

GWAS META-ANALYSIS AND REPLICATION IDENTIFIES THREE NEW SUSCEPTIBILITY LOCI FOR OVARIAN CANCER

P.D.P. Pharoah et al

Supplementary Table 1: Description of individual OCAC studies and case-control sets

| Study Name | Study abbreviation | Study Location | Study Type | Total Number of Subjects ^b | | % of subjects of European Ancestry | Number of invasive cases of European ancestry |
|---|--------------------|---|-------------------------------|---------------------------------------|----------|------------------------------------|---|
| | | | | Cases | Controls | | |
| Australian Ovarian Cancer Study/Australian Cancer Study (Ovarian Cancer) | AUS | Australia | Population based/case-control | 949 | 1011 | 95.1 | 886(561) |
| Bavarian Ovarian Cancer Cases and Controls | BAV | Southeast Germany | Population based/case-control | 98 | 143 | 100 | 93(56) |
| Belgian Ovarian Cancer Study | BEL | Belgium, University Hospital Leuven | Hospital based/Case-control | 277 | 1352 | 99.6 | 274(194) |
| Diseases of the Ovary and their Evaluation | DOV | USA: 13 counties in western Washington state | Population based/case-control | 1350 | 1606 | 92 | 903(525) |
| Oregon Ovarian Cancer Registry | ORE | Portland, Oregon | Case only | 70 | 0 | 92.9 | 54(39) |
| German Ovarian Cancer Study | GER | Germany: two geographical regions in the states of Baden-Württemberg and Rhineland-Palatinate in southern Germany | Population based/case-control | 216 | 413 | 99.5 | 189(95) |
| Dr. Horst Schmidt Kliniken | HSK | Germany | Case only | 155 | 0 | 98.7 | 144(107) |
| Hawaii Ovarian Cancer Case-Control Study | HAW | USA: Hawaii; studies using only epi variables: Hawaii and Southern California, USA | Population based/case-control | 338 | 601 | 25.4 ^a | 60(38) |
| Hannover-Jena Ovarian Cancer Study | HJO | Germany | Hospital based/Case-control | 286 | 274 | 99.5 | 271(140) |
| Hannover-Minsk Ovarian Cancer Study | HMO | Belarus | Case-control | 144 | 140 | 98.6 | 142(50) |
| Helsinki Ovarian Cancer Study | HOC | Helsinki, Finland | Case-control | 226 | 447 | 99.9 | 217(113) |
| Novel Risk Factors and Potential Early Detection Markers for Ovarian Cancer | HOP | Western PA, Northeastern Ohio, Western New York | Population based/Case-control | 759 | 1501 | 97 | 654(377) |
| Gilda Radner Familial Ovarian Cancer Registry | GRR | USA | Familial cancer/Case only | 115 | 0 | 97.4 | 112(74) |
| Hospital-based Research Program at Aichi Cancer Center | JPN | Japan: Nagoya City | Case-control | 76 | 81 | 0 ^a | - |

| Study Name | Study abbreviation | Study Location | Study Type | Total Number of Subjects ^b | | % of subjects of European Ancestry | Number of invasive cases of European ancestry |
|--|--------------------|---|---------------------------------------|---------------------------------------|----------|------------------------------------|---|
| | | | | Cases | Controls | | |
| Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute | LAX | USA: Southern California | Case only | 330 | 0 | 84.2 | 278(217) |
| Los Angeles County Case-control studies of Ovarian Cancer-1 | USC | Los Angeles County | Population based/ Case-control | 1260 | 1370 | 71.8 | 660(424) |
| <u>MALignant OVarian Cancer</u> | MAL | Denmark | Population based/ Case-control | 571 | 829 | 99.9 | 440(272) |
| Danish Pelvic Mass Study | PVD | Denmark | Case only | 172 | 0 | 98.2 | 169(128) |
| Malaysia Ovarian Cancer Study | MAS | Malaysia | Need info | 106 | 106 | 0 ^c | - |
| Mayo Clinic Ovarian Cancer Case-Control Study | MAY | USA: North Central (MN, SD, ND, IL, IA, WI) | Clinic based/ Case-control | 791 | 753 | 98.5 | 697(507) |
| Melbourne Collaborative Cohort Study | MCC | Melbourne, Australia | Cohort/ Nested case-control | 64 | 68 | 99.2 | 63(34) |
| MD Anderson Ovarian Cancer Study | MDA | USA: Texas | Hospital based/ Case-control | 323 | 385 | 98.5 | 313(190) |
| Memorial Sloan-Kettering Cancer Center | MSK | USA: New York City | Case-control | 556 | 697 | 84.6 | 467(382) |
| North Carolina Ovarian Cancer Study | NCO | USA: Central and eastern North Carolina (48 counties) | Population based/ Case-control | 1063 | 984 | 82.8 | 729(410) |
| New England Case Control Study | NEC | USA: New Hampshire and Eastern Massachusetts | Population based/ Case-control | 949 | 1049 | 95.8 | 672(381) |
| Nurses' Health Study | NHS | USA | Population based/ Nested case-control | 147 | 429 | 99.3 | 127(68) |
| New Jersey Ovarian Cancer Study | NJO | USA: New Jersey (six counties) | Case-control | 190 | 194 | 91.2 | 169(100) |
| University of Bergen, Haukeland University Hospital, Norway | NOR | Norway | Case-control | 248 | 371 | 99.5 | 234(135) |
| Nijmegen Ovarian Cancer Study | NTH | Eastern part of the Netherlands | Case-control | 265 | 323 | 98.6 | 255(116) |
| Ovarian Cancer in Alberta and British Columbia | OVA | Alberta and British Columbia, Canada | Case-control | 855 | 810 | 91.5 | 621(344) |

| Study Name | Study abbreviation | Study Location | Study Type | Total Number of Subjects ^b | | % of subjects of European Ancestry | Number of invasive cases of European ancestry |
|---|--------------------|--|--------------------------------|---------------------------------------|---------------|------------------------------------|---|
| | | | | Cases | Controls | | |
| Polish Ovarian Cancer Study | POC | Poland: Szczecin, Poznan, Opole, Rzeszów | Case-control | 423 | 417 | 100 | 423(200) |
| Polish Ovarian Cancer Case Control Study | POL | Poland, Warszaw and Lodz | Population based/ Case-control | 236 | 223 | 100 | 236(106) |
| Study of Epidemiology and Risk Factors in Cancer Heredity | SEA | UK: East Anglia and West Midlands | Population based/ Case-control | 1496 | 6067 | 99.1 | 1372(572) |
| Family Registry for Ovarian Cancer and Genetic Epidemiology of Ovarian Cancer | STA | USA: Six counties in the San Francisco Bay area | Population based/ Case-control | 293 | 404 | 88.4 | 257(160) |
| Shanghai Women's Health Study | SWH | Shanghai, China | Cohort/Nested case-control | 135 | 891 | 0 ^c | - |
| Toronto Ovarian Cancer Study | TOR | Canada: Province of Ontario | Population based | 559 | 443 | 99.5 | 557(339) |
| University of California Irvine Ovarian Study | UCI | USA: Southern California (Orange and San-Diego, Imperial Counties) | Population based/ Case-control | 507 | 425 | 84.2 | 277(166) |
| United Kingdom Ovarian Cancer Population Study | UKO | United Kingdom (England, Wales and Northern Ireland) | Population based/ Case-control | 718 | 1123 | 98.1 | 702(353) |
| Royal Marsden Hospital Ovarian Cancer Study | RMH | UK: London | Hospital based/ Case only | 152 | 0 | 95.4 | 144 (49) |
| UK Familial Ovarian Cancer Registry | UKR | UK: National | Case only/ Familial Register | 48 | 0 | 97.9 | 47(23) |
| Southampton Ovarian Cancer Study | SOC | United Kingdom, Wessex region | Case only/ Hospital based | 295 | 0 | 97.2 | 266(102) |
| Scottish Randomised Trial in Ovarian Cancer | SRO | Coordinated through clinical trials unit, Glasgow UK from patients recruited worldwide | Case only from clinical trial | 159 | 0 | 98.7 | 157(92) |
| Warsaw Ovarian Cancer Study | WOC | Poland: Warsaw and central Poland | Case-control | 204 | 204 | 100 | 202(132) |
| TOTAL | | | | 18,174 | 26,134 | 89.8 | 14,533(8,371) |

a Studies combined into single case controls sets indicated by contiguous shading (AOCS+ACS; DOV+ORE; HOP+GRR; GER+HSK; LAX+USC; MAL+PVD; UKO+RMH+UKR+SOC+SRO)

b Totals represent the number of subjects passing genotyping quality control criteria. The numbers of subjects for these studies included in the GWAS analyses may differ.

c All or the majority of subjects from this site are of Asian ancestry: HAW (59.2%), JPN (100%), MAS (80.3%), SWH (100%)

Supplementary table 2: Number of cases by study and histological subtype

| Study Site | Controls | Cases | | | | | | |
|------------|----------|------------|----------|-----------|----------|--------------|------------|--------|
| | | Behavior | | Histology | | | | |
| | | Borderline | Invasive | Serous | Mucinous | Endometrioid | Clear cell | Other* |
| AUS | 1,011 | 0 | 949 | 592 | 40 | 123 | 64 | 130 |
| BAV | 143 | 5 | 93 | 60 | 9 | 13 | 6 | 10 |
| BEL | 1,352 | 0 | 277 | 195 | 25 | 22 | 23 | 12 |
| DOV | 1,606 | 360 | 990 | 780 | 162 | 174 | 75 | 159 |
| GER | 413 | 24 | 192 | 109 | 30 | 21 | 6 | 50 |
| GRR | 0 | 0 | 115 | 76 | 5 | 19 | 11 | 4 |
| HAW | 601 | 72 | 266 | 166 | 62 | 61 | 35 | 14 |
| HJO | 274 | 13 | 273 | 148 | 14 | 26 | 4 | 94 |
| HMO | 140 | 0 | 144 | 50 | 7 | 12 | 1 | 74 |
| HOC | 447 | 8 | 218 | 120 | 45 | 29 | 14 | 18 |
| HOP | 1,501 | 77 | 682 | 433 | 55 | 91 | 45 | 135 |
| HSK | 0 | 9 | 146 | 118 | 1 | 16 | 0 | 20 |
| JPN | 81 | 10 | 66 | 36 | 9 | 7 | 17 | 7 |
| LAX | 0 | 0 | 330 | 248 | 15 | 26 | 13 | 28 |
| MAL | 829 | 131 | 440 | 331 | 109 | 56 | 33 | 42 |
| MAS | 106 | 0 | 106 | 44 | 17 | 25 | 12 | 8 |
| MAY | 753 | 83 | 708 | 561 | 42 | 99 | 34 | 55 |
| MCC | 68 | 0 | 64 | 34 | 7 | 7 | 6 | 10 |
| MDA | 385 | 0 | 323 | 194 | 29 | 29 | 4 | 67 |
| MSK | 697 | 0 | 556 | 450 | 0 | 25 | 22 | 59 |
| NCO | 984 | 213 | 850 | 631 | 100 | 133 | 86 | 113 |
| NEC | 1,049 | 252 | 697 | 549 | 124 | 133 | 104 | 39 |
| NHS | 429 | 20 | 127 | 78 | 17 | 14 | 6 | 32 |
| NJO | 194 | 0 | 190 | 111 | 7 | 30 | 23 | 19 |
| NOR | 371 | 12 | 236 | 139 | 18 | 27 | 13 | 51 |
| NTH | 323 | 2 | 263 | 120 | 35 | 67 | 21 | 22 |

| Study Site | Controls | Cases | | | | | | | |
|--------------|---------------|--------------|---------------|---------------|-------------|--------------|--------------|--------------|--|
| | | Behavior | | Histology | | | | | |
| | | Borderline | Invasive | Serous | Mucinous | Endometrioid | Clear cell | Other* | |
| ORE | 0 | 11 | 59 | 50 | 5 | 4 | 4 | 7 | |
| OVA | 810 | 167 | 688 | 457 | 107 | 116 | 73 | 102 | |
| POC | 417 | 0 | 423 | 200 | 33 | 39 | 9 | 142 | |
| POL | 223 | 0 | 236 | 106 | 17 | 37 | 10 | 66 | |
| PVD | 0 | 0 | 172 | 130 | 11 | 14 | 8 | 9 | |
| RMH | 0 | 1 | 151 | 52 | 17 | 29 | 17 | 37 | |
| SEA | 6,067 | 101 | 1,395 | 614 | 194 | 232 | 147 | 309 | |
| SOC | 0 | 21 | 274 | 112 | 48 | 64 | 11 | 60 | |
| SRO | 0 | 0 | 159 | 93 | 3 | 17 | 9 | 37 | |
| STA | 404 | 11 | 282 | 174 | 22 | 45 | 23 | 29 | |
| SWH | 891 | 0 | 135 | 0 | 0 | 0 | 0 | 135 | |
| TOR | 443 | 0 | 559 | 341 | 39 | 132 | 34 | 13 | |
| UCI | 425 | 176 | 331 | 308 | 89 | 58 | 29 | 23 | |
| UKO | 1,123 | 0 | 718 | 357 | 76 | 116 | 68 | 101 | |
| UKR | 0 | 0 | 48 | 23 | 3 | 6 | 2 | 14 | |
| USC | 1,370 | 282 | 978 | 793 | 163 | 124 | 58 | 122 | |
| WOC | 204 | 2 | 202 | 132 | 10 | 20 | 17 | 25 | |
| Total | 26,134 | 2,063 | 16,112 | 10,316 | 1821 | 2,338 | 1,197 | 2,503 | |

Supplementary Table 3: Characteristics of six previously reported and three novel ovarian cancer susceptibility SNPs

| SNP | Region | Risk allele frequency | All invasive | | | | Serous invasive | | | |
|---------------------------------|--------|-----------------------|--------------|-----------|-----------------------|-----------|-----------------|-----------|-----------------------|-----------|
| | | | OR | 95% CI | P-value | Variance* | OR | 95% CI | P-value | Variance* |
| Previously reported loci | | | | | | | | | | |
| rs3814113 | 9p22 | 0.68 | 1.21 | 1.17-1.25 | 3.8×10^{-29} | 0.016 | 1.28 | 1.23-1.33 | 3.9×10^{-32} | 0.027 |
| rs10088218 | 8q24 | 0.87 | 1.18 | 1.13-1.24 | 2.9×10^{-12} | 0.0062 | 1.29 | 1.21-1.36 | 1.2×10^{-17} | 0.015 |
| rs2072590 | 2q31 | 0.32 | 1.11 | 1.08-1.15 | 5.4×10^{11} | 0.0047 | 1.13 | 1.09-1.18 | 2.5×10^{-10} | 0.0065 |
| rs7651446 | 3q25 | 0.05 | 1.44 | 1.35-1.54 | 1.6×10^{-29} | 0.013 | 1.59 | 1.48-1.71 | 1.5×10^{-34} | 0.020 |
| rs8170 | 19p13 | 0.19 | 1.11 | 1.07-1.15 | 1.5×10^{-7} | 0.0034 | 1.19 | 1.14-1.25 | 2.9×10^{-14} | 0.0093 |
| rs9303542 | 17q21 | 0.27 | 1.12 | 1.08-1.16 | 6.0×10^{-9} | 0.0051 | 1.14 | 1.09-1.18 | 3.0×10^{-10} | 0.0068 |
| Novel loci | | | | | | | | | | |
| rs11782652 | 8q21 | 0.07 | 1.19 | 1.12-1.26 | 5.5×10^{-9} | 0.0039 | 1.24 | 1.16-1.33 | 7.0×10^{-10} | 0.0069 |
| rs1243180 | 10p12 | 0.31 | 1.10 | 1.06-1.13 | 1.8×10^{-8} | 0.0039 | 1.11 | 1.07-1.15 | 1.4×10^{-7} | 0.0047 |
| rs757210 | 17q12 | 0.37 | 1.05 | 1.02-1.09 | 0.00090 | 0.0011 | 1.12 | 1.08-1.17 | 8.1×10^{-10} | 0.0060 |

* Variance of distribution of the natural logarithm of the odds ratio = $2.\ln(\text{OR})^2.p.(1-p)$, where p = risk allele frequency

