**Supplement S.1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Reference** | Our  data | | Riegman 20011 | | Varis 20012 | | Albrecht 20043 | Nancarrow 20084 | | Li 20085 | Wiech  20096 | | Gu  20107 | | Beroukhim  20108 | | Isinger-Ekstrand  20109 |
| **Methods** | aCGH/SNP  expression | | CGH | | CGH | | aCGH | SNP/aCGH | | SNP | SNP | | SNP | | SNP | | aCGH,  expression |
| **Validation** | PCR/FISH | | - | | - | | FISH | - | | - | FISH | | - | | FISH/PCR | | - |
| **n=** | 56 | | 30 | | 18 | | 18 | 23 | | 42 | 27 | | 42 | | 73 | | 16 |
| 1p | (34) | |  | |  | | (33) |  | |  |  | |  | | (ND) | (ND) | (44) |
| 1q | (11) | (11) |  | |  | | (28) | (52) | |  |  | |  | | (ND) | | (69) |
| 2p | (21) | |  | | (11) | | (44) |  | |  | \*(44) | |  | |  | |  |
| 2q | (5) | |  | |  | |  | (ND) | |  |  | |  | | (ND) | | (38-50) |
| 3p | (25) | (39) | (30) | |  | | (33) | (74) | (74) | (46) | (48) | | (48) | | (ND) | |  |
| 3q | (14) | | (27) | |  | | (33) | (25) | |  |  | | (21) | | (ND) | (ND) |  |
| 4p | (23) | | (47) | |  | | (39) | (ND) | |  |  | | (38) | |  | |  |
| 4q |  | | (40) | | (22) | |  | \*(22) | (22) |  | (20) | | (43) | | (ND) | |  |
| 5p |  | |  | |  | | (33) | (25) | | (ND) |  | | (20) | |  | | (69) |
| 5q |  | | (27) | | (22) | |  | \*(61) | (61) | (ND) | (36) | | (38) | | (ND) | |  |
| 6p | (27) | (14) |  | | (11) | |  |  | |  | (36) | | (17) | (41) | (ND) | (ND) | (44) |
| 6q |  | |  | |  | |  |  | |  |  | |  | | (ND) | (ND) |  |
| 7p | (11) | | (37) | | (28) | |  | (ND) | |  |  | | (26) | | (ND) | |  |
| 7q | (13) | | (60) | (30) | (17) | | (39) |  | | (ND) | (48) | (36) | (31) | | (ND) | (ND) |  |
| 8p | (11) | | (27) | |  | |  |  | |  | \*(26) | | (32) | (38) | (ND) | |  |
| 8q | (29) | | (47) | | (22) | | (50) | (56) | | (ND) | (30) | | (37) | | (ND) | | (75) |
| 9p | (11) | | (53) | | (17) | | (28) | (26) | (26) | (ND) | (59) | (59) | (38) | | (ND) | |  |
| 9q |  | |  | | (ND) | |  |  | |  |  | |  | |  | |  |
| 10p |  | |  | |  | |  |  | |  |  | |  | |  | |  |
| 10q |  | |  | | (22) | |  |  | | (ND) |  | |  | |  | |  |
| 11p | \*(11) | |  | |  | |  | (61) | |  | (36) | | (38) | | (ND) | |  |
| 11q |  | |  | |  | | (39) |  | |  |  | | (17) | (36) | (ND) | (ND) |  |
| 12p |  | | \*(27) | |  | |  | (ND) | |  | (ND) | | (20) | | (ND) | |  |
| 12q |  | |  | |  | |  |  | |  |  | | (20) | (36) |  | | (56) |
| 13p |  | |  | |  | |  |  | | (ND) |  | |  | | (ND) | |  |
| 13q |  | | (33) | | (22) | |  | (57) | | (ND) | (44) | (36) | (23) | | (ND) | | (62) |
| 14p |  | |  | |  | |  |  | |  |  | |  | | (ND) | |  |
| 14q |  | | (37) | |  | |  |  | |  | (44) | |  | | (ND) | |  |
| 15p |  | |  | |  | |  |  | |  |  | |  | | (ND) | |  |
| 15q | (11) | | (30) | | (11) | | (61) |  | |  |  | |  | | (ND) | (ND) |  |
| 16p |  | |  | |  | |  |  | |  |  | |  | | (ND) | |  |
| 16q |  | | (40) | |  | |  | (35) | (35) | (24) |  | | (43) | | (ND) | | (38) |
| 17p | (34) | | (30) | |  | | (39) | (70) | | (ND) | (22) | | (48) | |  | |  |
| 17q | (46) | | (27) | | (39) | | (44) | (43) | |  | (52) | (19) | (20) | | (ND) | |  |
| 18p |  | |  | |  | |  | (29) | |  |  | |  | |  | |  |
| 18q |  | | (63) | |  | | (39) | (61) | (78) | (ND) | (41) | | (18) | (36) | (ND) | (ND) | (56) |
| 19p |  | |  | |  | |  | (ND) | |  |  | | (38) | |  | |  |
| 19q | (11) | |  | |  | |  |  | |  | (ND) | |  | | (ND) | |  |
| 20p |  | |  | |  | |  | (87) | |  | (41) | |  | | (ND) | |  |
| 20q | (9) | | (53) | | (56) | | (39) | (52) | |  | (44) | | (29) | |  | | (88) |
| 21p |  | |  | |  | |  |  | |  |  | |  | |  | |  |
| 21q | \*(11) | |  | |  | |  | (87) | |  |  | | (36) | |  | |  |
| 22p |  | |  | |  | | (ND) |  | |  |  | |  | |  | |  |
| 22q |  | | (30) | |  | | (44) | (ND) | |  |  | | (36) | | (ND) | |  |
| Xp |  | |  | |  | |  | (26) | (26) |  |  | |  | | (ND) | |  |
| Xq |  | |  | |  | |  |  | |  |  | |  | |  | |  |
| Y |  | | (60) | |  | |  |  | |  |  | |  | |  | | (69) |
|  |  | |  | |  | |  |  | |  |  | |  | |  | |  |
| Colour | Gains | |  | | | \* | | Novel region | | | | | | | | | |
| key: | LOH | |  | | | | aCGH | Array comparative genomic hybridization | | | | | | | | | |
|  | HD | |  | | | | SNP | Single-nucleotide polymorphism array | | | | | | | | | |
|  | (not stated) | |  | | | | (ND) | Not disclosed | | | | | | | | | |

