Twin Research and Human Genetics

Differences in DNA methylation–based age prediction within twin pairs discordant for cancer

Running title: DNAm-based aging in cancer-discordant twins

Bode HF[1], Heikkinen A[1], Lundgren S[1], Kaprio J[1], Ollikainen M[1]

[1]Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland

Corresponding author:
Hannes Frederik Bode
Institute for Molecular Medicine Finland FIMM
P.O. BOX 20
FI-00014 University of Helsinki
Finland
ORCID ID: 0000-0003-3232-6518
Phone +491773641995
Email hannes.bode@helsinki.fi

Supplementary Materials and Methods

DNA methylation–based age prediction models

Each DNA methylation-based age prediction models used in the current study are described below and in Supplementary Table S2.

Horvath and Hannum

The DNA methylation–based age prediction model by Horvath was developed by elastic-net regression of chronological age over 21,369 CpG sites for 7844 samples across multiple tissues (Horvath, 2013). This resulted in a prediction model for chronological age based on 353 CpG sites. The DNA methylation–based age prediction model by Hannum was developed by elastic-net regression of chronological age over 473,034 CpG sites, age adjusted body mass index, sex, diabetes status, ethnicity and batch for 656 individuals blood samples (Hannum et al., 2013). This resulted in a prediction model for chronological age based on 71 CpG sites.

Horvath IEAA and Hannum IEAA

The Horvath IEAA (intrinsic epigenetic age acceleration) and the Hannum IEAA models are DNA methylation–based age prediction models based on the initial Horvath and Hannum models. The initially predicted DNA methylation–based ages are regressed over the chronological ages while correcting for blood cell counts (B. H. Chen et al., 2016). Blood cell counts were also estimated from DNA methylation using the Houseman method (Houseman et al., 2012). The residuals of these regressions resulted in the Horvath IEAA or the Hannum IEAA DNAmAA estimates. The goal for this approach was to have the DNAmAA estimate independent of the blood cell count (B. H. Chen et al., 2016).

Hannum EEAA

The Hannum EEAA (extrinsic epigenetic age acceleration) model is based on the Hannum model and estimated blood cell counts associated with aging (B. H. Chen et al., 2016). These blood cell counts up-weight the original Hannum model to describe aging-related haematological changes.

PhenoAge

While the Hannum and the Horvath models use an approach that directly predicts chronological age by a set of CpG sites, PhenoAge predicts a biological age surrogate based on further biomarkers (Levine et al., 2018). The PhenoAge model was developed by performing Cox-penalised regression on 42 aging-related biomarkers and chronological age over the hazard of mortality for 9.926 individuals. Then, nine best performing biomarkers, and chronological age were used to build the final PhenoAge model. Using elastic-net regression, such biological age biomarkers were regressed over 20,169 individual CpG sites to build the DNA methylation–based age prediction model. This resulted in a 513 CpG site - prediction model for phenotype-based biological age ('PhenoAge').

By incorporating blood-derived biomarkers, the PhenoAge model approaches biological aging by a combination of biological factors rather than chronological age alone.

GrimAge

Besides the PhenoAge model, the GrimAge model was also developed to predict a surrogate for biological age rather than directly predicting chronological age. Here, biomarkers and time-to-death data was used (Lu et al., 2019). Elastic-net regression for 88 individual plasma-protein levels and self-reported smoking pack-years were fitted over chronological age, sex and more than 485,000 CpG sites, for 6.935 individual blood samples. The 12 best-correlated plasma-protein DNA-methylation surrogates, DNA methylation–based smoking pack-years,

chronological age and sex were further used to build a model that predicts individuals' time to death by penalised Cox regression. The resulting prediction model includes 7 of the 12 plasma protein–describing DNA-methylation surrogates, DNA methylation–based smoking pack-years, chronological age, and sex. In total, the model includes 1030 unique CpG sites summed across all variables.

By fitting the DNA methylation data, chronological age, and sex on this model, time-to-death estimates are calculated, which are subsequently scaled to fit estimates for chronological age distribution resulting in the GrimAge age estimates.

