**Supplemental Data**



**Supplementary Table S2.** Guidelines for reporting recommendations for tumour marker prognostic studies (REMARK).

|  |  |
| --- | --- |
| **Item to be reported** | |
| **INTRODUCTION** | |
| 1 | *State the marker examined, the study objectives, and any pre-specified hypotheses.*  The study objective and hypothesis are stated in the introduction section.  Examined markers:  18-marker panel: *ACAA1*, *ADGRG6*, *BORCS6*, *CCNA2*, *CDCA5*, *CDKN2A*, *FAM91A1*, *HJURP*, *HSPA14*, *KIAA0494*, *LRRCC1*, *MTURN*, *NEIL3*, *PRR11*, *SKA2*, *SNX8*, *STAM*, *TRIP13*  Established markers: Patient age at diagnosis [years], histological grade [I, II, III], number of positive axillary lymph nodes, pathological tumour size [mm], HER2, PR, and ER status [positive, negative] |
| **MATERIALS AND METHODS** | |
| *Patients* | |
| 2 | *Describe the characteristics (e.g., disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.*  The characteristics of study patients are detailed in “Patients and clinicopathological data” in Material and Methods section and Table 1.  Previous study GEO: GSE20462, GSE97177 |
| 3 | *Describe treatments received and how chosen (e.g., randomized or rule-based).*  Treatment protocols were based on regional treatment guidelines and are listed in Table 1. Patients were treated with surgery by either mastectomy (69 cases) or lumpectomy (44 cases), with axillary node dissection followed by post-operative breast irradiation (40 cases). Adjuvant therapy with chemotherapy (35 cases) and/or hormone therapy (21 cases) was decided according to node status and hormone receptor status. The patients were treated in the years from 1991-1999 before the introduction of Trastuzumab. |
| *Specimen characteristics* | |
| 4 | *Describe type of biological material used (including control samples) and methods of preservation and storage.*  Primary tumour specimen were frozen after surgery and stored at -80 °C. |
| *Assay methods* | |
| 5 | *Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.*  Details of protocols can be found in “Gene expression microarray” in Material and Methods section. The samples were given a pseudo number. |
| *Study design* | |
| 6 | *State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.*  The patients were selected retrospectively through a search of tumour material at the Sahlgrenska University Hospital Tumour Bank, corresponding information in the Sympathy (Department of Pathology) and Melior (Department of Oncology) data bases, and the date of diagnosis between the years 1991-1999 to ensure sufficient follow-up time. The gene expression measurements for each marker were extracted retrospectively from existing microarray profiling data. The microarray was part of previous studies performed in 2009-2010. All patients in the study (n = 136) had complete information for date of diagnosis, date and cause of death. Patients without complete information for clinical parameters were removed from the training cohort of the established and combined marker model (n = 79).  Time period: 1991-1999  End of follow-up period: 01/01/2008  Median follow-up time (all patients): 2683.5 days  Median follow-up time (patients with DSS-event): 1380 days  Median follow-up time (patients without DSS-event): 4494 days  Patients with event (DSS): 67  Patients with event (OS): 78 |
| 7 | *Precisely define all clinical endpoints examined.*  The clinical endpoints used are defined in the Materials and Methods section. |
| 8 | *List all candidate variables initially examined or considered for inclusion in models.*  All candidate variables considered and included in the statistical models are specified in the Materials and Methods section. |
| 9 | *Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.*  Gene expression data was available for 150 tumours of which 136 were primary tumours with clinical information. The sample size provided sufficient statistical power. |
| *Statistical analysis methods* | |
| 10 | *Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.*   1. *Preliminary data preparation*   The preliminary data preparation is described in “Gene expression microarray” in the Material and Methods section.   1. *Association of marker values with other variables*   Possible confounding factors are described in Table 1 using the R-package {tableone} (v0.8.1.)   1. *Methods to evaluate a marker’s univariable association with clinical outcome*   Table 2 describes the univariable Cox proportional hazard models of the individual markers.   1. *Multivariable analysis*   Table 3 and the Material and Methods section describe the multivariable Cox proportional hazard models of the markers.   1. *Missing data*   Item 6 and the Material and Methods section describe the handling of missing data.   1. *Variable selection*   “Selection of prognostic panel” in the Material and Methods section describes the selection of the 18-marker panel.   1. *Checking model assumptions*   The proportional hazards assumption of the Cox regression was tested using the dedicated functions of the R-packages {survival} (v2.40-1) and {survminer} (v0.2.2) confirming the assumption of proportional hazards.   1. *Model validation*   For internal validation of the multivariate models, 1,000 bootstrap samples were created. Cox regression and AUC(t) analyses were applied to each sample. External validation was performed in three independent cohorts. |
| 11 | *Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.*  Details of how marker values were handled and cutpoints determined are provided in “Statistical analysis” in the Material and Methods section. |
| **RESULTS** | |
| *Data* | |
| 12 | *Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.*  The number of events and patients can be found in Item 6, Table 1 and Supplementary Table S1. |
| 13 | *Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumour marker, including numbers of missing values.*  Characteristics of study participants including numbers of missing values are detailed in Table 1 and Supplementary Table S1. |
| *Analysis and presentation* | |
| 14 | *Show the relation of the marker to standard prognostic variables.*  The relationship of the 18-marker panel to established prognostic variables and the Oncotype-based gene panel is detailed in the results section (Figure 3, Table 3, Supplementary Figure S2 and S3). |
| 15 | *Present univariable analyses showing the relation between the marker and outcome, with the estimated effect (e.g., hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analysed. For the effect of a tumour marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.*  Univariable analyses for the 18 markers including their Cox coefficient are stated in the results and Table 2. |
| 16 | *For key multivariable analyses, report estimated effects (e.g., hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.*  Multivariable analyses are stated in Figure 1 and Table 3. |
| 17 | *Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.*  Multivariable analyses including established clinical variables (Patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumour size, HER2, PR, and ER status) are outlined in Table 3. |
| 18 | *If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.*  The results of the internal and external validation are detailed in the Results section. |
| **DISCUSSION** | |
| 19 | *Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.*  The study results were discussed in the context of pre-specified hypotheses and other relevant studies in the Discussion section. |
| 20 | *Discuss implications for future research and clinical value.*  Reverse-transcriptase-polymerase-chain-reaction (RT-PCR) for the specific genes instead of whole-transcriptome gene expression microarrays could improve the clinical utility and be more cost-effective. The 18-marker panel needs to be studied in other population cohorts. |

