights on the epigenetic drivers differentiating breast cancer subtypes.

Results

Primary methylation profiles differentiating breast cancer subtypes

The overall methylation profile of each breast cancer subtype, computed as the average of all DNA methylation sites in each subtype, showed that DNA methylation status decreases in the order of HER2p, luminal, and TN subtypes, and the methylation pattern of luminal cancers is more dispersedly distributed than that of the other subtypes (as the heterogeneity of luminal cancers is higher than the other subtypes that can be further divided into the luminal A and B subtypes) (Fig. 1a). The difference on overall methylation patterns across breast cancer subtypes does not reach statistical significance (p value from ANOVA is 0.0675), due to stochastic gains and losses of cellular processes such as senescence at the population level. However, the results are informative
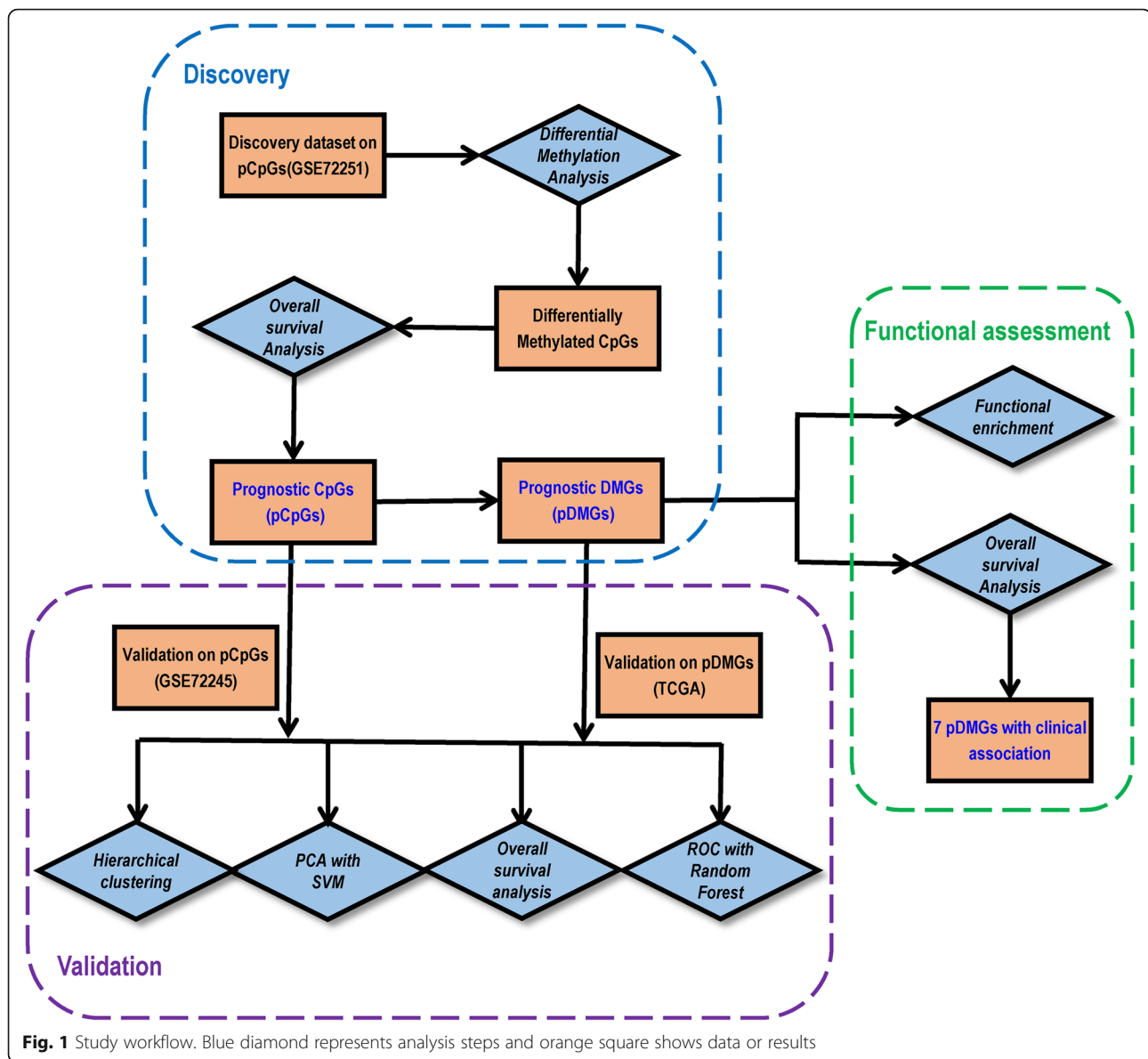
* Correspondence: xiaofeng.dai@jiangnan.edu.cn
†Xiao Chen and Jianying Zhang contributed equally to this work.
3Wuxi School of Medicine, Jiangnan University, Wuxi, Jiangsu, China
Full list of author information is available at the end of the article

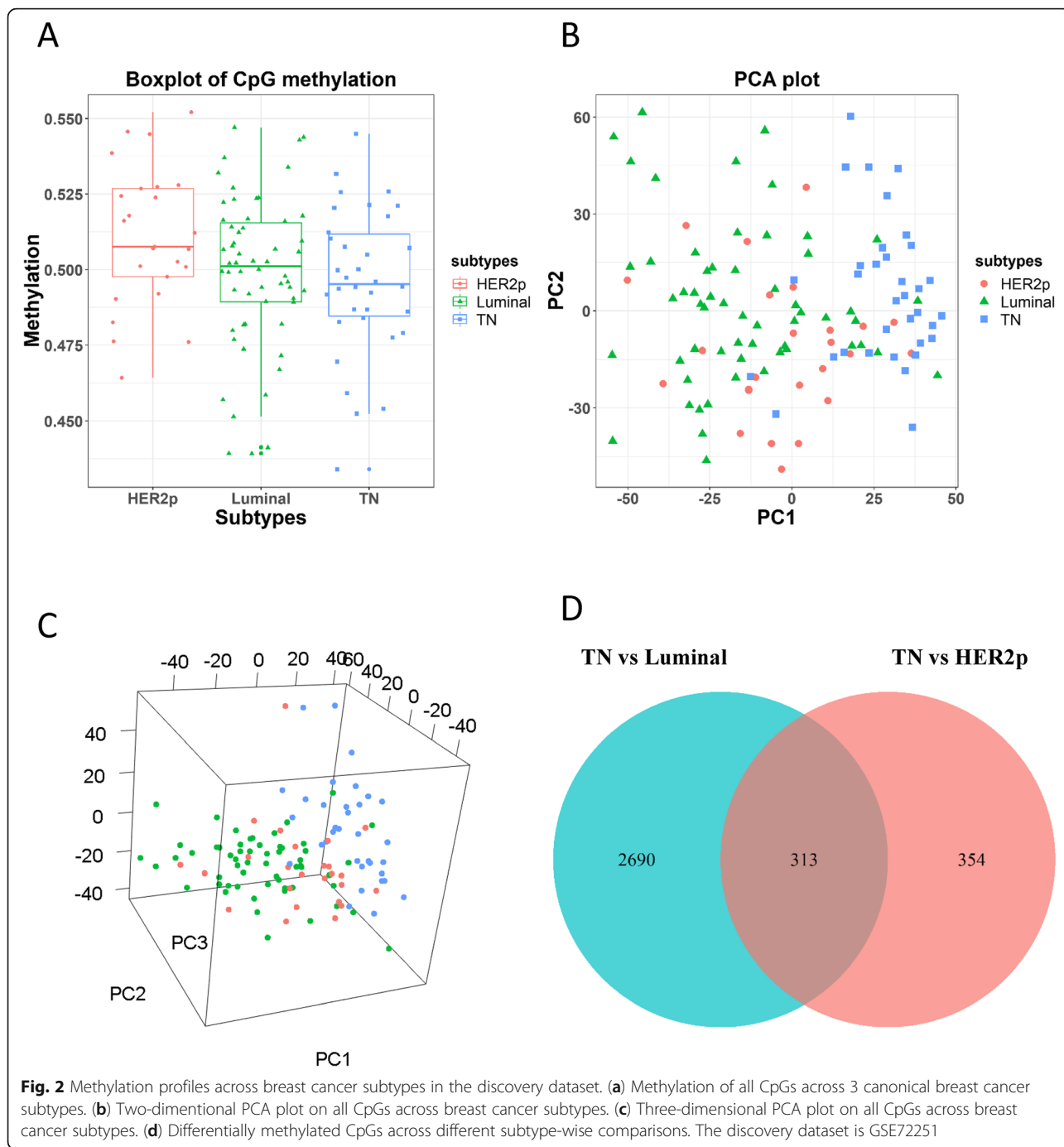**Fig. 1** Study workflow. Blue diamond represents analysis steps and orange square shows data or results

