**Materials and methods**

**Datasets**

Three breast cancer datasets, labeled the Big Series, METABRIC MB09 and METABRIC Expression, and one normal breast dataset NormalBreast13 were collected and analyzed at the British Columbia Cancer Agency after ethics committee approval (Fig. 1).

All of the three breast cancer datasets consisted of primary invasive breast carcinomas before treatment, along with clinical and pathological annotations. The Big Series dataset was derived from analysis of a tissue microarray (TMA) constructed from formalin-fixed paraffin-embedded (FFPE) samples from 3,992 breast carcinomas collected in Vancouver, Canada [Turashvili et al., 2011; Cheang et al., 2006]. The METABRIC MB09 dataset was derived from analysis of a TMA constructed from FFPE samples from 1,537 breast carcinomas collected from two different sites in Canada, Vancouver and Manitoba. The METABRIC Expression dataset was derived from fresh frozen samples from 1,992 breast carcinomas collected from five different sites in Canada and the United Kingdom [Curtis et al., 2012]. These three breast cancer datasets had some overlap of cases, so that analyses on the full case sets for these datasets yielded results not fully independent. To verify results using disjoint independent sets, separate analyses were conducted with independent subsets, assigning the overlapping cases to one set only among the Big Series dataset and the two METABRIC datasets. For independent analyses, 528 cases were excluded from the Big Series dataset and the remaining 3,464 cases were analyzed; 831 cases were excluded from the METABRIC Expression dataset and the remaining 1161 cases analyzed; the excluded cases being part of the 1537 cases of the METABRIC MB09 dataset.

The normal breast dataset consisted of FFPE samples of normal breast tissue from 537 cancer-free women who undertook reduction mammoplasty surgery at several hospitals in Vancouver, along with the information of patient age at surgery.

Additionally, publicly available data was obtained from The Cancer Genome Atlas (TCGA) for breast cancer (TCGA-BRCA, 1098 cases), prostate cancer (TCGA-PRAD, 500 cases), renal clear cell kidney cancer (TCGA-KIRC, 537 cases), lung adenocarcinoma (TCGA-LUAD, 585 cases) and thyroid cancer (TCGA-THCA, 507 cases). The TCGA results herein are in whole or part based upon data generated by the TCGA Research Network: [http://cancergenome.nih.gov/](https://cancergenome.nih.gov/).

A small series of skin samples was assessed [Borman et al., 2016]. Data was obtained from the ArrayExpress database under the accession numbers E-MTAB-4385 (methylation data) and E-MTAB-4382 (expression data). With only 59 evaluable cases, after adjustment for multiple comparisons, this data set yielded little information primarily due to the small sample size, and no results are presented herein.

ChIP-Seq data identifiying estrogen receptor binding sites was obtained from the Zwart Lab [Severson et al., 2018]. This data identifies ER binding sites with corresponding Ensembl gene IDs allowing identification of ER-binding gene targets in other annotated data sets described above.

**Construction of tissue microarrays**

The Big Series, the METABRIC MB09, and the normal breast datasets were used to construct tissue microarrays (TMAs) for protein expression analysis using immunohistochemistry. For the Big Series and the METABRIC MB09 datasets, a representative area of invasive breast cancer was marked on a haematoxylin-and-eosin-stained slide of a representative FFPE tissue block, and a single 0.6-mm core was taken from each of the sample blocks and placed in a recipient blank paraffin block using a manual tissue arrayer (Pathology Devices, Westminster, MD, USA). For the normal breast dataset, dual 1.5-mm cores were taken from representative areas of normal breast epithelium (lobules and/or ducts) and the TMAs were constructed using the above methods.

The Big Series:

This study cohort included 3992 female patients with primary invasive breast carcinoma diagnosed in 1986-1992 and referred to the British Columbia Cancer Agency for treatment. The clinicopathological characteristics and treatment strategies of the patients included in this study have previously been reported [Cheang et al., 2008; Turashvili et al., 2011]. The median follow-up was 15 years and median age at diagnosis was 60 years. Abstracted clinical information included age, menstrual status, histological type and grade of tumour, clinical and pathological TNM stage, status of final surgical margin at diagnosis, tumour size, number of positive axillary lymph nodes, type of local and initial adjuvant systemic therapy, dates of diagnosis and first loco-regional or distant recurrence and death. HER2 fluorescent *in situ* hybridization data and immunohistochemistry scores were also available for the following biomarkers: ER, PR, Ki-67, Bcl-2, HER2, EGFR, CK5/6, keratin 5 (KRT5), p53, YB-1, P-cadherin and E-cadherin, as previously published [Turashvili et al., 2011; Cheang et al., 2006; Chia et al., 2008; Habibi et al., 2008; Jensen et al., 2008; Cheang et al., 2009; Turashvili et al., 2009; Liu et al., 2010 ] . Ethical approval for the study was obtained from the Clinical Research Ethics Board of the University of British Columbia and the British Columbia Cancer Agency.