Supplementary Table 4: Summary of methylation data in tumor and normal tissues

| Gene | High grade serous (n=106) | | | | Normal (n=7) | | | | Methylation (mean beta), HGS Cases (N=227) | | | |
|------------------|-----------------------------------|--------------------------|------------------------------|------|--------------|------|---------|------------------------------|--|--------------|------------------------|--------------|
| | Methylation Probe ID ^a | Correlation ^b | P-value | Mean | Std Dev | Mean | Std Dev | P-value | Common allele homozygote | Heterozygote | Rare allele homozygote | P-trend |
| 8q21 | | | | | | | | | n=187 | n=37 | n=3 | |
| <i>FABP5</i> | cg10563714 | -0.38 | 0.012 | 0.22 | 0.12 | 0.32 | 0.13 | 0.004 | 0.19 | 0.19 | 0.14 | 0.95 |
| <i>PMP2</i> | cg11537619 | -0.73 | 3.4 x 10⁻⁸ | 0.72 | 0.11 | 0.79 | 0.04 | 0.03 | 0.73 | 0.73 | 0.7 | 0.42 |
| <i>FABP9</i> | | | | | | | | | | | | |
| <i>FABP4</i> | cg19422565 | -0.10 | 0.53 | 0.64 | 0.14 | 0.30 | 0.07 | 4.1 x 10⁻⁵ | 0.62 | 0.63 | 0.67 | 0.7 |
| <i>FABP12</i> | | | | | | | | | | | | |
| <i>IMPA1</i> | cg04491064 | -0.24 | 0.12 | 0.76 | 0.1 | 0.82 | 0.03 | 0.24 | 0.73 | 0.75 | 0.81 | 0.87 |
| <i>SLC10A5</i> | | | | | | | | | | | | |
| <i>ZFAND1</i> | cg07583137 | -0.30 | 0.053 | 0.44 | 0.15 | 0.88 | 0.03 | 1.4 x 10⁻⁵ | 0.39 | 0.47 | 0.32 | 0.03 |
| <i>CHMP4C</i> | cg00450487 | -0.32 | 0.036 | 0.25 | 0.06 | 0.24 | 0.07 | 0.96 | 0.19 | 0.24 | 0.28 | 0.001 |
| <i>SNX16</i> | cg23324259 | -0.27 | 0.086 | 0.06 | 0.02 | 0.08 | 0.01 | 0.01 | 0.06 | 0.07 | 0.07 | 0.001 |
| 10p12 | | | | | | | | | n=115 | n=92 | n=20 | |
| <i>NEBL</i> | cg00833352 | -0.39 | 0.01 | 0.19 | 0.06 | 0.19 | 0.04 | 0.61 | 0.17 | 0.18 | 0.17 | 0.12 |
| <i>C10orf113</i> | | | | | | | | | | | | |
| <i>C10orf114</i> | cg06613783 | -0.40 | 0.009 | 0.87 | 0.08 | 0.90 | 0.01 | 0.12 | 0.45 | 0.5 | 0.56 | 0.027 |
| <i>C10orf140</i> | cg24324815 | -0.47 | 0.0016 | 0.48 | 0.25 | 0.11 | 0.07 | 1.9 x 10⁻⁵ | 0.86 | 0.85 | 0.82 | 0.1 |
| <i>MLLT10</i> | cg10900703 | -0.26 | 0.092 | 0.3 | 0.13 | 0.22 | 0.06 | 0.18 | 0.28 | 0.27 | 0.2 | 0.05 |
| <i>DNAJC1</i> | cg03076324 | -0.46 | 0.002 | 0.02 | 0.01 | 0.02 | 0.01 | 0.58 | 0.02 | 0.02 | 0.02 | 0.05 |
| 17q12 | | | | | | | | | n=73 | n=111 | n=43 | |
| <i>ACACA</i> | cg01359274 | -0.42 | 0.006 | 0.88 | 0.07 | 0.90 | 0.04 | 0.57 | 0.86 | 0.88 | 0.88 | 0.7 |
| <i>C17orf78</i> | cg01156990 | -0.31 | 0.04 | 0.91 | 0.02 | 0.91 | 0.01 | 0.42 | 0.9 | 0.9 | 0.9 | 0.78 |
| <i>TADA2A</i> | cg06899192 | -0.30 | 0.05 | 0.05 | 0.01 | 0.04 | 0.01 | 0.06 | 0.04 | 0.04 | 0.04 | 0.67 |
| <i>DUSP14</i> | cg06191580 | -0.28 | 0.07 | 0.91 | 0.06 | 0.94 | 0.01 | 0.29 | 0.91 | 0.91 | 0.91 | 0.95 |
| <i>SYNRG</i> | cg07306685 | -0.62 | 1.5 x 10⁻⁵ | 0.92 | 0.01 | 0.92 | 0.01 | 0.22 | 0.91 | 0.92 | 0.92 | 0.4 |
| <i>DDX52</i> | cg07306685 | -0.50 | 0.001 | 0.92 | 0.01 | 0.92 | 0.01 | 0.22 | 0.91 | 0.92 | 0.92 | 0.4 |
| <i>HNF1B</i> | cg14487292 | -0.67 | 2.1 x 10⁻⁶ | 0.41 | 0.26 | 0.09 | 0.04 | 0.003 | 0.35 | 0.42 | 0.47 | 0.009 |
| <i>TBC1D3F</i> | | | | | | | | | | | | |
| <i>TBC1D3</i> | | | | | | | | | | | | |
| <i>MRPL45</i> | cg17519037 | -0.40 | 0.008 | 0.82 | 0.08 | 0.78 | 0.05 | 0.01 | 0.73 | 0.74 | 0.74 | 0.77 |
| <i>GPR179</i> | | | | | | | | | | | | |
| <i>SOCS7</i> | cg02937763 | -0.22 | 0.15 | 0.67 | 0.12 | 0.80 | 0.05 | 0.006 | 0.62 | 0.63 | 0.68 | 0.1 |
| <i>ARHGAP23</i> | | | | | | | | | | | | |

a CpG with the most significant negative correlations between CpG and RNA expression; build 37

b N=43; adjusted for age and histology

Supplementary Table 5: Summary of gene expression data in tumor and normal tissue and cell lines

| Gene | RNAseq ^a | | Relative expression ^b | | | | | Gene expression in tumor v normal ^d | | | Expression QTL ^e | | | Somatic mutation | |
|------------------|--------------------------|---------|----------------------------------|------------------|-----------------------|----------------------------------|-------------|--|------------------|--------------|-----------------------------|-----------------------------|------------|------------------|--|
| | Normal | Cancer | Normal | All EOC v normal | All EOC v normal | Serous EOC v normal ^c | Cell lines | TCGA ^e | MDA ^f | OSE/FTSE | LCL | TCGA tumors | OvCa | Any cancer | |
| | Average normalised reads | | | fold change | P-value | P-value | | direction | direction | direction | P-value | P-value | P-value | | |
| 8q21 | | | | | | | | | | | | | | | |
| <i>FABP5</i> | 12.9 | 32.1 | 12.9 | 1.5 | 0.002 | 0.002 | Up | Up | Inconsistent | 0.52 | 0.037 | 0.93 | No | No | |
| <i>PMP2</i> | no reads | N/A | N/A | N/A | N/A | N/A | N/A | Down | Up | N/A | N/A | 0.93 | No | No | |
| <i>FABP9</i> | no reads | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | No | No | |
| <i>FABP4</i> | 0.00 | 0.030 | 0.026 | 0.02 | 4.6x10 ⁻⁷ | 3.1x10 ⁻⁴ | Down | Down | Down | 0.49 | 0.9 | 0.13 | No | No | |
| <i>FABP12</i> | no reads | 0.00045 | 0.00015 | 8.9 | 9.7x10 ⁻⁵ | 0.004 | Up | N/A | N/A | 0.77 | 0.45 | N/A | No | No | |
| <i>IMPA1</i> | 3.8 | 6.4 | 4.8 | 1.5 | 0.486 | 0.6 | Up | Down | Down | 0.77 | 0.98 | 0.49 | No | Yes | |
| <i>SLC10A5</i> | no reads | 0.15 | 0.0097 | 17.7 | 2.2x10 ⁻¹⁶ | 5.5x10 ⁻¹⁴ | Up | N/A | N/A | 0.45 | 0.65 | 0.93 | No | No | |
| <i>ZFAND1</i> | 2.0 | 1.5 | 0.55 | 2.6 | 1.6x10 ⁻⁶ | 3.6x10 ⁻⁶ | Up | Same | Down | 0.42 | 0.32 | 0.52 | Yes | Yes | |
| <i>CHMP4C</i> | 1.2 | 2.9 | 0.31 | 9.1 | 4.8x10 ⁻⁹ | 4.7x10 ⁻¹⁴ | Up | N/A | Up | 0.5 | 0.012 | 3.9x10⁻¹⁴ | Yes | Yes | |
| <i>SNX16</i> | 0.94 | 1.3 | 0.38 | 3.9 | 5.9x10 ⁻⁴ | 1.7 x10 ⁻⁴ | Up | Up | Same | 0.29 | 0.088 | 0.53 | Yes | Yes | |
| 10p12 | | | | | | | | | | | | | | | |
| <i>NEBL</i> | 2.2 | 0.059 | 0.042 | 0.3 | 3.3x10 ⁻⁴ | 0.005 | Down | Down | Inconsistent | 0.043 | 0.5 | 0.19 | Yes | Yes | |
| <i>C10orf113</i> | no reads | 0.18 | 0.050 | 1.1 | 0.905 | 0.464 | Same | N/A | N/A | 0.38 | 0.58 | N/A | No | Yes | |
| <i>C10orf114</i> | 1.1 | 0.22 | 0.028 | 3.1 | 0.001 | 0.0148 | Up | N/A | Up | 0.027 | 0.79 | 0.02 | No | No | |
| <i>C10orf140</i> | no reads | 1.1 | 0.050 | 6.7 | 5.2x10 ⁻⁷ | 3.0 x10 ⁻⁴ | Up | N/A | Same | 0.46 | 0.96 | 0.02 | No | No | |
| <i>MLLT10</i> | 1.1 | 9.0 | 3.6 | 2.1 | 1.5x10 ⁻⁶ | 0.002 | Up | Up | Inconsistent | 0.13 | 0.79 | 0.13 | No | Yes | |
| <i>DNAJC1</i> | 14 | 1.4 | 2.5 | 0.6 | 3.4x10 ⁻⁶ | 3.9x10 ⁻⁵ | Down | Down | Down | 0.11 | 0.62 | 0.8 | No | Yes | |
| 17q12 | | | | | | | | | | | | | | | |
| <i>ACACA</i> | 1.4 | 25.6 | 4.4 | 4.4 | 5.3x10 ⁻¹⁴ | 8.3x10 ⁻¹¹ | Up | Same | Up | 0.77 | 0.33 | 0.32 | No | No | |
| <i>C17orf78</i> | no reads | N/A | N/A | N/A | N/A | N/A | N/A | N/A | Up | N/A | 0.074 | 0.86 | No | No | |
| <i>TADA2A</i> | 0.94 | 2.6 | 0.77 | 3.2 | 8.9x10 ⁻¹¹ | 4.4X10 ⁻⁶ | Up | Same | Inconsistent | 0.9 | 0.57 | N/A | No | Yes | |
| <i>DUSP14</i> | 21.7 | 32.2 | 11.3 | 2.1 | 2.2x10 ⁻⁴ | 0.021 | Up | Down | Same | 0.72 | 0.61 | 0.97 | No | No | |
| <i>SYNRG</i> | 0.86 | 6.3 | 2.3 | 2.0 | 7.9x10 ⁻⁴ | 0.077 | Up | N/A | Inconsistent | 0.5 | 0.34 | N/A | Yes | Yes | |
| <i>DDX52</i> | 7.0 | 36.5 | 5.9 | 3.0 | 3.6.x10 ⁻⁵ | 0.006 | Up | Same | Same | 0.98 | 0.7 | 0.68 | No | Yes | |
| <i>HNF1B</i> | no reads | 17.4 | 0.0011 | 2290.0 | 0.001 | 0.18 | Up | Down | Up | 0.36 | 0.77 | 0.69 | Yes | Yes | |
| <i>TBC1D3F</i> | no reads | 0.0032 | 0.00085 | 7.4 | 5.4x10 ⁻⁴ | 0.002 | Up | Same | Down | 0.97 | 0.8 | N/A | No | No | |
| <i>TBC1D3</i> | 0.40 | 3.0 | 0.81 | 2.1 | 0.003 | 0.41 | Up | Same | Down | 0.23 | 0.17 | N/A | No | Yes | |

| Gene | RNAseq ^a | | Relative expression ^b | | | | | Gene expression in tumor v normal ^d | | | Expression QTL ^g | | | Somatic mutation | |
|----------|--------------------------|--------|----------------------------------|------------------|-----------------------|----------------------------------|------------|--|------------------|----------|-----------------------------|-------------|------|------------------|--|
| | Normal | Cancer | Normal | All EOC v normal | All EOC v normal | Serous EOC v normal ^c | Cell lines | TCGA ^e | MDA ^f | OSE/FTSE | LCL | TCGA tumors | OvCa | Any cancer | |
| | Average normalised reads | | | fold change | P-value | P-value | direction | direction | direction | P-value | P-value | P-value | | | |
| MRPL45 | 4.9 | 26.9 | 7.0 | 3.0 | 8.8x10 ⁻¹⁰ | 1.2x10 ⁻⁵ | Up | N/A | Down | 0.73 | 0.38 | 0.91 | No | No | |
| GPR179 | no reads | 0.012 | 0.00080 | 1.6 | 1.7x10 ⁻⁷ | 4.5 x10 ⁻⁴ | Up | N/A | N/A | 0.47 | 0.63 | N/A | Yes | Yes | |
| SOCS7 | 1.1 | 2.2 | 0.73 | 2.3 | 1.8x10 ⁻⁷ | 0.002 | Up | Same | Inconsistent | 0.79 | 0.23 | 0.49 | No | No | |
| ARHGAP23 | 1.9 | 2.3 | 2.4 | 0.5 | 6.4x10 ⁻⁴ | 0.014 | Down | N/A | Inconsistent | 0.93 | 0.62 | 0.85 | No | Yes | |

a average of two OSEC lines

b Gene expression relative to β-actin and GAPDH n 50 EOC cell lines and 73 normal OSEC/FTSEC lines

c enriched for serous (removed known clear cell and mucinous cell lines)

d Up or down regulation in tumor samples, red bold text indicates significant differences in one or more probes and remaining probes in same direction, black text indicates if there is a non significant trend in direction for all probes for gene

e TCGA data: gene expression in 568 serous EOC and 8 fallopian tube (normal) samples

f MDA data: gene expression in 53serous EOC and 10 ovary (normal) samples

g 8q21 rs11782652 OSE/FTSE AA=52, AG/GG=11, LCL AA=84, AG/GG=10, TCGA AA=322 AG/GG=72

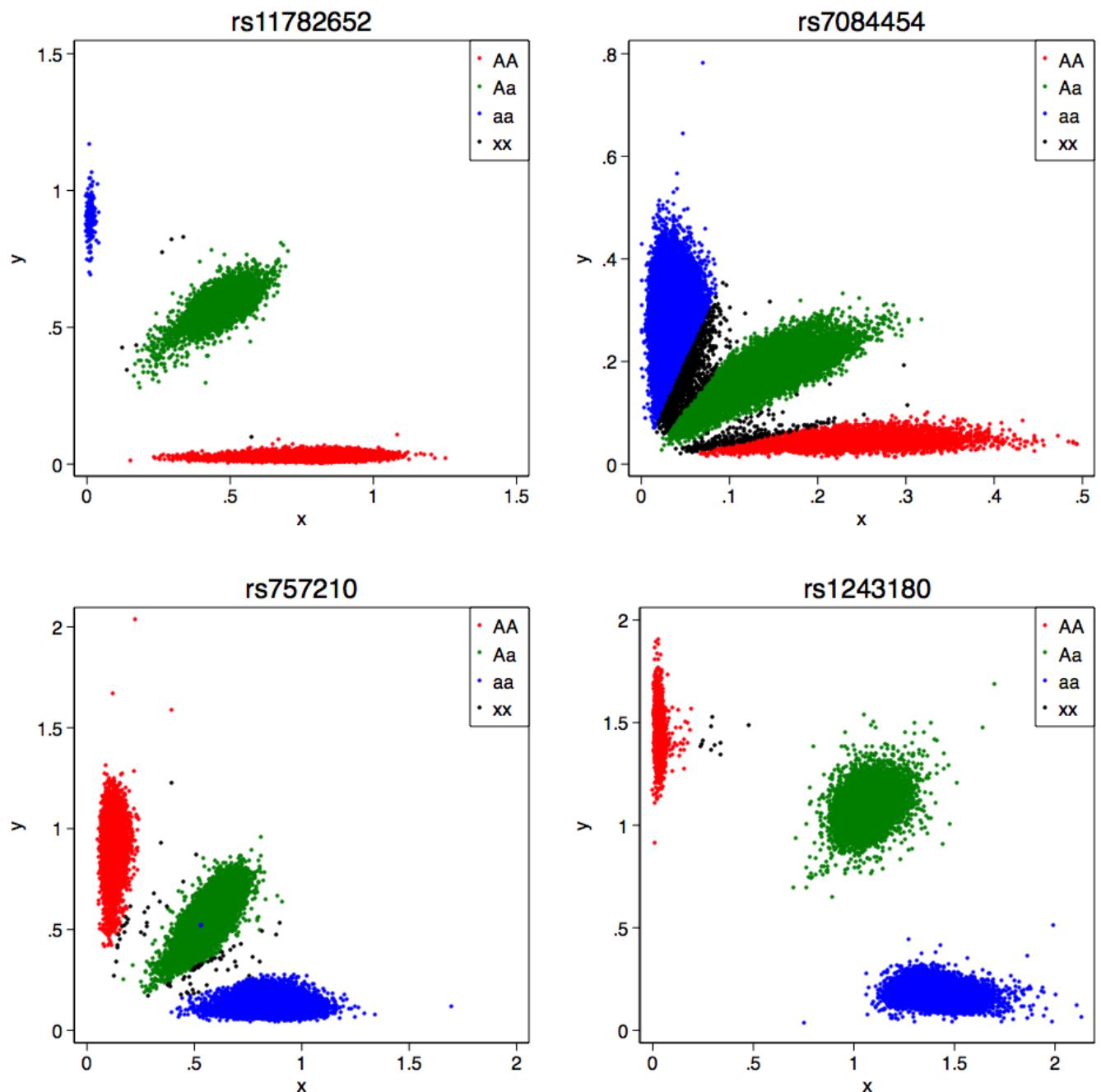
g 10p12 rs1243180 OSE/FTSE TT=37, AT/TT=25, LCL TT=39, AT/AA=55, rs7098100 with r2=0.86 in TCGA GG=156 AG/AA=238

g 17cen-q21 rs757210 OSE/FTSE GG=32, AG/GG=30, LCL GG=35, AG/AA=59, TCGA GG=134 AG/GG=260

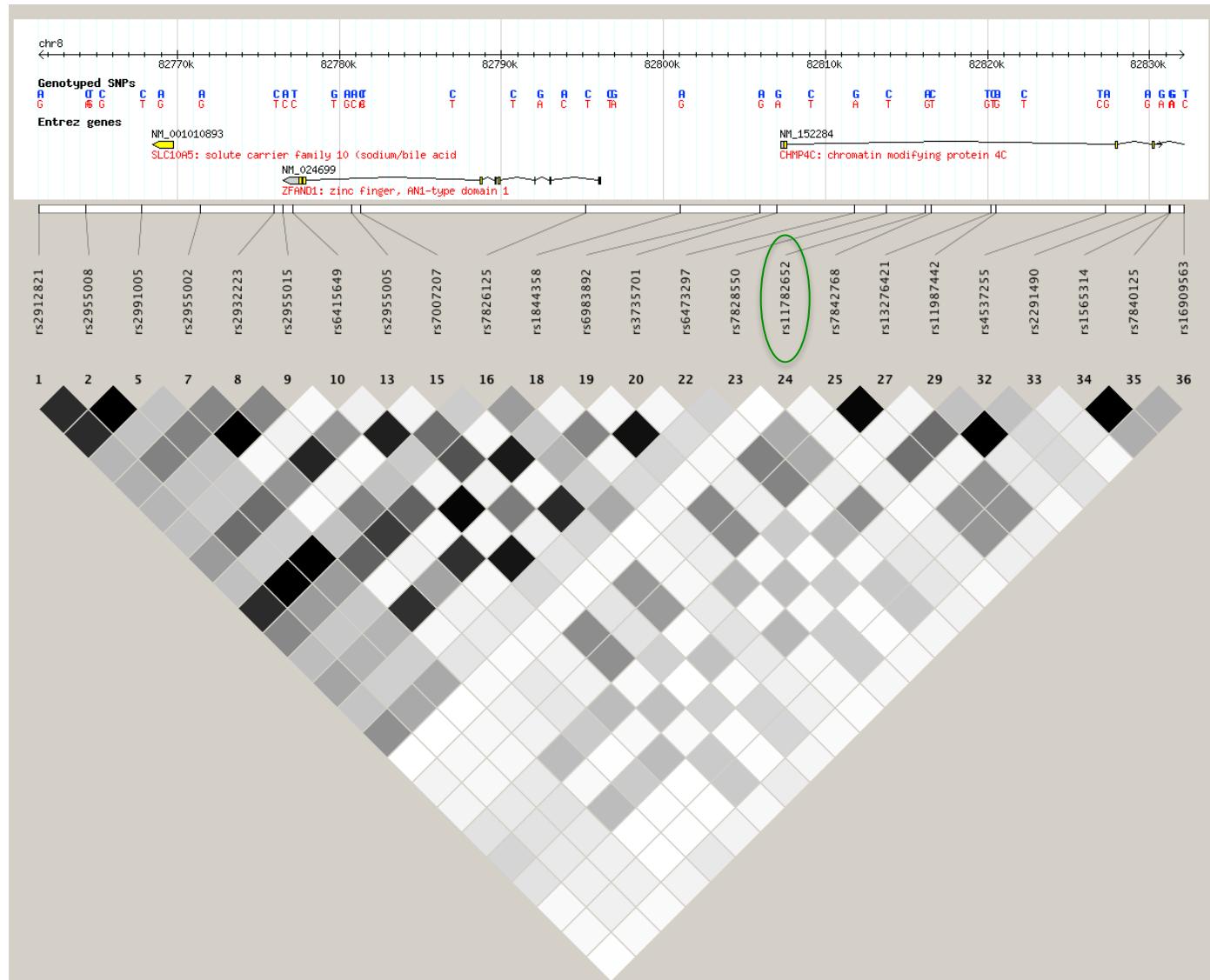
Supplementary Table 6: Primer sequences used for analysis of regulatory regions