Supplementary tables

Supplementary table S1: Frequencies of the diagnostic groups	
Diagnostic group	**Frequency in the dataset (total n = 95)**
Breast cancer	24
Cancer of the female reproductive tract	18
Haematological malignancy	8
Non-melanoma skin cancer	7
Stomach and upper gastrointestinal tract cancer	7
Colorectal cancer	6
Melanoma	5
Bladder cancer	4
Kidney cancer	3
Lung cancer	3
Prostate cancer	3
Thyroid cancer	3
Others	4

Supplementary table S2 Comparing the different DNAmAA models

DNAmAA model	DNAm platforms	Number of CpGs	Sample size	Tissues	Variables	Prediction	Phenotypes/Conditions shown to be associated based on the original paper	reference
Horvath	27k[a], 450k[b]	353	7844	51 different tissues and cell types	Chronological age	Chronological age	Cancer, mutations in hormone receptor in breast cancer, mutations in proto-oncogenes in colorectal cancer, mutations glioblastoma multiform	Horvath, 2013
Hannum	450k[b]	71	656	Whole blood samples	Chronological age, age adjusted body mass index, sex, diabetes status, ethnicity	Chronological age	Sex, BMI, methylome heterogeneity, epigenetic drift, gene expression	Hannum et al., 2013
Horvath IEAA	450k[b]	353	7844	Whole blood samples	Chronological age, blood cell types by Houseman's estimation method	Chronological age adjusted for blood cell counts	all-cause mortality, time to death	B.H.Chen et al., 2016 Horvath, 2013
Hannum IEAA	450k[b]	71	656	Whole blood samples	Chronological age, age adjusted body mass index, sex, diabetes status, ethnicity, blood cell types by Houseman's estimation method	Chronological age adjusted for blood cell counts	none tested	B.H.Chen et al., 2016 Hannum et al., 2013
Hannum EEAA	450k[b]	71	656	Whole blood samples	Chronological age, age adjusted body mass index, sex, diabetes status, ethnicity, blood cell types by Houseman's estimation method	Chronological age up weighted for age related haematological changes	all-cause mortality, time to death	B.H.Chen et al., 2016 Hannum et al., 2013
PhenoAge	27k[a], 450k[b], EPIC[c]	513	9926	Whole blood samples	Chronological age, albumin, creatinine, serum glucose, c-reactive protein, lymphocyte percent, mean cell volume, red blood cell distribution width, alkaline phosphatase, white blood cell count	Biological age	all-cause mortality, aging related morbidity, ethnicity, diet, health related clinical measurements, socio-behavioural factors	Levine et al., 2018
GrimAge	450k[b]	1030	6935	Whole blood samples	Time to death, smoking, adrenomedullin, beta-2-microglobulin, cystatin-c, growth differentiation factor 15, leptin, plasminogen activator inhibitor 1, tissue Inhibitor Metalloproteinases 1	Morbidity and mortality	all-cause mortality, coronary heart disease, hypertension, type 2 diabetes, physical functioning, age at menopause, nutrition, C-reactive protein measures, socio economic status, measures of adiposity	Lu et al., 2019

a: Infinium HumanMethylation27, (Illumina, San Diego, CA, USA); b: Infinium HumanMethylation450 (Illumina, San Diego, CA, USA); c: Infinium HumanMethylationEPIC, (Illumina, San Diego, CA, USA)