*Table S.1:* A summary of genome-wide studies showing data from this study and 9 others (2001-2010). The percentages of cases involved per study are quoted in parentheses.

**Supplement S.2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **Total** **(%)** | **Male** **(%)** | **Female** **(%)** |
| Number of patients | 56 (100) | 38 (67.9) | 18 (32.1) |
| Median age (years, range) | 69 (46 – 89) | 65.5 (46 – 82) | 77.5 (51 – 89) |
| Histological differentiation   * Well * Moderate * Poorly * Undifferentiated * Not available | 4 (7.1)  21 (37.5)  22 (39.3)  0  9 (16.1) | 4 (10.5)  15 (39.5)  17 (44.7)  0  2 (5.3) | 0  6 (33.3)  5 (27.8)  0  7 (38.9) |
| TNM staging   * Stage I & II   (T1-3, N0, M0)   * Stage III   (T3/4, N1, M0)   * Stage IV   (any T, any N, M1)   * No information | 14 (25.0)  29 (51.8)  3 (5.4)  10 (17.9) | 10 (26.3)  23 (60.5)  3 (7.9)  2 (5.3) | 4 (22.2)  6 (33.3)  0  8 (44.4) |
| Treatment   * Surgery * Surgery + chemotherapy | 53 (94.6)  3 (5.4) | 35 (92.1)  3 (4.8) | 18 (100%)  0 |
| Median survival from date of diagnosis (years, range) | 2.29  (0.11 – 8.54) | 2.16  (0.11 – 7.99) | 2.77  (0.64 – 8.54) |

*Table S.2:* Sample characteristics and survival information of 56 oesophageal adenocarcinoma patients used in this study.

**Supplement S.3**

***DNA extraction.*** DNA extraction was done following a modification of the method of Sambrook *et. al* 10, cleaned using phenol:chloroform:isoamyl alcohol (Ambion Inc., Huntingdon, UK) and precipitated using ethanol and sodium acetate, then quantified with a NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA).

