Epigenetic clock detected a breast cancer mitosis subtype with improved immunotherapy

Shipeng Shang

College of Bioinformatics Science and Technology

Xin Li

College of Bioinformatics Science and Technology

Yue Gao

College of Bioinformatics Science and Technology

Shuang Guo

College of Bioinformatics Science and Technology

Hanxiao Zhou

College of Bioinformatics Science and Technology

Dailin Sun

College of Bioinformatics Science and Technology

Hongjia Liu

College of Bioinformatics Science and Technology

Hui Zhi

College of Bioinformatics Science and Technology

Peng Wang

College of Bioinformatics Science and Technology

Jing Bai

College of Bioinformatics Science and Technology

Shangwei Ning (✉ ningsw@ems.hrbmu.edu.cn)

Harbin Medical University

Xia Li

College of Bioinformatics Science and Technology

Epigenetic clock detected a breast cancer mitosis subtype with improved immunotherapy

Shipeng Shang[1], Xin Li[1], Gao Yue[1], Guo Shuang[1], Hanxiao Zhou[1], Dailin Sun[1], Hongjia Liu[1], Hui Zhi[1], Peng Wang[1], Jing Bai[1,*], Shangwei Ning[1,*] and Xia Li[1,*]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilong jiang 150081, China.

*Corresponding Author:

Xia Li, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Email: lixia@hrbmu.edu.cn

Shangwei Ning, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Email: ningsw@ems.hrbmu.edu.cn

Jing Bai, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Email: baijing@hrbmu.edu.cn

Abstract

Background

Epigenetic clock based on DNA methylation can estimate the epigenetic age of tissue and cell that can describe the process of aging. However, the exploration of diseases by the epigenetic clock is still an uncharted territory. Our objective was to assess the role of the epigenetic clock in breast cancer.

Methods

In this study, DNA methylation data of breast tissue sample was download from TCGA and GEO database. DNA methylation level of CpG sites and age of samples was calculated by using pearson correlation test. Differentially expressed genes were identified using the limma package and Kruskal-Wallis test was used for the difference between cancer subtypes.

Results

We developed a workflow to construct the Breast Epigenetic Clock (BEpiC) that could accurately predict the chronological age of normal breast tissue. Furthermore, the BEpiC was applied to breast cancer to identify three breast cancer subtypes (including development, homeostasis, and mitosis) by using the deviation between epigenetic age and chronological age. Interestingly, the prognosis of the three breast cancer subtypes is significantly different. In addition, the three breast cancer subtypes had distinct differences in multiple immune cells and the mitosis subtype had the highest tumor mutation burden that was used to estimate response to checkpoint inhibitors.

Conclusion

Our model highlights that epigenetic age of breast cancer samples had an important impact on immunotherapy. We constructed a BEpiC web server (http://bio-bigdata.hrbmu.edu.cn/BEpiC/) where users submit DNA methylation data and age information to predict the epigenetic age of breast tissue and breast cancer subtypes.

Trial registration

Not applicable

Keywords: DNA methylation; epigenetic clock; epigenetic age deceleration; breast cancer subtype; immunotherapy

Background

DNA methylation is an important epigenetic marker and plays critical roles during mammalian development, including X chromosome inactivation, cell differentiation and parental imprinting(1, 2). DNA methylation, especially methylation of CpG island (CGI), can suppress gene expression(3) and inactivate tumor suppressor genes in human cancers(4). DNA methylation exhibits dynamic changes(5) and are more stable than RNA during aging(6). Therefore, some researchers have built epigenetic clock that can predict human epigenetic age based on methylation levels of multiple human tissues and blood(7-11). However, DNA methylation pattern is specific in human tissues and shows a negative correlation with aging in some tissues(12, 13). A more accurate epigenetic clock needs to be developed for individual tissue. Epigenetic clock was also used to predict the biological age based on the methylation level of multiple tissues in mouse(14). Daniel et al. found that epigenetic clock could be used to evaluate

62   interventions that alter the rate of aging, such as calorie restricted diet(15). In addition,

63   the mitotic clock based on DNA methylation found that the epigenetic age acceleration

64   in normal tissues increases the risk of cancer(16, 17). In brief, epigenetic clock could

65   measure aging of tissues that are independent of chronological age.

66   Cancer is the most common type of malignancy which is currently one of the leading

67   causes of death worldwide(18). The development of cancer has been shown to be

68   associated with global hypomethylation and local hypermethylation(19). Previous

69   studies found that DNA methylation loss in late-replicating partial methylation domains

70   promotes immune evasion of tumors, especially in domains with CGI

71   hypermethylation(20). Hypermethylation in the promoter region of important gene can

72   lead to gene expression disorders that promote cancer(21, 22). DNA methylation occurs

73   at an early stage of cancer, so it plays an important role in cancer screening and

74   prognosis prediction(23, 24). Methylated DNA has been shown to be specific in cancers,

75   so it can be used as a potential tumor marker(25, 26). Breast cancer is a highly

76   heterogeneous cancer with a higher incidence in women. Studies have also found that

77   the increased risk of breast cancer may be associated with DNA methylation of

78   promoter region of BRCA1(27). Aging is associated with molecular, cellular, and

79   physiological changes that affect carcinogenesis and cancer growth(28, 29). Epigenetic

80   clock based on methylation that describes the aging process may be useful for the study

81   of cancer closely related to aging.

82   The response triggered by immunotherapy intervention will clearly target and eliminate

83   tumor cells while retaining normal cells(30). At present challenge facing

84   immunotherapy is selecting biomarkers that predict clinical responses to CTLA-4 and

85   PD-1 blockade. Multiple studies have shown that immunotherapy is associated with

86   breast cancer molecular subtypes that can provide prognostic information for breast

87   cancer patients(31, 32). However, some studies suggested that existing subtypes did not

88   predict treatment effects of patients(33).

89   We developed a breast epigenetic clock to identify cancer subtypes that respond

90   differently to immunotherapy based on the deviation between epigenetic age and

91   chronological age. Eventually, a web server of epigenetic clock constructed based on

92   DNA methylation that was used to calculate the epigenetic age and identify cancer

93   subtypes in breast tissue. Cancer subtypes based on epigenetic clock may provide a

94   theoretical basis for early cancer diagnosis and immunotherapy (Figure1).

95

Methods

DNA methylation data collection and preparation

98   DNA methylation datasets of breast cancer based on Illumina Infinium

99   HumanMethylation450 BeadChip array were obtained from the TCGA portal

100   (http://xena.ucsc.edu/). We preprocessed DNA methylation data by removing probes

101   that were on the sex chromosomes and chondriosome. In addition, the CpG sites with

102   single nucleotide polymorphism (SNP) and missing value were also removed. Finally,

103   351,427 CpG sites passed quality control. We removed the sample without age

104   information, and the remaining samples contained 788 samples of breast cancer and 96

105   samples of paracancerous tissue. In the follow-up analysis, we considered the

106 paracancerous samples as normal samples.

107 To avoid outlier samples, we used principal component analysis (PCA)(34) for normal

108 samples. Firstly, we calculated the squared distance from the first principal component

109 of each sample to the population mean as the z-score of the sample. Then, we converted

110 z-score(7) to a false-discovery rate using the Gaussian cumulative distribution and the

111 Benjamini-Hochberg procedure.

112
$$Z_i = (PC_i - \frac{1}{n}\sum_{i}^{n} PC_i)^2$$

113 Where Z_i

114 Samples with FDR below 0.05 are considered outliers and removed. Finally, four

115 samples were removed.