**Supplementary Table S3.** Overview of the 18 genes, their biological functions and association with cancer and clinical outcome.

|  |  |  |
| --- | --- | --- |
| **Gene** | **Alias** | **Function** |
| *ACAA1* | Acetyl-CoA Acyltransferase 1 | Encodes an enzyme operative in the beta-oxidation system of the peroxisomes. Associated with early stages of cervical carcinogenesis and absent in the later cancer stages (1). Cross-talk between PPARα and ER has been reported to influence the effect of PPARα on obesity and lipid metabolism (2). |
| *ADGRG6* | Adhesion G Protein-Coupled Receptor G6; *GPR126* | G-protein coupled receptor which is activated by type IV collagen, a major constituent of the basement membrane. The adhesion G-protein coupled receptor G6 is associated with angiogenesis by regulating endothelial cell proliferation, migration, and tube formation (3). *ADGRG6* is part of the Mammaprint® gene signature (4). |
| *BORCS6* | BLOC-1 Related Complex Subunit 6; *C17orf59* | Component of the BLOC-one-related complex (BORC) which may play a role in lysosomes movement and localization at the cell periphery. |
| *CCNA2* | Cyclin A2 | Binds and activates cyclin-dependent kinase 2 and thus promotes transition through G1/S and G2/M. Overexpression has been associated with unfavourable survival in ER-positive breast cancer (5). |
| *CDCA5* | Cell Division Cycle Associated 5 | Regulator of sister chromatid cohesion in mitosis stabilizing cohesin complex association with chromatin. May antagonize the action of WAPL which stimulates cohesin dissociation from chromatin. Cohesion ensures that chromosome partitioning is accurate in both meiotic and mitotic cells and plays an important role in DNA repair. Required for efficient DNA double-stranded break repair (6). |
| *CDKN2A* | Cyclin Dependent Kinase Inhibitor 2A | Capable of inducing cell cycle arrest in G1 and G2 phases. Acts as a tumor suppressor. Binds to MDM2 and blocks its nucleocytoplasmic shuttling by sequestering it in the nucleolus. Also induces G2 arrest and apoptosis in a p53-independent manner by preventing the activation of cyclin B1/CDC2 complexes. |
| *FAM91A1* | Family With Sequence Similarity 91 Member A1 | *FAM91A1* is a protein coding gene. Among its related pathways are Gastric cancer network 2. |
| *HJURP* | Holliday Junction Recognition Protein; *DKFZp762E1312*; *FAKTS*; *URLC9*; *hFLEG1* | Centromeric protein that plays a central role in the incorporation and maintenance of histone H3-like variant CENPA at centromeres. Acts as a specific chaperone for CENPA and is required for the incorporation of newly synthesized CENPA molecules into nucleosomes at replicated centromeres. Prevents CENPA-H4 tetramerization and prevents premature DNA binding by the CENPA-H4 tetramer. Directly binds Holliday junctions. Elevated expression levels have been connected to unfavourable prognosis in breast cancer (7,8). Involved in homologous recombination within the DNA double-strand break repair pathway and might be an important factor for chromosomal stability in cancer cells (9). |
| *HSPA14* | Heat Shock Protein Family A (Hsp70) Member 14 | Component of the ribosome-associated complex (RAC), a complex involved in folding or maintaining nascent polypeptides in a folding-competent state. In the RAC complex, binds to the nascent polypeptide chain, while DNAJC2 stimulates its ATPase activity. The *HSPA14* encoded heat shock protein is connected to chromosomal instability in Nijmegen breakage syndrome as a part of a pathway to induce migration, invasion, and transformation (10). |