| Primer | Forward primer sequence | Reverse primer sequence |
|-----------------------------|---|---|
| chr8:82798344 | AGCCTCCATAGCTTTCTT | CCATGATGATGCTGTGAAC |
| chr8:82800064 | CTCCATCATGAGAGGAACAT | TTTGCTTGACACCCTTATT |
| chr8:82802020 | AAATAAGGGTGTCAAGCAA | ATAACTCCAGGCTTCCTTT |
| chr8:82803822 | ATCTGGGTGAGAGATGACAG | AGTGGCCTGAAAAGTGATA |
| chr8:82805644 | GTAAGGCTCCAAAATGAATG | GACTTCTGGAGTGTGTTGGT |
| chr8:82807123 | GGGTAACGTTGCTCAAATAC | AAATCCACGATGAACAAATC |
| chr8:82809053 | GATTGTTCATCGTGGATT | AACCAGCCTTCTTACTCC |
| chr8:82810904 | GGAGTAAGAAGAGGCTGGTT | GATGATGGTAACTGCCAGAT |
| chr8:82812621 | ATCTGGCAGTTACCATCATC | GTTATATCTCCCAGGGGAAC |
| chr8:82814551 | TACTCAGATCACCTGCATCA | GAACCCCAGGTGTGAAATA |
| chr8:82816369 | TTCCTAACAGGACATGAACC | GGCCTGTTAACAAACGTTAC |
| chr8:82818090 | TTACAGGCAAGGAAAAATAGG | TGGGAACTATAAGCCTAAA |
| chr8:82819912 | TCTGCAATCTCATTCTTT | GGATCAGGTCTAGGATACC |
| chr8:82821909 | GCCTCCTGTCTTAATGTTG | TCTTTTCTCAGCATGGAAT |
| chr8:82823586 | CACACGTTGGTAGGATAGGT | TCCAATATCCAACAAGTTCC |
| chr8:82825197 | ATGCTGCTCTTTCTCAA | TCTTAACAATTGCCCTAAGC |
| chr8:82827011 | GCTTAGGGCAATTGTTAAGA | GAGTGAGCTGTTCTCGAAC |
| chr8:82828794 | GTTCGAGAACAGCTCACTC | TGAAGCTAACACTGGGTCT |
| chr8:82830672 | AGACCCAGTGTAGCTTCA | CAAAACAATCAGGGACAAAT |
| chr8:82832568 | GATGAAAATGATGGAGGAGA | CAAGTAGCTGGGATTACAGG |
| chr8:82834567 | CCTGTAATCCCAGCTACTTG | GTTATGAATGAGCTCCCTTG |
| chr8:82836058 | TTTCAGACATTGGACAACA | TTCACATACCAAGGGAGTTC |
| chr8:82837825 | ACACATACCTAGCCGGTTA | TTGCTCAGATTCTGTGTTG |
| KPNI M13 Fwd / SAC1 M13 Rev | 5'ATGGTACCGTAAACGACGCCAG | 5'ATGAGCTCCAGGAAACAGCTATGAC |
| 2KB TEST GC enh | 5'GGGGACAAGTTGTACAAAAAAGCAGGCTCTGAACGATGGAGCGGAGAATGG | 5'GGGGACCACTTGTACAAGAAAGCTGGGTCTACCGAAGAACGTTTCAATGAG |

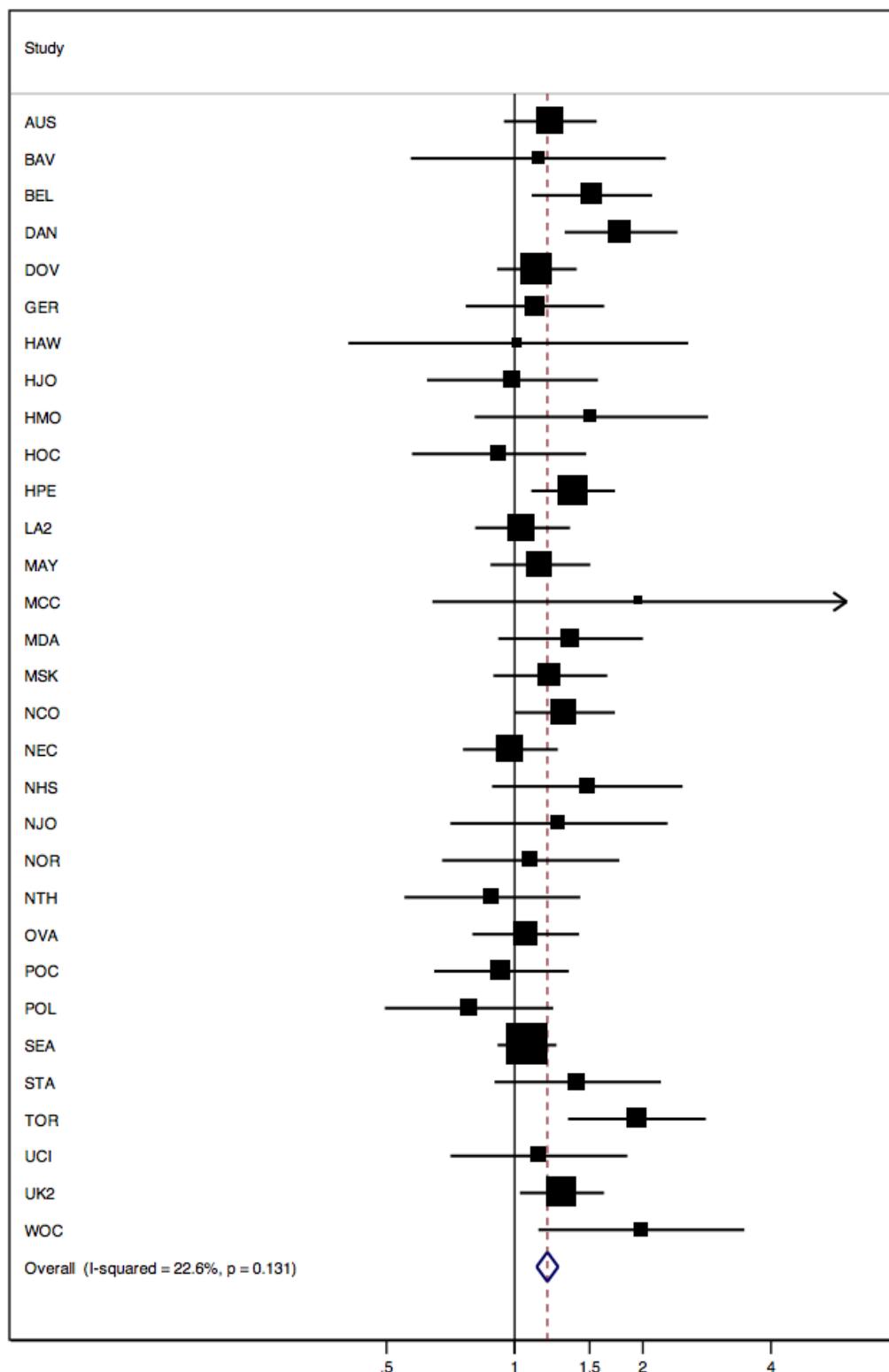
Supplementary Figure 1: Signal intensity cluster plots for most strongly associated SNPs at 8q21 (rs11782652), 10p12 (rs7084454 and rs1243180) and 17q12 (rs757210)



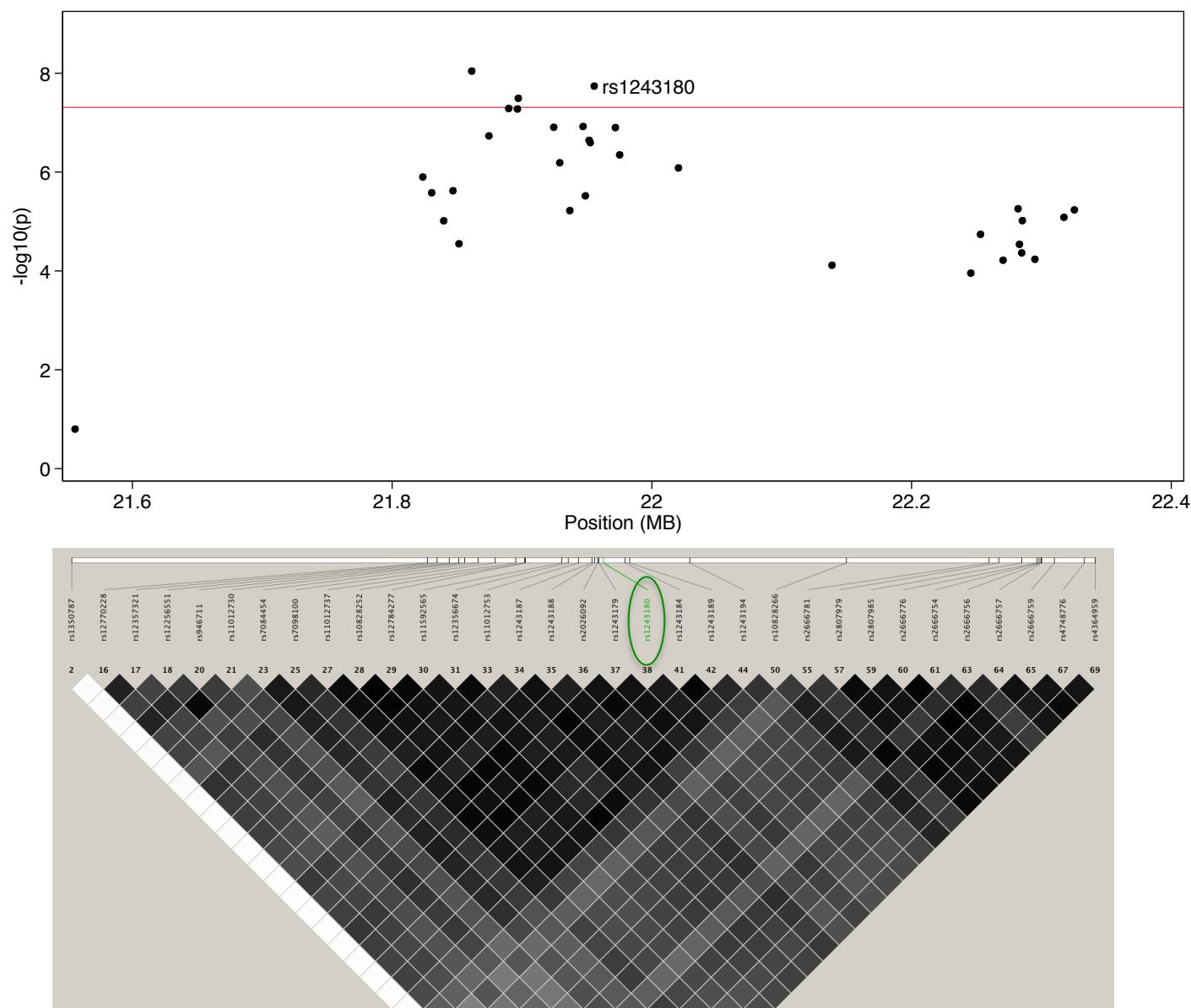
Supplementary Figure 2: LD plot for 8q21 region based on HapMap data for SNPs in 1MB region centred on rs11782652



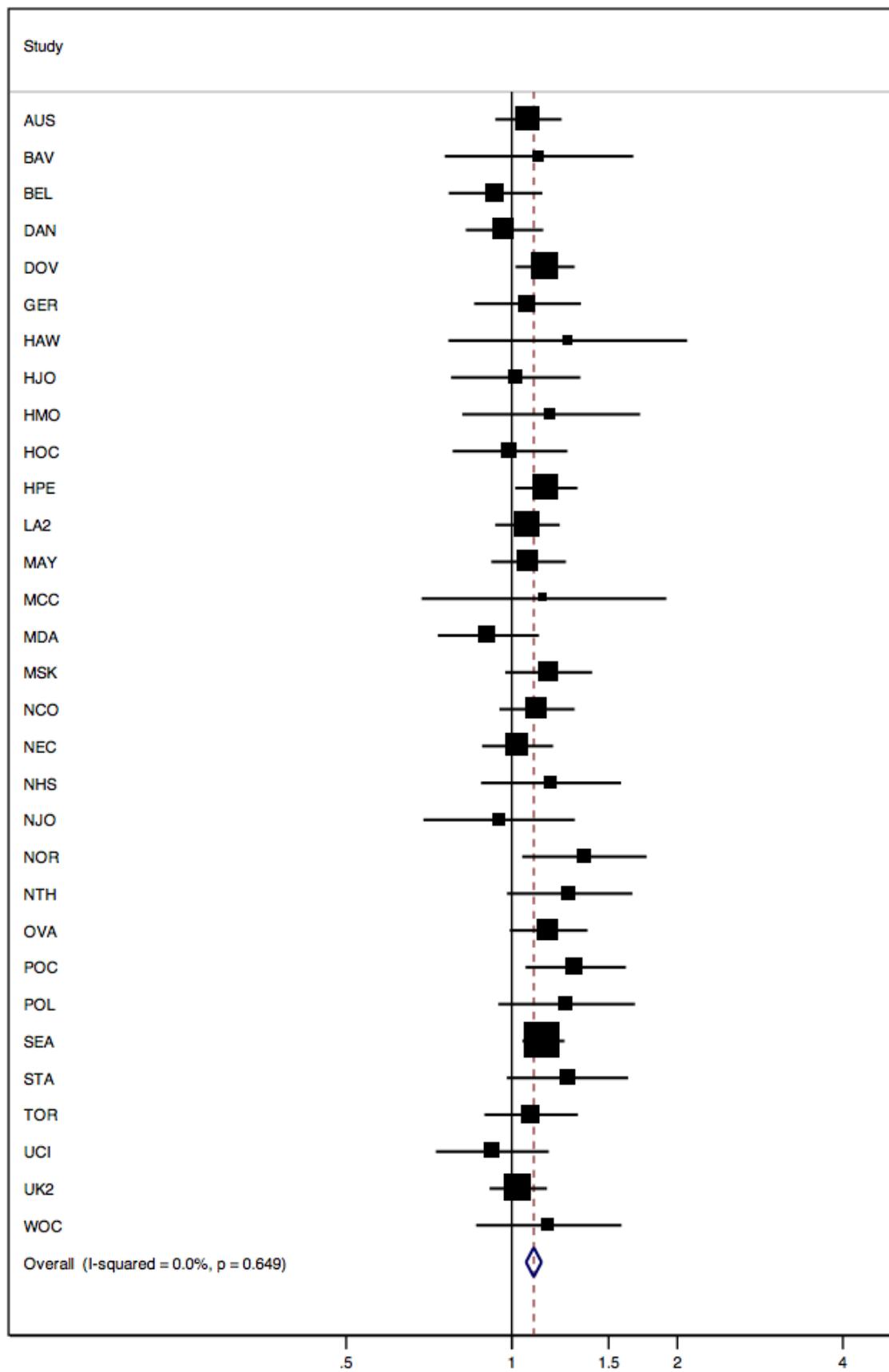
Supplementary Figure 3: Study specific odds ratios for association of rs11782652 with all invasive EOC