***PCR and multiplex PCR.*** Primers/STS markers for genes within the potential HD region on 9p21.3 and genes flanking the HD region were identified from the University of California at Santa Cruz (UCSC) online genomic database (http://genome.ucsc.edu/cgi-bin/hgGateway) and/or designed using PubMed primer designing tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Reactions were set up using 25 ng of genomic DNA in a 20 µl reaction containing 1x RDA buffer 11 (16.6 mM (NH4)2SO4, 67 mM Tris pH 8.8, 10 mM 2-mercaptoethanol (Sigma Aldrich, Poole, UK), 6.7 mM MgCl2), 25 mM dNTPs (GE Healthcare, Little Chalfont, UK), 1 mM of each primer (Sigma Aldrich, Haverhill, UK), 1 µl of DMSO (100%, Sigma Aldrich, Poole, UK) and Red Hot Taq Polymerase (Thermo Scientific, UK). PCR conditions: initial denaturation and enzyme activation at 94 ºC (2 minutes); 30 cycles of denaturation at 94 ºC (30 seconds), annealing (30 seconds) and elongation at 72 ºC (30 seconds); followed by a final elongation step at 72 ºC (2 minutes). PCR products were visualised following agarose gel electrophoresis with a UV transilluminator (synGene G: Box, Cambridge, UK) after staining with ethidium bromide (10 mg/ml; Invitrogen, Carlsbad, CA) (1 µl ethidium bromide/100 ml TBE buffer). Initial gel images were taken with GeneSnap software (version 7.0.8, SynGene, Cambridge, UK) under exposure time 150-300 ms. After initial images were recorded, the settings for gamma an contrast were adjusted on whole image so that the final images have black bands on a light greyish background. Figures in the paper were not cropped from the original images and were presented as they were taken.

***PCR primers/STS markers.*** Primer pairs are named according to the genes they are representing – left (For) and right Rev) for the potential HD genes and 4 adjacent genes located on the 9p chromosome arm. PCR-1 primers are primers used in first round PCR in the nested reactions. PCR-2 primers are primers used in the second round PCR in the nested reactions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PCR-1** | **Sequence** | | **Tm (⁰C)** | **Band size (bp)** |
| IFNA14 | For | GGGAGGTTGTCAGAGCAGAA | 63 | 376 |
| Rev | AACCACAATGTAAAGGTACACATGA |
| CDKN2A | For | CGTGAGTGCTCACTCCAGAA | 60 | 359 |
| Rev | TGCCACACATCTTTGACCTC, |
| DMRTA1 | For | CAAAGACTTGACTGCGACCA | 59 | 502 |
| Rev | CCCTGAAAAGGCATAATCCA |
| TUSC1 | For | CCGACTTGAGAAGCTGGAAG | 64 | 458 |
| Rev | TAAAATCCGTCCCTTCGTTG |
| AQP3 | For | GCCACAGCTTAGGTTTGGAG | 64 | 330 |
| Rev | AGAAAAGCAGCAGAGCAAGC |
| **PCR-2** | **Sequence** | | **Tm (⁰C)** | **Band size (bp)** |
| IFNA14 | For | CATCATGGAAATGATTCTCATTG | 60 | 200 |
| Rev | GATGAAAGGAGACATCAGCATG |
| CDKN2A | For | AGGCAGAAGCGGTGTTTTTC | 57 | 197 |
| Rev | AAGGTCCTACAGGGCCACAA |
| DMRTA1 | For | GGGGATGTGGTCCAAGCCAT | 61 | 276 |
| Rev | TCCTGCAGAAGAATATGCCA |
| TUSC1 | For | TCGGAGCCCTCTGGGCCCTG | 62 | 267 |
| Rev | CAAGAGGCAGGGGAACTGTA |
| AQP3 | For | AGAGGATTAAAGGAGTGACGACC | 60 | 128 |
| Rev | TCAGGTCATAAGTTTCATGTTTGC |

*Table S.3:* Information on primer pairs used in this study. Note: Tm: Annealing temperature.

**Supplement S.4**

A.

Supp 3-1.tifSupp 3-2.tif

B.

60 70 80 90 100 110

AGANNTATAANTGNCCTGCCTTTTAACGTNNATATATGCCTTCCCCCACTACCGTAAATT

AGAGCTTTAAATGTCCTGCCTTTTAACGTAGATATATGCCTTCCCCCACTACCGTAAATT

31389

120 130 140 150 160 170

CCATTTATATCATTTTTTATATATTCTTATAAAAATGTAAAAAAGAAAAACACCGCTTCTG

CCATTTATATCATTTTTTATATATTCTTATAAAAATGTAAAAAAGAAAAACACCGCTTCTG

31449

CC

Sequenced product

CC

Genetic sequence of *CDKN2A* obtained from NCBI

31510

*Figure S.4:* A. Sequencing traces of the nested PCR gene product of primer pair *p16/CDKN2A* (D37 Del) using microdissected DNA from sample D037, viewed with Chromas Lite version 2.01 (Technelysium Pty Ltd). The sequences enclosed by the green box are enlarged and shown in B to compare with genetic sequence of *p16/CDKN2A*. B. PCR product of primers of p16/CDKN2A shows identical sequence with the *p16/CDKN2A* transcript obtained from NCBI Nucleotide Basic Local Alignment Search Tool (BLAST).