Supplementary figures

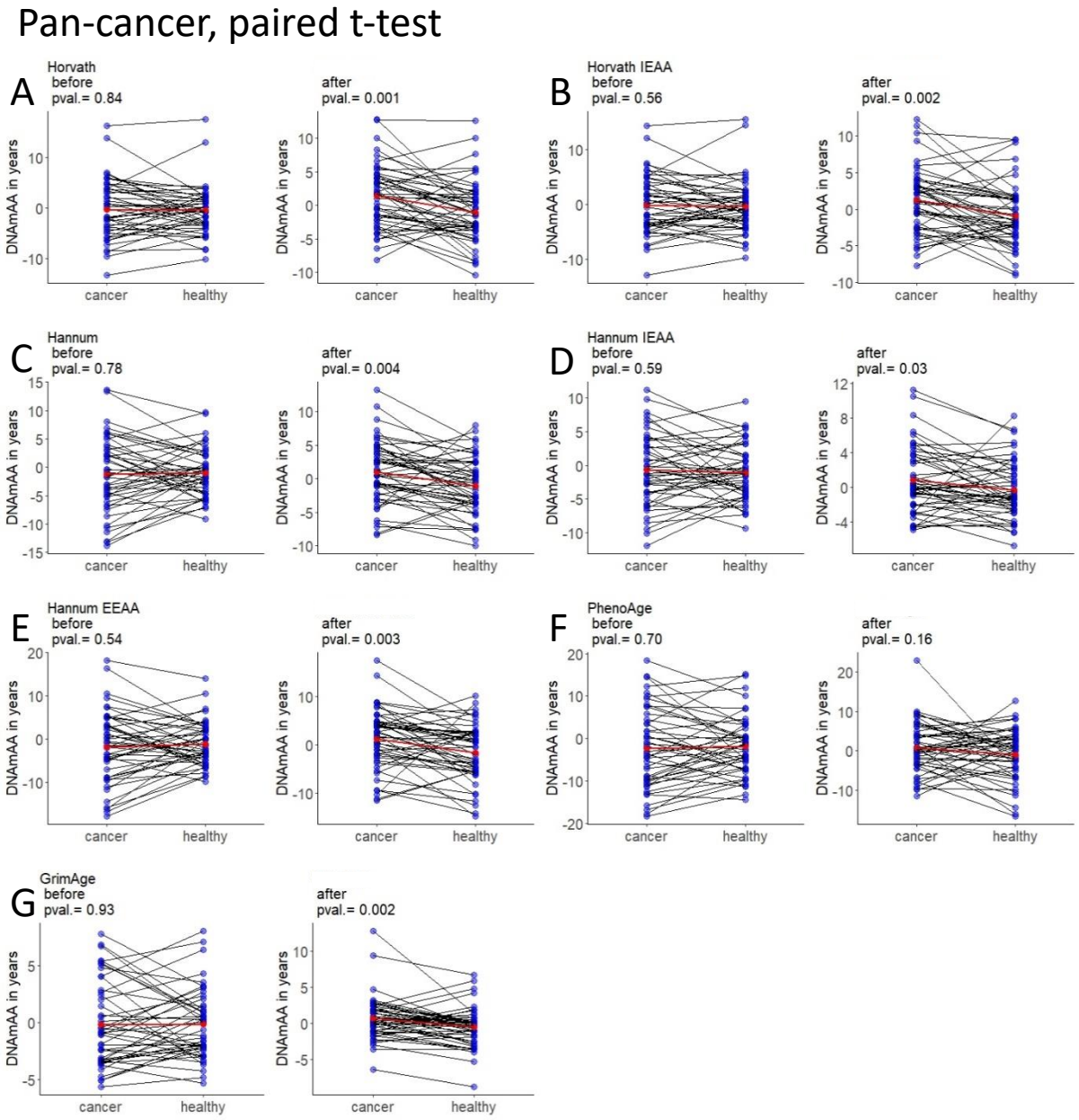

Figure S1:
Within-pair differences in DNAmAA for the twin pairs (n=95) discordant for any cancer (pan-cancer group). Each twin in a pair (blue) is linked to the co-twin, with the mean values presented in red. Within-pair differences in DNAmAA are shown separately for pairs sampled before and after the diagnosis, and for each DNAmAA model. (A) Horvath, (B) Horvath IEAA, (C) Hannum, (D) Hannum IEAA, (E) Hannum EEAA, (F) PhenoAge, (G) GrimAge. Paired t-test p-values are given for mean within-pair differences in DNAmAA different from zero.