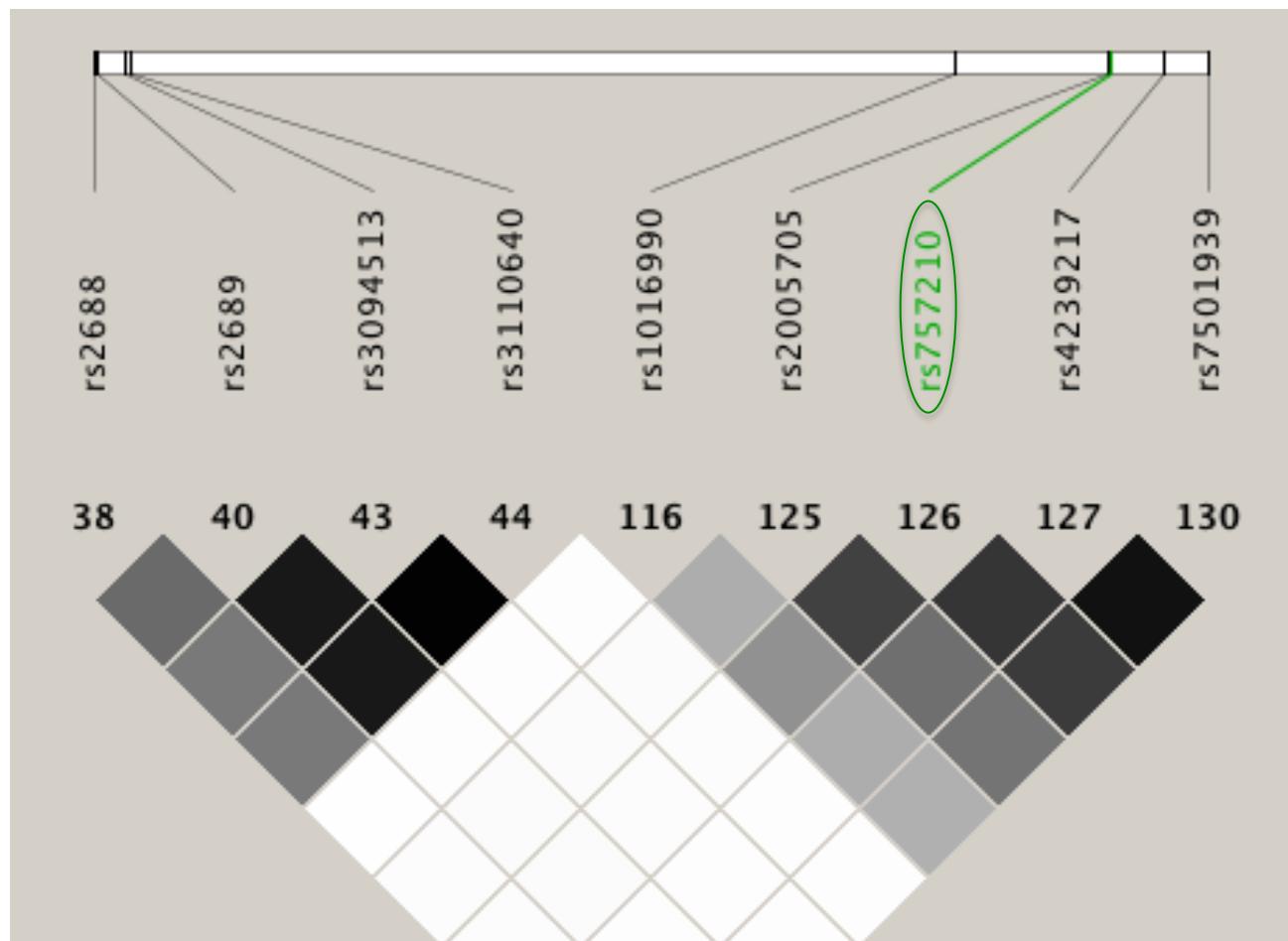
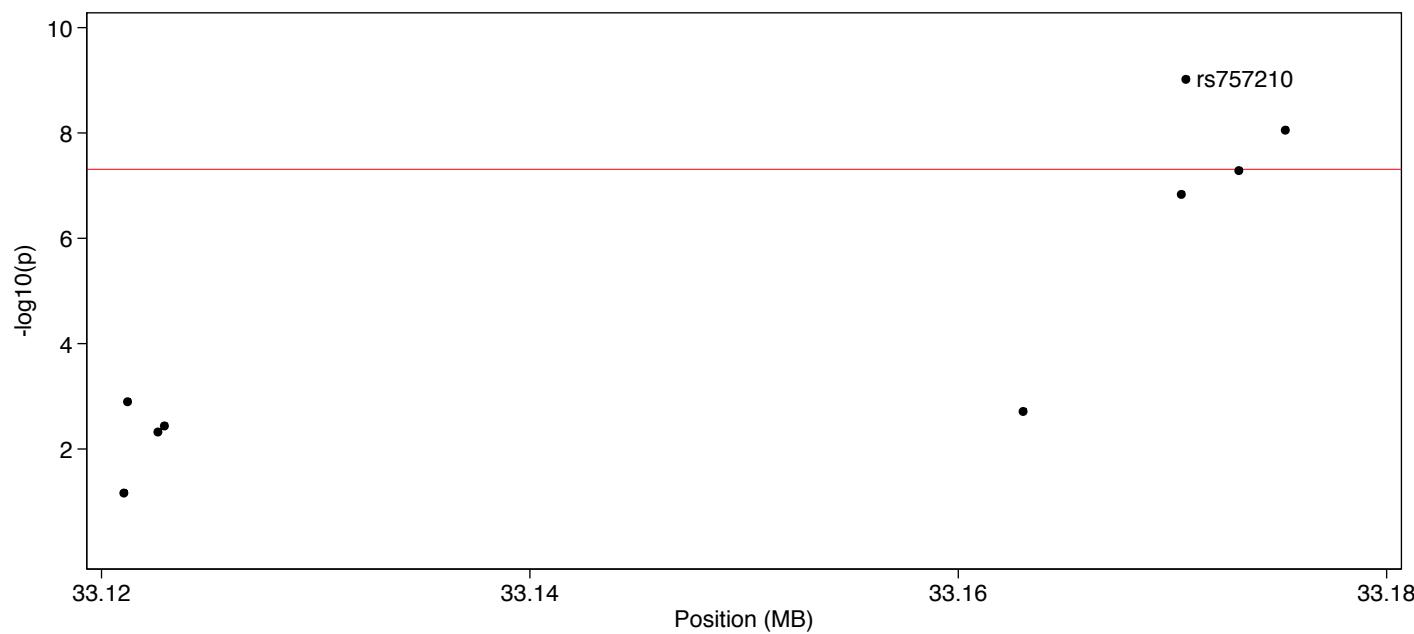
Supplementary Figure 4: Regional LD and association plots for genotyped SNPs in 1MB region centred on rs1243180



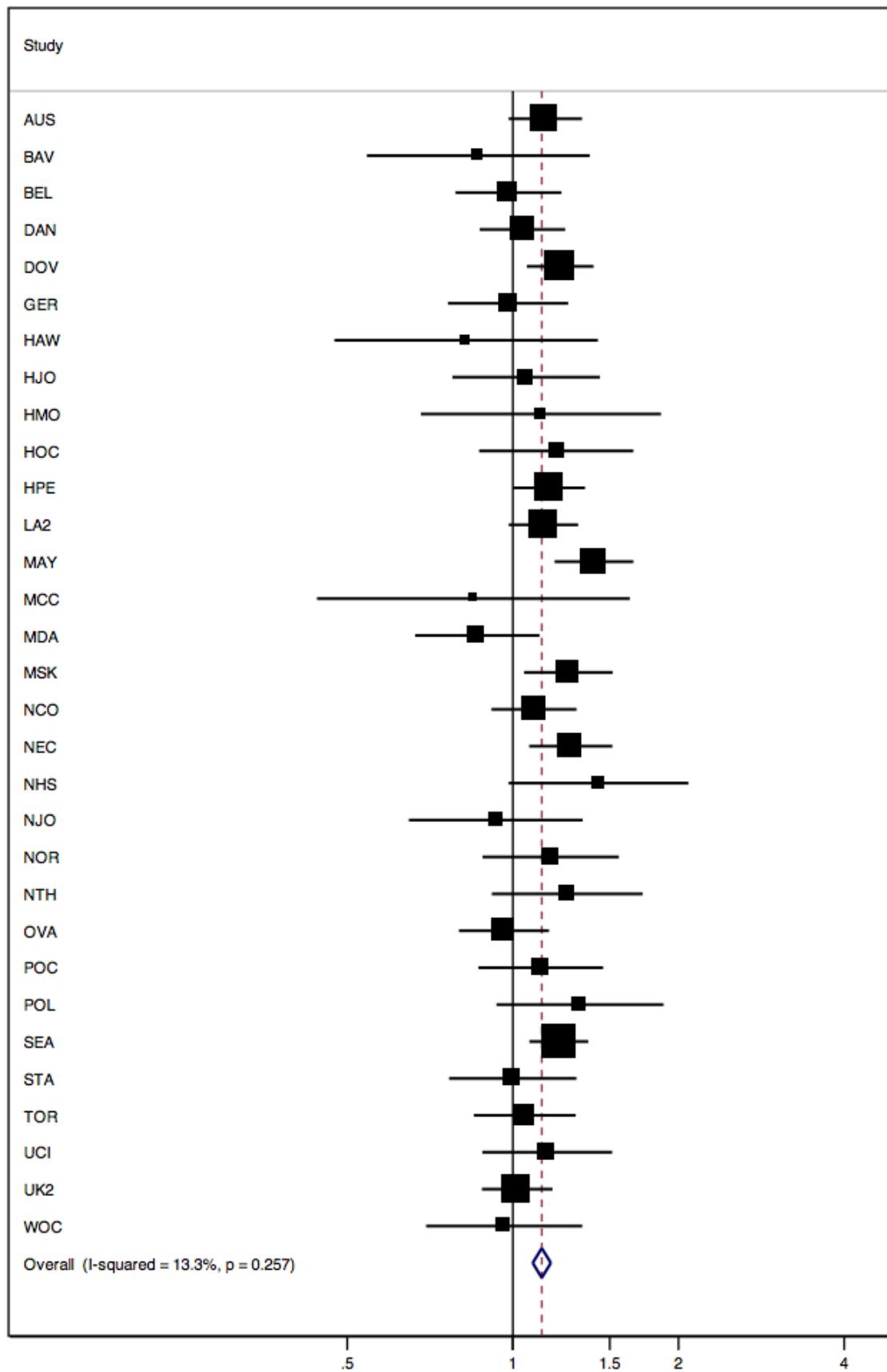
Supplementary Figure 5: Study specific odds ratios for association of rs1243180 with all invasive EOC



Supplementary Figure 6: Regional LD and association plots for genotyped SNPs in 1MB region centred on rs757210



Supplementary Figure 7: Study specific odds ratios for association of rs757210 with serous EOC



Supplementary figure 8i: Gene expression and methylation analyses of all protein coding genes within a one megabase region centered around rs11782652 at 8q21; (ii) 10p12, centered around rs1243180; (iii) 17q12, centered around rs757210

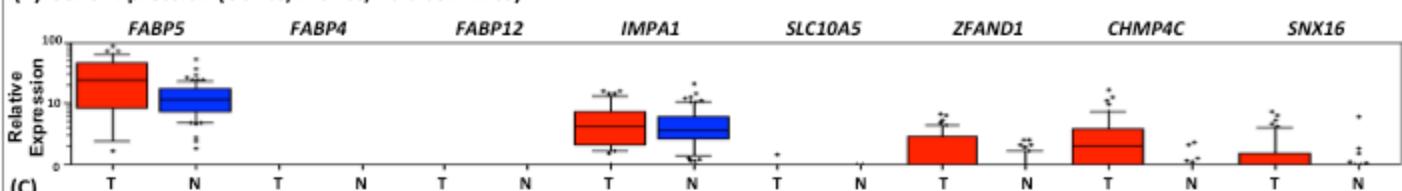
(A) Genomic map of a one-megabase region centred on the most statistically significant SNP at each locus, relative to the chromosomal location at each locus. The location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region are shown relative to the most significant SNP (red dashed line) at each locus. For 10q12 data for *PMP2* and *FABP9* are not included as no expression was detected; additional data for these genes are in Supplementary Tables 3 and 4. **(B)** Boxplots of semi-quantitative PCR based relative gene expression (normalized to β -actin and GAPDH) in epithelial ovarian cancer (EOC) cell lines ($n=50$) and normal ovarian surface epithelial cell (OSEC) lines and fallopian tube secretory epithelial cell (FTSEC) lines ($n=73$). **(C)** Boxplots of semi-quantitative PCR based gene expression in EOC (T) and normal (N; OSEC/FTSEC) cell lines. These are the same data as in panel B rescaled to emphasise the contrast between tumour and normal. **(D)** Boxplots of TCGA array based gene expression data in primary high-grade serous ovarian tumors (T; $n=568$) and normal fallopian tube tissues (N; $n=8$) for all genes for which probe sets were available. For genes not represented in TCGA, expression data from the GEO dataset of EOC tumors (T; $n=53$) and normal ovarian tissue (N; $n=10$) are shown (indicated by a red asterisk). **(E)** Gene expression by DNA copy number alteration in 316 high-grade serous ovarian tumours based on data from TCGA. Graphs show mRNA expression Z-scores by putative copy number alterations: homozygous deletion (0), heterozygous loss (1), diploid (2), copy number gain (3), and amplification (4). **(F)** Boxplots of methylation of 227 high-grade serous ovarian tumours compared to normal ovarian tissues ($n=7$). **(G)** Boxplots of gene methylation by genotype for the most strongly associated SNP at each locus in 227 high-grade serous ovarian tumors. Probes that show significant cis negative correlations with gene expression are indicated by a red open circle. **(H)** Boxplot of qPCR based gene expression by genotype for the most strongly associated SNP for 73 normal OSE/FTSE cell lines. **(I)** Boxplot of array based gene expression by germline genotype in high-grade serous tumors based on TCGA data. * Adjusted $P<0.05$, ** $P<0.01$, *** $P<0.001$.

Supplementary figure 8.i

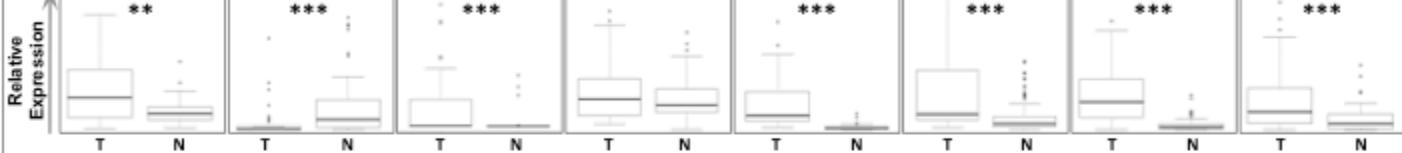
(A) Genomic Map and LD Structure



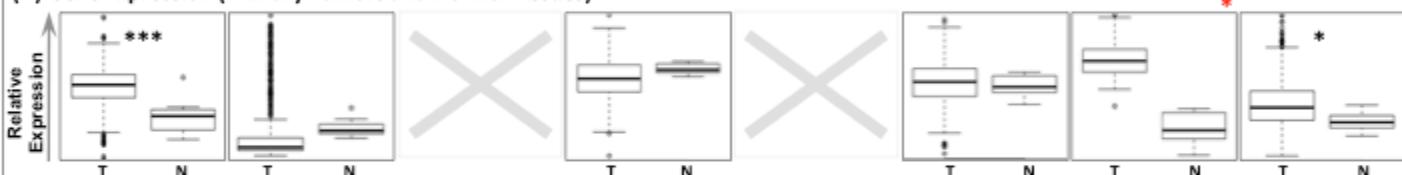
(B) Gene Expression (OSECs/FTSECs/EOC Cell Lines)



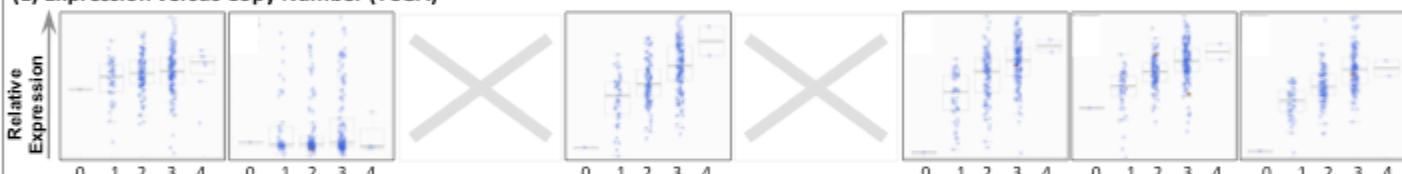
(c)



(D) Gene Expression (Primary Tumors and Normal Tissues)



(E) Expression versus Copy Number (TCGA)



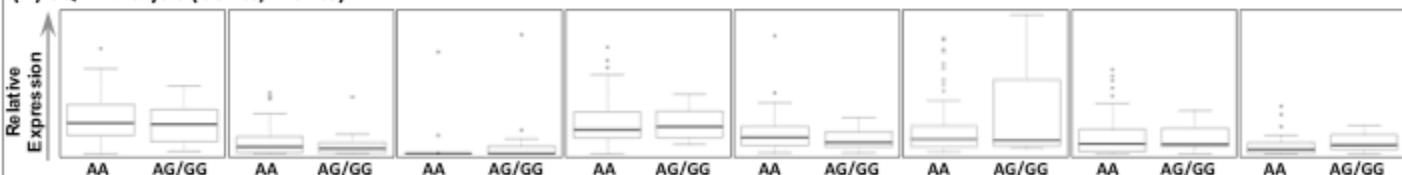
(E) Methylation (Primary EOCs and Normal Tissues)



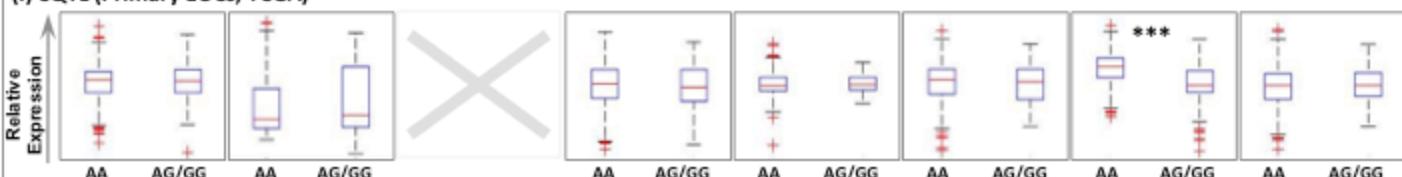
(G) mOTI Analysis (Primary EOFs)



(H) eQTL Analysis (OSECs/FTSECs)



(I) eQTL (Primary EOCs, TCGA)

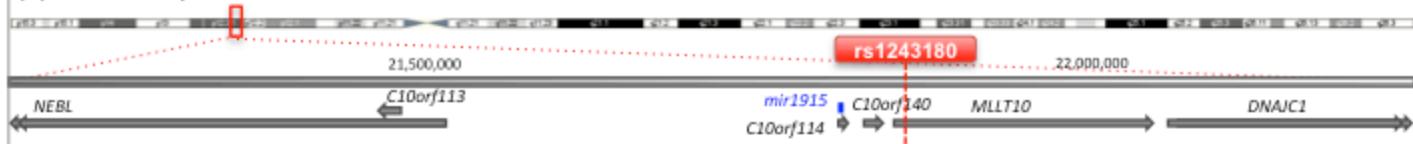


Supplementary figure 8ii: Gene expression and methylation analyses of all protein coding genes within a one megabase region centered around rs1243180 at 10p12

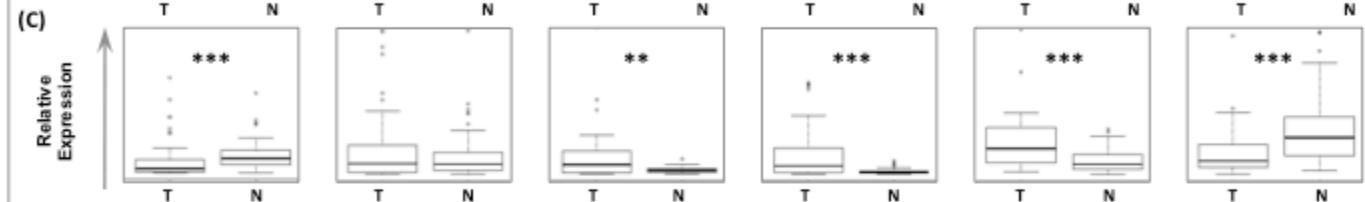
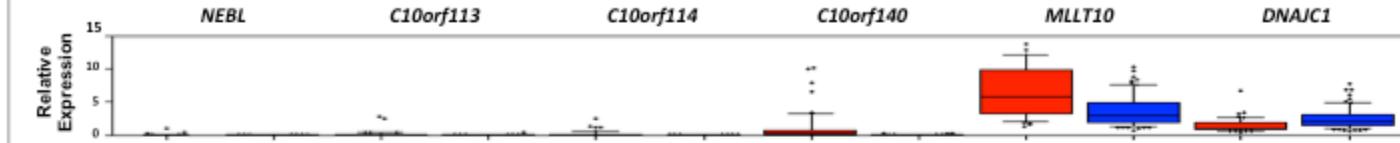
(A) Genomic map of a one-megabase region centred on the most statistically significant SNP at each locus, relative to the chromosomal location at each locus. The location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region are shown relative to the most significant SNP (red dashed line) at each locus. For 10q12 data for *PMP2* and *FABP9* are not included as no expression was detected; additional data for these genes are in Supplementary Tables 3 and 4. **(B)** Boxplots of semi-quantitative PCR based relative gene expression (normalized to β -actin and GAPDH) in epithelial ovarian cancer (EOC) cell lines ($n=50$) and normal ovarian surface epithelial cell (OSEC) lines and fallopian tube secretory epithelial cell (FTSEC) lines ($n=73$). **(C)** Boxplots of semi-quantitative PCR based gene expression in EOC (T) and normal (N; OSEC/FTSEC) cell lines. These are the same data as in panel B rescaled to emphasise the contrast between tumour and normal. **(D)** Boxplots of TCGA array based gene expression data in primary high-grade serous ovarian tumors (T; $n=568$) and normal fallopian tube tissues (N; $n=8$) for all genes for which probe sets were available. For genes not represented in TCGA, expression data from the GEO dataset of EOC tumors (T; $n=53$) and normal ovarian tissue (N; $n=10$) are shown (indicated by a red asterisk). **(E)** Gene expression by DNA copy number alteration in 316 high-grade serous ovarian tumours based on data from TCGA. Graphs show mRNA expression Z-scores by putative copy number alterations: homozygous deletion (0), heterozygous loss (1), diploid (2), copy number gain (3), and amplification (4). **(F)** Boxplots of methylation of 227 high-grade serous ovarian tumours compared to normal ovarian tissues ($n=7$). **(G)** Boxplots of gene methylation by genotype for the most strongly associated SNP at each locus in 227 high-grade serous ovarian tumors. Probes that show significant cis negative correlations with gene expression are indicated by a red open circle. **(H)** Boxplot of qPCR based gene expression by genotype for the most strongly associated SNP for 73 normal OSE/FTSE cell lines. **(I)** Boxplot of array based gene expression by germline genotype in high-grade serous tumors based on TCGA data. * Adjusted $P<0.05$, ** $P<0.01$, *** $P<0.001$).