116 **The construction of breast epigenetic clock (BEpiC)**

117 Initially, correlation coefficient between age and DNA methylation level (beta values)

118 of CpG sites was calculated by using pearson correlation test and multiple testing

119 correction was performed using false discovery rate (FDR). CpG sites with FDR less

120 than 0.05 were defined as significant age-related CpG sites.

121 We randomly divided 92 samples into training set and testing set according to the ratio

122 of 1:1. We constructed the breast epigenetic clock using the elastic-net generalized

123 linear model as implemented in the GLMNET package(35). By using 10-fold cross-

124 validation, optimal lambda (2.296) was identified. Finally, we obtained 21 CpG sites

125 and corresponding weights through elastic regression network. The epigenetic age

126 EpiAge was defined as

$$\mathrm{EpiAge} = \left(\sum_{i}^{n} w_i * M_i \right) + c$$

127

128 where w_i is the weight of CpG_i, M_i is the methylation level of CpG_i and c= -86.729.

129 Eventually, we then used the validation set and a public dataset which at NCBI Gene

130 Expression Omnibus (GEO) under accession number GSE108213(20) and

131 GSE67919(36) to verify the predict performance of BEpiC.

132 **Genomic enrichment analysis of significant aging-associated CpG sites**

133 To understand the enrichment of aging-associated CpG sites on the genome.

134 Normalized degree of enrichment(14) was calculated as:

$$\mathrm{N}_i = \left(\frac{s}{b} \times \frac{B}{S} \right) - 1$$

135

136 Where s, b are the number of significant sites and the number of background sites at a

137 given region i, such as CpG island. S, B is total number of significant sites and total

138 number of background sites, respectively.

139 **Breast cancer subtypes based on epigenetic age**

140 We calculated the epigenetic age of breast cancer samples by using BEpiC. The mean

141 absolute error (MAE) between chronological age and epigenetic age of normal breast

142 sample from TCGA and GSE108213 is 5.3 years. We use twice the MAE (5.3 years) of

143 BEpiC in normal samples as the standard for dividing subtypes. Breast cancer samples

144 were divided into three subtypes, which were samples with deviation between

145 epigenetic age and chronological age higher than 10.6 years, samples with deviation

146 between epigenetic age and chronological age lower than 10.6 years, and the remaining

147 samples. Up-regulated genes for each breast cancer subtype were screened in a "one vs.

148 rest" fashion by using the Bioconductor package limma, with false-discovery rate-

149   corrected p-value less than 0.05 and fold change greater than 1.5 considered significant.

150   **Statistical Analysis**

151   CpG sites that differ in methylation between tumor and normal samples were defined

152   by t test and multiple testing correction was performed using FDR. CpG sites with both

153     $FDR<0.05$ and a minimum change of ±0.3 in DNA methylation level between tumor

154   samples and normal samples were defined as differential methylation CpG sites. The

155   Kaplan-Meier survival plots were used to estimate OS and disease-free survival (DFS),

156   and the difference in OS or DFS between the Development, Homeostasis and Mitosis

157   subtype was determined using log-rank tests with R package 'survival'. The mitotic

158   index could reflect the proportion of dividing cells in the population. Studies have

159   confirmed that the mitotic index can be represented by the average expression level of

160   mRNA of CDKN3, ILF2, KDELR2, RFC4, TOP2A, MCM3, KPNA2, CKS2 and

161   CDC2(16). We used the average FPKM expression values of the nine genes in the

162   cancer samples to represent the mitotic indexes of each sample. To calculate TMB of

163   breast cancer samples, synonymous variant and intron variant was excluded(13).

164   Finally, the number of mutations on the million-base was calculated as the TMB of the

165   sample. Tumor-infiltrating immune cells in breast cancer samples was estimated by

166   CIBERSORT(37). The R package "estimate"(38) was used to evaluate stromal score

167   and immune score of breast cancer sample. All the statistical analyses were performed

168   in R versions 3.5.3.

169   # Results

170   **DNA methylation is associated with age in human breast tissue**

171 DNA methylation data of 884 breast samples were downloaded from The Cancer

172 Genome Atlas (TCGA) and tissues adjacent to cancer were extracted as normal breast

173 tissues, aged 28 to 90. Aging-associated CpG sites were identified by using pearson

174 correlation analysis between age and DNA methylation levels of CpG sites in normal

175 breast tissues. 4,330 CpG sites were associated with age, of which 4,150 (95.8%) CpG

176 sites were significantly positively correlated with age (Figure 2A). It shows that DNA

177 methylation levels of most of aging-associated CpG sites increases with age in normal

178 breast tissues (Figure 2B). This phenomenon has also been demonstrated in other

179 human tissues(39). Some CpG sites regulated the expression of age-associated gene

180 (Supplementary Figure 1). It proves that aging-associated DNA methylation may cause

181 gene expression to change with aging.

182 We corrected the number of all probes on each genomic region in Illumina Infinium

183 HumanMethylation450 BeadChip to test the enrichment of aging-related CpG sites on

184 the genomic region. The significantly positive aging-associated CpG sites mainly

185 enriched in CpG island shore, exon and 2Kb downstream regions of the gene. CpG sites

186 which significantly negatively correlated with age mainly enriched in the CpG island

187 desert, introns and intergenic regions. We found that changes of DNA methylation

188 levels of aging-associated CpG sites rarely occur on CpG islands and promoter regions

189 (Figure 2C). It can be inferred that these two regions may protect the DNA methylation

190 levels of aging-associated CpG sites and this phenomenon was also found in multiple

191 tissues of mouse(14). Next, the relationship between aging and cancer in DNA

192 methylation level was explored. We identified 5,192 differential methylation CpG sites

9

193 (DMC) between breast cancer samples and normal samples. There was significant

194 correlation (cor: -0.4457) between DNA methylation difference level of DMC and

195 aging-correlation coefficient of DMC by pearson correlation test (Figure 2D). This

196 suggested that cancer and aging were opposite processes(40) in breast tissue(12). This

197 phenomenon is also confirmed at the level of gene expression(41). Hypermethylated

198 CpG sites in cancer were demethylated during aging, conversely, Hypomethylated CpG

199 sites in cancer were methylated during aging.

200

201 **Epigenetic clock could predict the chronological age of normal breast tissue**

202 A dataset consisting of 92 normal breast samples and 351427 CpG sites was constructed.

203 The dataset was randomly divided into training set and testing set according to the ratio

204 of 1:1. We used an elastic-net regression model to predict the age of normal breast tissue

205 samples and the final model was based on 21 CpG sites. BEpiC was constructed using

206 the DNA methylation level and weight of 21 CpG sites and performed well across

207 training set and testing set. In training set, epigenetic age is highly correlated with

208 chronological age (cor: 0.9807), with mean absolute error (MAE) is 3.62 years (Figure

209 3A). In testing set, the correlation coefficient between epigenetic age and chronological

210 age is 0.8815, with MAE is 6.10 years (Figure 3B). GSE108213(42) dataset was

211 downloaded from Gene Expression Omnibus (GEO https://www.ncbi.nlm.nih.gov/geo/)

212 as an independent validation set, including 85 normal breast samples. Epigenetic age of

213 samples from GSE108213 were calculated by BEpiC and is relevant to chronological

214 age (cor: 0.657), with MAE is 5.77 years (Figure 3C). The weight of 21 CpG sites is

215 closely related to correlation coefficients between DNA methylation level and age

216 (Figure 3D). In addition, the 21 CpG sites are mainly located in the gene body and

217 neighboring genes were enriched in the known functions including DNA replication,

218 mitotic cell cycle and lipid transport.