**Supplement S.5**

***Fluorescence in situ hybridization (FISH).*** Visualisation of FISH was carried out using a Nikon E800 epifluorescence microscope (Nikon, Tokyo, Japan) with a 100W mercury lamp and a charge-coupled device camera interfaced to CytoVision (Version 3.93 Build 215, © Applied Imaging 2007, Newcastle, UK). CytoVisionTM was then used to apply pseudocolours to raw images captured by the camera.

|  |  |  |
| --- | --- | --- |
| **BAC/Plasmid** | **Probing for (genes)** | **Chromosome** |
| RP11-149I2 | *p16/CDKN2A, CDKN2B* | 9 |
| RP11-113D19 | *IFNA21, IFNB1, IFNW1* | 9 |
| RP11-781C22 | *EGFR* | 7 |
| RP11-589N15 | *GATA4, NEIL2, FDFT1, CTSB* | 8 |
| RP11-347P10 | *MTMR9* | 8 |
| RP1-74J1 | *WT1* | 11 |
| pZ8.4 | Centromere | 8 |
| pZ7.5 | Centromere | 7 |
| pRB11 | Centromere | 11 |

*Table S.5:* Details of FISH probes used in this study and their chromosomal locations.

**Supplement S.6**

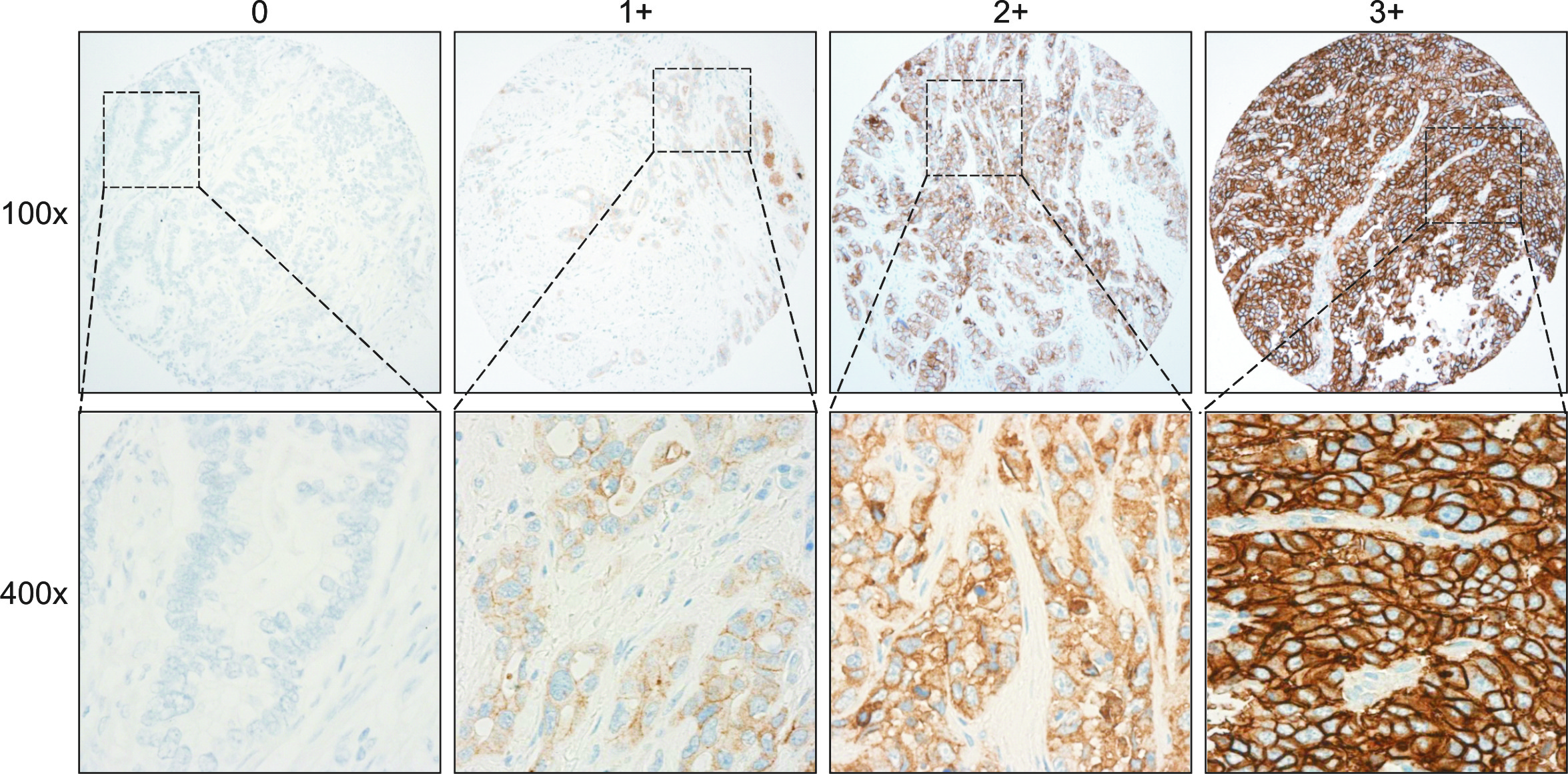
***Immunohistochemistry.*** Slides cut from TMA blocks were first dewaxed in xylene (3 x 15 minutes) and rehydrated in ethanol (15 minutes) before putting them into the BondMax autostainer (Leica Microsystems, Milton Keynes, UK). Initial antigen retrieval was performed using heat-induced epitope retrieval (HIER) solution 1 (citrate-based solution at pH 6.0) or 2 (EDTA-based solution at pH 9.0) for x minutes, depending on results of optimisation procedure. Both positive and negative controls were included alongside each staining procedure to ensure specificity of staining. For each antibody, 3 slides from TMA with sections of 3 different regions from each tumor were tested and scored accordingly to test for over-expression of genes of interest.