Supplementary figure 8.ii

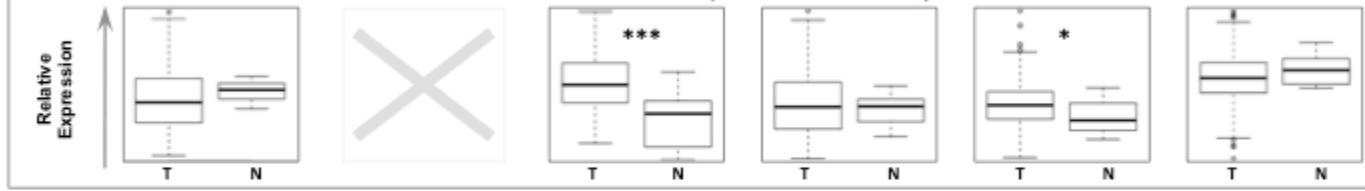
(A) Genomic Map and LD Structure



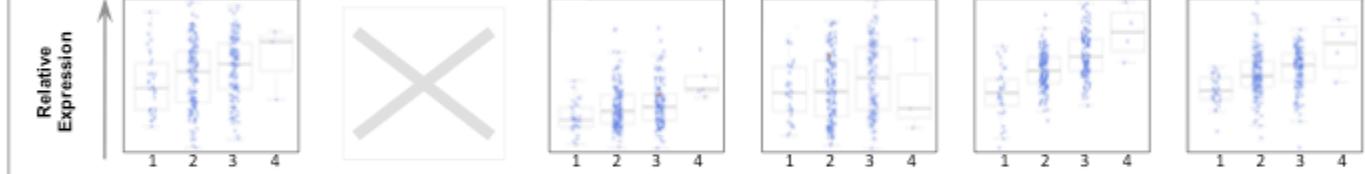
(B) Gene Expression (OSECs/FTSECs/EOC Lines)



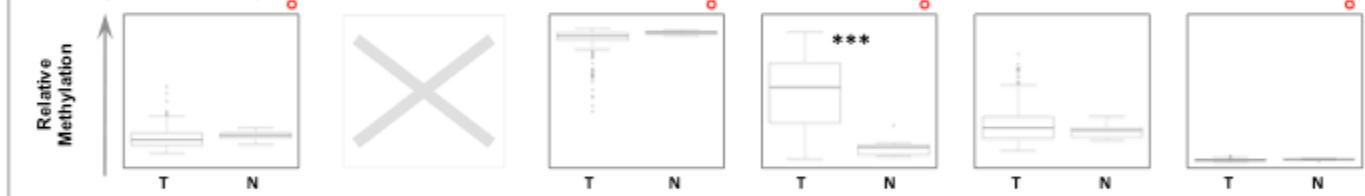
(D) Gene Expression (Primary EOCs and Normal Tissues)



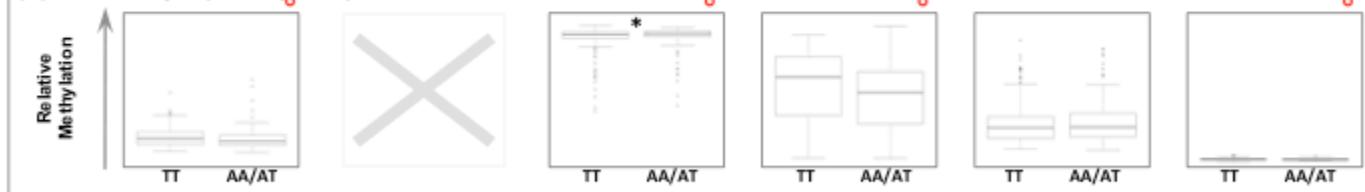
(E) Expression versus Copy Number (TCGA)



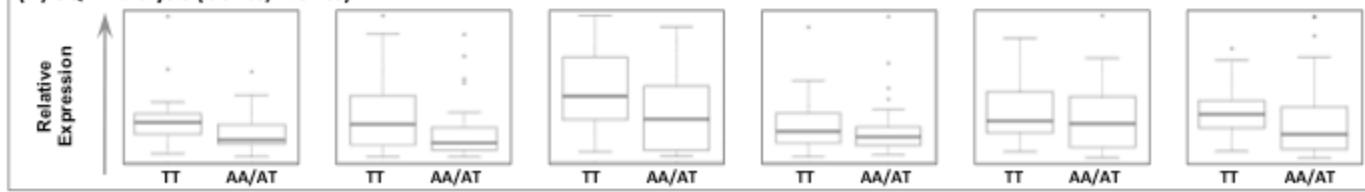
(F) Methylation (Primary EOCs and Normal Tissues)



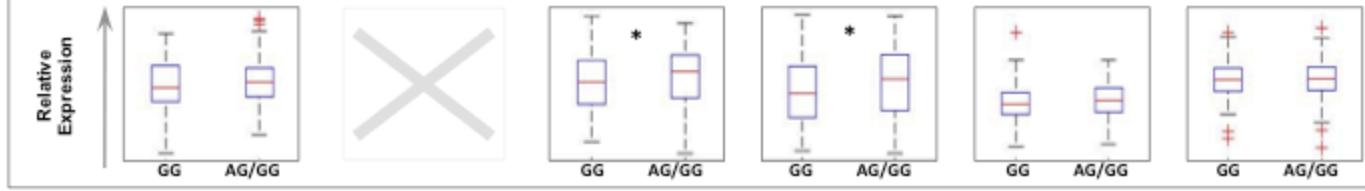
(G) mQTL Analysis (Primary EOCs)



(H) eQTL Analysis (OSECs/FTSECs)



(I) eQTL (Primary EOCs, TCGA)



Supplementary figure 8iii: Gene expression and methylation analyses of all protein coding genes within a one megabase region centered around rs757210 at 17q12

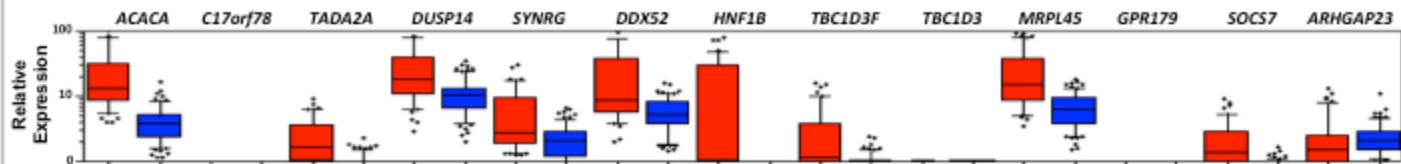
(A) Genomic map of a one-megabase region centred on the most statistically significant SNP at each locus, relative to the chromosomal location at each locus. The location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region are shown relative to the most significant SNP (red dashed line) at each locus. For 10q12 data for *PMP2* and *FABP9* are not included as no expression was detected; additional data for these genes are in Supplementary Tables 3 and 4. **(B)** Boxplots of semi-quantitative PCR based relative gene expression (normalized to β -actin and GAPDH) in epithelial ovarian cancer (EOC) cell lines (n=50) and normal ovarian surface epithelial cell (OSEC) lines and fallopian tube secretory epithelial cell (FTSEC) lines (n=73). **(C)** Boxplots of semi-quantitative PCR based gene expression in EOC (T) and normal (N; OSEC/FTSEC) cell lines. These are the same data as in panel B rescaled to emphasise the contrast between tumour and normal. **(D)** Boxplots of TCGA array based gene expression data in primary high-grade serous ovarian tumors (T; n=568) and normal fallopian tube tissues (N; n=8) for all genes for which probe sets were available. For genes not represented in TCGA, expression data from the GEO dataset of EOC tumors (T; n=53) and normal ovarian tissue (N; n=10) are shown (indicated by a red asterisk). **(E)** Gene expression by DNA copy number alteration in 316 high-grade serous ovarian tumours based on data from TCGA. Graphs show mRNA expression Z-scores by putative copy number alterations: homozygous deletion (0), heterozygous loss (1), diploid (2), copy number gain (3), and amplification (4). **(F)** Boxplots of methylation of 227 high-grade serous ovarian tumours compared to normal ovarian tissues (n=7). **(G)** Boxplots of gene methylation by genotype for the most strongly associated SNP at each locus in 227 high-grade serous ovarian tumors. Probes that show significant cis negative correlations with gene expression are indicated by a red open circle. **(H)** Boxplot of qPCR based gene expression by genotype for the most strongly associated SNP for 73 normal OSE/FTSE cell lines. **(I)** Boxplot of array based gene expression by germline genotype in high-grade serous tumors based on TCGA data. * Adjusted $P<0.05$, ** $P<0.01$, *** $P<0.001$).

Supplementary figure 8.iii

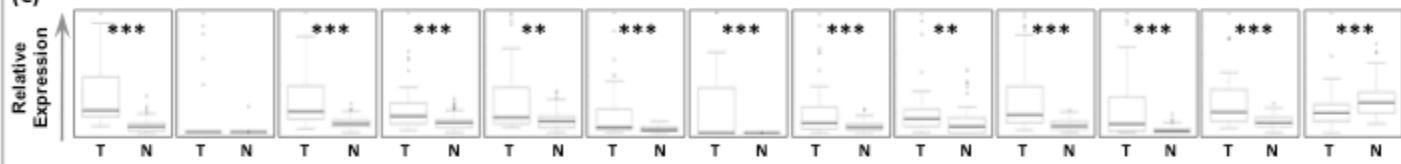
(A) Genomic Map and LD Structure



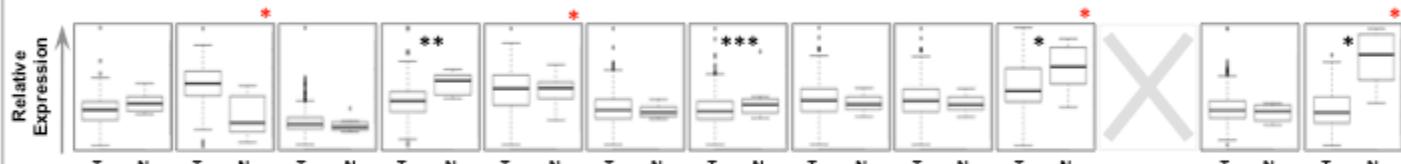
(B) Gene Expression (OSECs/FTSECs/EOC Cell Lines)



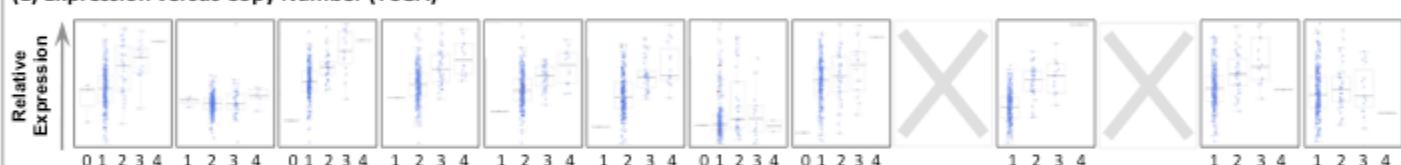
(C)



(D) Gene Expression (Primary EOCs and Normal Tissues)



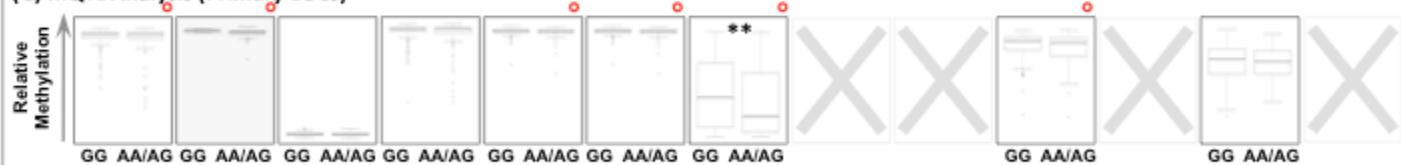
(E) Expression versus Copy Number (TCGA)



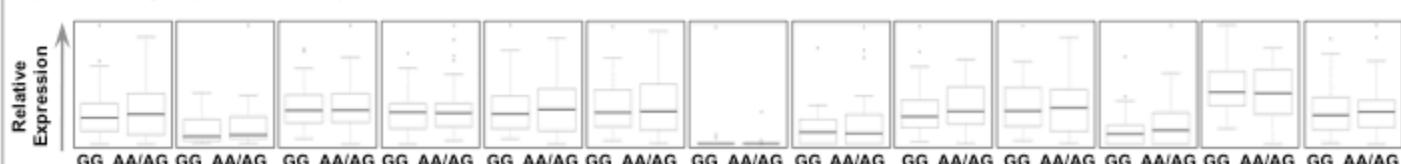
(F) Methylation (Primary EOCs and Normal Tissues)



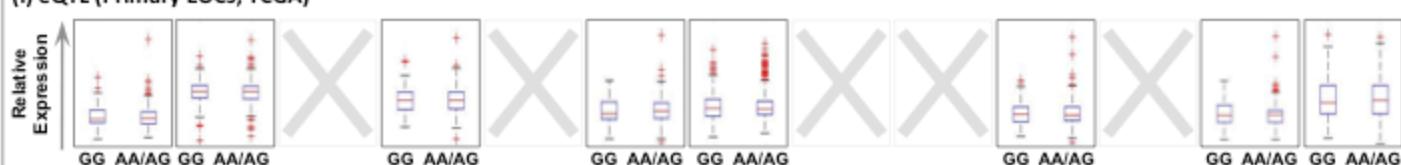
(G) mQTL Analysis (Primary EOCs)



(H) eQTL Analysis (OSECs/FTSECs)



(I) eQTL (Primary EOCs, TCGA)

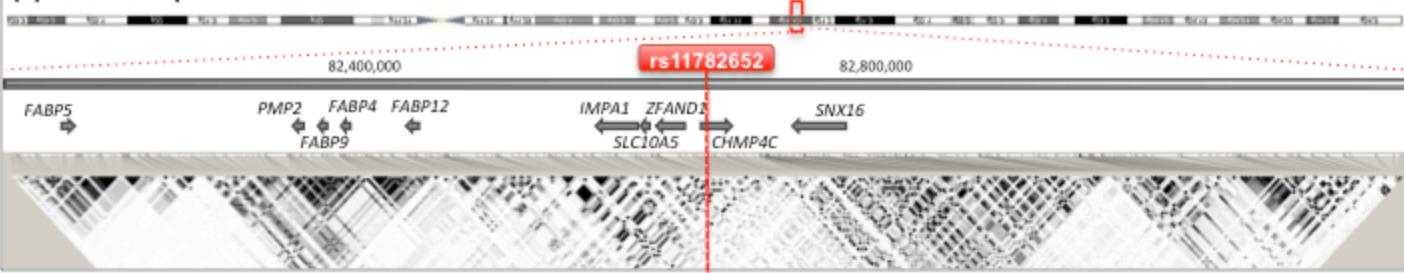


Supplementary Figure 9i: Genomic architecture, tumor copy number variation, regulatory elements and RNAseq analysis of a 1 megabase region around rs11782652 at 8q21.

(A) Genomic map of a one-megabase region, centred on the most statistically significant SNP at each locus (red dashed line), relative to the chromosomal location. Linkage disequilibrium structure showing r^2 between SNPs (in grey scale – darker color for higher r^2) based on HapMap 2 data for European population. Also shown is the location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region. (B) Analysis of somatic DNA copy number variation across the region generated by analysis of Affymetrix 6.0 SNP array data from 481 high-grade serous ovarian tumors (TCGA). Red = proportion of tumors showing loss; green = proportion of tumors showing gain/amplification; yellow = proportion of tumors that are diploid (no loss or gain). (C) Analysis of ENCODE data to evaluate putative regulatory DNA or non-coding RNA elements. Data were generated for non-ovarian cancer associated cell lines and show H3K27Ac regulatory marks, DNasel Hypersensitivity elements, transcription factor binding site information from chromatin immunoprecipitation (ChIP) analysis, and taretscan analysis of microRNA binding sites. (D) Formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) performed in normal ovarian surface epithelial cells (OSEC) and fallopian tube secretory epithelial cells (FTSEC). From genome wide FAIRE-seq, the profile of peaks of open chromatin (and putative regulatory sites) are illustrated for each region. (E) ENCODE RNA sequencing analysis (RNA-seq) of non-ovarian cancer associated cell lines shows the spectrum of coding transcripts across each region. (F) RNA-seq analysis of normal OSEC, FTSEC and epithelial ovarian cancer cell lines shows the spectrum of poly-adenylated transcripts (protein-coding and non-protein-coding) across each region in ovarian cancer associated tissues.