| *KIAA0494* | *EFCAB14*; EF-Hand Calcium Binding Domain 14 | GO annotations related to this gene include calcium ion binding. |
| *LRRCC1* | Leucine Rich Repeat And Coiled-Coil Centrosomal Protein 1 | Required for the organization of the mitotic spindle. Maintains the structural integrity of centrosomes during mitosis. |
| *MTURN* | Maturin; *C7orf41* | May be involved in early neuronal development. |
| *NEIL3* | Nei Like DNA Glycosylase 3 | NEIL3 belongs to a class of DNA glycosylases homologous to the bacterial Fpg/Nei family. These glycosylases initiate the first step in base excision repair by cleaving bases damaged by reactive oxygen species and introducing a DNA strand break via the associated lyase reaction. The *NEIL3* gene is associated with increased levels of somatic mutations and a poorer outcome in breast cancer (11). |
| *PRR11* | Proline Rich 11 | Plays a critical role in cell cycle progression. |
| *SKA2* | Spindle And Kinetochore Associated Complex Subunit 2; *FAM33A* | Component of the SKA1 complex, a microtubule-binding subcomplex of the outer kinetochore that is essential for proper chromosome segregation. Required for timely anaphase onset during mitosis, when chromosomes undergo bipolar attachment on spindle microtubules leading to silencing of the spindle checkpoint. |
| *SNX8* | Sorting Nexin 8 | May be involved in several stages of intracellular trafficking. May play a role in intracellular protein transport from early endosomes to the trans-Golgi network. |
| *STAM* | Signal Transducing Adaptor Molecule | Involved in intracellular signal transduction mediated by cytokines and growth factors. Upon IL-2 and GM-CSL stimulation, it plays a role in signaling leading to DNA synthesis and MYC induction. May also play a role in T-cell development. |
| *TRIP13* | Thyroid Hormone Receptor Interactor 13 | Plays a key role in chromosome recombination and chromosome structure development during meiosis. Required at early steps in meiotic recombination that leads to non-crossovers pathways. Also needed for efficient completion of homologous synapsis by influencing crossover distribution along the chromosomes affecting both crossovers and non-crossovers pathways. Promotes early steps of the DNA double-strand breaks (DSBs) repair process upstream of the assembly of RAD51 complexes. Functions as a mitotic checkpoint-silencing protein and has been associated with unfavourable clinical outcome and chromosome instability of breast cancer patients (12,13). |

References:

1. Gius D, Funk MC, Chuang EY, Feng S, Huettner PC, Nguyen L*, et al.* Profiling Microdissected Epithelium and Stroma to Model Genomic Signatures for Cervical Carcinogenesis Accommodating for Covariates. Cancer Research **2007**;67(15):7113.

2. Jeong S, Yoon M. Inhibition of the Actions of Peroxisome Proliferator-activated Receptor α on Obesity by Estrogen. Obesity **2007**;15(6):1430-40 doi 10.1038/oby.2007.171.

3. Cui H, Wang Y, Huang H, Yu W, Bai M, Zhang L*, et al.* GPR126 Protein Regulates Developmental and Pathological Angiogenesis through Modulation of VEGFR2 Receptor Signaling. The Journal of Biological Chemistry **2014**;289(50):34871-85 doi 10.1074/jbc.M114.571000.

4. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M*, et al.* Gene expression profiling predicts clinical outcome of breast cancer. Nature **2002**;415(6871):530-6 doi 10.1038/415530a.

5. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 Is a Prognostic Biomarker for ER+ Breast Cancer and Tamoxifen Resistance. PLoS ONE **2014**;9(3):e91771 doi 10.1371/journal.pone.0091771.

6. Rankin S, Ayad NG, Kirschner MW. Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. Mol Cell **2005**;18(2):185-200 doi 10.1016/j.molcel.2005.03.017.

7. Hu Z, Huang G, Sadanandam A, Gu S, Lenburg ME, Pai M*, et al.* The expression level of HJURP has an independent prognostic impact and predicts the sensitivity to radiotherapy in breast cancer. Breast Cancer Research : BCR **2010**;12(2):R18-R doi 10.1186/bcr2487.

8. Montes de Oca R, Gurard-Levin ZA, Berger F, Rehman H, Martel E, Corpet A*, et al.* The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. Molecular oncology **2015**;9(3):657-74 doi 10.1016/j.molonc.2014.11.002.

9. Kato T, Sato N, Hayama S, Yamabuki T, Ito T, Miyamoto M*, et al.* Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. Cancer Res **2007**;67(18):8544-53 doi 10.1158/0008-5472.can-07-1307.

10. Wu CY, Lin CT, Wu MZ, Wu KJ. Induction of HSPA4 and HSPA14 by NBS1 overexpression contributes to NBS1-induced in vitro metastatic and transformation activity. Journal of biomedical science **2011**;18:1 doi 10.1186/1423-0127-18-1.

11. Shinmura K, Kato H, Kawanishi Y, Igarashi H, Goto M, Tao H*, et al.* Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer. Oxidative Medicine and Cellular Longevity **2016**;2016:1546392 doi 10.1155/2016/1546392.

12. Maurizio E, Wiśniewski JR, Ciani Y, Amato A, Arnoldo L, Penzo C*, et al.* Translating Proteomic Into Functional Data: An High Mobility Group A1 (HMGA1) Proteomic Signature Has Prognostic Value in Breast Cancer. Molecular & Cellular Proteomics : MCP **2016**;15(1):109-23 doi 10.1074/mcp.M115.050401.

13. Wang K, Sturt-Gillespie B, Hittle JC, Macdonald D, Chan GK, Yen TJ*, et al.* Thyroid Hormone Receptor Interacting Protein 13 (TRIP13) AAA-ATPase Is a Novel Mitotic Checkpoint-silencing Protein. The Journal of Biological Chemistry **2014**;289(34):23928-37 doi 10.1074/jbc.M114.585315.





**Supplemental Data: Figure legends**

**Supplementary Figure S1.** AUC(t) functions of multivariable models.

The lines represent the time-dependent area under the ROC curve (AUC(t)) for the 18-marker panel (grey), the established markers (blue) and the combined model with the 18-marker panel and the established markers (red). (A) Estimated performance of the TCGA validation cohort for overall survival. Established clinical variables contain number of positive axillary lymph nodes, tumour size, age, ER and PR status for all 720 patients. (B) Estimated performance of the GSE4922 validation cohort for disease-free survival. Established clinical variables contain age, ER status, tumour size, axillary lymph node status and histological grade for all 237 patients. (C-D) Estimated performance of the GSE1456 validation cohort for overall and recurrence-free survival. Established clinical variables contain histological grade and subtype for all 128 patients.

**Supplementary Figure S2.** AUC(t) functions of multivariable models.

The lines represent the time-dependent area under the ROC curve (AUC(t)) for the 18-marker panel (black) and the Oncotype Dx-based 16-marker model (blue). (A) Estimated performance of the 18-marker panel in comparison to the Oncotype Dx-based 16-marker panel in the complete training cohort (n = 136) cohort for disease-specific survival (DSS). (B) Estimated performance of the 16-marker panel in comparison to the Oncotype Dx-based 16-marker panel in the ER-positive training cohort (n = 107) cohort for disease-specific survival (DSS).

**Supplementary Figure S3.** Overview of molecular network within different parts of the cell as determined by IPA.

Magenta-coloured molecules show association with DNA replication, recombination, and repair while green molecules are connected to the cell cycle. Blue-coloured molecules summarize the remaining categories.

**Supplemental Data: Figures**

U:\12 genes Gene expression microarray validation\Supplementary Figure S1.tif

U:\12 genes Gene expression microarray validation\Supplementary Figure S2.tif

U:\12 genes Gene expression microarray validation\Supplementary Figure S3.tif