|  |  |  |  |
| --- | --- | --- | --- |
| **Antibody** | **Available from** | **Titration used** | **HIER (minutes)** |
| EGFR | Novocastra, Leica, Milton Keynes, UK | 1:10 | 1 (30) |
| WT1 | DAKO, Denmark (Clone 6F-H2) | 1:800 | 2 (30) |
| NEIL2 | Sigma-Aldrich, Dorset, UK (Clone 1B7) | 1:50 | 2 (30) |
| MTMR9 | Novus Biologicals, Littleton, USA (Clone 4A9) | 1:350 | 2 (30) |

*Table S.6A:* Details of antibodies used in this study and their optimal conditions.

The immunohistochemistry staining and its subsequent scoring of EGFR were done following the scoring system for the clinically licensed HercepTest ™ (DAKO, HercepTest™ Interpretation Manual) and according to the maximum intensities of membranous staining in tumor cells, giving scores of 0 (no membranous staining); 1+ (weak incomplete membranous staining); 2+ (weak/moderate complete membranous staining); or 3+ (strong complete membranous staining). Seefigure S.6A.

The scoring for WT1 was done following a similar procedure according to the intensities of cytoplasmic staining in tumor cells, giving scores of 0 (no cytoplasmic staining); 1+ (weak/little cytoplasmic staining, <25% tumor cells); 2+ (moderate cytoplasmic staining, 25-50% tumor cells); or 3+ (strong cytoplasmic staining, >50% tumor cells). The IHC staining for NEIL2 and MTMR9 was scored according to the scale: 0 (no staining, or <5% tumor cells with very weak cytoplasmic stayining); 1+ (<50% tumor cells with weak cytoplasmic staining); 2+ (>50% tumor cells with moderate cytoplasmic staining); or 3+ (>50% tumor cells with strong cytoplasmic staining). See figures S.6B-D.



*Figure S.6A:* Representative images from EGFR immunohistochemistry on OAC TMA section slides at 100x and 400x magnifications showing how scoring was carried out.

*Figure S.6B:* Representative images from WT1 immunohistochemistry on OAC TMA section slides at 100x and 400x magnifications showing how scoring was carried out.

*Figure S.6C:* Representative images from NEIL2 immunohistochemistry on OAC TMA section slides at 100x and 400x magnifications showing how scoring was carried out.

*Figure S.6D:* Representative images from MTMR9 immunohistochemistry on OAC TMA section slides at 100x and 400x magnifications showing how scoring was carried out.

Table *S.6B* shows breakdown of samples with staining scores as a result of analyses of IHC assays. Protein over-expression was defined by samples having IHC scores of 2+ or 3+.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **Immunohistochemistry scores** | | | | **Total number of samples (n=)** |
|  | **0** | **1+** | **2+** | **3+** |
| ***EGFR*** | 235 (65.5%) | 88 (24.5%) | 23 (6.4%) | 13 (3.6%) | 359 |
| ***WT1*** | 142 (39.7%) | 145 (40.5%) | 52 (14.5%) | 19 (5.3%) | 358 |
| ***NEIL2*** | 159 (44.5%) | 147 (41.1%) | 41 (11.5%) | 10 (2.8%) | 357 |
| ***MTMR9*** | 142 (39.9%) | 126 (35.4%) | 65 (18.3%) | 23 (6.5%) | 356 |

*Table S.6B*: Immunohistochemistry analysis of OAC samples from independent datasets represented on TMAs.