Supplementary figure 9.i

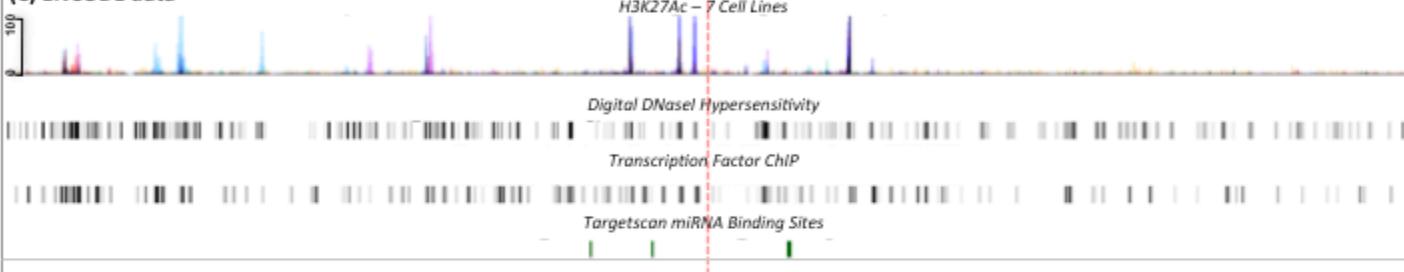
(A) Genomic Map and LD Structure



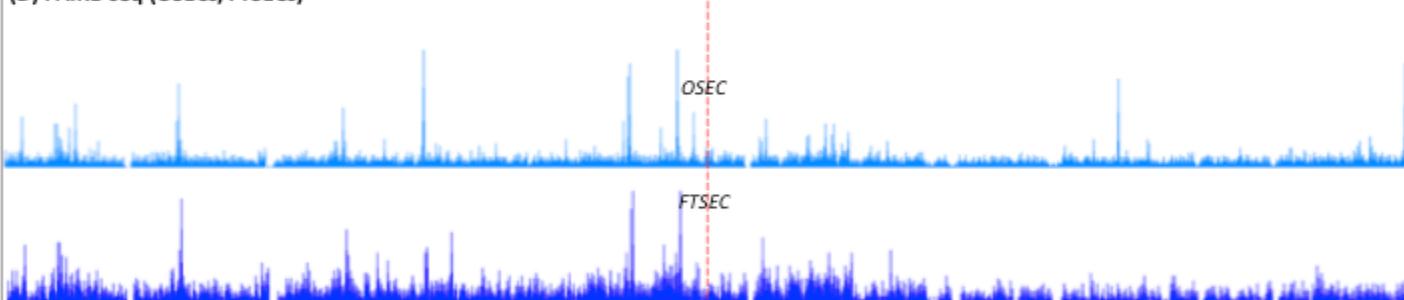
(B) Tumor Copy Number Variation (Primary EOCs, TCGA)



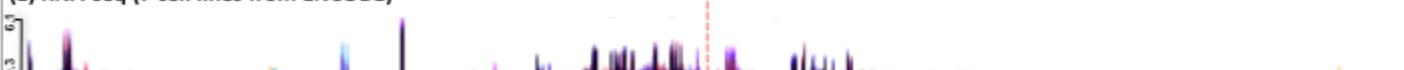
(C) ENCODE data



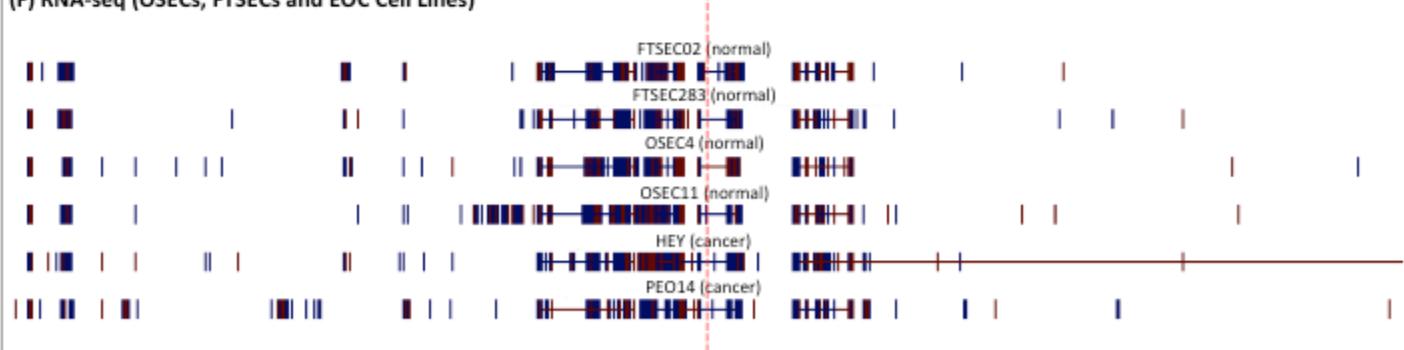
(D) FAIRE-seq (OSECs/FTSECs)



(E) RNA-seq (7 cell lines from ENCODE)



(F) RNA-seq (OSECs, FTSECs and EOC Cell Lines)



Supplementary Figure 9ii: Genomic architecture, tumor copy number variation, regulatory elements and RNAseq analysis of a 1 megabase region around rs1243180 at 10p12.

(A) Genomic map of a one-megabase region, centred on the most statistically significant SNP at each locus (red dashed line), relative to the chromosomal location. Linkage disequilibrium structure showing r^2 between SNPs (in grey scale – darker color for higher r^2) based on HapMap 2 data for European population. Also shown is the location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region. (B) Analysis of somatic DNA copy number variation across the region generated by analysis of Affymetrix 6.0 SNP array data from 481 high-grade serous ovarian tumors (TCGA). Red = proportion of tumors showing loss; green = proportion of tumors showing gain/amplification; yellow = proportion of tumors that are diploid (no loss or gain). (C) Analysis of ENCODE data to evaluate putative regulatory DNA or non-coding RNA elements. Data were generated for non-ovarian cancer associated cell lines and show H3K27Ac regulatory marks, DNasel Hypersensitivity elements, transcription factor binding site information from chromatin immunoprecipitation (ChIP) analysis, and taretscan analysis of microRNA binding sites. (D) Formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) performed in normal ovarian surface epithelial cells (OSEC) and fallopian tube secretory epithelial cells (FTSEC). From genome wide FAIRE-seq, the profile of peaks of open chromatin (and putative regulatory sites) are illustrated for each region. (E) ENCODE RNA sequencing analysis (RNA-seq) of non-ovarian cancer associated cell lines shows the spectrum of coding transcripts across each region. (F) RNA-seq analysis of normal OSEC, FTSEC and epithelial ovarian cancer cell lines shows the spectrum of poly-adenylated transcripts (protein-coding and non-protein-coding) across each region in ovarian cancer associated tissues.

Supplementary figure 9.ii

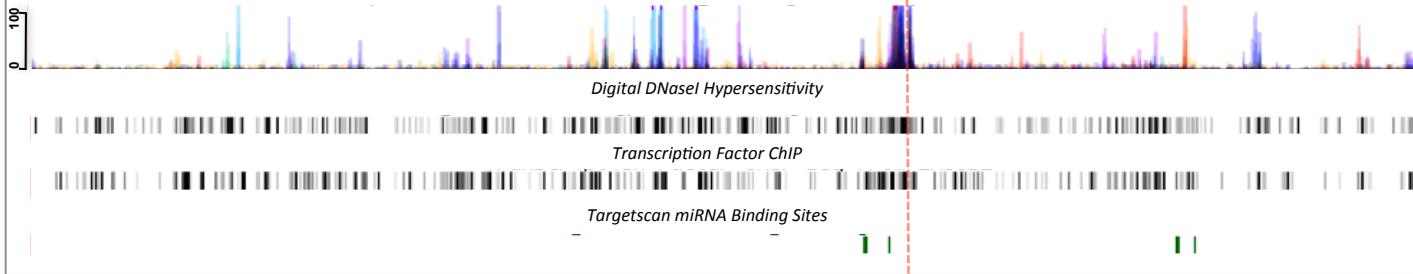
(A) Genomic Map and LD Structure



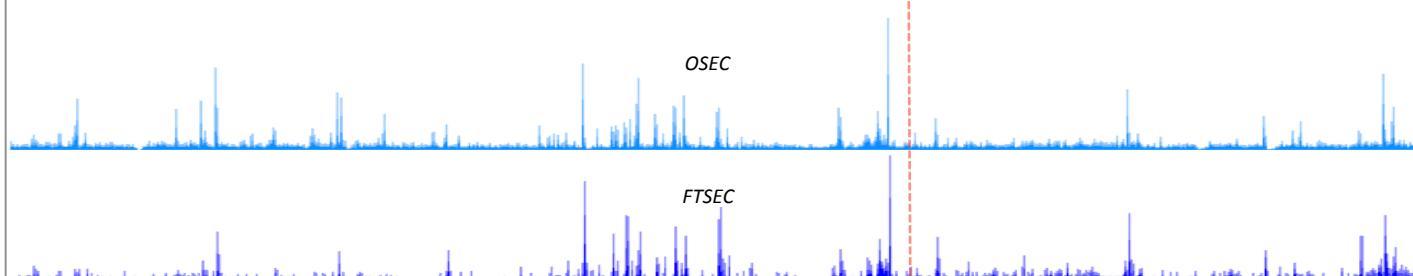
(B) Tumor Copy Number Variation (Primary EOCs, TCGA)



(C) ENCODE data



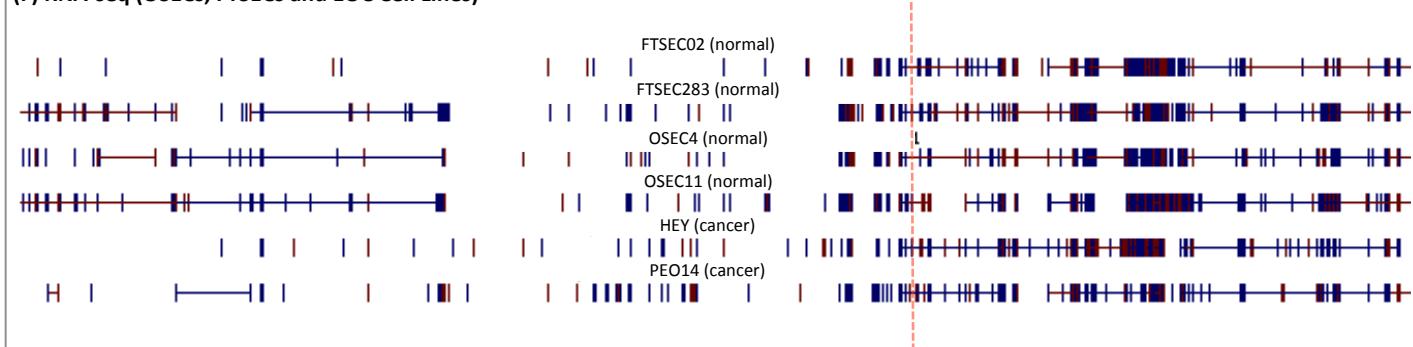
(D) FAIRE-seq (OSECs/FTSECs)



(E) RNA-seq (7 cell lines from ENCODE)



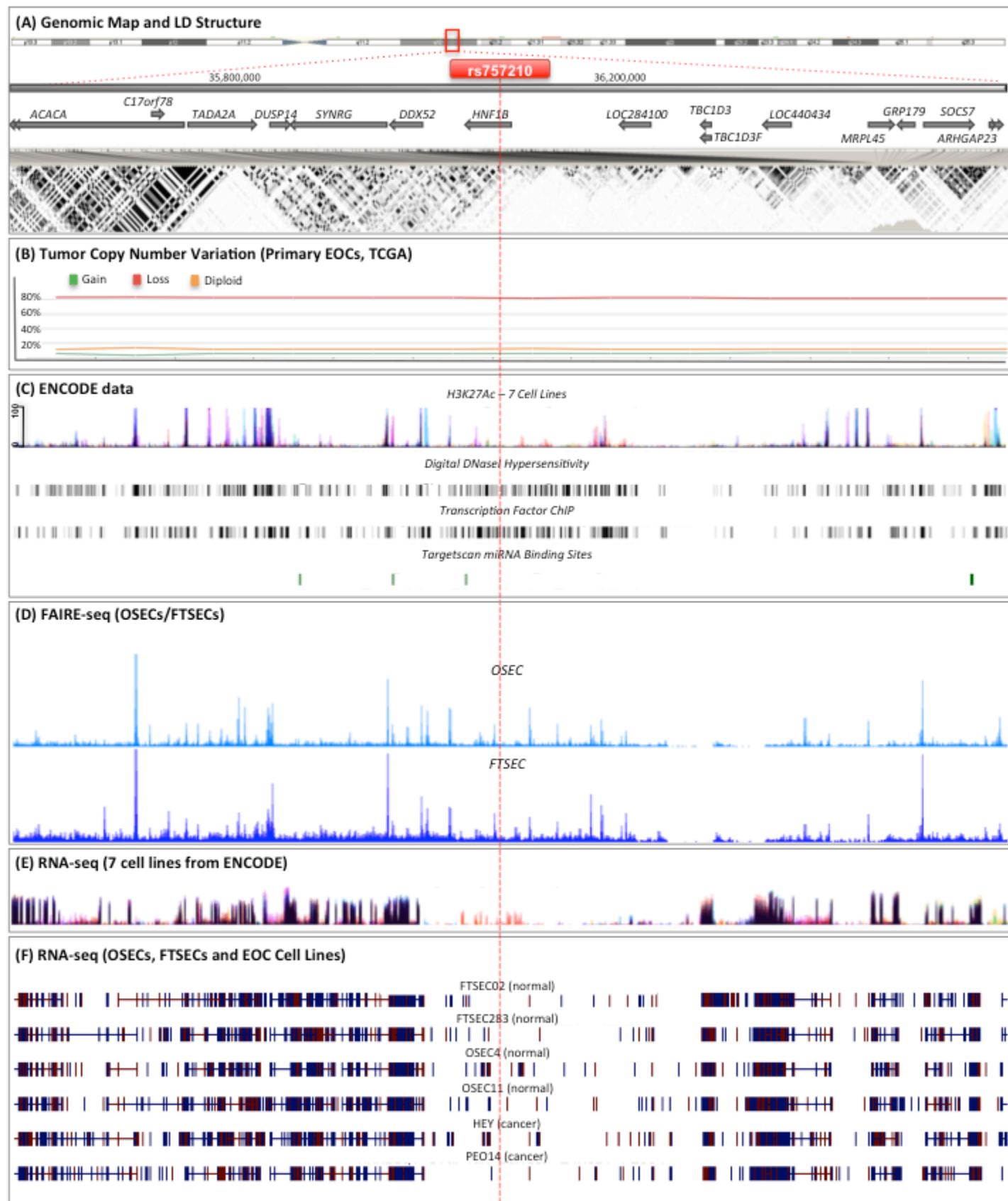
(F) RNA-seq (OSECs, FTSECs and EOC Cell Lines)



Supplementary Figure 9iii: Genomic architecture, tumor copy number variation, regulatory elements and RNAseq analysis of a 1 megabase region around rs757210at 17q12.

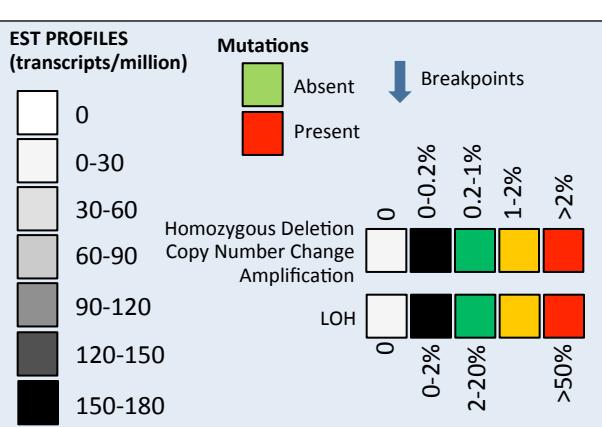
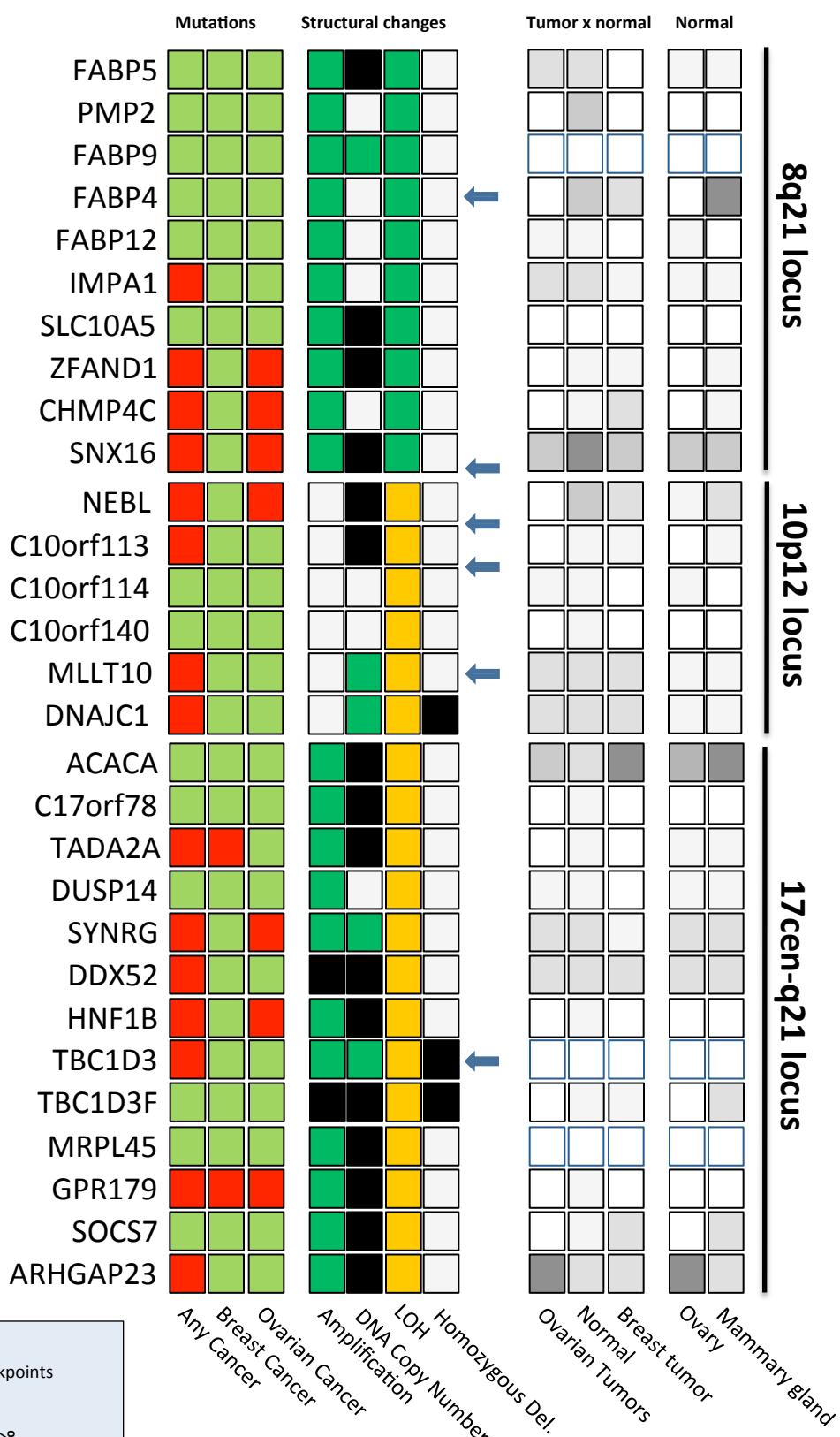
(A) Genomic map of a one-megabase region, centred on the most statistically significant SNP at each locus (red dashed line), relative to the chromosomal location. Linkage disequilibrium structure showing r^2 between SNPs (in grey scale – darker color for higher r^2) based on HapMap 2 data for European population. Also shown is the location and approximate size of all known protein coding genes (grey) and non-coding RNA sequences (blue) at each region. (B) Analysis of somatic DNA copy number variation across the region generated by analysis of Affymetrix 6.0 SNP array data from 481 high-grade serous ovarian tumors (TCGA). Red = proportion of tumors showing loss; green = proportion of tumors showing gain/amplification; yellow = proportion of tumors that are diploid (no loss or gain). (C) Analysis of ENCODE data to evaluate putative regulatory DNA or non-coding RNA elements. Data were generated for non-ovarian cancer associated cell lines and show H3K27Ac regulatory marks, DNasel Hypersensitivity elements, transcription factor binding site information from chromatin immunoprecipitation (ChIP) analysis, and taretscan analysis of microRNA binding sites. (D) Formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-seq) performed in normal ovarian surface epithelial cells (OSEC) and fallopian tube secretory epithelial cells (FTSEC). From genome wide FAIRE-seq, the profile of peaks of open chromatin (and putative regulatory sites) are illustrated for each region. (E) ENCODE RNA sequencing analysis (RNA-seq) of non-ovarian cancer associated cell lines shows the spectrum of coding transcripts across each region. (F) RNA-seq analysis of normal OSEC, FTSEC and epithelial ovarian cancer cell lines shows the spectrum of poly-adenylated transcripts (protein-coding and non-protein-coding) across each region in ovarian cancer associated tissues.