Previously frozen breast cancer tissue samples were fixed in 10% neutral buffered formalin and embedded in paraffin. Representative areas of invasive carcinoma were selected and marked on the hematoxylin and eosin stained slides, and on their corresponding tissue blocks to be sampled for the tissue microarray. Seventeen single core (0.6 mm size) tissue microarray blocks were then assembled using a manual arrayer (Beecher Instruments, Inc., Silver Springs, Maryland, USA) as previously described [[35](#_ENREF_35)]. From each tissue microarray block, 4 μm thick sections were cut and immunostained on a Ventana Discovery XT staining system (Ventana Medical Systems, AZ, USA). Sections were deparaffinized in xylene, dehydrated through three alcohol changes, and transferred to Ventana Wash solution. Endogenous peroxidase activity was blocked in 3% hydrogen peroxide. Antigen retrieval was performed in cell conditioner 1 and slides were incubated with anti-EZH2 mouse monoclonal antibody (1:10 dilution, clone 11, BD Transduction Laboratories, ON, Canada) for 32 minutes. Finally, sections were incubated with the pre-diluted Ventana Universal Secondary Antibody and DAB Map detection system, counterstained with hematoxylin, dehydrated, cleared and mounted. Optimization of the immunohistochemical protocol involved three different antigen-retrieval conditions and a serial dilution of the antibody to establish the optimal staining concentration. Although freezing of the tissue samples prior to formalin fixation could have potentially affected the immunoreactivity for EZH2, appropriate negative and positive controls were performed to ensure the quality and adequacy of staining. The negative control was performed by omission of the primary antibody, and tonsil tissue was used as a positive control.

EZH2 expression was scored visually based on the determination of staining intensity (0 – negative, 1 – weak, 2 – moderate, 3 – strong) and percentage of cells with nuclear staining (0-100%). Scores were entered into a standardized Excel worksheet with a sector map matching each tissue microarray section. Biomarker information was considered uninterpretable if there were no tumour cells in the cores or the cores were missing. Original scoring grids were converted to tables using Deconvoluter 1.10 [[36](#_ENREF_36)] and combined in a single text file with TMA-Combiner 1.00 [[37](#_ENREF_37)]. The resulting text files were imported into SPSS 17.0 for Windows [SPSS 2008].

The hematoxylin and eosin and EZH2 immunohistochemistry images and scores of all cores used in this study are publicly available at the companion site (<http://www.gpecimage.ubc.ca>). This site was constructed using the Genetic Pathology Evaluation Centre database and image viewers provided by Olympus America, Inc. All slides were scanned with the BLISS scanner (Olympus America, Inc., Center Valley, PA, USA; Bacus Laboratories, Inc., Lombard, IL, USA).

The expanded surrogate immunopanel of ER, PR, HER2, Ki-67, EGFR, and CK5/6 was used to define five major biologically distinct immunohistochemical subtypes of breast cancer- 1. Luminal A: ER+ or PR+ and HER2− and Ki-67−; 2. Luminal B: ER+ or PR+ and HER2− and Ki-67+. Ki-67 positivity was defined based on a 14% cut point; 3. Luminal/HER2+: HER2+ and ER+ or PR+; 4. HER2+: HER2+ and ER− and PR−; 5. Basal: this subtype was defined differently by two classification schemes: the triple negative phenotype and the five-biomarker method as published previously [Cheang et al., 2008]. Using the triple negative phenotype method, basal-like carcinoma is triple-negative (ER−/PR−/HER2−). Using the five-biomarker method, the triple negative phenotype can be divided into two groups: (a) Core basal: triple negative subset, which also expresses EGFR or CK5/6; and (b) Five-marker negative phenotype: ER−/PR−/HER2−/EGFR−/CK5/6−. Tumours missing any of ER, PR, or HER2 data were categorized as unassigned.