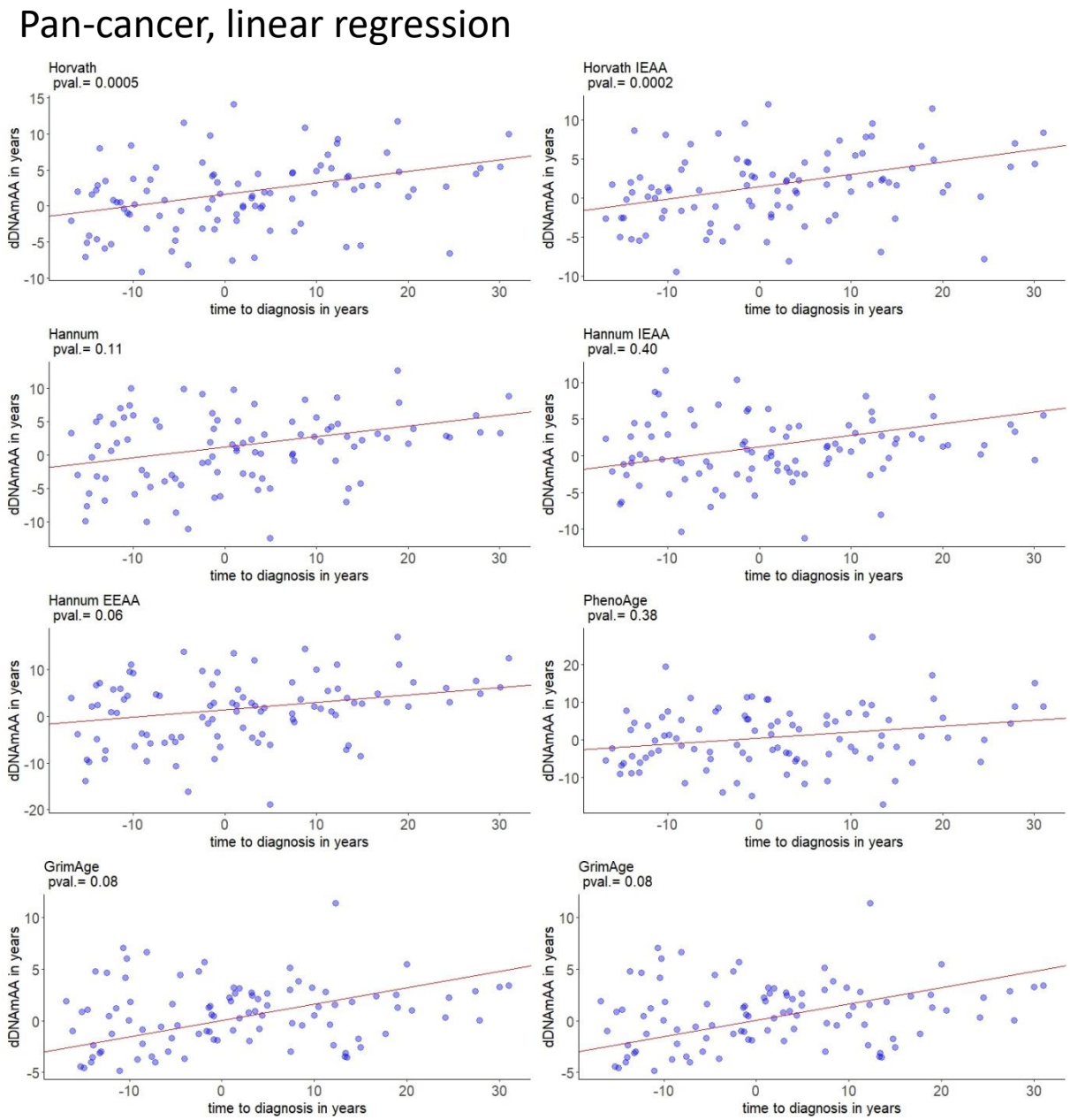

Figure S2:
The within-pair difference in DNAmAA (dDNAmAA) over time to cancer diagnosis for the twin pairs (n=95) discordant for any cancer (pan-cancer group). The individual data points (blue) each represent the differences in DNAmAA within one pair. Negative values for time to diagnosis mean that a pair was sampled before diagnosis, positive values that the pair was sampled after diagnosis. Each DNAmAA model is presented in separate graphs with the regression line (red) and F-statistic's p-value.