219 Then, we compared the prediction performance of BEpiC and the Horvath clock for

220 epigenetic age of breast tissues. Epigenetic age identified by BEpiC had significant

221 correlation with that identified by Horvath clock in Dataset1 and GSE108213

222 (Supplementary Figure 2). And MAE between epigenetic age and chronological age

223 from BEpiC and Horvath clock were 4.86 years and 7.59 years in normal breast samples

224 from TCGA, respectively. In GSE108213, MAE between epigenetic age and

225 chronological age from BEpiC and Horvath clock were 5.77 years and 9.60 years,

226 respectively (Figure 3E). GSE67919(36) was also downloaded to compare the

227 predictive performance of BEpiC and Horvath model for epigenetic age of normal

228 breast tissue. The correlation between chronological age and epigenetic age by BEpiC

229 and Horvath was 0.703 and 0.679, respectively (Figure 3F-G). The MAE between

230 epigenetic age and chronological age of BEpiC was significantly less than that of

231 Horvath in GSE67919 (Figure 3H). Thus, BEpiC was more accurate than Horvath clock

232 for predicting epigenetic age for breast normal samples.

233

234 **Epigenetic clock identified three breast cancer subtypes**

235 The DNA methylation pattern of breast cancer samples had changed compared to

236 normal tissue. To test whether epigenetic age could predict the chronological age of

237   breast tumor sample, BEpiC was used to calculate epigenetic age in breast cancer

238   samples. We found a large difference between the epigenetic age and chronological age

239   of breast cancer samples (MAE: 17.55 years). This is because the aging process of

240   breast cancer tissue was significantly different from that of normal tissue(43).

241   It has not yet been determined whether there is heterogeneity in breast cancer tumor

242   samples with different aging patterns. So, breast cancer samples were divided into three

243   subtypes based on the deviation between epigenetic age and chronological age (Figure

244   4A). A sample whose epigenetic age was 10.6 years (twice the MAE of BEpiC in

245   normal samples) older than chronological age was defined as epigenetic age

246   acceleration. Then, a sample whose epigenetic age was 10.6 years younger than

247   chronological age was defined as epigenetic age deceleration. To obtain the biological

248   characteristics of each subtype, we identified genes which were significantly up-

249   regulated in a subtype compared with other subtypes (Figure 4B). High expression

250   genes in the epigenetic age acceleration samples mainly enriched in mammary gland

251   epithelium development, mammary gland epithelial cell proliferation, mammary gland

252   development, branching involved in mammary gland duct morphogenesis and

253   reproductive structure development, so we termed this subtype as development

254   (Supplementary Figure 3A). High expression genes in the epigenetic age deceleration

255   samples mainly enriched in mitotic nuclear division, mitotic spindle assembly, spindle

256   assembly, nuclear division and organelle fission, so we termed this subtype as mitosis

257   (Supplementary Figure 3C). High expression genes in remaining samples mainly

258   enriched in retina homeostasis, antibacterial humoral response, tissue homeostasis,

12

receptor-mediated endocytosis and regulation of respiratory burst, so we termed this

subtype as homeostasis (Supplementary Figure 3B).

We compared the overall survival and disease-free survival of breast cancer patients

with three subtypes (Figure 4C). Patients of these three subtypes have significant

differences in overall survival (p=0.00047) and disease-free survival (Supplementary

Figure 3D). Patients with development subtype had the best prognosis and patients with

mitosis subtype had the worst prognosis. We inferred that breast cancer patients whose

epigenetic age is decelerated may have worse survival prognosis. Similarly, samples

with epigenetic age deceleration were also found to have higher cancer risk in normal

colon tissue(44). This may suggest that epigenetic age deceleration is danger signal in

both normal and cancer tissues. Mitotic index of breast cancer samples was calculated

and negatively correlated with epigenetic age (cor: -0.288). Patients with mitosis

subtype had a higher tumor mitotic rate than patients in other subtypes (Figure 4D).

This indicates that tumor cells in mitosis subtype were more active than other

subtypes(18), so breast cancer patients with mitosis subtype had a worse survival

prognosis.

Molecular and clinical characteristics analysis of mitosis subtype.

To explain the differences of the three subtypes at survival, we examined the somatic

mutation of the three subtypes. TP53 is a tumor suppressor gene, and its mutation occurs

in many breast cancer patients(45, 46). In 788 samples, TP53 mutations occurred 223

(28.3%) breast cancer samples (Supplementary Figure 4). We found that the mutation

frequency of TP53 is lowest in development subtype (24%), the mutation frequency of

TP53 is highest in mitosis subtype (32%), and the mutation frequency of TP53 is 28%

in homeostasis subtype. TTN that proved to be associated with breast cancer

prognosis(47) had different mutation frequency in three subtypes. The mutation

frequencies of TTN were 18%, 21% and 31% in development, homeostasis, and mitosis

subtype, respectively (Figure 5A-C). Next, copy number variation in breast cancer

patients was analyzed. It was found that the three subtypes had significant differences

in copy number amplification (p<2.2e-16) and copy number deletion(p=1.3e-14). The

mitosis subtype was significantly higher than the other two subtypes in terms of copy

number amplification and copy number deletion (Figure 5D-E). This may reveal the

reasons for the poor survival prognosis of the mitosis subtype.

The clinical information of breast cancer patients was obtained from TCGA. The

proportion of patients with three subtypes was calculated in stage T, stage N, stage M

and pathological stage. It was found that the proportion of patients with mitosis subtype

in T4, N3, stage iii-iv was higher than the other two subtypes (Supplementary Table 2).

It also demonstrated that most breast cancer patients with mitosis subtype were

advanced cancer patients with poor prognosis.

We analyzed the relationship between PAM50 signature and BEpiC and found that

Basal subtypes has a large overlap with mitosis subtypes (Figure 5F). This indicates

that the epigenetic age of breast cancer tissues was decelerate in the Basal subtypes, so

the prognosis of Basal subtypes patients was worse.

Immunotherapy has a better therapeutic effect for mitotic subtype

303

304 We focus on the treatment of patients with three subtypes, and whether the patients with

305 three subtypes are different in terms of medication. We compared the prognosis of three

306 breast cancer subtypes taking tamoxifen as a commonly used drug in the treatment of

307 breast cancer (Supplementary Figure 5). There was no significant difference in

308 prognosis of the three subtypes taking tamoxifen (p=0.57).

309 Since the patients of the three subtypes were significantly different at the molecular

310 level and survival prognosis, we wanted to know if the patients of the three subtypes

311 are different for immunotherapy. The proportion of immune infiltrating cells in breast

312 cancer samples was calculated using CIBERSORT (https://cibersort.stanford.edu/), and

313 the difference in the proportion of immune infiltrating cells of samples of the three

314 subtypes was compared (Supplementary Figure 6A). Activated T cells CD4 memory, T

315 cells follicular helper, Macrophages M0 and Macrophages M1 differed significantly

316 among the three subtypes and had a higher proportion in mitosis subtype. While naive

317 B cells, plasma cells, resting memory CD4 T cells, resting dendritic cells and resting

318 mast cells differed significantly among the three subtypes and had a lower proportion

319 in mitosis subtype (Figure 6). The different infiltration patterns indicated that the three

320 subtypes might have different immune system.

321

322 Previous studies have shown that stromal cells have a huge impact on immunotherapy

323 (48, 49). Stromal scores of breast cancer samples were evaluated by using R Package

324 "estimate". Stromal score of mitosis subtype was significantly higher than stromal score

325   of development subtype (p = 0.011) and homeostasis subtype (p = 0.00046,

326   Supplementary Figure 6B). This may indicate that stromal environment of patients with

327   mitosis subtype is significantly different from the other two subtypes. We also

328   compared the tumor mutation burden among the three subtypes and found that tumor

329   mutation burden (TMB) of patients with mitosis subtypes had significantly higher than

330   the other two subtypes (Figure 7A). Basal subtypes and Her2 subtypes also had higher

331   TMB than the LumA subtypes and LumB subtypes. The proved that Basal subtypes and

332   Her2 subtypes might have a better response to immunotherapy (Supplementary Figure

333   6C). It shows that mitosis subtype had a better immunotherapy response than the other

334   two subtypes(50).