**Supplement S.7**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Patient ID** | SCAMP2 | ZMYND15 | SYCP2L | PMP2 | MDH2 | PCBP1 | MAPK8IP2 | BC040153 | CLN8 | SH3GL1 | SMG6 | TSC22D4 | LYPD6 | HPTP1E | CLDN11 | MEXD3 | **TOTAL** |
| GSM266705-7834 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 11 |
| GSM265809-5855 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 9 |
| GSM266707-7836 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| GSM265791-5801 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 6 |
| GSM265500-5247 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 8 |
| GSM266715-7845 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 6 |
| GSM265501-5275 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 8 |
| GSM266659-7757 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 10 |
| GSM266660-7758 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 8 |
| GSM266074-6524 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| GSM265789-5799 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 9 |
| GSM266119-6630 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 8 |
| GSM265786-5796 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 8 |
| GSM266708-7837 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 11 |
| GSM265787-5797 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 6 |
| GSM266703-7832 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 5 |
| GSM265788-5798 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 8 |
| GSM265790-5800 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 9 |
| GSM266706-7835 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 7 |
| GSM265808-5854 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 8 |
| GSM266586-7268 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 5 |
| GSM266075-6525 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 10 |
| GSM266661-7759 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 9 |

*Table S.7:* LogR values were taken as the average of logR values of all probes covering the gene regions. Genes with genomic gains (logR >0.2) or losses (logR <-0.35) that give poor prognosis were given a score of 1. Number of genes dysregulated per patient, a summation of scores given to each gene, was used for subsequent survival analysis.