in showing observable alterations and patterns of the mode expression or trend of breast cancer subtypes that does not necessarily need to be statistically significant. The TN subtype has, on average, lower methylation level than the other two subtypes, suggesting that more genes are activated in TN breast cancers, a more malignant state of breast cancers, than the other subtypes [8, 9]. PCA results showed that TN and luminal cancers can be well separated along the first principle component (Fig. 1b, c).

Prognostic methylation profiles differentiating breast cancer subtypes

There are 2690 differentially methylated CpGs between luminal and TN subtypes ('Luminal vs TN', 80.1% of all differentially methylated CpGs, Fig. 1d), 354 between HER2p and TN subtypes ('HER2p vs. TN', 10.5% of all differentially methylated CpGs, Fig. 1d). There are 183 hypermethylated and 130 hypo-methylated CpGs shared between 'Luminal vs TN' and 'HER2p vs. TN' comparisons, which correspond to 191 differentially methylated genes (Additional file 1: Table S1, S2, S4). Therefore, we obtained 313 CpGs and 191 pDMGs.

Applying the pCpGs to the discovery dataset through hierarchical clustering showed 4 distinct patterns across sample subtypes, which correspond to the TN, HER2p and two luminal sample cohorts (Fig. 2a). Purity test revealed that our pCpGs performs better in identifying TN tumors (over 90% purity maximum) than differentiating all the three subtypes (around 80% purity maximum); and

Fig. 2 Methylation profiles across breast cancer subtypes in the discovery dataset. (**a**) Methylation of all CpGs across 3 canonical breast cancer subtypes. (**b**) Two-dimentional PCA plot on all CpGs across breast cancer subtypes. (**c**) Three-dimentional PCA plot on all CpGs across breast cancer subtypes. (**d**) Differentially methylated CpGs across different subtype-wise comparisons. The discovery dataset is GSE72251

the purity reaches the plateau when the number of clusters reached 3 or above in differentiating TN and non-TN tumors, implicating the existence of at least three distinct sample cohorts (excluding HER2) regarding their methylation profiles (Fig. 2b). PCA analysis revealed that the first principle component could distinguish TN and non-TN breast cancers into separate groups, suggesting that the identified pCpGs could capture the primary molecular differences between these

subtypes (Fig. 2c). Breast cancer 10-year OS using the pCpGs suggest patients could be stratified into two distinct groups regarding their outcome (p = 0.0199, HR = 11.53, Fig. 2d).

Performance validation of prognostic methylation profiles
We used the GSE72245 dataset to validate our pCpGs (Fig. 3a, d, g, j) and TCGA data to validate pDMGs regarding both the methylation (Fig. 3b, e, h, k) and
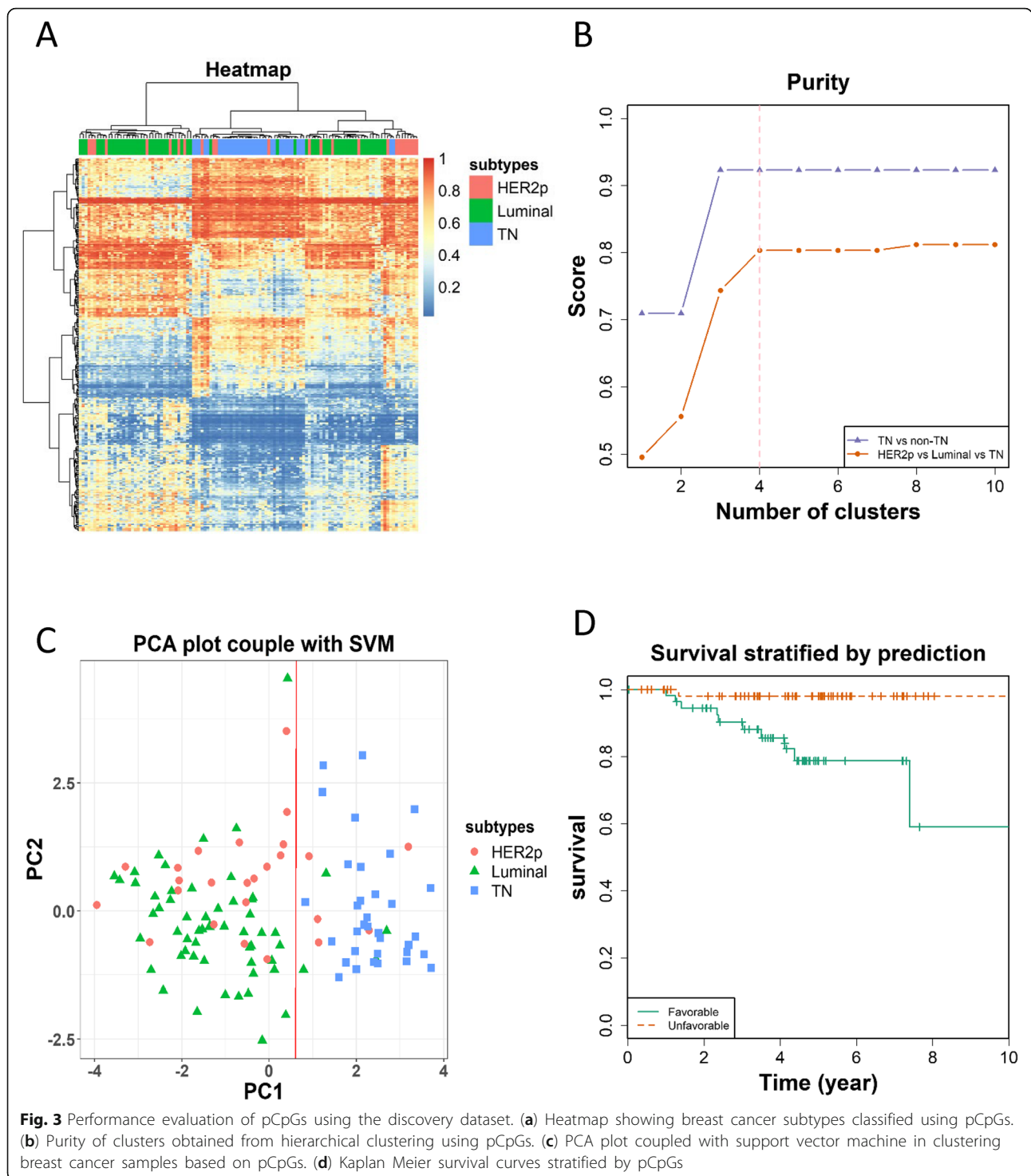
Fig. 3 Performance evaluation of pCpGs using the discovery dataset. (**a**) Heatmap showing breast cancer subtypes classified using pCpGs. (**b**) Purity of clusters obtained from hierarchical clustering using pCpGs. (**c**) PCA plot coupled with support vector machine in clustering breast cancer samples based on pCpGs. (**d**) Kaplan Meier survival curves stratified by pCpGs