335   Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) is a key immune checkpoint

336   which maintain the immune system(51). Expression level of CTLA4 was compared in

337   three subtypes and significantly different between development subtype and mitosis

338   subtype (Figure 7B). Methylation of CTLA4 can be used to assess response of

339   immunotherapy(52), so we focus on whether expression of CTLA4 is regulated by

340   methylation. Four CpG sites were found in the promoter region of CTLA4 and their

341   methylation levels were lower in the mitosis subtype compared to the development

342   subtype. It suggests difference in methylation of CpG results in difference of expression

343   levels of immune genes and ultimately might affect the response to immunotherapy(53).

344

345   **Online tool for breast epigenetic clock**

346   To help researchers get epigenetic age of breast tissue and breast cancer subtype, we

347 developed a BEpiC web server (http://bio-bigdata.hrbmu.edu.cn/BEpiC/). The platform

348 (Figure 8) provides a friendly interface that provides computing and download

349 functions. To obtain the epigenetic age of breast tissue, users need to submit

350 methylation data containing 21 CpG sites. If users want to get the breast cancer subtype,

351 they need to add chronological age information.

352

Discussion

354 Horvath and Hannum have built epigenetic clock in multiple human tissues and blood(8,

355 14, 54), but DNA methylation exhibits tissue-specific pattern in mammals that regulates

356 tissue-specific gene transcription(55). Therefore, it is necessary to establish a tissue-

357 specific age prediction model, which can more accurately predict the chronological age

358 of tissues. In this study, we systematically examined the associations between age and

359 DNA methylation in breast tissue. BEpiC was constructed to predict epigenetic age

360 significantly associated with chronological age in normal breast tissue. Epigenetic age

361 that can describe the degree of aging was associated with mortality independently of

362 chronological age(44). BEpiC was also used to identify breast cancer subtypes based

363 on the difference between epigenetic age of breast cancer samples and chronological

364 age. The three subtypes had significant differences in prognosis and the mitosis subtype

365 had the worst prognosis. We also found that epigenetic age had a negative correlation

366 with cell mitosis. Epigenetic age deceleration is a risk factor not only for normal

367 tissues(44, 56), but also for cancer tissues. It might confirm that breast tissue tissues

368 with epigenetic age deceleration had higher mitotic index. Mitotic cell division

17

369 increases tumor mutation burden and copy number variation, making tumor cells easier

370 to recognize by immune cells(28).

371 So, we studied the response to immunotherapy among the three subtypes. The mitosis

372 subtype had higher TMB and it might lead to that mitosis subtype had a better

373 immunotherapy response than the other two subtypes(50).

Conclusions

375 Using DNA methylation data, we constructed an apparent clock of breast tissue that

376 could be used to classify breast cancer subtypes. Our results suggest that breast cancer

377 samples presenting epigenetic age deceleration had a poor survival prognosis and a

378 better response to immunotherapy. The findings may contribute to the diagnosis and

379 treatment of breast cancer.

380

List of abbreviations

382 BEpiC: Breast Epigenetic Clock

383 CGI: CpG island

384 TCGA: The Cancer Genome Atlas

385 DMC: Differential methylation CpG sites

386 GEO: Gene Expression Omnibus

387 CIBERSORT: Cell-type Identification By Estimating Relative Subsets Of RNA

388 Transcripts

389 TMB: Tumor mutation burden

390 CTLA-4: Cytotoxic T-lymphocyte-associated protein 4

18

391 SNP: Single nucleotide polymorphism

392 PCA: Principal component analysis

393 MAE: Mean absolute error

394 FDR: False-discovery rate

395

Declarations

396

Ethics approval and consent to participate

397

Not applicable

398

Consent for publication

399

Not applicable

400

Availability of data and materials

401

All data relevant to the study are included in the article or as supplementary material.

402

Competing interests

403

No potential conflicts of interest were disclosed.

404

Funding

405

408

Author's contributions

409

Xia Li, Shangwei Ning and Jing Bai designed and directed all the research. Shipeng

410

Shang, Xin Li, Gao Yue, Guo Shuang, Hanxiao Zhou, Dailin Sun, Hongjia Liu, Hui

411

Zhi and Peng Wang performed the data processing and experimental

412

413   analysis. Shipeng Shang, Xin Li and Gao Yue drafted the manuscript. All authors

414   reviewed and approved the final version of the manuscript.

415   **Acknowledgments**

418

References

420   1.   Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet.
421   2013;14(3):204-20.

422   2.   Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development
423   and disease. Nat Rev Mol Cell Biol. 2019;20(10):590-607.

424   3.   He XJ, Chen T, Zhu JK. Regulation and function of DNA methylation in plants and animals. Cell Res.
425   2011;21(3):442-65.

426   4.   Levine JJ, Stimson-Crider KM, Vertino PM. Effects of methylation on expression of TMS1/ASC in
427   human breast cancer cells. Oncogene. 2003;22(22):3475-88.

428   5.   Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging
429   (Albany NY). 2011;3(10):1018-27.

430   6.   Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and
431   early recurrence in stage I lung cancer. N Engl J Med. 2008;358(11):1118-28.

432   7.   Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles
433   reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359-67.

434   8.   Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.

435   9.   Voisin S, Harvey NR, Haupt LM, Griffiths LR, Ashton KJ, Coffey VG, et al. An epigenetic clock for
436   human skeletal muscle. J Cachexia Sarcopenia Muscle. 2020;11(4):887-98.

437   10.   Shireby GL, Davies JP, Francis PT, Burrage J, Walker EM, Neilson GWA, et al. Recalibrating the
438   epigenetic clock: implications for assessing biological age in the human cortex. Brain. 2020.

439   11.   Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, et al. Improved precision of
440   epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med.
441   2019;11(1):54.

442   12.   Dmitrijeva M, Ossowski S, Serrano L, Schaefer MH. Tissue-specific DNA methylation loss during
443   ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates.
444   Nucleic Acids Res. 2018;46(14):7022-39.

445   13.   Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine.
446   Lancet. 2018;392(10149):777-86.

447   14.   Stubbs TM, Bonder MJ, Stark AK, Krueger F, Team BIAC, von Meyenn F, et al. Multi-tissue DNA
448   methylation age predictor in mouse. Genome Biol. 2017;18(1):68.

449   15.   Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation

450   Profiling to Evaluate Biological Age and Longevity Interventions. Cell Metab. 2017;25(4):954-60 e6.

451   16.   Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, et al. Correlation of an epigenetic mitotic
452   clock with cancer risk. Genome Biol. 2016;17(1):205.

453   17.   Youn A, Wang S. The MiAge Calculator: a DNA methylation-based mitotic age calculator of human
454   tissue types. Epigenetics. 2018;13(2):192-206.

455   18.   Collaborators GBDCoD. Global, regional, and national age-sex-specific mortality for 282 causes of
456   death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease
457   Study 2017. Lancet. 2018;392(10159):1736-88.

458   19.   Shukla V, Coumoul X, Lahusen T, Wang RH, Xu X, Vassilopoulos A, et al. BRCA1 affects global DNA
459   methylation through regulation of DNMT1. Cell Res. 2010;20(11):1201-15.