**Supplement S.8**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Genes | CGH log2ratios  (median) | 95% CI | Expression fold-change (median) | 95% CI |
| RPL22 | 0.406 | 0.343 – 0.481 | 4.506 | 3.153 – 5.185 |
| TNFRSF25 | 0.010 | -0.102 – 0.061 | 1.197 | 0.933 – 1.457 |
| ERRFI1 | -0.233 | -0.605 – 0.128 | 1.697 | -1.007 – 9.133 |
| PIK3CD | 0.197 | 0.028 – 0.630 | 1.253 | 0.908 – 2.594 |
| RBP7 | 0.315 | -0.105 – 0.766 | 1.163 | 0.889 – 1.807 |
| APITD1 | 0.369 | -0.002 – 0.626 | 1.398 | 0.937 – 2.056 |
| RPL5 | -0.264 | -0.307 – -0.176 | 2.177 | 0.693 – 3.263 |
| F3 | 0.347 | 0.218 – 0.571 | 1.947 | 0.431 – 3.024 |
| PDE4DIP | -0.298 | -0.490 – -0.168 | 0.8425 | 0.014 – 1.671 |
| NOTCH2N | -0.372 | -1.311 – 0.411 | 0.5945 | 0.249 – 0.940 |
| PEX11B | -0.024 | -0.280 – 0.363 | 0.7266 | -2.396 – 3.850 |
| POLR3C | -0.372 | -1.311 – 0.411 | 0.4583 | -0.351 – 1.267 |
| TXNIP | -0.386 | -0.621 – -0.187 | 0.3566 | 0.055 – 0.657 |
| PDZK1 | -0.104 | -1.470 – 1.072 | 0.7625 | -3.717 – 5.242 |
| FMO5 | -0.018 | -0.524 – 0.333 | 0.8594 | 0.488 – 1.230 |
| DARS2 | 0.119 | -0.212 – 0.508 | 1.676 | 0.705 – 2.371 |
| SERPINC1 | -0.100 | -0.422 – 0.174 | 1.532 | -0.020 – 3.293 |
| MEIS1 | 0.292 | 0.142 – 0.409 | 1.342 | 0.782 – 2.692 |
| EPHA4 | -0.221 | -0.707 – 0.361 | 0.6685 | 0.197 – 1.300 |
| SCG2 | -0.272 | -0.838 – 0.187 | 0.8746 | 0.778 – 1.018 |
| FHIT | -0.182 | -0.291 – -0.181 | 0.8061 | 0.652 – 0.805 |
| SERPINE2 | -0.488 | -0.755 – -0.168 | 0.4882 | -0.508 – 1.948 |
| EIF2A | -0.329 | -0.482 – -0.231 | 1.094 | 0.452 – 1.536 |
| MBNL1 | -0.562 | -0.616 – -0.242 | 0.7782 | -4.378 – 8.603 |
| DGKQ | 0.417 | 0.336 – 0.685 | 1.609 | 0.930 – 2.419 |
| PDCD6 | -0.284 | -0.797 – 0.079 | 0.4838 | -0.455 – 1.908 |
| GTPBP2 | 0.079 | -0.128 – 0.194 | 1.702 | 0.514 – 3.129 |
| MRPL4 | -0.210 | -0.373 – -0.128 | 1.725 | 0.332 – 3.439 |
| RUNX2 | 0.395 | 0.305 – 0.493 | 2.131 | 0.416 – 3.819 |
| CLIC5 | 0.451 | 0.425 – 0.704 | 4.421 | -2.796 – 12.530 |
| CYP39A1 | 0.022 | -0.191 – 0.264 | 1.531 | 0.150 – 2.619 |
| CD2AP | 0.009 | -0.098 – 0.231 | 2.089 | 0.244 – 4.343 |
| SEC61G | 1.233 | -0.379 – 2.746 | 2.141 | 0.436 – 6.059 |
| EGFR | 1.344 | -0.262 – 3.282 | 11.48 | 1.140 – 19.70 |
| GBAS | 0.781 | 0.290 – 1.652 | 2.530 | 0.153 – 7.181 |
| PSPH | 0.439 | 0.280 – 0.812 | 1.580 | 0.864 – 2.803 |
| CCT6A | 0.046 | -0.286 – 0.492 | 1.294 | 0.435 – 2.367 |
| Genes | CGH log2ratios  (median) | 95% CI | Expression fold-change (median) | 95% CI |
| BAIAP2L1 | 0.306 | 0.179 – 0.498 | 2.142 | 0.899 – 3.799 |
| CYP3A5 | -0.013 | -0.454 – 0.157 | 3.316 | 1.037 – 5.315 |
| PPP1R3B | 0.726 | 0.082 – 2.137 | 1.737 | -4.566 – 24.890 |
| MSRA | 1.863 | 0.403 – 2.815 | 6.503 | 0.297 – 12.640 |
| SOX7 | 0.472 | -0.371 – 2.254 | 3.290 | -0.086 – 7.807 |
| XKR6 | 0.285 | -0.586 – 2.340 | 4.890 | -3.925 – 29.020 |
| MTMR9 | 0.928 | 0.091 – 1.963 | 6.900 | 1.570 – 18.360 |
| NEIL2 | 0.480 | 0.074 – 1.113 | 8.604 | 3.153 – 14.000 |
| FDFT1 | 1.818 | 0.571 – 2.706 | 6.584 | 1.794 – 11.480 |
| CMYC | 0.431 | 0.322 – 0.701 | 2.168 | 1.847 – 2.529 |
| CDKN2A | -0.459 | -0.636 – -0.314 | 0.6668 | 0.179 – 1.253 |
| FBXO10 | -0.188 | -0.275 – 0.119 | 1.271 | 0.789 – 1.825 |
| WT1 | 0.535 | 0.432 – 0.677 | 9.920 | -2.049 – 21.740 |
| CD59 | 0.239 | -0.053 – 0.560 | 1.287 | 0.919 – 1.505 |
| RPL23 | 0.173 | -0.123 – 0.496 | 5.968 | 1.701 – 12.35 |
| ERBB2 | 0.780 | 0.309 – 1.422 | 1.560 | 0.524 – 5.654 |
| GRB7 | 0.780 | 0.102 – 1.230 | 2.300 | 1.123 – 4.049 |
| CSF3 | 0.337 | -0.082 – 0.987 | 1.643 | 0.653 – 2.761 |
| CASC3 | 0.119 | -0.543 – 1.228 | 1.656 | 0.760 – 2.902 |
| CDC6 | 0.170 | --0.085 – 0.688 | 2.165 | 0.561 – 4.675 |
| RARA | 0.342 | 0.061 – 0.704 | 1.386 | 0.936 – 2.944 |
| IGRBP4 | 0.300 | 0.037 – 0.523 | 1.810 | 0.916 – 1.970 |
| SMARCE1 | 0.401 | -0.440 – 2.029 | 1.372 | -0.350 – 6.320 |
| TOMM34 | 0.153 | -0.907 – 1.075 | 1.500 | 0.774 – 2.046 |
| MATN4 | 0.525 | 0.044 – 1.210 | 1.945 | -0.305 – 4.243 |
| SDC4 | 0.618 | 0.232 – 0.844 | 2.255 | -0.209 – 5.639 |

*Table S.8:* Statistical summary of genes with differential expressions plotted in figure 3A.