transcriptional profiles (Fig. 3c, f, i, l), with results being consistent with those obtained using the discovery dataset. The alteration directions of methylation and transcriptional profiles are opposite in pDMGs (Additional file 1: Table S3).

The ROCs constructed from random forest classification using our validation datasets revealed an AUC of

0.88, 0.82, 0.95, respectively, for pCpGs, methylation and gene expression of pDMGs (Fig. 4).

Six genes (including genes encoding ER, PR, HER2 and their transcription factors MYC, FOXA1, MYBL2) were removed from PAM50 (the new panel is named PAM50–6) to exclude their confounding effect on the classifier, as the ground truth was based on tumor
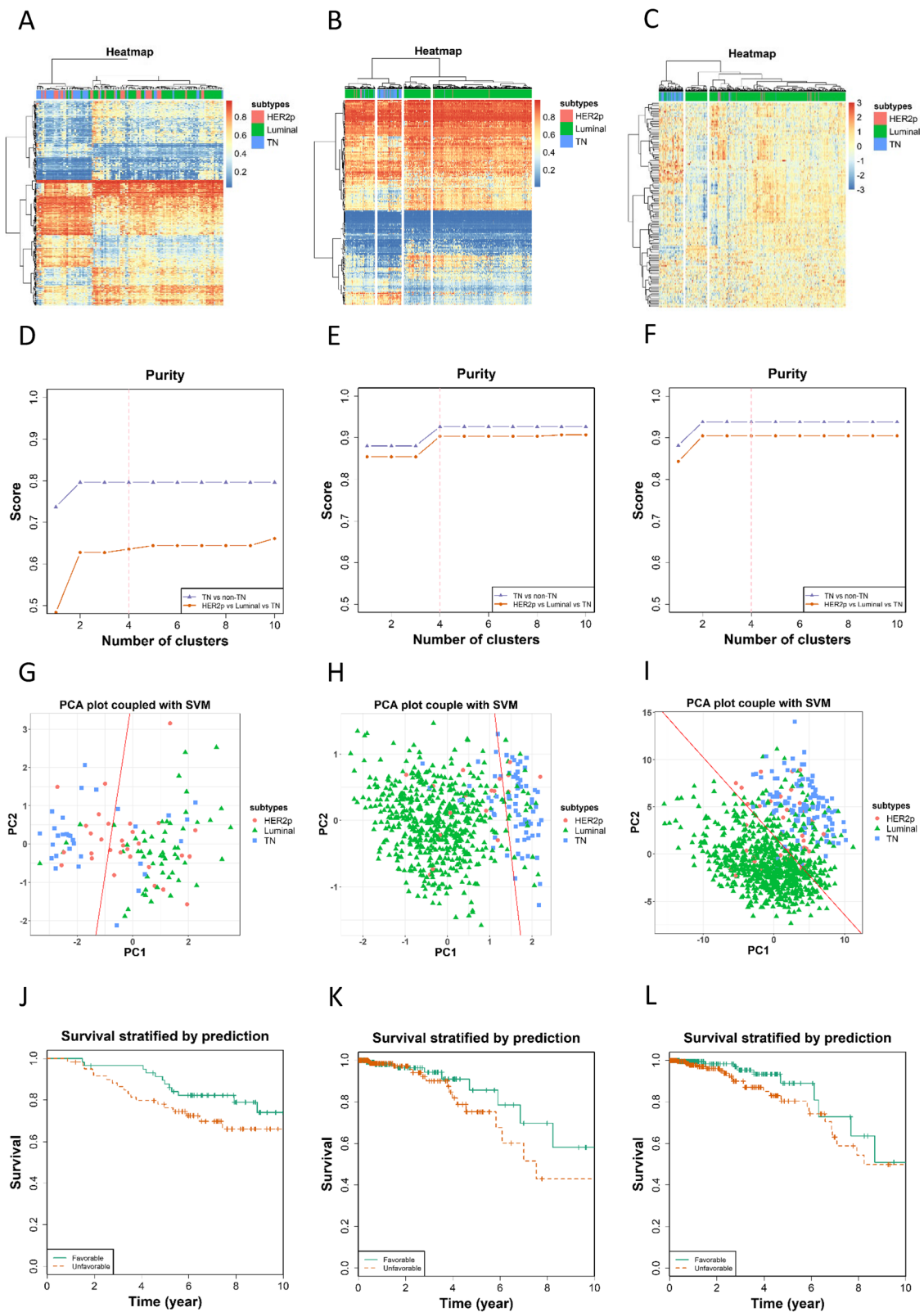
Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Performance validation of pCpGs and pDMGs using the validation datasets. Heatmaps showing breast cancer subtypes classified using (**a**) pCpGs of GSE72245 on gene methylation, pDMGs of TCGA on (**b**) gene methylation and (**c**) gene expression. Purity of clusters obtained from hierarchical clustering using (**d**) pCpGs of GSE72245 on gene methylation, pDMGs of TCGA on (**e**) gene methylation, and (**f**) gene expression. PCA plot coupled with support vector machine in clustering breast cancer samples using (**g**) pCpGs of GSE72245 on gene methylation, pDMGs of TCGA on (**h**) gene methylation, and (**i**) gene expression. Kaplan Meier survival curves stratified using (**j**) pCpGs of GSE72245 on gene methylation, pDMGs of TCGA on (**k**) gene methylation and (**l**) gene expression

classification stratified by ER, PR and HER2. The confusion matrix showed that our pDMGs and the PAM50 each has 563 and 577 samples correctly classified using the TCGA mRNA data (with the F1_nonTN score being 0.97 for pDMGs and PAM50–6; F1_TN score being 0.70 and 0.79, respectively, for pDMGs and PAM50–6) (Table 1).

