**Supplementary Tables**

**Supplementary Table 1:** Clinicopathological characteristics of both METABRIC and Nottingham series.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **METABRIC series**  **Number of cases (%)** | **Nottingham series**  **Number of cases (%)** |
| **Patient Age**  ≤ 50 years  > 50 years | 383 (19.3)  1556 (78.5) | 398 (30.7)  897 (69.3) |
| **Tumour size**  ≤ 2cm  > 2cm | 622 (32)  1331 (68) | 796(61.5)  499 (38.5) |
| **Tumour grade**  1  2  3 | 170 (9)  770 (41)  952 (50) | 199 (15.4)  493(38)  603(46.6) |
| **Nodal Stage**  1  2  3 | 1053 (53)  622 (31)  316 (16) | 804 (62)  351 (27)  140 (11) |
| **Vascular invasion**  Definite  Negative/Probable | 1207(61)  773 (39) | 361 (28)  934 (72) |
| **Oestrogen receptor status**  Negative  Positive | 474 (24)  1506 (76) | 263 (20)  1032 (80) |
| **Progesterone receptor status**  Negative  Positive | 940 (48)  1040 (52) | 540 (42)  755 (58) |
| **Human epidermal growth factor receptor 2 (HER2+) status**  Negative  Positive | 1733 (87)  247 (13) | 1116(86)  179 (14) |
| **Nottingham prognostic index (NPI)**  Good  Moderate  Poor | 680 (34)  1101 (56)  199 (10) | 433(33.5)  642 (49.5)  220 (17) |
| **Survival status**  Alive  Dead | 1071 (68)  505 (32) | 936 (72.3)  359 (27.7) |
| **Chemotherapy**  Treated  Non treated | 418 (21)  1562 (79) | 446 (34.4)  849 (65.6) |
| **Endocrine therapy**  Treated  Non treated | 1215 (61.4)  765 (38.6) | 1022 (79)  273 (21) |
| **Median follow-up (months)**  **Interquartile range (IQR)** | 86  (49-143) | 133  (89-167) |

**Supplementary Table 2:** The REMARK checklist

|  |  |  |
| --- | --- | --- |
| **Items to be reported** | | Page no. |
| INTRODUCTION | |  |
| 1 | State the marker examined, the study objectives, and any pre-specified hypotheses. | 3,4 |
| 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. | 4 |
| 3 | Describe treatments received and how chosen (e.g., randomized or rule-based). | 4,5 |
| Specimen characteristics | |  |
| 4 | Describe type of biological material used (including control samples) and methods of preservation and storage. | 5,6 |
| 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. | 5,6 |
| 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 from which cases were taken, the end of the follow-up period, and the median follow-up time. | 4 |
| 7 | Precisely define all clinical endpoints examined. | 4 |
| 8 | List all candidate variables initially examined or considered for inclusion in models. |  |
| 9 | Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size. |  |
| 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. | 6,7 |
| 11 | Clarify how marker values were handled in the analyses; if relevant, describe methods used for cut- point determination. | 6,7 |
| 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. | 7,8 |
| 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. | 7,8 |
| Analysis and presentation | |  |
| 14 | Show the relation of the marker to standard prognostic variables. | 7 |
| 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. | 8,9 |
| 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. | 9 |
| 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. | 9 |
| 18 | If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation. |  |
| 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. | 10-13 |
| 20 | Discuss implications for future research and clinical value. | 10-13 |