**Supplement S.9**

****

D4

D23

D25

D26

D40

D43

D47

D52

D54

D84

D97

D101

D115

D117

D121

D126

D128

D129

D143

D154

D155

D160

D161

D167

D8

D10

D13

D24

D29

D38

D96

D105

D118

D166

D18

D27

D35

D36

D37

D50

D51

D53

D58

D86

D107

D120

D133

O134

D136

D138

D139

D156

D2

D3

D5

D141

Cluster 5 (n=24)

Cluster 4 (n=10)

Cluster 3 (n=3)

Cluster 2 (n=1)

Cluster 1 (n=18)

*Figure S.9*: Unsupervised K-means clustering analysis (K=5, 50 iterations, reproducibility >50%) of aCGH data, generated from Gene Cluster 3.0, C Clustering Library version 1.47. The 56 OAC samples were segregated into give clusters according to their genomic profiles.

**Supplement S.10**

|  |  |  |
| --- | --- | --- |
| **Genes** | **Cytoband** | **Functions, if known** |
| SCAMP2 | 15q24.1 † | Post-Golgi vesicle-mediated transport |
| ZMYND15 \* | 17p13.2 | Zinc-ion binding |
| SYCP2L \* | 6p24.2 | - |
| PMP2 \* | 8q21.13 | Lipid-binding, transporter activity |
| MDH2 | 7q11.23 | Catalyses oxidation of malate to oxaloacetate |
| PCBP1 | 2p14 † | Multifunctional: RNA-binding, translational control |
| MAPK8IP2 | 22q13.33 | Anti-apoptosis, signal transduction (MAPK/JNK cascades) |
| BC040153 | 22q13.2 | (Hypothetical protein) |
| CLN8 | 8p23.3 | Lipid transport, apoptosis , nervous system development |
| SH3GL1 | 19p13.3 | Central nervous development, endocytosis |
| SMG6 | 17p13.3 | Telomere maintenance, mRNA export from nucleus |
| TSC22D4 | 7q22.1 † | Transcription regulator, response to osmotic stress |
| LYPD6 \* | 2q23.2 | - |
| HPTP1E | 4q21.3 | (mRNA) |
| CLDN11 | 3q26.2 | Cell adhesion, spermatogenesis, axon ensheathment |
| MEXD3 \* | 19p13.3 | - |

*Table S.10:* Summary of genes representative of the clones that have significantly different log2 ratios (p<4x10-7) between K-means cluster 1 and the other 4 clusters combined.

\* indicates a novel gene

† indicates a common region with copy number gains previously described.

**References**

1. Riegman PH, Vissers KJ, Alers JC, et al. Genomic alterations in malignant transformation of Barrett's esophagus. *Cancer Res* 2001;61(7):3164-70.

2. Varis A, Puolakkainen P, Savolainen H, et al. DNA copy number profiling in esophageal Barrett adenocarcinoma: comparison with gastric adenocarcinoma and esophageal squamous cell carcinoma. *Cancer Genet Cytogenet* 2001;127(1):53-8.

3. Albrecht B, Hausmann M, Zitzelsberger H, et al. Array-based comparative genomic hybridization for the detection of DNA sequence copy number changes in Barrett's adenocarcinoma. *J Pathol* 2004;203(3):780-8.

4. Nancarrow DJ, Handoko HY, Smithers BM, et al. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. *Cancer Res* 2008;68(11):4163-72.

5. Li X, Galipeau PC, Sanchez CA, et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. *Cancer Prev Res (Phila Pa)* 2008;1(6):413-23.

6. Wiech T, Nikolopoulos E, Weis R, et al. Genome-wide analysis of genetic alterations in Barrett's adenocarcinoma using single nucleotide polymorphism arrays. *Lab Invest* 2009;89(4):385-97.

7. Gu J, Ajani JA, Hawk ET, et al. Genome-wide catalogue of chromosomal aberrations in barrett's esophagus and esophageal adenocarcinoma: a high-density single nucleotide polymorphism array analysis. *Cancer Prev Res (Phila)* 2010;3(9):1176-86.

8. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463(7283):899-905.

9. Isinger-Ekstrand A, Johansson J, Ohlsson M, et al. Genetic profiles of gastroesophageal cancer: combined analysis using expression array and tiling array--comparative genomic hybridization. *Cancer Genet Cytogenet* 2010;200(2):120-6.

10. Sambrook J, Russell D. *Rapid isolation of mammilian DNA*. 3rd ed. Cold Spring Harbour: Cold Spring Harbour Press, 2001.

11. Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993;259(5097):946-51.