460   20.   Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, et al. DNA methylation loss promotes immune
461   evasion of tumours with high mutation and copy number load. Nat Commun. 2019;10(1):4278.

462   21.   Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter
463   hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst.
464   2000;92(7):564-9.

465   22.   Lee DD, Leao R, Komosa M, Gallo M, Zhang CH, Lipman T, et al. DNA hypermethylation within TERT
466   promoter upregulates TERT expression in cancer. J Clin Invest. 2019;129(1):223-9.

467   23.   Muller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, et al. Methylation changes in
468   faecal DNA: a marker for colorectal cancer screening? Lancet. 2004;363(9417):1283-5.

469   24.   Watanabe Y, Kim HS, Castoro RJ, Chung W, Estecio MR, Kondo K, et al. Sensitive and specific
470   detection of early gastric cancer with DNA methylation analysis of gastric washes. Gastroenterology.
471   2009;136(7):2149-58.

472   25.   Liu H, Liu X, Zhang S, Lv J, Li S, Shang S, et al. Systematic identification and annotation of human
473   methylation marks based on bisulfite sequencing methylomes reveals distinct roles of cell type-specific
474   hypomethylation in the regulation of cell identity genes. Nucleic acids research. 2016;44(1):75-94.

475   26.   Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, et al. Detection in fecal DNA of colon cancer-
476   specific methylation of the nonexpressed vimentin gene. J Natl Cancer Inst. 2005;97(15):1124-32.

477   27.   Prajzendanc K, Domagala P, Hybiak J, Rys J, Huzarski T, Szwiec M, et al. BRCA1 promoter
478   methylation in peripheral blood is associated with the risk of triple-negative breast cancer. Int J Cancer.
479   2020;146(5):1293-8.

480   28.   Balducci L, Ershler WB. Cancer and ageing: a nexus at several levels. Nat Rev Cancer. 2005;5(8):655-
481   62.

482   29.   Wang Y, Zhang J, Xiao X, Liu H, Wang F, Li S, et al. The identification of age-associated cancer
483   markers by an integrative analysis of dynamic DNA methylation changes. Sci Rep. 2016;6:22722.

484   30.   Jia H, Truica CI, Wang B, Wang Y, Ren X, Harvey HA, et al. Immunotherapy for triple-negative breast
485   cancer: Existing challenges and exciting prospects. Drug Resist Updat. 2017;32:1-15.

486   31.   Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-
487   infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771
488   patients treated with neoadjuvant therapy. The Lancet Oncology. 2018;19(1):40-50.

489   32.   Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of
490   breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A.
491   2001;98(19):10869-74.

492   33.   Pogue-Geile KL, Song N, Jeong JH, Gavin PG, Kim SR, Blackmon NL, et al. Intrinsic subtypes, PIK3CA
493   mutation, and the degree of benefit from adjuvant trastuzumab in the NSABP B-31 trial. J Clin Oncol.

494 2015;33(12):1340-7.

495 34.   Svante, Wold, and, Kim, Esbensen, and, et al. Principal component analysis.

496 35.   Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via
497 Coordinate Descent. J Stat Softw. 2010;33(1):1-22.

498 36.   Hair BY, Xu Z, Kirk EL, Harlid S, Sandhu R, Robinson WR, et al. Body mass index associated with
499 genome-wide methylation in breast tissue. Breast Cancer Res Treat. 2015;151(2):453-63.

500 37.   Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets
501 from tissue expression profiles. Nat Methods. 2015;12(5):453-7.

502 38.   Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring
503 tumour purity and stromal and immune cell admixture from expression data. Nat Commun.
504 2013;4:2612.

505 39.   Perez RF, Tejedor JR, Bayon GF, Fernandez AF, Fraga MF. Distinct chromatin signatures of DNA
506 hypomethylation in aging and cancer. Aging Cell. 2018;17(3):e12744.

507 40.   Campisi J. Aging, cellular senescence, and cancer. Annu Rev Physiol. 2013;75:685-705.

508 41.   Chatsirisupachai K, Palmer D, Ferreira S, de Magalhaes JP. A human tissue-specific transcriptomic
509 analysis reveals a complex relationship between aging, cancer, and cellular senescence. Aging Cell.
510 2019;18(6):e13041.

511 42.   Harlid S, Xu Z, Kirk E, Wilson LE, Troester MA, Taylor JA. Hormone therapy use and breast tissue
512 DNA methylation: analysis of epigenome wide data from the normal breast study. Epigenetics.
513 2019;14(2):146-57.

514 43.   Calcinotto A, Kohli J, Zagato E, Pellegrini L, Demaria M, Alimonti A. Cellular Senescence: Aging,
515 Cancer, and Injury. Physiological Reviews. 2019;99(2):1047-78.

516 44.   Wang T, Maden SK, Luebeck GE, Li CI, Newcomb PA, Ulrich CM, et al. Dysfunctional epigenetic aging
517 of the normal colon and colorectal cancer risk. Clin Epigenetics. 2020;12(1):5.

518 45.   Zhang Y, Xiong S, Liu B, Pant V, Celii F, Chau G, et al. Somatic Trp53 mutations differentially drive
519 breast cancer and evolution of metastases. Nat Commun. 2018;9(1):3953.

520 46.   Patocs A, Zhang L, Xu Y, Weber F, Caldes T, Mutter GL, et al. Breast-cancer stromal cells with TP53
521 mutations and nodal metastases. The New England journal of medicine. 2007;357(25):2543-51.

522 47.   Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, et al. Genetic variants associated
523 with breast size also influence breast cancer risk. BMC Med Genet. 2012;13:53.

524 48.   Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour
525 microenvironment. Nat Rev Immunol. 2015;15(11):669-82.

526 49.   Mao M, Yu Q, Huang R, Lu Y, Wang Z, Liao L. Stromal score as a prognostic factor in primary gastric
527 cancer and close association with tumor immune microenvironment. Cancer Med. 2020;9(14):4980-90.

528 50.   Huang RSP, Haberberger J, Severson E, Duncan DL, Hemmerich A, Edgerly C, et al. A pan-cancer
529 analysis of PD-L1 immunohistochemistry and gene amplification, tumor mutation burden and
530 microsatellite instability in 48,782 cases. Mod Pathol. 2020.

531 51.   Xiaolei L, Wenhui S, Changshun S, Yufang S, Weidong H. Emerging predictors of the response to
532 the blockade of immune checkpoints in cancer therapy. Cellular & Molecular Immunology. 2018.

533 52.   Goltz D, Gevensleben H, Vogt TJ, Dietrich J, Dietrich D. CTLA4 methylation predicts response to
534 anti–PD-1 and anti–CTLA-4 immunotherapy in melanoma patients. jci insight. 2018.

535 53.   Goltz D, Gevensleben H, Vogt TJ, Dietrich J, Golletz C, Bootz F, et al. CTLA4 methylation predicts
536 response to anti-PD-1 and anti-CTLA-4 immunotherapy in melanoma patients. JCI Insight. 2018;3(13).

537 54.   Knight AK, Craig JM, Theda C, Baekvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, et al. An

538 epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol.
539 2016;17(1):206.
540 55. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology.
541 2013;38(1):23-38.
542 56. McKenna BG, Hendrix CL, Brennan PA, Smith AK, Stowe ZN, Newport DJ, et al. Maternal prenatal
543 depression and epigenetic age deceleration: testing potentially confounding effects of prenatal stress
544 and SSRI use. Epigenetics. 2020:1-11.

545

Figures legends

Figure 1. The Construction and application of breast epigenetic clock.