Functional analysis of prognostic methylation genes

The 191 pDMGs were enriched in 32 GO terms and 3 KEGG pathways. The top 10 GO terms fell into 3 categories, which are 'extracellular matrix organization and cell movement', 'kinase signaling and cell proliferation', 'morphogenesis and cell differentiation', with 'extracelluar structure organization' and 'extracellular matrix organization' being the top 2 (Fig. 5a). The 3 top KEGG pathways are 'focal adhesion', 'Wnt signaling pathway' and 'Hippo signaling pathway' (Fig. 5b), which correspond to 'cell movement', 'cell proliferation' and 'cell differentiation' processes, respectively.

Among the 191 pDMGs, 7 genes, *MATK, IFI35, FAM150B, LBXCOR1, WNT10A, ABLIM1* and *CPT1A* were found to be significantly differentially methylated and expressed with opposite clinical associations (Fig. 6). The hypo-methylation and higher expression of *MATK, IFI35, FAM150B, LBXCOR1* and *WNT10A* are associated with favorable patient survival, while those of *ABLIM1* and *CPT1A* are associated with poor patient outcome (Table 2).

Discussion

We identified 313 pCpGs which correspond to 191 pDMGs capable of distinguishing breast cancer subtypes and especially identifying TN breast cancers. The performance of the 191 pDMGs is similar with that of PAM50–6 in identifying non-TNBCs, which has been clinically used for subtyping and prognosis as BioClassifier™ [10] and ProSigna® [11]; however, as a classifier, pDMGs does not outperform PAM50–6 due to the small size of TNBC cohort. This suggests that these differentially methylated genes could effectively capture the molecular heterogeneity of breast cancers, but the power in identifying small size of tumors may be largely compromised by the restriction that genes need to be under epigenetic regulation in the epigenetic classifier.
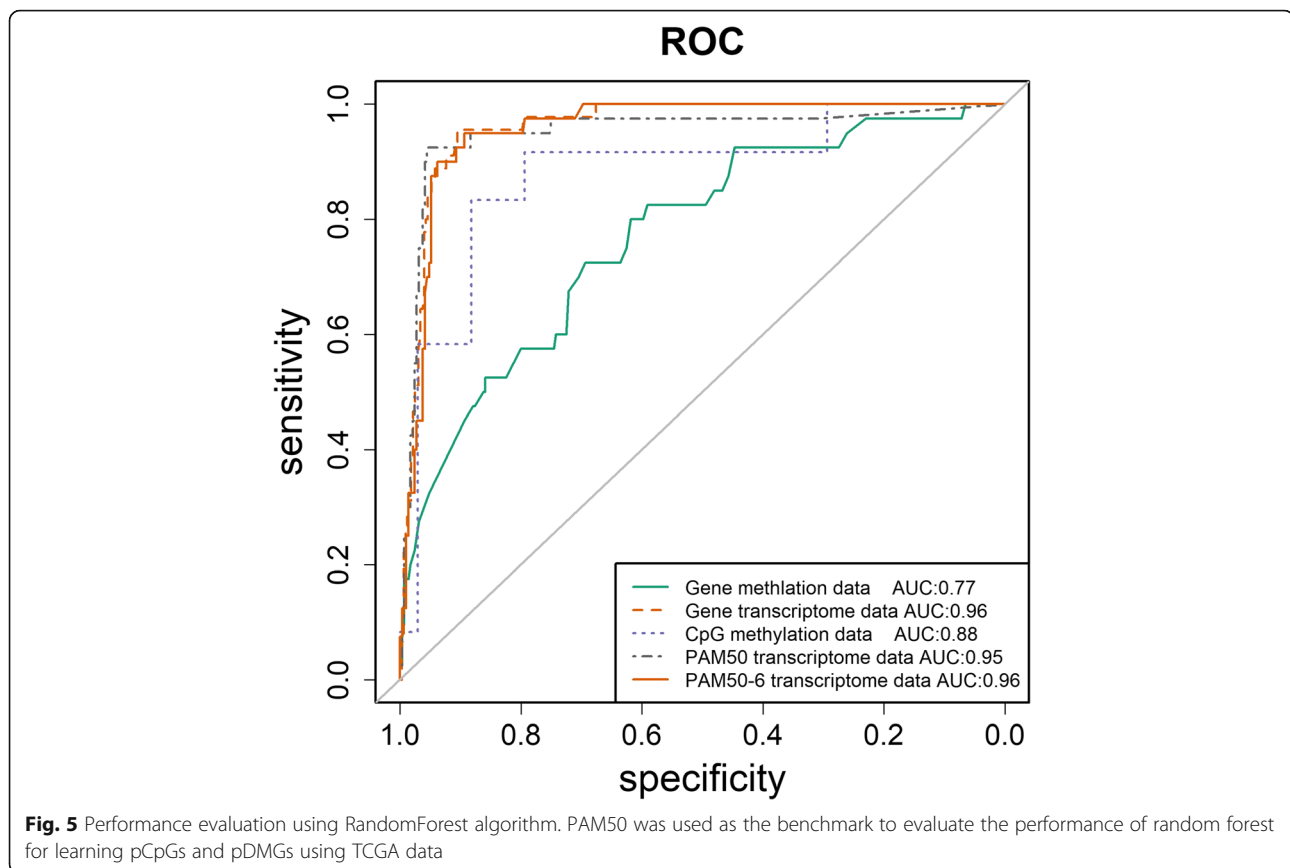
Among these 191 pDMGs, 23 are associated with breast cancer outcome and driven by differential methylation status, given the opposite clinical associations and subtype distributions between the methylation and gene expression levels. Out of the 23 pDMGs, 7 have reached statistical significance on clinical associations at both methylation and gene expression levels, which are *MATK, IFI35, FAM150B, LBXCOR1, WNT10A, ABLIM1, CPT1A*.