**Supplementary Table 3. Association of INCENP mRNA expression with clinicopathological parameters in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and in the Cancer Genome Atlas (TCGA) breast cancer series**.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **METABRIC INCENP mRNA** | | | **TCGA INCENP mRNA** | | |
| **Low**  **No (%)** | **High**  **No (%)** | **x2**  ***P-*value** | **Low**  **No (%)** | **High**  **No (%)** | **x2**  ***P-*value\*** |
| **Patient age (years)**  < 50  ≥ 50 | 261 (68)  1160(75) | 121 (32)  386 (25) | 7.11  **0.008** | 47  159 | 184  464 | 2.466  **0.116** |
| **Tumour size (cm)**  < 2  ≥ 2 | 490 (79)  940 71() | 131 (21)  381 (29) | 13.05  **<0.001** | 75 (57)  131 (46 | 164(43)  484 (54) | 9.55  **0.002** |
| **Tumour grade**  1  2  3 | 156 (92)  670(87)  550 (58) | 14 (8)  98 (13)  393 (42) | 213.15  **<0.001** | 41 (43)  119(42)  34 (67) | 48 (57)  256(57)  318 (33) | 29.56  **<0.001** |
| **Molecular subtypes**  Luminal A  Luminal B  Basal like  HER2  Normal | 671 (93)  336(70)  131(70)  119 (49.5)  184 (93) | 47 (7)  147(30)  194 (30)  121(51.5)  13(7) | 443.9  **<0.001** | 90 (52)  343 (52)  690 (54) | 8 (48)  57 (47)  153 (46) | 8.117  **0.017** |
| **Mitotic count**  1  2  3 | **N/A** | **N/A** | **N/A** | 560 (43)  232 (52)  331 (64) | 73 (57)  32 (48)  113 (36) | 41.26  **<0.001** |
| **Histological subtypes**  Non-specific type (NST)  Lobular  Mixed NST and lobular  Mixed NST and special type  Other special types\* | 1074(70)  128(87.6)  82 (92)  2 (67)  10 (83) | 460(30)  18 (12.4)  7 (8)  1(33)  2 (7) | 107.9  **<0.001** | 702(80.4)  94(92)  66(87)  27(79)  17(94) | 172(19.6)  8(8)  10(13)  7(21)  1(6) | 26.25  **<0.001** |
| **Axillary nodal stage**  Stage 1  Stage 2  Stage 3 | 781 (75.6)  433 (70)  226 (80) | 251 (24.4)  184 (30)  88 (20) | 6.34  **0.042** | 693 (49)  310 (56)  10 (72) | 141 (51)  56 (44)  21 (28) | 0.697  **0.706** |
| **Nottingham Prognostic Index**  Good  Moderate  Poor | 589 (86.6)  732 (67)  125 (63.5) | 91 (44)  360 (34)  72(36.5) | 93.59  **<0.001** | **N/A** | **N/A** | **N/A** |
| **Lympho-vascular invasion**  Negative  Positive | 400 (55)  1017 (84.5) | 329 (35)  186 (37) | 204.37  **<0.001** | 148 (26.5)  58 (20) | 411 (73.5)  237(80) | 4.90  **0.027** |
| **Oestrogen receptor**  Negative  Positive | 220(47)  1226(82) | 249(53)  274(18) | 222.13  **<0.001** | 15(8)  187(29) | 170(92)  452(71) | 37.061  **<0.001** |
| **Progesterone receptor**  Negative  Positive | 564(60.4)  882(85) | 369(39.6)  154(15) | 153.35  **<0.001** | 31(11.4)  170(31) | 241(88.6)  376(69) | 40.797  **<0.001** |
| **Human epidermal growth factor receptor 2 (HER2) status**  Negative  Positive | 1299(75)  147(60.7) | 428(25)  95(39.3) | 22.79  **<0.001** | 141(24.8)  22(16.5) | 426(75.2)  111(83.5) | 5.563  0.062 |
| **Triple negative status**  Non-triple negative  Triple negative | 1309(79)  137(43) | 342(21)  181(57) | 179.164  **<0.001** | 954(84.6)  155(78) | 174(15.4)  43(22) | 4.87  **0.02** |
| **P53 mutation Status**  Mutation  Wild type | 48(49.5)  570(79.5) | 49(50.5)  147(20.5) | 43.84  **<0.001** | **N/A** | **N/A** | **N/A** |
| **Ki67 index**  Low  High | 886(92)  244(43) | 76(8)  342(57) | 470.44  **<0.001** | 190(44.5)  16(3.7) | 237(55.5)  411(96.3) | **193.69**  **<0.001** |

\*p value in bold: significant

\*INCENP Inner centromere protein,

\*Other special types Medullary, Mucinous, Tubular.