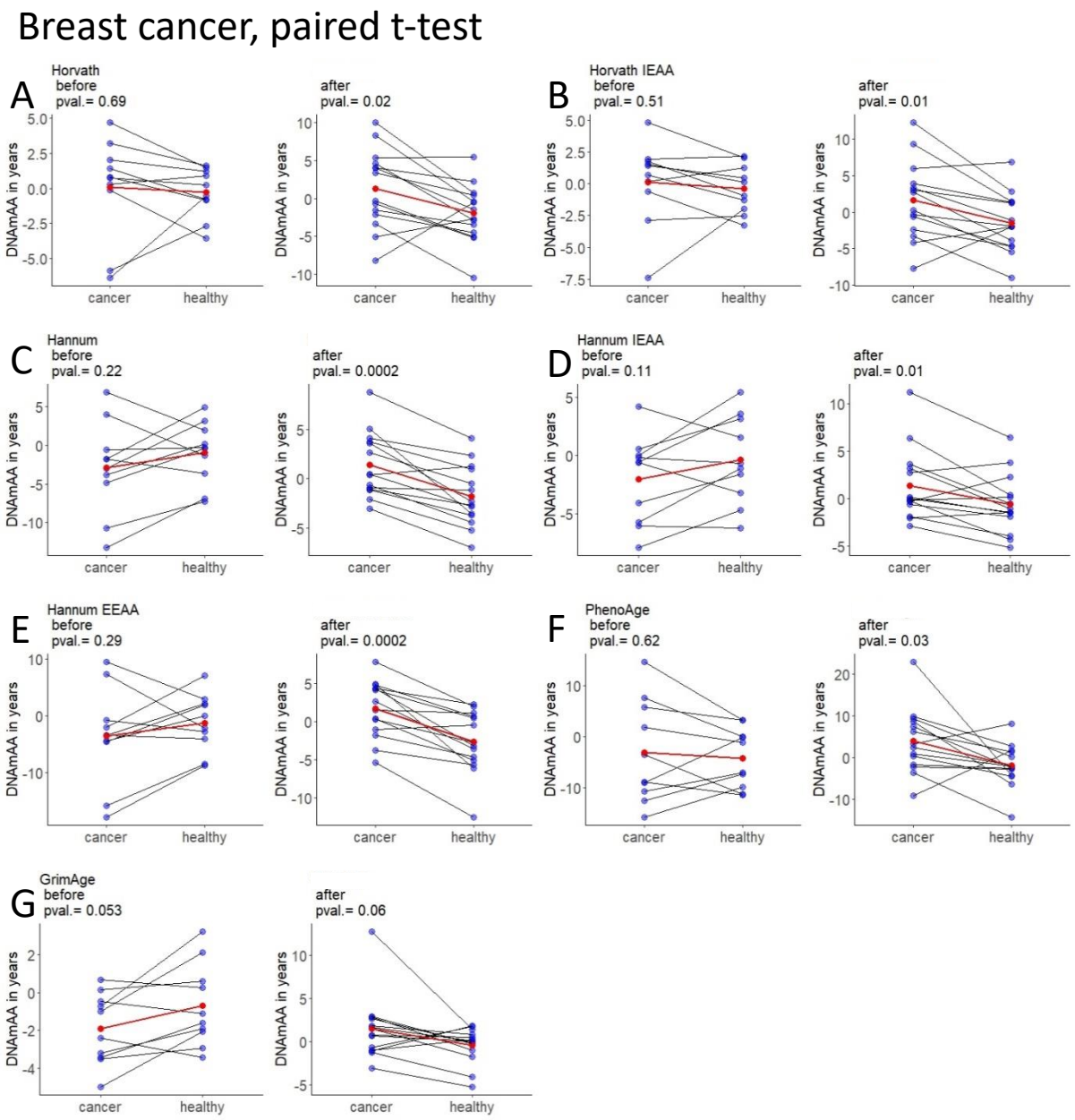

Figure S3:
Within-pair differences in DNAmAA for the twin pairs (n=24) discordant for breast cancer. Each twin in a pair (blue) is linked to the co-twin, with the mean values presented in red. Within-pair differences in DNAmAA are shown separately for pairs sampled before and after the diagnosis, and for each DNAmAA model. (A) Horvath, (B) Horvath IEAA, (C) Hannum, (D) Hannum IEAA, (E) Hannum EEAA, (F) PhenoAge, (G) GrimAge. Paired t-test p-values are given for mean within-pair differences in DNAmAA different from zero.