Though not many, several evidence exist to support the roles of these 7 pDMGs played in carcinogenesis. *MATK* encodes the megakaryocyte-associated tyrosine-protein kinase that can phosphorylate and inactivate the SRC protein, which is one of the 5 markers used in

Table 1 The confusion matrix of Random Forest classification using the proposed methylation signature

Signature	Data type	Dataset		TN	nonTN	F1 score		MCC
						F1_TN	F1_nonTN	
pCpGs	Methylation	GSE72245	TN	7	2	0.67	0.90	0.58
			non-TN	5	32			
pDMGs	Methylation	TCGA	TN	11	43	0.29	0.93	0.26
			non-TN	12	384			
pDMGs	mRNA	TCGA	TN	44	25	0.70	0.97	0.67
			non-TN	12	519			
PAM50	mRNA	TCGA	TN	60	9	0.84	0.98	0.82
			non-TN	14	517			
PAM50–6	mRNA	TCGA	TN	55	15	0.79	0.97	0.76
			non-TN	15	515			

The values show the number of consistent and inconsistent samples clustered using each signature panel and identified using immunohistochemistry staining. F1 score that captures both false positives and false negatives is used to assess the classification accuracy. 'TN' and 'non-TN' are triple negative and non-triple negative breast cancers, respectively. PAM50 and PAM50–6 (with ER, PR, HER2, FOXA1, MYC, MYBL2 excluded) are used as a benchmark. 'F1_TN' and 'F1_nonTN' each refers to F1 scores in identifying TN and non-TN tumors

Fig. 5 Performance evaluation using RandomForest algorithm. PAM50 was used as the benchmark to evaluate the performance of random forest for learning pCpGs and pDMGs using TCGA data

ProEx™Br for breast cancer prognosis, in vitro [12]. The *IFI35* gene is located in the centromeric region and 500 kb away from the *BRCA1* gene in the genome, and suppresses NFkB signaling that plays a promotive role in carcinogenesis [13]. *LBXCOR1* encodes a transcriptional corepressor of *LBX1* and inhibits BMP signaling which predisposes colorectal cancers [14, 15]. *ABLIM1* encodes the actin binding LIM protein 1, whose over-activation promotes tumorigenesis in, e.g., brain and pancreas [16, 17]. *CPT1A* is involved in the fatty acid oxidation pathway [18] and has been proposed as a target of cancers such as nasopharyngeal [19] and prostate [18] carcinomas. It might be possible that malignant cells have accelerated metabolism to meet their up-regulated requirements on biomass production, and targeting *CPT1A* could kill cancers cells through disrupting their fast and efficient fatty acid oxidation. TNBCs have lower *CPT1A* expression than non-TNBCs, suggesting that while accelerated fatty acid metabolism is a characteristic feature of non-TNBCs, the malignancy of TNBCs is driven by other mechanisms such as cell migration and cancer stemness. It was also reported that dietary fat can perturbate genomic structure by reducing DNA methylation at *CPT1A* gene [20], suggesting an over-dose of *CPT1A* expression on high fat dietary exposure that contributes to cancer cell malignancy

and warranting our attention to adopting low fat dietary in reducing the risk of developing cancers.

Several of these 7 pDMGs may be novel players or have novel roles during carcinogenesis and deserve further investigations. *FAM150B* hyper-methylation was shown to suppress its expression and be associated with poor clinical outcome (Fig. 7c), where direct evidence between *FAM150B* methylation and cancer has not been reported according to our knowledge. *WNT10A* functions as an oncogene in renal cell carcinoma, whose depletion was reported to prevent tumor growth in vitro and in vivo in melanoma [21]; however, it shows tumor suppressive roles in breast cancers in our study which worth further investigations.

The pDMGs identified are largely involved in extracellular structure organization, the regulation of actin filament-based process, transmembrane receptor protein serine/threonine kinase signaling pathway, and connective tissue development, which are all indispensable during cell movement and known to play critical roles in breast cancer progression [22]. These pDMGs are enriched in focal adhesion according to our KEGG pathway analysis (Fig. 5b), which promotes breast cancer initiation and progression once deregulated [23]. MAPK and Wnt pathways are the second most enriched GO terms or pathways of these pDMGs following extracellular structure organization and local
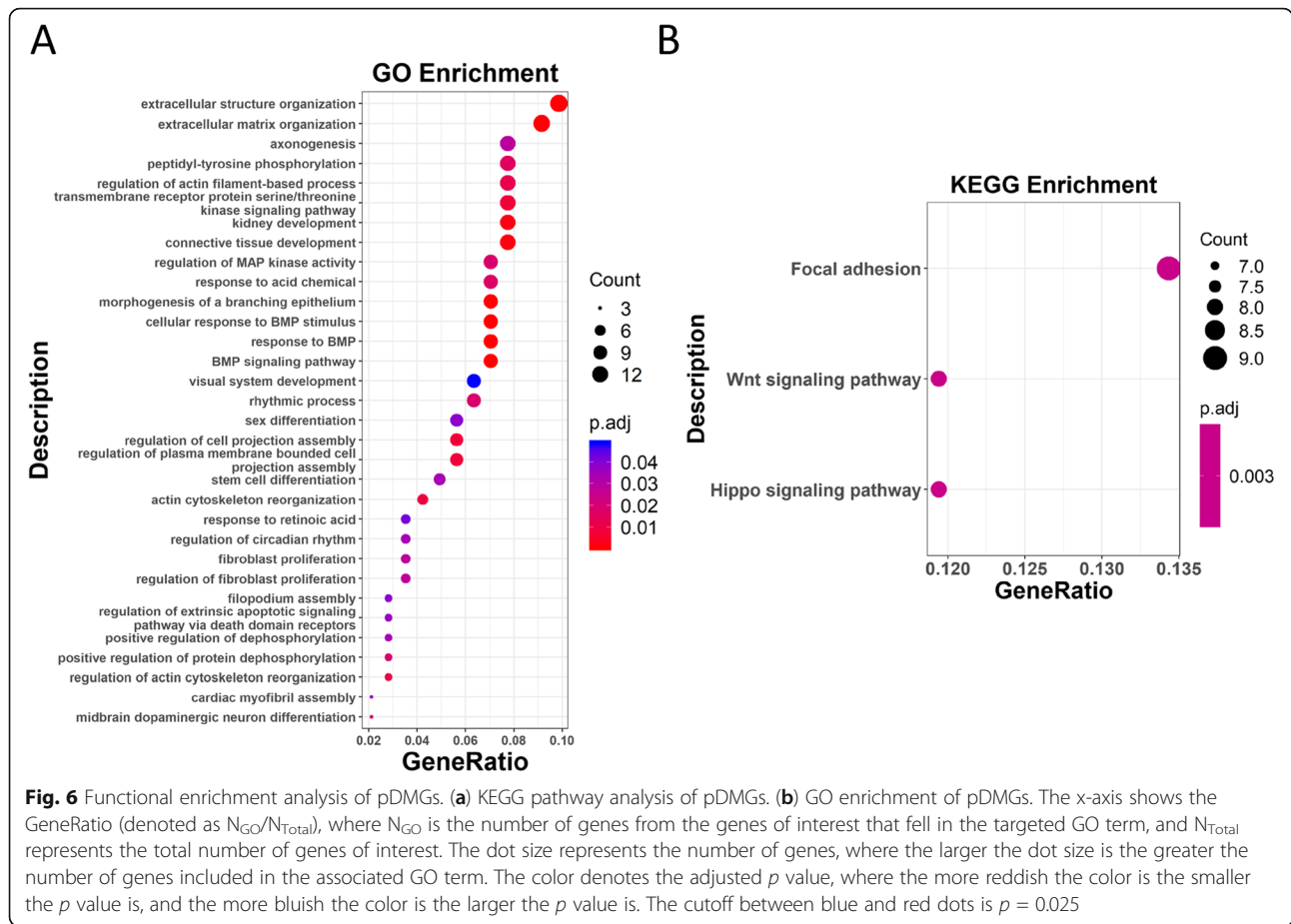
Fig. 6 Functional enrichment analysis of pDMGs. (**a**) KEGG pathway analysis of pDMGs. (**b**) GO enrichment of pDMGs. The x-axis shows the GeneRatio (denoted as N_{GO}/N_{Total}), where N_{GO} is the number of genes from the genes of interest that fell in the targeted GO term, and N_{Total} represents the total number of genes of interest. The dot size represents the number of genes, where the larger the dot size is the greater the number of genes included in the associated GO term. The color denotes the adjusted *p* value, where the more reddish the color is the smaller the *p* value is, and the more bluish the color is the larger the *p* value is. The cutoff between blue and red dots is *p* = 0.025