Figure 2. Aging-associated CpG sites in normal breast tissue. (A) Overview of pearson correlation calculated in normal breast tissue. Nominal correlations are highlighted (p value < 0.05) in wathet (positive) and light red (negative). Significant correlations are highlighted (FDR < 0.05) in blue (positive) and red (negative). **(B)** Heatmap of DNA methylation of aging-associated CpG sites. A white to green gradient represents the gradual increase in age. **(C)** Genome enrichment of aging-associated CpG sites (tested using chi-square test; *p value < 0.05). **(D)** Correlation between DNA methylation difference of DMC and aging-correlation coefficient of DMC. The horizontal axis represents the correlation coefficient between age and DNA methylation of DMC in normal samples. The vertical axis represents DNA methylation difference between cancer samples and normal samples.

Figure 3. Chronological age was predicted based on DNA methylation in human breast tissue. (A-C) Epigenetic age vs. chronological age from training set **(A)**, testing set **(B)** and GSE108213 dataset **(C)**. **(D)** Scatter plot depicting the relationship between weight and aging-correlation coefficient of CpG sites. Blue spots are negatively aging-associated CpG sites and orange spots are positively aging-associated CpG sites. **(E)**

Comparison of predictive performance between BEpiC and the Horvath clock. The Y-axis represents the mean absolute error between epigenetic age and chronological age (tested using t test; * represent p<0.05, ** represent p<0.01, *** represent p<0.001, **** represent p<0.0001). **(F-G)** Chronological age vs. epigenetic age estimated by BEpiC **(F)** and Horvath clock **(G)** in GSE67919. **(H)** Comparison of predictive performance between BEpiC and the Horvath clock in GSE67919 (tested using t test; * represent p<0.05, ** represent p<0.01, *** represent p<0.001, **** represent p<0.0001).

Figure 4. Three breast cancer subtypes identified by BEpiC based on epigenetic age. (A) Subtypes of breast cancer were classified based on the deviation between methylated age and chronological age. **(B)** Gene signature expression profiler and GO annotation for the three subtypes. The red dot, the yellow dot, and the blue dot represent development subtype, homeostasis subtype and mitosis subtype, respectively. **(C)** Kaplan–Meier curve for patients of the three subtypes. **(D)** Mitotic index in three subtypes (p values by Wilcoxon test).

Figure 5. Molecular differences among the three subtypes. (A-C) Gene mutation frequency in development subtype **(A)**, homeostasis subtype **(B)** and mitosis subtype **(C)**. **(D-E)** Number of copy number amplification **(D)** and copy number deletion **(E)** among the three subtypes (P values by Kruskal-Wallis test). **(F)** Association between breast cancer subtypes defined by BEpiC and pam50 molecular subtypes.

Figure 6. Immune infiltrating cells that differ significantly between the three subtypes. (A-J) Comparison of proportion of T cell CD4 memory activated, T cell CD4

586 memory resting, T cell follicular helper, Macrophages M0, Macrophages M1, B cells

587 naive, Plasma cells, Dendritic resting, Mast cells resting among the three subtypes (P

588 values by Kruskal-Wallis test).

589 **Figure 7. Comparison of immunotherapeutic response among the three subtypes.**

590 **(A)** Tumor mutation burden in each subtype (P values by Kruskal-Wallis test). **(B)**

591 Expression of CTLA4 in each subtype (P values by Kruskal-Wallis test). **(C-F)**

592 Methylation level of cg05074138 **(C)**, cg08460026 **(D)**, cg22572158 **(E)** and

593 cg26091609 **(F)** in development subtype and mitosis subtype.

594 **Figure 8. Schematic overview of the breast epigenetic clock.** The web server could

595 use the data submitted by the user to calculate the epigenetic age and breast cancer

596 subtypes.

597

Figures

Figure 1

The Construction and application of breast epigenetic clock.

Figure 2

Aging-associated CpG sites in normal breast tissue. (A) Overview of pearson correlation calculated in normal breast tissue. Nominal correlations are highlighted (p value < 0.05) in wathet (positive) and light red (negative). Significant correlations are highlighted (FDR < 0.05) in blue (positive) and red (negative). (B) Heatmap of DNA methylation of aging-associated CpG sites. A white to green gradient represents the gradual increase in age. (C) Genome enrichment of aging-associated CpG sites (tested using chi-square test; *p value < 0.05). (D) Correlation between DNA methylation difference of DMC and aging-correlation coefficient of DMC. The horizontal axis represents the correlation coefficient between age and DNA

methylation of DMC in normal samples. The vertical axis represents DNA methylation difference between cancer samples and normal samples.

Figure 3

Chronological age was predicted based on DNA methylation in human breast tissue. (A-C) Epigenetic age vs. chronological age from training set (A), testing set (B) and GSE108213 dataset (C). (D) Scatter plot depicting the relationship between weight and aging-correlation coefficient of CpG sites. Blue spots are

negatively aging-associated CpG sites and orange spots are positively aging-associated CpG sites. (E) Comparison of predictive performance between BEpiC and the Horvath clock. The Y-axis represents the mean absolute error between epigenetic age and chronological age (tested using t test; * represent p<0.05, ** represent p<0.01, *** represent p<0.001, **** represent p<0.0001). (F-G) Chronological age vs. epigenetic age estimated by BEpiC (F) and Horvath clock (G) in GSE67919. (H) Comparison of predictive performance between BEpiC and the Horvath clock in GSE67919 (tested using t test; * represent p<0.05, ** represent p<0.01, *** represent p<0.001, **** represent p<0.0001).

Figure 4

Three breast cancer subtypes identified by BEpiC based on epigenetic age. (A) Subtypes of breast cancer were classified based on the deviation between methylated age and chronological age. (B) Gene