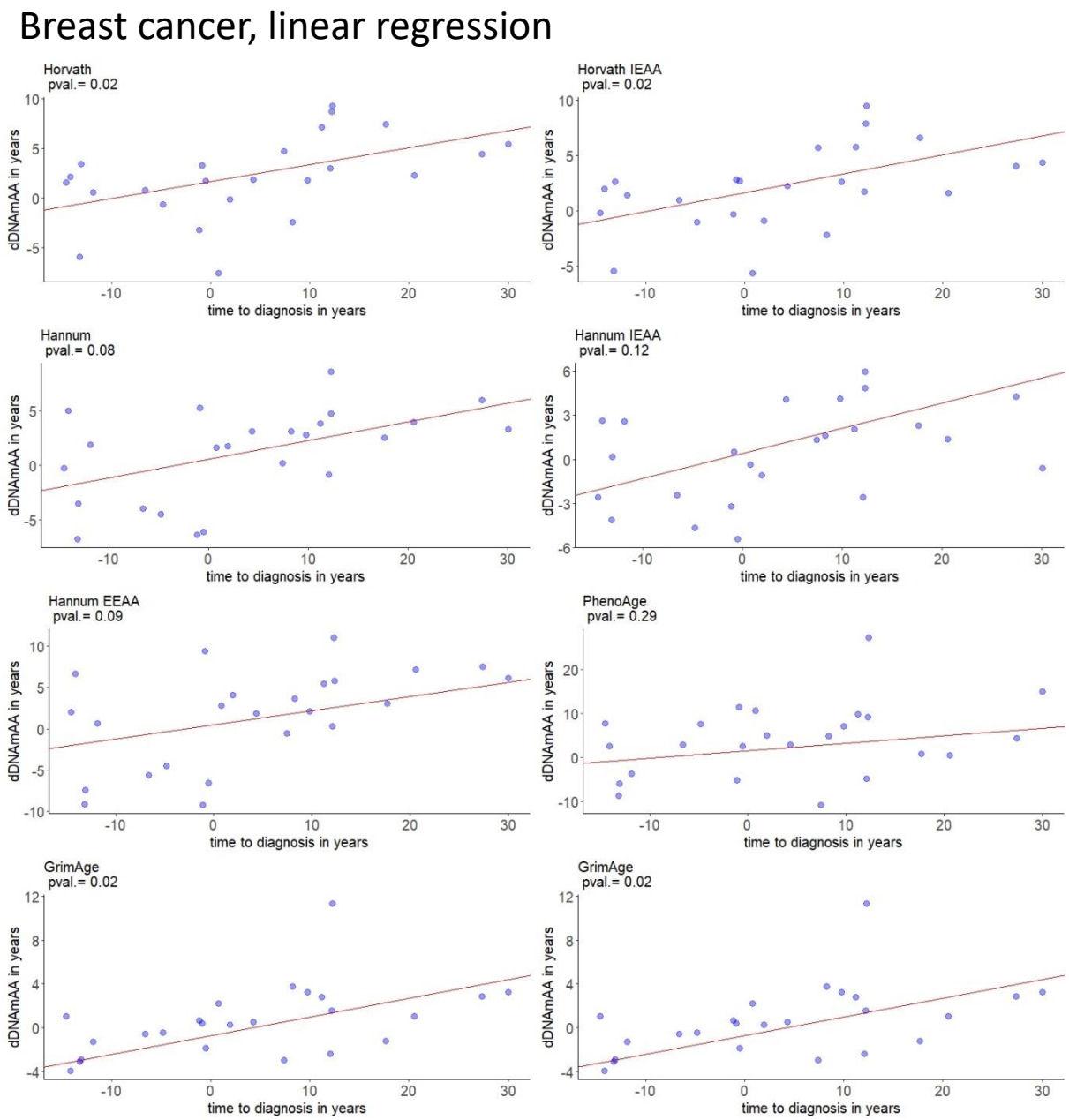

Figure S4:
The within-pair difference in DNAmAA (dDNAmAA) over time to cancer diagnosis for the twin pairs (n=24) discordant for breast cancer. The individual data points (blue) each represent the differences in DNAmAA within one pair. Negative values for time to diagnosis mean that a pair was sampled before diagnosis, positive values that the pair was sampled after diagnosis. Each DNAmAA model is presented in separate graphs with the regression line (red) and F-statistic's p-value.