Table 2 Ten years overall survival of breast cancer patients using the 7 pDMGs associated with clinical outcome

Gene	Data Type	Hazard Ratio	P value
ABLIM1	Methylation	0.40	6.91E-03
	Gene expression	2.03	1.23E-02
CPT1A	Methylation	0.35	3.64E-03
	Gene expression	2.45	8.72E-04
FAM150B	Methylation	1.85	4.84E-02
	Gene expression	0.49	1.53E-02
IFI35	Methylation	2.29	4.22E-02
	Gene expression	0.57	3.53E-02
LBXCOR1	Methylation	2.65	8.55E-03
	Gene expression	0.53	2.18E-02
MATK	Methylation	2.12	2.82E-02
	Gene expression	0.44	4.77E-03
WNT10A	Methylation	2.96	1.11E-02
	Gene expression	0.55	3.23E-02

Cox regression analysis was used. Both methylation status and gene expression level were used for patient stratification

adhesion. Both MAPK and Wnt signalings have known connections with carcinogenesis, whose aberration enables cells with uncontrolled proliferation abilities. MAPK and Wnt signalings have cross-talks through TGFβ signaling via a Smad-independent manner [24], and can suppress or promote each other under different circumstances. For example, increased MAPK signaling could down-regulate the Wnt pathway by stabilizing Axin in melanoma, and Wnt signaling activates the MAPK pathway through Ras stabilization in colorectal cancers [25, 26].

Conclusion

We identified 313 CpGs, corresponding to 191 differentially methylated genes, which capture the molecular differences among breast cancer subtypes with accuracy equivalent to that of PAM50. 'Cell migration' as represented by extracellular matrix organization and 'cell proliferation' as mediated via MAPK and Wnt signalings were identified as primary factors stratifying breast cancer subtypes that are modulated via aberrant methylations. Our study provides DNA methylation profiles with prognostic values and clinical translation potential, and offers novel insights on the driving force orchestrating breast cancer heterogeneity from the epigenetic perspective.
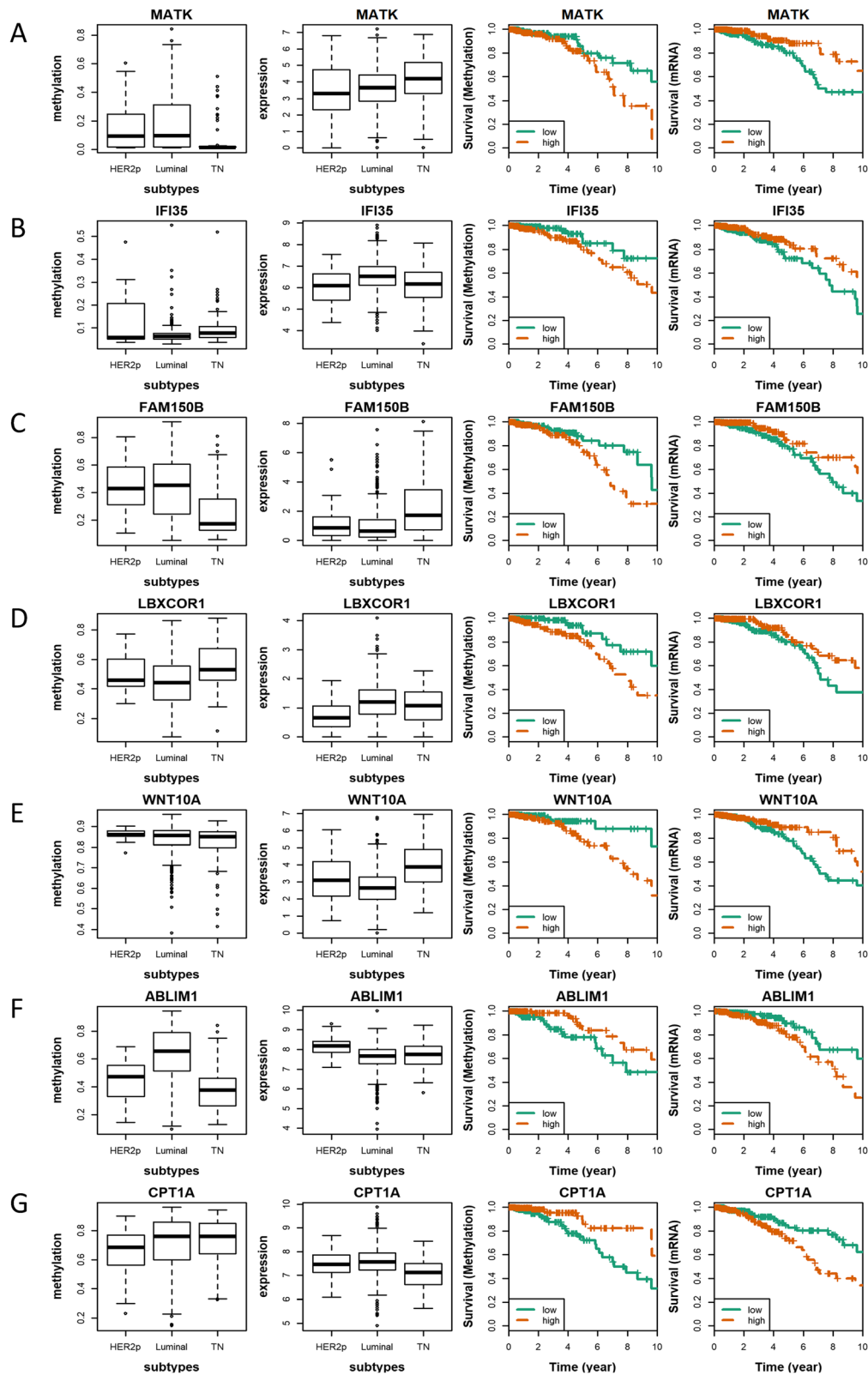
Fig. 7 Differentially methylated genes with prognostic values. (**a**) Methylation and gene expression levels, together with the Kaplan Meier plot of each gene at both methylation and gene expression levels are presented for (**a**) MATK, (**b**) IFI35, (**c**) FAM150B, (**d**) LBXCOR1, (**e**) WNT10A, (**f**) ABLIM1, (**g**) CPT1A

Material and methods

Data

The GEO dataset, GSE72251, performed using the Illumina Infinium Human Methylation Beadchip (450 k array), was retrieved from the NCBI Gene Expression Omnibus (GEO) database [27] and used as the discovery dataset, which is consisted of 119 breast cancer samples and 415,080 CpGs.