Statistical analysis was performed using SPSS 17.0 and R-2.10.0 [CRAN 2017]. Clinical covariate and biomarker associations were assessed using contingency tables, and significance of associations was determined using the likelihood ratio chi-square or Fisher’s exact test. Survival and relapse interval associations were graphically assessed using Kaplan-Meier plots. Significance of time-to-event associations was assessed using logrank and Breslow tests and Cox proportional hazards regression analysis. Survival and relapse interval end-point groups include overall survival, disease-specific survival, distant relapse-free interval, and loco-regional relapse-free interval. Cox proportional hazards models were used to calculate adjusted hazard ratios to account for covariates of known clinical relevance.

Complete survival data for this tissue microarray series was obtained as of June 30, 2004. One of the complexities of working with such a large series with long-term follow-up data, noted in previous Cox modeling exercises using this series, was the changing structure in the hazard functions for breast cancer subtypes and subsets of other variables, between 5 and 10 years post-diagnosis. This is readily seen in the hazard function plots in Figures 1 and 2 of [Turashvili et al., 2011]. Hazards are proportional over the first 5 to 8 years, then exhibit a shift and tend to converge by 15 to 20 years. This long-term shift in hazard structure yields a decrease in power in testing for associations with a Cox model, as its basic proportional hazards requirement is violated. To mitigate this violation of the proportional hazards assumption, we formed an additional survival record, with data followed completely through June 30, 1999. Censoring data at an earlier time point to restrict analysis to a period in which hazards exhibit proportional structure represents a simple and effective strategy for handling this issue of changing hazard structure across long periods of time [[39](#_ENREF_39)]. Survival analysis results thus reveal short-term and long-term associations that can differ, and both should be considered when assessing associations. Shorter term (0-10 years) associations will be better assessed with the Breslow statistic in Kaplan-Meier analysis and with statistics from Cox models fitted to the data censored in 1999. Longer term associations are better assessed via the logrank statistic in Kaplan-Meier analysis, and with Cox models fitted to the data censored in 2004. This strategy was proposed and implemented prior to any data analysis.

For initial analyses of the Big Series TMA data, we used a split-sample validation technique for statistical analysis, as described previously [Turashvili et al., 2011; Rajput et al., 2008; Hastie et al., 2001]. In brief, a large data collection (n = 3992) was randomly split into a 'training' set (n = 2003) and a 'validation' set (n = 1989). Though it is common practice in studies of prognostic markers to conduct exploratory analyses with the training set, and take only a selected subset of analyses onward to the validation set, such strategies tend to result in the over-reporting of positive findings and an under-reporting of negative findings. To avoid the resultant reporting bias associated with this analysis strategy [Rifai et al., 2008], we pre-specified a set of hypotheses, all of which were evaluated in the training and validation sets.

**Protein expression analysis using immunohistochemistry**

Deparaffinized 4-µm sections of the TMA blocks were immunostained with specific antibodies under their staining protocols (Suppl Table S1–3). A total of 44 protein markers were analyzed, of which 42, 19 and 12 protein markers were analyzed using the Big Series, the METABRIC MB09 and the normal breast datasets, respectively. EZH2, H3k27me3 and other clinically or biologically important markers in breast cancer (e.g. estrogen receptor α [ERα], progesterone receptor [PgR], human epidermal growth factor receptor 2 [HER2], Ki67, epidermal growth factor receptor [EGFR], cytokeratin 5 [CK5] and p53) were analyzed in all of the three datasets. The percentage of positive cells, the staining intensity, or both were scored for each of the markers by pathologists at the BC Cancer Agency or University of British Columbia. Invasive cancer cells were evaluated for the Big Series and the METABRIC MB09 datasets. Non-malignant breast epithelium (lobules and ducts) were evaluated for the normal breast dataset.

**Global gene expression analysis**

The METABRIC Expression dataset was used for global gene expression, and the detailed methods have been described previously [Curtis et al., 2012]. Briefly, RNA was extracted from each of the fresh frozen samples. The RNA was assayed using the HumanHT-12 v3 Expression BeadChip Kit (Illumina, San Diego, CA), which targeted approximately 25,000 annotated genes with 48,803 probes. Each case was classified into one of the five PAM50-based intrinsic subtypes using the array-derived gene expression data.