**Supplementary Table 4:** Associations between INCENPmRNA expression and the mRNA expression levels of other related biomarkers.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **INCENP mRNA expression (METABRIC)** | | | | **INCENP mRNA expression**  **(TCGA)** | | | |
| **Biomarker status** | **Low**  **No (%)** | **High**  **No (%)** | **X2**  ***P*-value** | **Pearson’s correlation *P*-value** | **Low**  **No (%)** | **High**  **No (%)** | **X2**  **P-value** | **Pearson’s correlation *P*-value** |
| **KI67**  Low  High | 886(92)  544(55) | 76(8)  441(45) | 339.25  **<0.001** | 0.60  **<0.001** | 190(44.5)  16(3.7) | 237(55.5)  411(96.3) | 193.69  **<0.001** | 0.669  **<0.001** |
| **Budding Uninhibited by Benzimidazoles 1 (BUB1)**  Low  High | 902 (92.4)  539 (54.6) | 74 (7.6)  447 (45.3) | 358.44  **<0.001** | 0.638  **<0.001** | 184(43)  22(5) | 243(57)  405(95) | 167.89  **<0.001** | 0.644  **<0.001** |
| ***Centromere Protein E (CENPE)***  Low  High | 896 (92.5)  537 (54.4) | 73(5)  450 (95) | 361.56  **<0.001** | 0.60  **<0.001** | 183(43)  23(5) | 244(57)  404(95) | 163.77  **<0.001** | 0.594  **<0.001** |
| ***Polo-likekinase1 (PLK1)***  Low  High | 829 (94.4)  588(55.6) | 49 (14)  469 (86) | 368.12  **<0.001** | 0.613  **<0.001** | 169(40)  37(9) | 258(60)  390(91) | 111.47  **<0.001** | 0.703  **<0.001** |
| ***Cell division cycle associated 8 (CDCA8)***  Low  High | 915 (93.7)  527 (53) | 48 (32)  474 (68) | 451.488  **<0.001** | 0.68  **<0.001** | 178(42)  27(6) | 247(58)  400(94) | 147.4  **<0.001** | 0.652  **<0.001** |
| ***Cell division cycle 20 (CDC20)***  Low  High | 920(87)  522(60) | 62  461 | 414.24  **<0.001** | 0.678  **<0.001** | 157(37)  49(11.5) | 270(63)  378(88.5) | 74.62  **<0.001** | 0.590  **<0.001** |
| ***Cyclin-dependent kinase 1 (CDK1)***  Low  High | 852(90.3)  588 (56.5) | 126 (79)  397 (21) | 188.8  **<0.001** | 0.426  **<0.001** | 159(37)  47(11) | 268(63)  380(89) | 80.25  **<0.001** | 0.425  **<0.001** |
| ***Kinesin Family Member 23 (KIF23)***  Low  High | 881 (90.4)  554 (56.4) | 94 (9.6)  427 (43.6) | 287.33  **<0.001** | 0.543  **<0.001** | 179(42)  27(6) | 248(58)  400(94) | 147.81  **<0.001** | 0.684  **<0.001** |
| ***Kinesin Family Member 20A (KIF20A)***  Low  High | 794(95)  569(55) | 40(5)  473(45) | 384.28  **<0.001** | 0.659  **<0.001** | 172(40)  34(8) | 255(60)  393(92) | 121.83  **<0.001** | 0.621  **<0.001** |
| ***Aurora kinase A (AURKA)***  Low  High | 911(93.2)  529(54) | 66(6.8)  455(46) | 391.75  **<0.001** | 0.602  **<0.001** | 167(39)  39(9) | 260(61)  388(91) | 104.81  **<0.001** | 0.527  **<0.001** |
| ***Aurora kinase B (AURKB)***  Low  High | 926(94)  515(52.7) | 57  463 | 434.2  **<0.001** | 0.675  **<0.001** | 149(35)  57(13) | 277(65)  370(87) | 54.45  **<0.001** | 0.552  **<0.001** |

p-value <0.05 significant

\*The cut-off values for dichotomizing mRNA expression in the MEATBRIC and TCG cohorts were determined using breast cancer survival using X-tile programme (Yale University, version 3.6.1).

\*Pearson’s correlation: used to test the association, between INCENP mRNA expression and other related genes involved in cell proliferation as continuous variables