signature expression profiler and GO annotation for the three subtypes. The red dot, the yellow dot, and the blue dot represent development subtype, homeostasis subtype and mitosis subtype, respectively. (C) Kaplan–Meier curve for patients of the three subtypes. (D) Mitotic index in three subtypes (p values by Wilcoxon test).
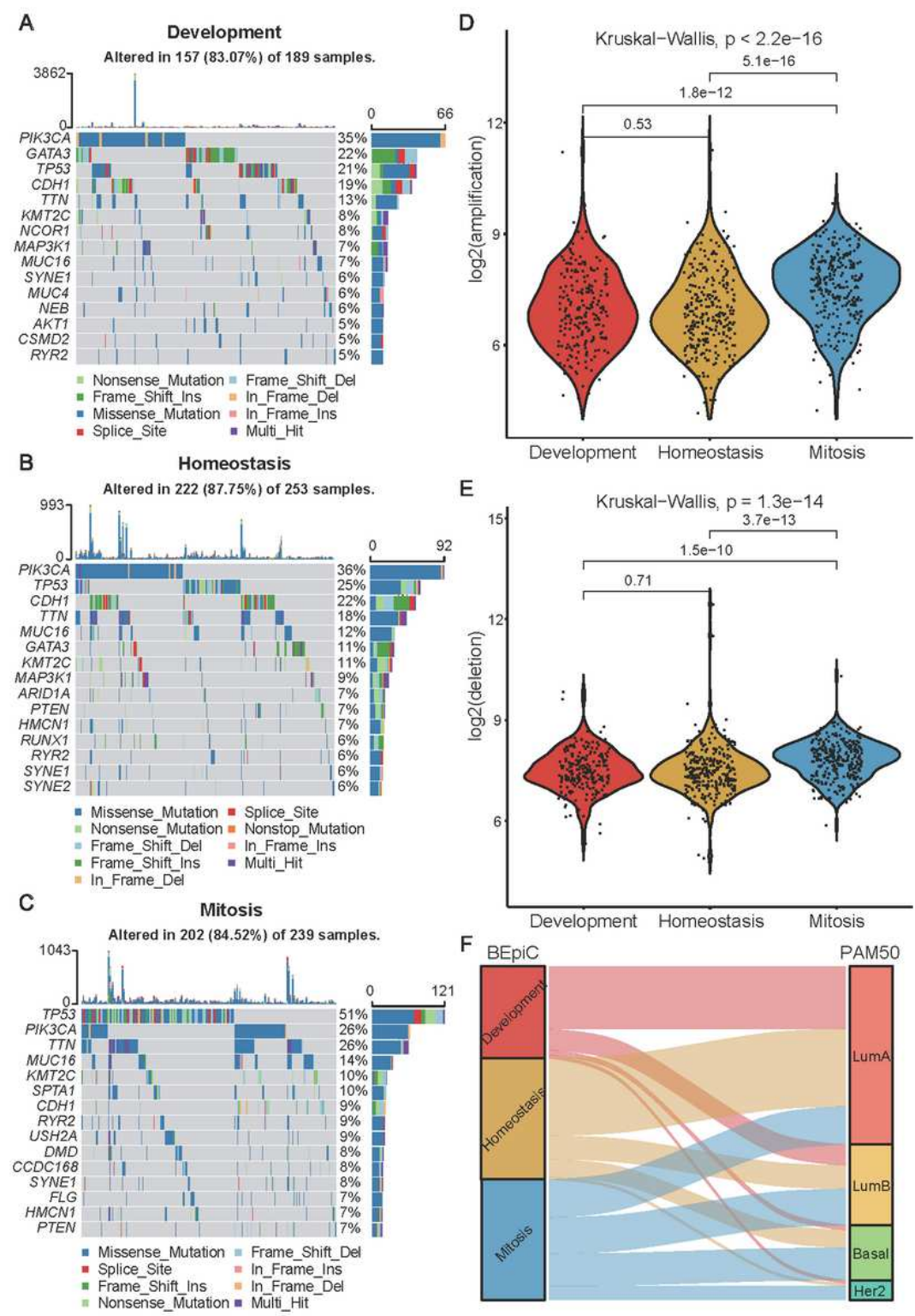

Figure 5

Molecular differences among the three subtypes. (A-C) Gene mutation frequency in development subtype (A), homeostasis subtype (B) and mitosis subtype (C). (D-E) Number of copy number amplification (D) and copy number deletion (E) among the three subtypes (P values by Kruskal-Wallis test). (F) Association between breast cancer subtypes defined by BEpiC and pam50 molecular subtypes.
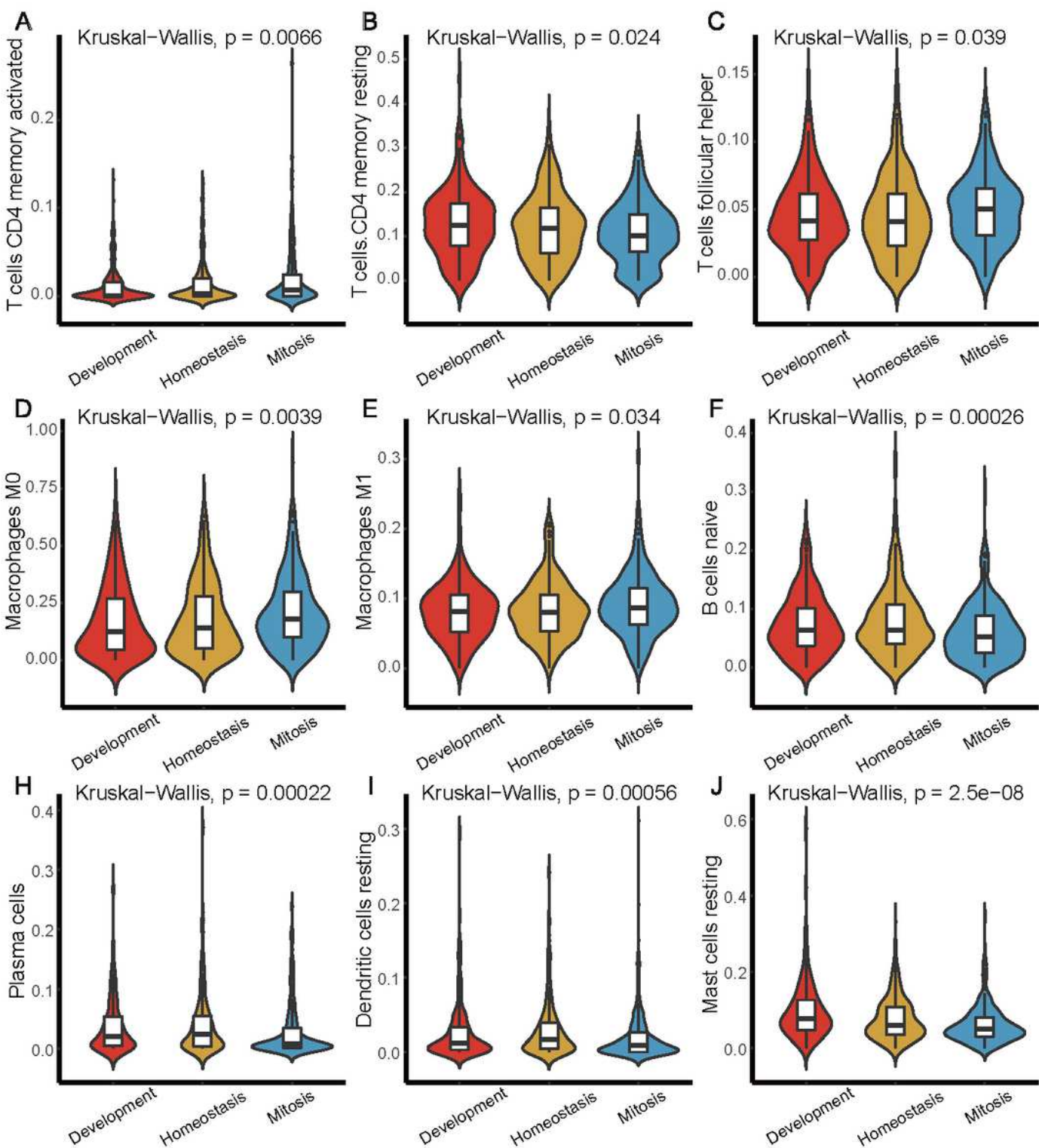