The GEO dataset, GSE72245, was retrieved and used as one validation dataset, which encompasses 118 samples and 415,080 CpGs.

Both GSE72251 and GSE72245 were preprocessed by removing high-detection *p*-values, SNP-containing, cross-reactive and heterochromosomic probes (which were replaced by 'null') and both datasets were normalized using the peak-based approach. Methylation and mRNA data as well as clinical information from The Cancer Genome Atlas (TCGA) [28] were downloaded from the cBioportal [29] and used as another validation dataset. This data is comprised of 550 samples and 16,474 genes. Based on the assumption that DNA methylation is a common epigenetic signaling tool that cells use to lock genes in the 'off' state [30], only the CpG probe showing the strongest negative correlation with gene expression was kept and used as the methylation probe of the gene when multiple probes existed to target one single gene. Besides, as pCpGs with gene regulatory roles typically occur in the promoter region of the targeted gene [30], it is unlikely to select a CpG that is functionally irrelevant to the gene in question using this function-based approach, e.g., if a CpG happened to be located in the 3'UTR of gene A and in the promoter region of gene B, its association could only be possibly found with gene B but not with gene A.

All samples were classified into triple negative (TN), HER2 positive (HER2p) and luminal subtypes according to estrogen receptor (ER), progestogen receptor (PR) and human epithelial receptor 2 (HER2) immunohistochemistry status (i.e., HER2p = ER-PR-HER2+, luminal = ER + |PR+, TN = ER-PR-HER2-). Although luminal cancers can be further divided to the A and B subtypes, they share similar molecular patterns and are considered as one large class in many studies [31] including this paper.

Differential methylation analysis

Differential methylation analysis was performed based on student T test and Bayes theorem using the 'limma' [32] package from the Bioconductor [33] package. Based on the empirical Bayes method (the 'eBayes' function), CpG sites specifically hypo-methylated or hyper-methylated in TN breast cancers were ranked in the order of the statistical significance of methylation difference, where the Benjamin-Hochberg adjusted *P*-values < 0.01 was used as the significance threshold.

Survival analysis

The 10-year breast cancer overall survival (OS) analysis of the selected differentially methylated CpGs or genes (DMGs) was performed using the methylation profiles and the clinical data. The analysis was conducted using the Cox proportional hazards model, with the logrank *p*-value less than 0.01 being considered statistically significant. We defined CpGs with prognostic significance as pCpG and genes where pCpG reside in as pDMGs. To interrogate the prognostic value of a panel of methylation sites, we implemented a multi-methylation survival analysis where each methylation site was assigned with 1 (favorable) or 0 (unfavorable), defined as the prognostic score, according to the univariate survival analysis; and an averaged prognostic score of each sample was calculated by averaging the prognostic scores over all methylation sites for each sample, and used for sample stratification in the survival analysis.

Hierarchical clustering and purity analysis

Hierarchical clustering, an unsupervised machine learning approach, was performed using the Euclidean or pair-wise sample correlation (1-r, where 'r' represents the correlation) distance and the Ward linkage [34]. The subtyping performance of pDMGs was assessed by the purity statistics at cutoffs ranged from 1 to 10. The purity of each cluster was computed by assigning it to the most represented breast cancer subtype by its encompassed nodes following the calculation of the fraction of nodes with correct assignment in each cluster.

Principle component analysis and support vector machine classification

Principle component analysis (PCA) was conducted using the `prcomp` function from the 'base' package in R. Support vector machine (SVM) was used to classify samples projected by PCA via the 'svm' function from the 'e1071' package in R.

Random forest and receiver operating characteristic curve construction

Random forest classification was conducted using the 'randomForest' function from 'randomForest' package in R. The number of nodes in a tree was determined through iterations from 1 to 'n-1' where 'n' represents the sample size, and the one with the minimum error was picked. Receiver operating characteristic curve (ROC) and the area under the curve (AUC) were computed using the 'roc' function from the 'pROC' package

Chen *et al. BMC Genomics*   (2019) 20:823

Page 11 of 12

in R to assess the clustering accuracy. To take both false positives and false negatives into account in the assessment, F1 score was calculated as below:

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$

Functional analysis

Functional enrichment analysis was performed based on Gene Ontology (GO) [35] and Kyoto Encyclopedia of Genes and Genomes database (KEGG [36]) using the R package 'clusterProfiler' [37]. Fisher's exact test was utilized to measure the significance of GO terms and biological pathways. The *p*-values were adjusted using Benjamini-Hochberg false discovery rate (FDR), and $p < 0.01$ was used as the threshold to assess the statistical significance of each test [38]. The overall workflow is demonstrated in Fig. 7.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-019-6142-y.

Additional file 1: Table S1. Annotations of prognostic CpGs. **Table S2.** Annotations of prognostic DMGs. **Table S3.** Correlation between the methylation and expression of pDMGs. **Table S4.** Learning matrix of the 313 CpG (with gene attribution) and outcome variables.

Abbreviations

AUC: Area under the curve; DMGs: Differentially methylated genes; ER: Estrogen receptor; FDR: False discovery rate; GEO: Gene expression omnibus; GO: Gene ontology; HER2: Epithelial receptor 2; HER2p: HER2 positive; KEGG: Kyoto encyclopedia of genes and genomes database; OS: Overall survival; PCA: Principle component analysis; pCpG: Prognostic differentially methylated CpG; pDMG: Prognostic differentially methylated gene; PR: Progestogen receptor; ROC: Receiver operating characteristic curve; SVM: Support vector machine; TCGA: The Cancer Genome Atlas; TN: Triple negative

Acknowledgements

Not applicable.

Authors' contributions

XFD designed, supervised and financed this study. XC and JYZ conducted computational analysis. XFD and XC prepared the manuscript, figures and tables. All authors have read and approved the content of this study.

Availability of data and materials

GSE72251 and GSE72245 datasets were retrieved from the NCBI Gene Expression Omnibus (GEO) database [27]. Methylation and mRNA data as well as clinical information of The Cancer Genome Atlas (TCGA) [28] dataset were downloaded from the cBioportal (http://download.cbioportal.org/brca_tcga.tar.gz) [29].

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Biotechnology, Jiangnan University, Wuxi, Jiangsu, China. [2]Henan Academy of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China. [3]Wuxi School of Medicine, Jiangnan University, Wuxi, Jiangsu, China.