**Statistical analysis**

For TMA protein expression data, curves showing association between marker expression and patient age were drawn using a smoother (supsmu or loess) to allow assessment of non-linear age-associated structure. Randomization test analysis was used to obtain confidence intervals for the smoother curves. If there were no association of measured biomarker level with age, then the patient age would be arbitrary with respect to biomarker values. This allows for a randomization test to assess for age associated trends. Under the hypothesis of no age associated trend, the patient ages can be randomly assigned to the patients, and a smoother curve fitted to the randomized data. This process of randomly assigning the N patient ages to the N patients without replacement and fitting a new randomization smoother was carried out thousands of times, and a (1 - α)\*100% confidence region determined as that region containing the proportion (1 - α) of all the randomization smoother curves. If any portion of the observed curve fell outside the (1 - α) confidence region, we rejected the null hypothesis of no age association for the biomarker (Type I error rate α). Randomization test repetitions of size B = 2,000; 10,000; 40,000 and 400,000 were used to obtain 95%, 99%, 99.9% and 99.99% confidence intervals respectively. Biomarkers showing Bonferroni-adjusted *P*-values <0.05 (smoother shows any portion outside the bootstrap confidence interval) were defined as age-related. For the Big Series, 43 biomarkers were assessed, so the Bonferroni-adjusted significance level is 0.05/43 = 0.0012, thus any biomarker whose smoother curve exceeds its bootstrap 99.9% confidence interval was declared as age-associated at the 5% significance level. Similarly, the 1% FDR Bonferroni-adjusted significance level is 0.01/43 = 0.00023, thus any biomarker whose smoother curve exceeded its bootstrap 99.99% confidence interval was declared as age-associated at the 1% adjusted significance level.

For the mRNA expression analysis, two regression lines, one showing the younger cases (age ≤60) and the other showing the older cases (age >60), were plotted for each of the markers. A cutoff at age 60 was chosen as this was generally close to the median age for the several breast cancer data series, and patient ages ranged from about 25 years to 95 years. Age 60 is unambiguously post-menopausal. The amount of data in the younger and older case groups was close to evenly balanced for the age 60 cutoff, so that statistically significant effects would not be artefactually elevated in either group due to unequal sample sizes. As about 35 years worth of data under and over age 60 were available, a change of log2 (1.25)/35 represents an absolute fold change of 1.25 in the younger and older age groups using the raw expression data. This fold change corresponds to a regression line slope of 0.0092 or approximately 0.01. Additionally, a third regression line for all cases was fitted. A change of log2 (1.25)/70 represents an absolute fold change of 1.25 in the whole cohort. Age associated non-linear trend structure was thus detectable via the two regression lines fitted to the younger and older cohorts, and linear trend structure detectable with the single regression line fitted to the whole cohort.

Probes showing fold changes >1.25 in absolute value with Benjamini-Hochberg-adjusted *P*-values <0.01 (FDR rate of 0.01) were defined as age-related. When a gene target had multiple probes, age-related genes were defined to have at least one probe showing age association. Fold changes of 2 and 4 and an FDR rate of 0.01 were also used to derive more stringent gene lists.

The number of genes in the human genome is still not clearly understood or defined, and gene annotation is still evolving. Thus it was not possible to determine quantities such as the proportion of genes showing age association. Instead the proportion of protein-coding genes was assessed, as the number and identity of protein coding genomic regions is currently well understood. Pathway analysis was undertaken, and age-associated gene targets identified, and proportions for protein coding genes were computed.

Using the age-related probes, Manhattan plots and volcano plots were depicted for the whole METABRIC Expression dataset and for each of the PAM50 and intClust subtypes. The Manhattan plots showed the distribution of genomic locations of age-related probes across the genome, and the volcano plots showed the association between the fold changes and the *P*-values of the regression curves for each of the transcripts. Using the age-related transcripts, a gene set enrichment analysis (GSEA) was performed for the whole dataset, the luminal A subtype and the luminal B subtype (PAM50) and the intClust groups (METABRIC intClust classification). The enriched pathways showing the *P*-values ≤0.001 and the false discovery rates ≤0.001 were selected using the Cytoscape (http://www.cytoscape.org/). Networks of the enriched pathways from the GSEA were visualized using the Reactome FI Cytoscape Plugin 4 (http://www.reactome.org/).