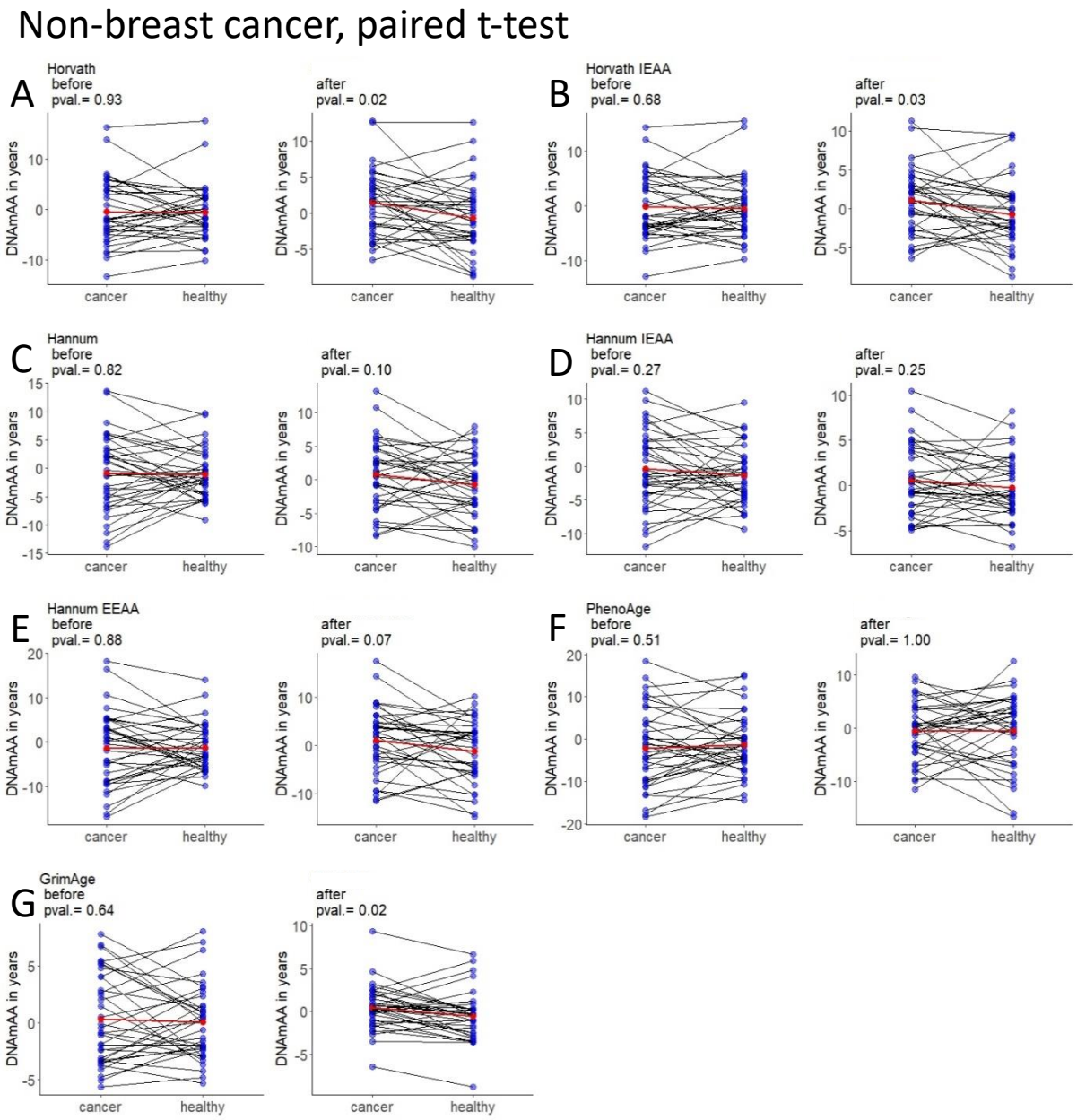

Figure S5:
Within-pair differences in DNAmAA for the twin pairs (n=71) discordant for non-breast cancer. Each twin in a pair (blue) is linked to the co-twin, with the mean values presented in red. Within-pair differences in DNAmAA are shown separately for pairs sampled before and after the diagnosis, and for each DNAmAA model. (A) Horvath, (B) Horvath IEAA, (C) Hannum, (D) Hannum IEAA, (E) Hannum EEAA, (F) PhenoAge, (G) GrimAge. Paired t-test p-values are given for mean within-pair differences in DNAmAA different from zero.