Supplementary figure 9.iii



Supplementary Figure 10: Somatic variation and expression sequence tags for genes in 1MB region around the lead SNPs at 8q21, 10p12 and 17q12 loci

COSMIC EST PROFILES

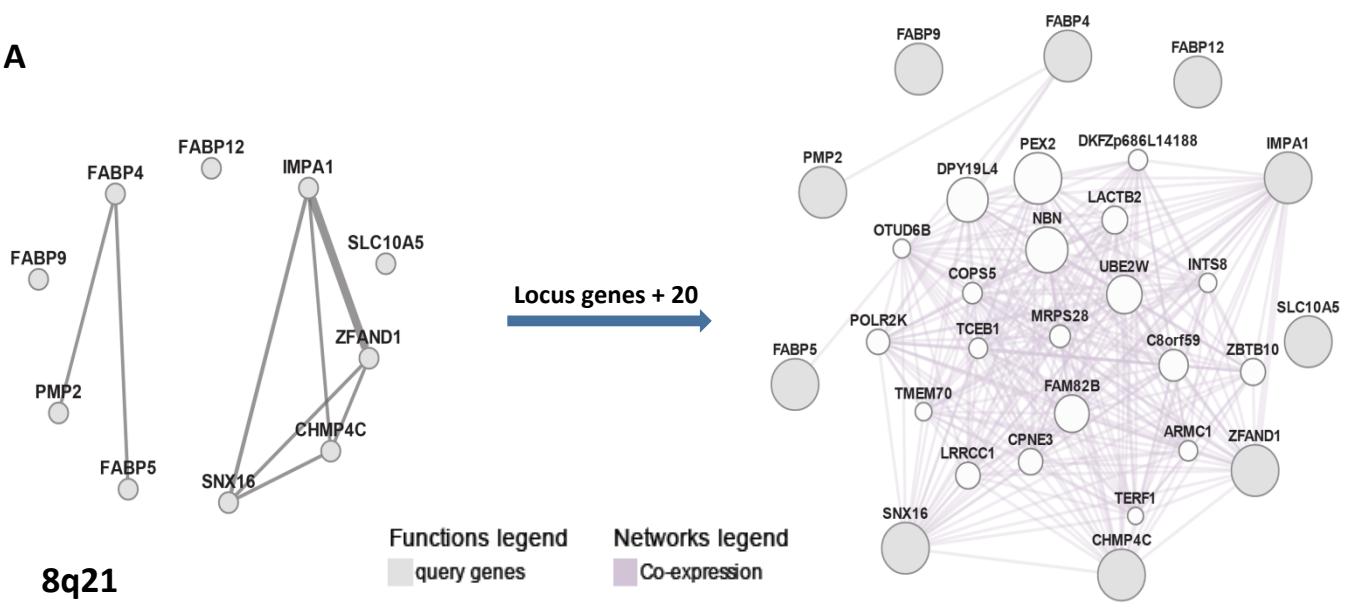


Supplementary Figure 11: Co-expression networks for genes in 1MB region around the lead SNPs at 8q21, 10p12 and 17q12 loci

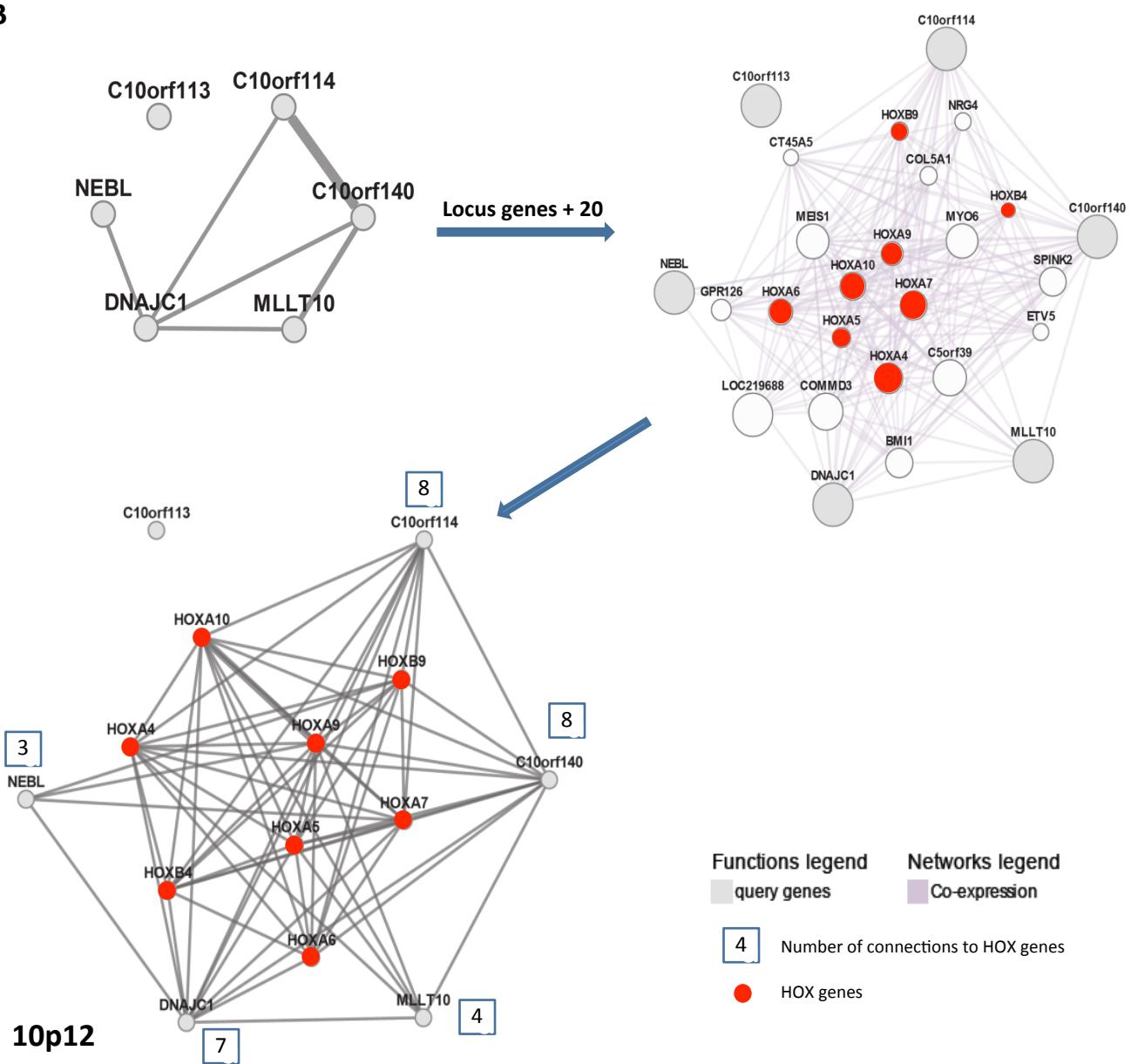
Gene co-expression networks were generated by Genemania enabling all co-expression networks ($n=154$) and using a query-dependent weighting assigned based on the query genes (the weights are chosen automatically using linear regression). **(A)** 8q21 locus; *left panel*, co-expression between genes in the locus; *right panel*, also shows the 20 most connected additional genes in the datasets. **(B)** 10p12 locus; *left panel*, co-expression between genes in the locus; *right panel*, also shows the 20 most connected additional genes in the datasets. Note enrichment of HOXA genes (red nodes); *Bottom panel*, in order to better visualize direct interactions only the genes in the locus and HOXA genes were retained. Boxes indicate the number of interactions of each gene in the locus to HOXA genes. **(C)** 17q12 locus; *left panel*, co-expression between genes in the locus; *right panel*, also shows the 20 most connected additional genes in the datasets.

Supplementary figure 11

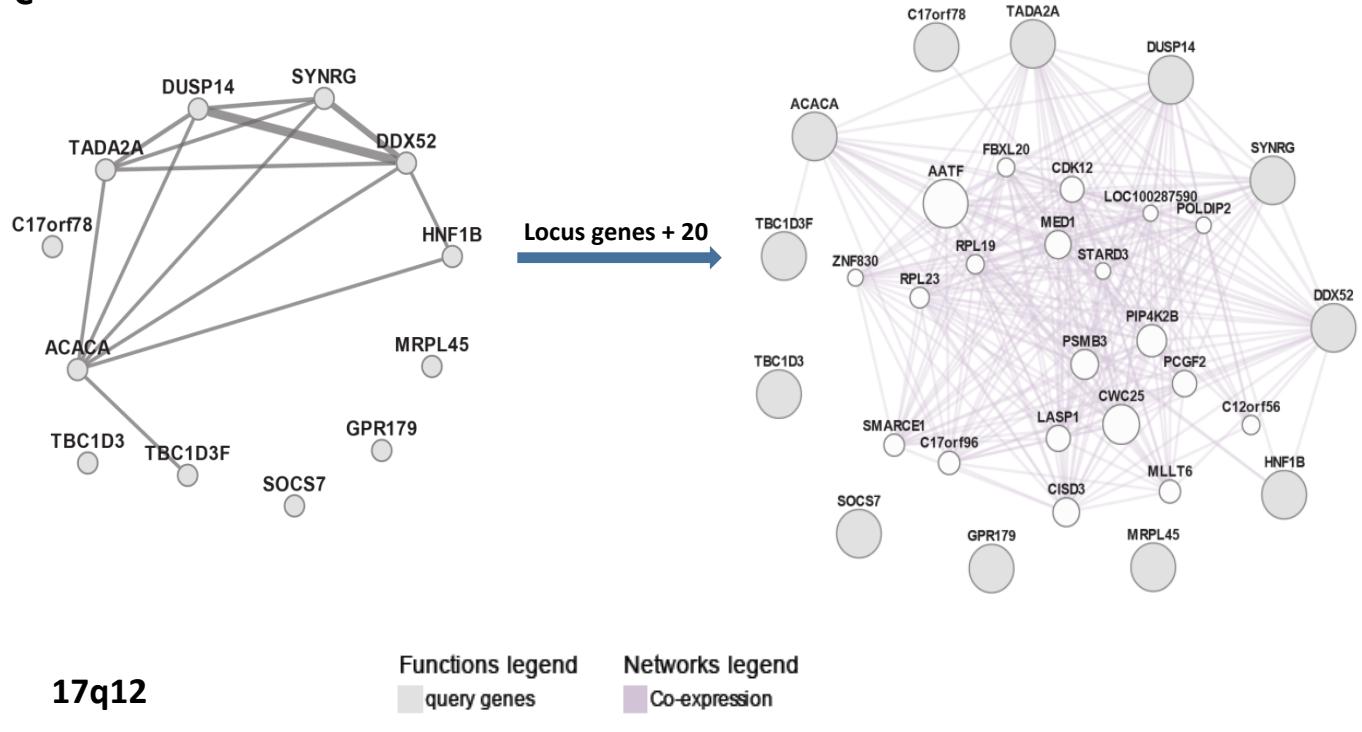
A



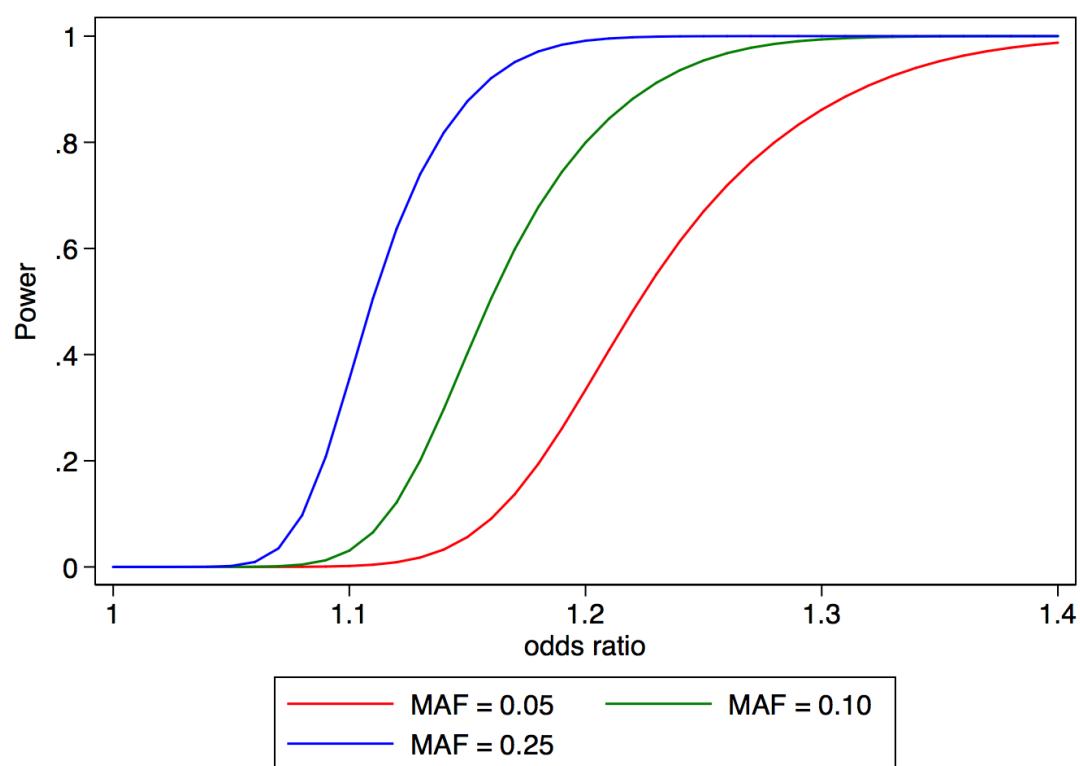
B



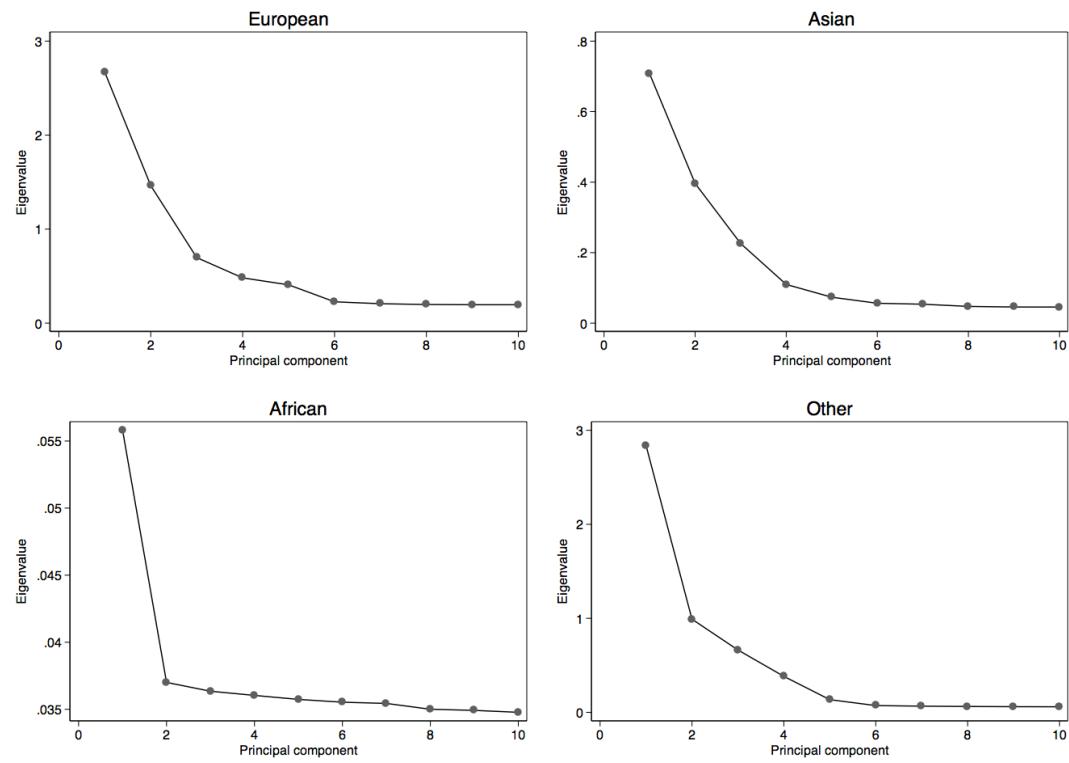
C



Supplementary Figure 12: Power of this two-phase study to detect alleles of varying risk and frequency at a Type I error rate of 10^{-8}



Supplementary Figure 13: Eigen value scree plot for subjects of European, Asian, African and Other ancestry



SUPPLEMENTARY NOTE

Combined GWAS analysis

North American GWAS

This study included 1,952 EOC cases and 2,052 controls from five case-control studies: MAY, NCO, TOR, TBO, and NEC (Supplementary Table 1). Samples from MAY, NCO, TOR, and TBO were genotyped with the Illumina 610-quad Beadchip Array™ at the Mayo Clinic Medical Genome Facility (Rochester, MN). Detailed QC procedures have been described elsewhere¹. In brief, samples and SNPs with call rates <95% were excluded; samples with less than 80% of European ancestry were excluded using 446 European ancestry informative markers. Genotyping for NEC was performed at the National Institute of Aging (Bethesda, Maryland) with the Illumina 317K and 370K arrays. The same QC criteria on samples and SNPs were applied. 288 eligible subjects from NEC were attempted; of these, 7 were excluded because of overall call rate <95%, ambiguous gender (n=1) or less than 80% European ancestry (n=1), leaving 137 cases and 142 controls. To account for different marker sets and improve genome coverage, imputation was performed using MACH version 1.0.16 with phased HapMap haplotypes (release 22) using 60 CEU founders as reference. 2,508,744 out of 2,543,887 SNPs (98.6%) passed the quality control.