Genomic regions containing estrogen receptor elements (EREs) were assessed via ChIP-Seq [Severson et al., 2018] and labeled in figures, including the Manhattan and volcano plots. Tier 1 genes were those for which ESR1 binding sites occurred directly at promotor regions, and Tier 2 genes were those for which ESR1 binding sites occurred within 20 kilobases downstream of the gene's promotor region.

Due to differences in sample size among the breast cancer subtypes, larger subtype groups will show more age-associated gene targets in regression analysis and Cytoscape analyses. Comparisons between breast cancer subtypes must thus be avoided as differences will include sample size effects in addition to effects related to subtype physiology. Comparisons between subtypes will require more balanced data sizes among the subtypes to be compared. This issue was assessed by performing a bootstrap analysis for the intClust subtypes wherein the same number of cases was chosen at random with replacement from all of the cases within each intClust subtype. Bootstrap samples of size 50, 75, 100, 150, 200 and 300 cases were drawn from each intClust subtype in the METABRIC expression data set. Assessment of the number of age-associated gene targets was carried out for each bootstrap sample and recorded.

**Age-associated enrichment analysis**

The number of age-associated gene targets was noted for each breast cancer subtype in order to assess whether age-associated trends occurred at a higher rate within some subgroups. This analysis was confounded by the issue of sample size, as detection of age association was more readily achieved in subtype groups with larger sample sizes. To mitigate this issue, results from the bootstrap analysis were used. A test for age-associated enrichment was performed using Fisher’s exact test, and Pearson’s chi-square test when sample sizes in all subgroups were adequate.

References:

[Cheang et al., 2008]

Cheang, M.C., et al., *Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype.* Clin Cancer Res, 2008. **14**(5): p. 1368-76.

[Turashvili et al., 2011]

Turashvili, G., et al., *P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer.* Mod Pathol, 2011. **24**(1): p. 64-81.

[Cheang et al., 2006]

Cheang, M.C., et al., *Immunohistochemical detection using the new rabbit monoclonal antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1D5 in predicting survival.* J Clin Oncol, 2006. **24**(36): p. 5637-44.

[Chia et al., 2008]

Chia, S., et al., *Human epidermal growth factor receptor 2 overexpression as a prognostic factor in a large tissue microarray series of node-negative breast cancers.* J Clin Oncol, 2008. **26**(35): p. 5697-704.

[Habibi et al., 2008]

Habibi, G., et al., *Redefining prognostic factors for breast cancer: YB-1 is a stronger predictor of relapse and disease-specific survival than estrogen receptor or HER-2 across all tumor subtypes.* Breast Cancer Res, 2008. **10**(5): p. R86.

[Jensen et al., 2008]

Jensen, K.C., et al., *New cutpoints to identify increased HER2 copy number: analysis of a large, population-based cohort with long-term follow-up.* Breast Cancer Res Treat, 2008. **112**(3): p. 453-9.

[Cheang et al., 2009]

Cheang, M.C., et al., *Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer.* J Natl Cancer Inst, 2009. **101**(10): p. 736-50.

[Turashvili et al., 2009]

Turashvili, G., et al., *Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis.* BMC Cancer, 2009. **9**: p. 165.

[Liu et al., 2010]

Liu, S., et al., *Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients.* Breast Cancer Res Treat, 2010. **119**(1): p. 53-61.

[SPSS 2008]

SPSS Inc. Released 2008. SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.

[CRAN 2017]

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2017 https://www.R-project.org

[Rajput et al., 2008]

Rajput, A.B., et al., *Stromal mast cells in invasive breast cancer are a marker of favourable prognosis: a study of 4,444 cases.* Breast Cancer Res Treat, 2008. **107**(2): p. 249-57.

[Hastie et al., 2001]

Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. 2001, New York: Springer. 533.

[Rifai et al., 2008]

Rifai, N., D.G. Altman, and P.M. Bossuyt, *Reporting bias in diagnostic and prognostic studies: time for action.* Clin Chem, 2008. **54**(7): p. 1101-3.

[Curtis et al., 2012]

Curtis C, et al., *The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups*. Nature 2012. **486**: p. 346–352.