References

1. Yassi M, Shams Davodly E, Mojtabanezhad Shariatpanahi A, Heidari M, Dayyani M, Heravi-Moussavi A, Moattar MH, Kerachian MA. DMRFusion: a differentially methylated region detection tool based on the ranked fusion method. Genomics. 2018;110(6):366–74.
2. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer. 2011;2(6):607–17.
3. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glockner S, Piantadosi S, Gabrielson E, Pridham G, et al. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med. 2008;358(11):1118–28.
4. Leygo C, Williams M, Jin H, Chan M, Chu W, Grusch M, Cheng Y. DNA methylation as a noninvasive epigenetic biomarker for the detection of Cancer. Dis Markers. 2017;2017(1):3726595.
5. Xiaoke H, Huiyan L, Michal K, Wei W, Wenqiu W, Juan W, Ken F, Jiayi H, Heng Z, Shaohua Y. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci U S A. 2017;114(28):7414–9.
6. Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert Rev Mol Diagn. 2012;12(5):473–87.
7. van Steensel B, Braunschweig U, Filion GJ, Chen M, van Bemmel JG, Ideker T. Bayesian network analysis of targeting interactions in chromatin. Genome Res. 2010;20(2):190–200.
8. Roll JD, Rivenbark AG, Sandhu R, Parker JS, Jones WD, Carey LA, Livasy CA, Coleman WB. Dysregulation of the epigenome in triple-negative breast cancers: basal-like and claudin-low breast cancers express aberrant DNA hypermethylation. Exp Mol Pathol. 2013;95(3):276–87.
9. Roll JD, Rivenbark AG, Jones WD, Coleman WB. DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. Mol Cancer. 2008;7:15.
10. Prat A, Lluch A, Turnbull AK, Dunbier AK, Calvo L, Albanell J, de la Haba-Rodriguez J, Arcusa A, Chacon JI, Sanchez-Rovira P, et al. A PAM50-based Chemoendocrine score for hormone receptor-positive breast Cancer with an intermediate risk of relapse. Clin Cancer Res. 2017;23(12):3035–44.
11. Prat A, Galvan P, Jimenez B, Buckingham W, Jeirianian HA, Schaper C, Vidal M, Alvarez M, Diaz S, Ellis C, et al. Prediction of response to Neoadjuvant chemotherapy using Core needle biopsy samples with the Prosigna assay. Clin Cancer Res. 2016;22(3):560–6.
12. Whitehead C, Nelson R, Hudson P. Selection and optimization of a panel of early stage breast cancer prognostic molecular markers. Mod Pathol. 2004; 17:50A.
13. Jian D, Wang W, Zhou X, Jia Z, Wang J, Yang M, Zhao W, Jiang Z, Hu X, Zhu J. Interferon-induced protein 35 inhibits endothelial cell proliferation, migration and re-endothelialization of injured arteries by inhibiting the nuclear factor-kappa B pathway. Acta Physiol (Oxf). 2018;223(3):e13037.

14. Mizuhara E, Nakatani T, Minaki Y, Sakamoto Y, Ono Y. Corl1, a novel neuronal lineage-specific transcriptional corepressor for the homeodomain transcription factor Lbx1. J Biol Chem. 2005;280(5):3645–55.

15. Tomlinson I. The BMP pathway and predisposition to colorectal cancer. Cancer Genet Cytogenet. 2010;203(1):44.

16. Chedotal A, Kerjan G, Moreau-Fauvarque C. The brain within the tumor: new roles for axon guidance molecules in cancers. Cell Death Differ. 2005; 12(8):1044–56.

17. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012;491(7424): 399–405.

18. Schlaepfer IR, Rider L, Rodrigues LU, Gijon MA, Pac CT, Romero L, Cimic A, Sirintrapun SJ, Glode LM, Eckel RH, et al. Lipid catabolism via CPT1 as a therapeutic target for prostate cancer. Mol Cancer Ther. 2014;13(10):2361–71.

19. Du Q, Tan Z, Shi F, Tang M, Xie L, Zhao L, Li Y, Hu J, Zhou M, Bode A, et al. PGC1alpha/CEBPB/CPT1A axis promotes radiation resistance of nasopharyngeal carcinoma through activating fatty acid oxidation. Cancer Sci. 2019;110(6):2050–62.

20. Moody L, Xu GB, Chen H, Pan YX. Epigenetic regulation of carnitine palmitoyltransferase 1 (Cpt1a) by high fat diet. Biochim Biophys Acta Gene Regul Mech. 2019;1862(2):141–52.

21. Kumagai M, Guo X, Wang KY, Izumi H, Tsukamoto M, Nakashima T, Tasaki T, Kurose N, Uramoto H, Sasaguri Y, et al. Depletion of WNT10A prevents tumor growth by suppressing microvessels and collagen expression. Int J Med Sci. 2019;16(3):416–23.

22. Manoharan R, Seong HA, Ha H. Dual roles of serine-threonine kinase receptor-associated protein (STRAP) in redox-sensitive signaling pathways related to Cancer development. Oxidative Med Cell Longev. 2018;2018:5241524.

23. Brennan K, Offiah G, McSherry EA, Hopkins AM. Tight junctions: a barrier to the initiation and progression of breast cancer? J Biomed Biotechnol. 2010; 2010:460607.

24. Bikkavilli RK, Malbon CC. Mitogen-activated protein kinases and Wnt/beta-catenin signaling: molecular conversations among signaling pathways. Commun Integr Biol. 2009;2(1):46–9.

25. Guardavaccaro D, Clevers H: Wnt/beta-catenin and MAPK signaling: allies and enemies in different battlefields. Sci Signal 2012, 5(219):pe15.

26. Cheruku HR, Mohamedali A, Cantor DI, Tan SH, Nice EC, Baker MS. Transforming growth factor-β, MAPK and Wnt signaling interactions in colorectal cancer. EuPA Open Proteomics. 2015;8:104–15.

27. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30(1):207–10.

28. The Cancer Genome Atlas (TCGA) [https://tcga-data.nci.nih.gov/tcga/].

29. cBioportal [http://www.cbioportal.org/about].

30. Maurano Matthew T, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos John A. Role of DNA methylation in modulating transcription factor occupancy. Cell Rep. 2015;12(7):1184–95.

31. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, Shi B. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res. 2015;5(10): 2929–43.

32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

33. Bioconductor [https://www.bioconductor.org/].

34. Wang S, Zhu J. Variable selection for model-based high-dimensional clustering and its application to microarray data. Biometrics. 2008;64(2):440–8.

35. The Gene Ontology(GO) [http://www.geneon¬tology.org].

36. Kyoto Encyclopedia of Genes and Genomes database (KEGG) [http://www.genome.jp/kegg/pathway.html].

37. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.

38. Wang Y, Zhao Q, Lan N, Wang S. Identification of methylated genes and miRNA signatures in nasopharyngeal carcinoma by bioinformatics analysis. Mol Med Rep. 2018;17(4):4909–16.

Publisher's Note