Figure 6

Immune infiltrating cells that differ significantly between the three subtypes. (A-J) Comparison of proportion of T cell CD4 memory activated, T cell CD4 memory resting, T cell follicular helper, Macrophages M0, Macrophages M1, B cells naive, Plasma cells, Dendritic resting, Mast cells resting among the three subtypes (P values by Kruskal-Wallis test).
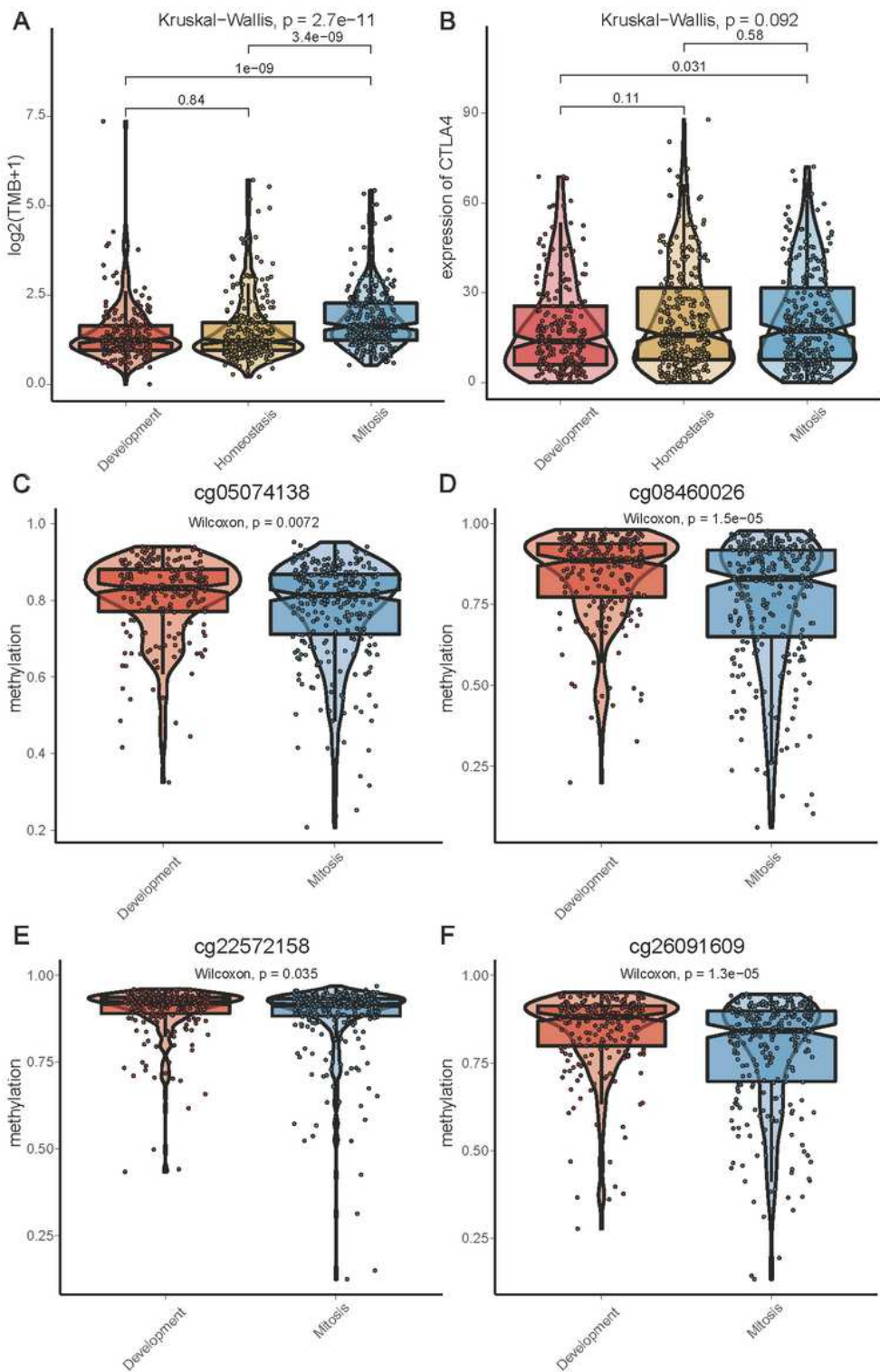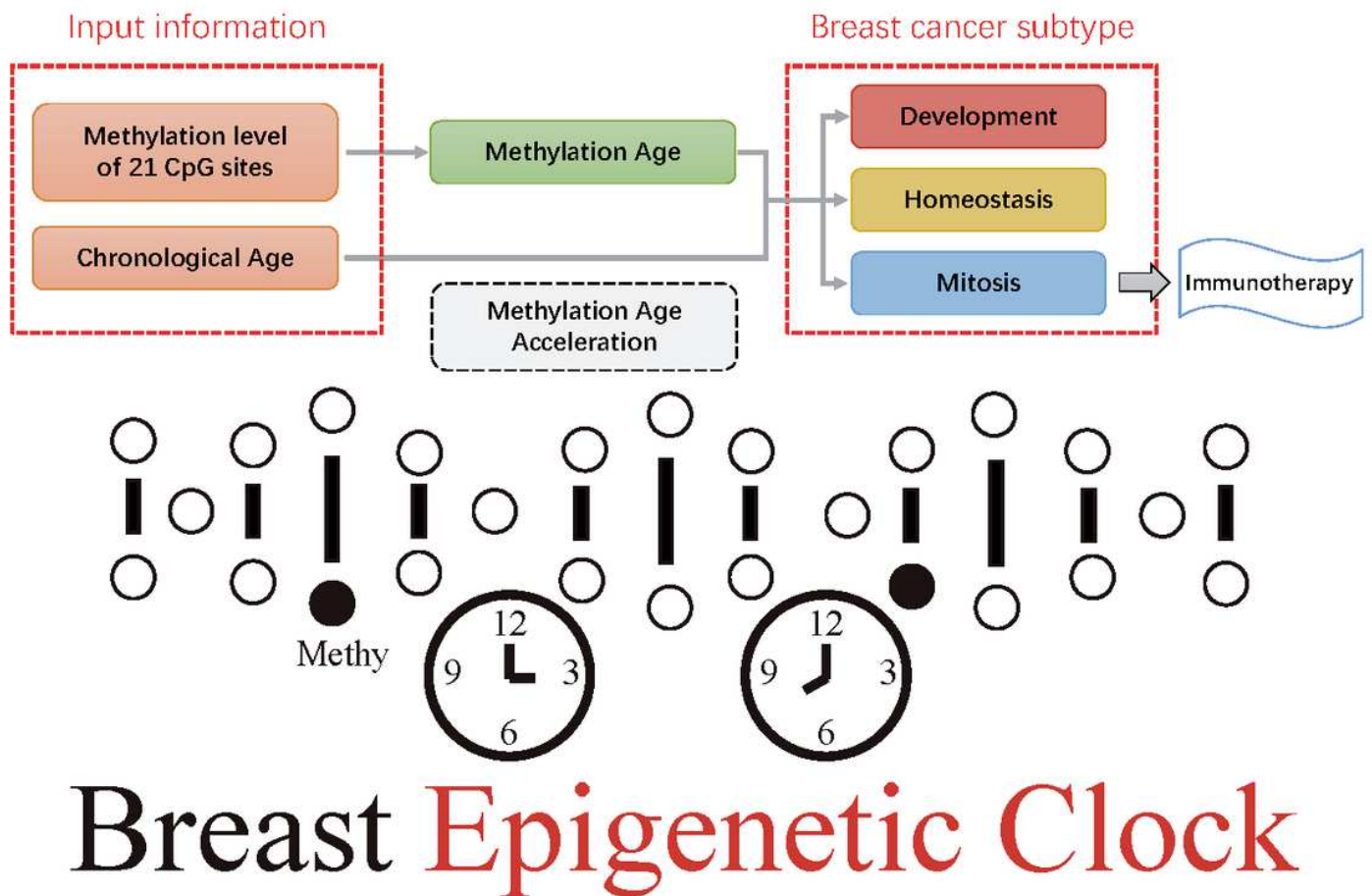

Figure 7

Comparison of immunotherapeutic response among the three subtypes. (A) Tumor mutation burden in each subtype (P values by Kruskal-Wallis test). (B) Expression of CTLA4 in each subtype (P values by Kruskal-Wallis test). (C-F) Methylation level of cg05074138 (C), cg08460026 (D), cg22572158 (E) and cg26091609 (F) in development subtype and mitosis subtype.

Figure 8

Schematic overview of the breast epigenetic clock. The web server could use the data submitted by the user to calculate the epigenetic age and breast cancer subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.pdf
- SupplementaryFigure2.pdf
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf

- SupplementaryFigure6.pdf
- Supplementarymaterial.pdf