[Borman et al., 2016]

Bormann F, et al. *Reduced DNA methylation patterning and transcriptional*

*connectivity define human skin aging.*  Aging Cell (2016) **15**, pp563–571. Doi: 10.1111/acel.12470

[Carroll et al., 2006]

Carroll J, et al. *Genome-wide analysis of estrogen receptor binding sites*. Nature Genetics (2006) **38**:11 pp 1289-1297 doi:10.1038/ng1901

[Severson et al., 2018]

Severson T, et al. *Characterizing steroid hormone receptor chromatin binding landscapes in male and female breast cancer*. Nature Communications (2018) **9**: article 482 DOI: 10.1038/s41467-018-02856-2

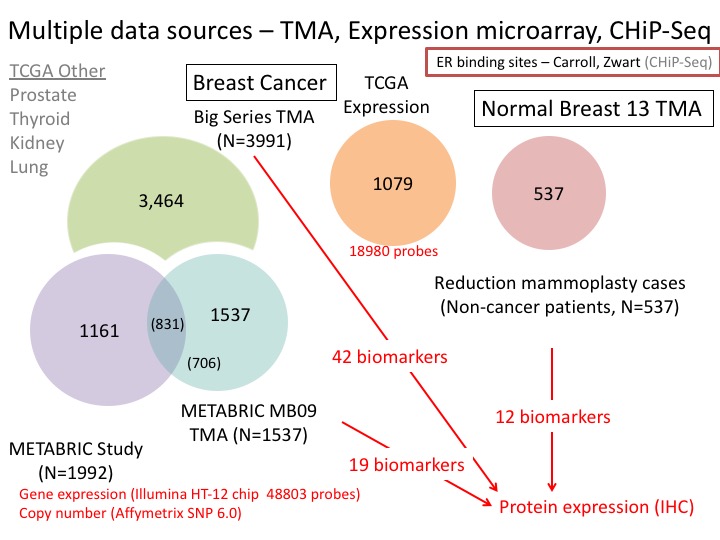


Figure 1: Sources of data

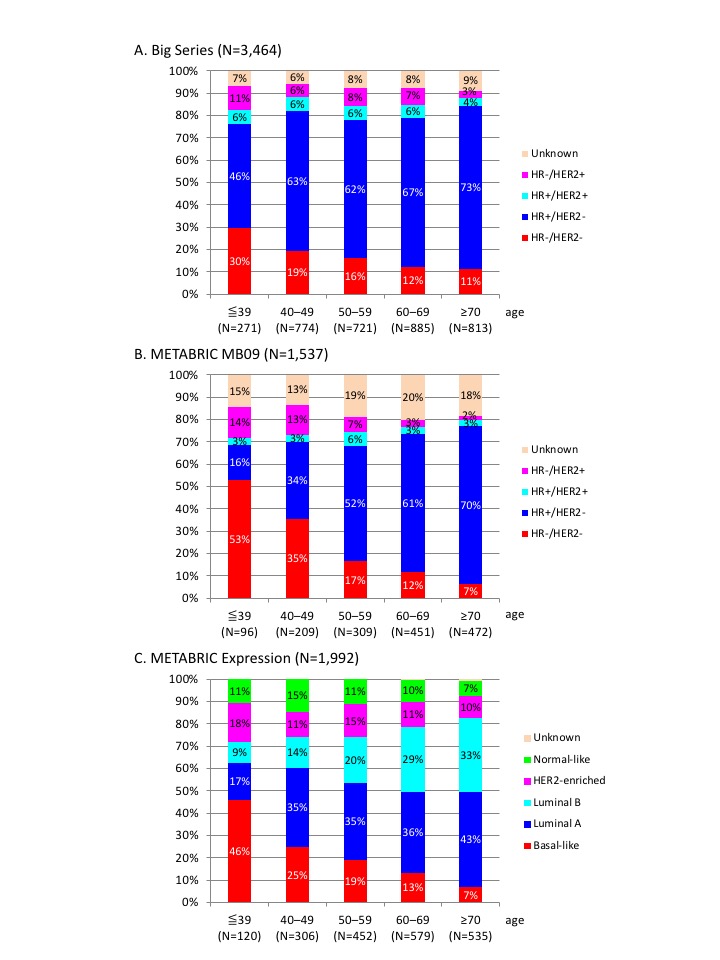


Figure 2: Age and subtype distribution characteristics