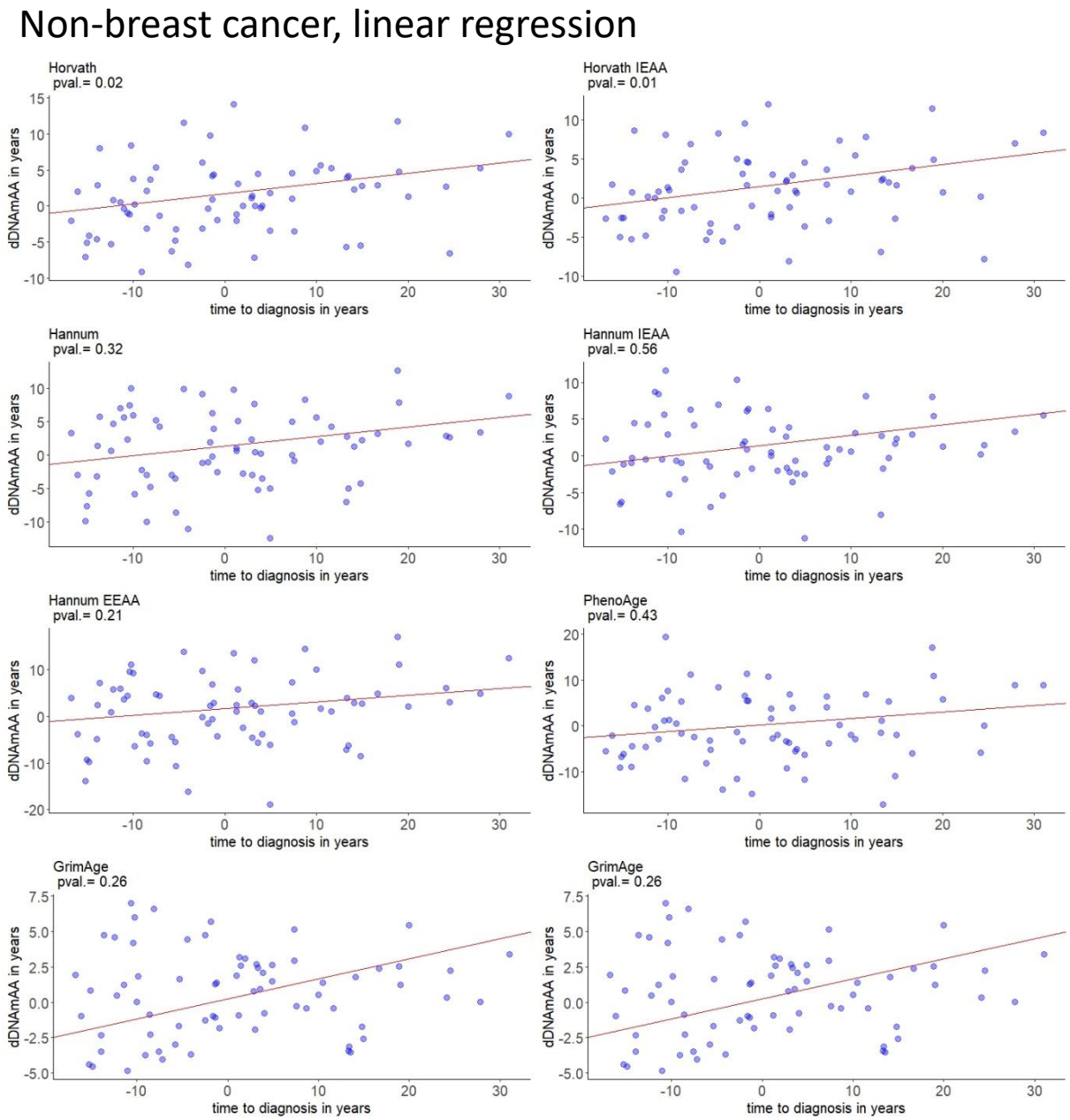

Figure S6:
The within-pair difference in DNAmAA (dDNAmAA) over time to cancer diagnosis for the twin pairs (n=71) discordant for non-breast cancer (other than breast cancer). The individual data points (blue) each represent the differences in DNAmAA within one pair. Negative values for time to diagnosis mean that a pair was sampled before diagnosis, positive values that the pair was sampled after diagnosis. Each DNAmAA model is presented in separate graphs with the regression line (red) and F-statistic's p-value.