UK GWAS

The UK GWAS Phase I comprised of 1,817 cases and 2,354 controls from the UK. The cases were from the RMH, SEA, UKR and UKO studies (**Supplementary table 1**) and were genotyped using the Illumina Infinium 610K at Illumina Corporation. Genotype data for UK controls came from the Wellcome Trust Case-Control Consortium 1958 Birth Cohort and a national colorectal study both genotyped using the Illumina 550K array. Detailed QC criteria are described elsewhere². Data were available for 507,094 genotyped SNPs and an in-house method that combines the features of fastPHASE³ and IMPUTE⁴ was used to impute ungenotyped SNPs and missing data on genotyped SNPs using the phased CEU haplotypes as reference. The program LAMP⁵ was used to assign intercontinental ancestry based on the HapMap (release 22) and subjects with less than 90 percent European ancestry were excluded from the analysis. In Phase 2, the top ranked 21,955 SNPs were genotyped in 4,162 cases and 4,810 controls from 10 studies in OCAC (AOCS/ACS, BAV, DOV, HOP, MAL, POC, POL, STA, USC and UKO, **Supplementary table 1**).

COGS iSelect genotyping array design

SNPs were selected for inclusion on the custom genotyping array separately by each consortium. Each consortium was given a share of the array: nominally 25 percent of the SNPs each for BCAC, PRACTICAL and OCAC; 17½ percent for CIMBA; and 7½ percent for SNPs of potential interest to multiple consortia. In

practice, the allocations were larger as a result of overlaps. For each consortium, the allocation was divided into three categories: “GWAS replication”, “fine-mapping” and “candidate SNPs”.

In general, we considered only SNPs with an Illumina design score of 0.8 or greater (some OCAC and CIMBA SNPs with lower design scores were included). Where possible, preference was given to SNPs previously genotyped by Illumina (design score 1.1). For each category, we defined a series of ranked lists of SNPs. For the GWAS SNPs, these were merged in the following way, in order to generate a single list. We selected SNPs in priority order from each list, according to predefined weightings. Where a SNP (or a surrogate) was selected on the basis of more than one list, the SNP counted towards the tally for each list. For each SNP, we preferentially accepted the SNP if it had a design score of 1.1 (i.e., had previously been genotyped on an Illumina platform). If not, we sought SNPs with $r^2=1$ with the selected SNP, and selected the SNP with the best design score. If no such SNP was available, we selected SNPs with $r^2>0.8$ with the chosen SNP, and selected the SNP with the best design score. We excluded SNPs that were in strong LD with a previously selected SNP ($r^2>0.9$). However, for SNPs that were highly significant in each list ($P<.00001$), we required two surrogate SNPs. The candidate lists were merged in the same way, giving equal weight to lists from each study. The only differences were (a) there was no provision for additional surrogates (b) SNPs were excluded if there was existing surrogate at $r^2=1$.

To merge the three categories, we first included all the selected fine-mapping SNPs, and then included SNPs from the merged GWAS and candidate lists in priority order. SNPs of common interest were selected in a similar way.

Finally, lists from each of the constituent consortia were merged, in priority order and in proportion to their allocated shares. SNPs selected by one consortium and subsequently selected by another counted towards both lists. The process continued until the maximum 240,000 attempted beadtypes had been reached. The final list comprised 220,123 SNPs. Of these, 211,155 were successfully manufactured on the array.

COGS studies

Forty-three individual participating OCAC studies contributed samples to the COGS project. Of these, nine studies were case-only (GRR, HSK, LAX, ORE, PVD, RMH, SOC, SRO, UKR). The cases from these studies were pooled with case-control studies from the same geographic region and the two national Australian case-control studies were combined into a single study to create 34 case-control sets. Study characteristics are summarized in **Supplementary table 1** and the number of cases by histological subtype is shown in **Supplementary table 2**. A subset of the samples (3,045 cases and 1,053 controls) included in the GWAS were also included in the samples genotyped using the iCOGS array.

Molecular studies

We used several different molecular assays to evaluate the putative role in ovarian cancer development of every known protein-coding gene in a one megabase region centered on the most significant SNP at the 8q21, 10p12 and 17q12 susceptibility loci (**Supplementary Figs. 8i-iii**). We also evaluated associations between genotype of the most strongly associated SNP at each locus and the expression of putative target genes. We used a combination of genome wide molecular assays and publicly available dataset to establish architectural maps of the 8q21, 10p12 and 17q12 susceptibility loci (**Supplementary Figs. 9i-iii**). This included analysis of linkage disequilibrium (LD) across each region; levels of somatic copy number variation in primary ovarian cancers to evaluate regional loss, gain and amplification across each region; analysis of regulatory regions, including DNA elements (e.g. putative enhancers, promoters and repressors) and RNA elements (e.g. genes transcripts and non coding RNAs such as micro-RNAs and lncRNA), that may coincide with risk associated SNPs.

mRNA expression analyses

Cell culture and RNA extraction for mRNA expression studies

Normal ovarian surface epithelial cells (OSEC) and fallopian tube secretory epithelial cells (FTSEC) (N=73) were collected with informed consent under the approval of the University College Hospital (UCH), Ethics Committee in London. Primary cell lines were established in culture and expression of lineage-specific markers confirmed by immunofluorescence. Epithelial ovarian cancer (EOC) cells lines (N=50) were obtained from the ATCC, or were a kind donation from Dr G Mills at MD Anderson. EOC cells were grown in the recommended media for each cell line. All cell lines used in these studies were confirmed to be free of mycoplasma.

RNA for qPCR was extracted from cell cultures harvested at ~80% confluence using the QIAgen RNeasy Kit, according to manufacturers instructions. On-column DNaseI digests were performed. RNA for RNAseq was extracted with a GE Healthcare Illustra RNAspin mini kit with no DNaseI digestion. Concentrations were determined with a nanodrop Spectrophotometer.

mRNA expression by qPCR

Expression analysis was performed on 73 normal OSE/FTSE cells, 108 LCL and 50 EOC cell lines. For each cell line 500 ng of RNA was reverse transcribed using SuperScript III First-Strand Synthesis System (Invitrogen). The cDNA was diluted to 10ng/ul and 12.5 ng was used in target specific amplification prior to real-time PCR using TaqMan PreAmp Master Mix Kit (Applied Biosystems) following Fluidigm's Specific Target Amplification Protocol. An aliquot of 1.25 ul of the 25 ul pre-amplified cDNA was added to each chip. Each cDNA sample was run in triplicate and each experiment included no template controls and no

template controls from the cDNA reactions. 96.96 Dynamic Array Integrated Fluidic Circuits (Fluidigm) were loaded with 96 pre-amplified cDNA samples and 96 TaqMan gene expression probes (Applied Biosystems) using the BioMark HD System (Fluidigm).

Expression data were analyzed using the comparative $\Delta\Delta Ct$ method. The Ct values were normalized to glyceraldehydes-3-phosphate dehydrogenase (GAPDH) and B-Actin controls, and these were then normalized to the highest expressing cancer cell line. The difference between the average ddCT values for cancers and normals of each gene was determined. The fold change was calculated as $2^{-(\Delta\Delta Ct \text{ cancer} - \Delta\Delta Ct \text{ normal})}$. To analyze relative expression across all genes at each locus, $\Delta\Delta Ct$ values were calculated based on the average Ct for two endogenous controls (β -actin and GAPDH), and expression of all genes at each locus was calculated relative to the first gene in the region with detectable expression in the EOC cell lines (**Supplementary Figs. 8i-iii panel B**).

Relative gene expression in EOC and normal tissue cell lines was compared using the Wilcoxon rank-sum test (**Supplementary Table 5, Supplementary Figures 8i-iii panel C**). Genotype specific gene expression in the normal tissue cell lines (eQTL analysis) was compared using the Jonckheere-Terpstra test (**Supplementary Table 5, Supplementary Figures 8i-iii panel H**).

mRNA expression by RNAseq

RNAseq was performed on six cell lines (two OSEC two FTSE and two serous EOC) using polyA selection and libraries were prepared with an Illumina Tru-Seq RNA sample prep kit. The libraries were barcoded with 4 samples multiplexed per lane of an Illumina HiSeq 2000 using 50bp paired end reads. The average number of aligned reads per sample was 46.5 million reads.

In silico analysis of mRNA expression in tumor and normal tissue

Affymetrix U133A based gene expression profiling data were obtain from the TCGA website. The data set relates to 568 primary serous ovarian tumor samples and 8 fallopian tube samples. Robust Multi-array analysis was used to normalize and calculate signal intensity for the entire set. The boxplot function in R was used to compare ovarian tumor samples to the fallopian tube for 24 probes mapping to 17 of 28 protein-coding genes at the three loci. A difference in relative expression between normal and EOC was carried out using the Wilcoxon rank-sum test (**Supplementary Figs. 8i-iii panel D**).

Affymetrix U133Plus based gene expression profiling data were obtained from the Gene Expression Omnibus (GEO) series GSE18520. The data set relates to 53 advanced stage serous ovarian tumors and 10 normal epithelial ovarian samples. Robust Multi-array analysis was used to normalize and calculate signal intensity for the entire set. The boxplot function in R was used to compare the ovarian tumor samples to normal ovary for 51 probes mapping to 24 unique genes.

FAIRE-seq analysis

For each OSE and FTSE cell line, pairs of biological replicate 15 cm culture dishes containing cells at 80 – 90 percent confluence were cross-linked in 1% formaldehyde. Cells were harvested and lysed in a Tris-buffered 1% SDS lysis buffer containing protease inhibitors. Lysates were sonicated using a QSONICA Model Q125 Ultra Sonic Processor to shear chromatin to 200bp-1kb fragments. Insoluble cell material was removed through centrifugation, and 4x 50 μ l aliquots from each supernatant were designated as INPUT and 4x 50 μ l aliquots for FAIRE processing. INPUT samples were incubated overnight at 65° C to reverse cross-linking. All samples were purified through 2 rounds of phenol-chloroform extraction followed by a round of chloroform extraction. DNA was recovered through ethanol precipitation and resulting material from the 4x aliquots of FAIRE and INPUT combined respectively for each biological replicate.

Genome wide sequencing was performed on the Illumina HiSeq platform using 50 bp single read runs. For the FAIRE samples, the 4 samples were barcoded and multiplexed in each lane and 6 lanes were run (equivalent to 1.5 lanes per sample) with an average number of aligned reads per sample of 127 million reads. For the INPUT samples, 2 samples were multiplexed in one lane and 2 lanes were run (equivalent to 1.0 lane per sample) with an average number of aligned reads per sample of 85.3 million reads. The raw data files were uploaded onto the UCSC browser, and a minimum quality score of q30 used.

Copy number analysis using TCGA data

Serous ovarian cancer samples for 481 tumors with log2 copy number data were analyzed using the cBio portal for analysis of TCGA data (<http://www.cbiportal.org/>). For each gene in a region the classes of copy number; homozygous deletion, heterozygous loss, diploid, gain, and amplification were queried individually using the advanced onco query language (OQL) option. The frequency of gain and amplification were combined as “gain”, and homozygous deletion and heterozygous loss were combined as “loss”.

Analysis of copy number vs mRNA expression using TCGA data

Serous ovarian cancer samples for 316 complete tumors (those with CNA, mRNA and sequencing data) were analyzed (**Supplementary Figs. 8i-iii panel E**). Graphs were generated using the cBio portal for analysis of TCGA data and the setting were mRNA expression data Z-score (all genes) with the Z-score threshold of 2 (default setting) and putative copy number alterations (GISTIC). The Z-score is the number of standard deviations away from the mean of expression in the reference population. GISTIC is an algorithm that attempts to identify significantly altered regions of amplification or deletion across sets of patients.

DNA Methylation analysis

Snap frozen tumor samples were obtained during surgery at the Mayo Clinic, reviewed by an experienced gynecologic pathologist (Dr. Gary Keeney), and stored in liquid nitrogen.

Infinium HumanMethylation450 BeadChip analyses were performed on extracted tumor DNA from fresh frozen samples with >70% tumor content (n=323 genotyped MAY cases) and normal ovarian tissue (n=7) by the Mayo Clinic Genotyping Shared Facility using the recommended Illumina protocol (Bibikova, 2009). Tumor DNA samples (1 ug) were bisulfite modified using the Zymo EZ96 DNA Methylation Kit (Zymo Research, Orange, CA) according to the manufacturer's protocol. BeadChips were imaged on an Illumina BeadArray iScan reader and analyzed by the GenomeStudio Methylation Module. Analysis included control probes for assessing sample-independent and -dependent performance. The methylation status of the target CpG sites was determined by comparing the ratio of fluorescent signal from the methylated allele to the sum from the fluorescent signal from both methylated and unmethylated alleles. Two batches of samples were processed (n=199 and n=137) with separate QC samples and analysis. The quality of bisulfite modification and the performance of the CpG probes were assessed using CEPH control, whole-genome amplified negative control and placental positive control samples. The mean intra-class correlation for the QC samples across the two batches of samples was 0.99, 0.96, and 0.90, respectively. The intra-class correlation for duplicate samples was > 0.99. Based on a plate effect observed within each batch of samples, a correction was applied by fitting models with a fixed plate effect for each maker with the unstandardized residual saved. The logit transformation of the probe mean was added back onto the residual before back transforming to get on 0 to 1 scale.

Numerous CpG probes were assessed within each 1 Mb region (112 in 8q21, 188 in 10p12, and 207 in 17q12). In order to prioritize CpG probes for analysis (tumor v normal and mQTL), we used a gene-based approach integrating data from Agilent whole human genome 4x44K expression arrays from subset of 43 cases. Expression data were in the form of log ratios of signal from each case to signal from a reference mix of 106 cases. For each gene (8 genes in 8q21, 6 genes in 10p12, and 10 genes in 17q12) we estimated correlations between RNA expression (log ratios) and all CpG probes within 1 Mb (beta values), adjusting for age and histology. We then identified the CpG probe within 20 kb of each gene region most strongly negatively correlated with each gene's RNA expression (**Supplementary Table 4** and **Supplementary Figs. 8i-iii panel F**).

Association between selected CpGs for each gene and genotypes at key SNPs in 8q21, 10p12 and 17q12 were completed for the high-grade serous histological subtype cases (n=227) (**Supplementary Table 4** and **Supplementary Fig 8i-iii panel G**). Analyses were adjusted for methylation batch and age treating SNP genotype was treated as a continuous variable (additive genetic model) coded in terms of the number of

minor alleles present. Age adjusted methylation beta values (% methylated) were compared between the high-grade serous tumor samples (n=106) and normal tissue (n=7) using the Kruskal-Wallis test.

COSMIC analysis

Mutation data was obtained from the Sanger Institute Catalogue Of Somatic Mutations In Cancer web site (COSMIC; <http://www.sanger.ac.uk/cosmic>) using COSMIC v58 release⁶. All mutations for each of the genes in a 1Mb region surrounding the most significant SNP were extracted from COSMIC. Definitive evidence that the mutation was somatically acquired is lacking for many mutations reported in the database and some mutations may not have a functional impact on the protein. We therefore applied a stringent filter for the mutations. All frameshift and truncating mutations were retained and all synonymous mutations were discarded. All missense mutations extracted were used to batch query PolyPhen-2 (v2.2.2r395)⁷ which uses multiple sequence alignments of sequences from UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011). Only those mutations scoring as “probably damaging” were retained and combined with truncating and frameshift mutations for a final analysis. There were very few mutational events per gene (1-3) and thus the data are represented as “absent” or “present” in any cancer, in ovarian cancer, and in breast cancer (**Supplementary fig. 10**).

In addition to point mutations we also extracted other structural alterations from COSMIC to annotate each locus including: copy number change, loss of heterozygosity, amplification, homozygous deletions (% range of samples with copy number gains from PICNIC analysis of Affymetrix SNP6.0 micro array data), and structural variant breakpoints.

Expression sequence tag profiles

Expression sequence tag (EST) profiles were obtained from NCBI Unigene (<http://www.ncbi.nlm.nih.gov/unigene>) for all genes in the 1Mb region around the most significant SNP. Approximate expression patterns are inferred from EST counts in cDNA library sources. Expression is depicted as transcripts per million (**Supplementary fig. 10**). Data were extracted for normal tissues (mammary gland and ovary) as well as for disease states (breast tumor, ovarian tumor).

Analysis of Regulatory Regions

General strategy

In order to identify potential regulatory regions active in ovarian epithelial cells we generate a series of contiguous fragments of ~2 kb size spanning approximately 40 kb surrounding the most significant SNP in each region. We defined the bounds of each region based on LD derived from HapMap CEU data populations of European ancestry (Rel 27 Phasell-III, Feb09, on NCBI B36 assembly, dbSNP b126). These

contiguous fragments, called tiling clones, were cloned in a vector designed to test for the presence of enhancer regions. Constructs containing tiling clones were then transfected into ovarian epithelial cells. We attempted amplification of each locus; however, the 10p12 and 17q21 loci present with repetitive regions or structures with tendency to form hairpin structures and were not amplified efficiently..

Primer Design and PCR

The hg18 human genome assembly provided the sequence for primer design. PCR Tiler (<http://pcrtiler.alaingervais.org:8080/PCRTiler/>) created primer sets separated at approximately every 1,800 base pairs. The sizes of amplicons ranged from 1800-2300 base pairs. Gateway cloning forward and reverse *attB* sites were added to the 5' end of the forward and reverse primers, respectively. Additionally, *Mlu*I and *Xho*I restriction sites were added to the 5' end of the forward and reverse primer respectively to allow screening of clones. PCR was performed using 0.02 Units/uL of KOD Polymerase, 1X KOD Polymerase Buffer (Stratagene), 1mM MgSO₄, 40 nM dNTPs (Stratagene), 0.4 pMoles/uL of forward and reverse primers, and 200 ng of the corresponding Bacterial Artificial Chromosome (BAC)(Empire Genomics) containing the region of interest as a template. PCR parameters were set to 1 cycle at 95°C for 5 min followed by 30 cycles at 95°C for 1 min, 50°C for 1 min, and 72°C for 3 min; 1 cycle at 72°C for 5 min. PCR products were separated on a 1% agarose gel, appropriate size bands were excised from the gel and were gel purified using Qiaquick Gel Purification Kit (Qiagen). The primers used to amplify the 8q21 region are given in **Supplementary table 6**.

Cloning and Plasmids

The pGL3-Promoter vector (Promega) which expresses firefly luciferase was modified to insert *attP* recombination sites flanking the *cddB* gene between the recombination sites. Primers KPNI M13 FWD and SAC1 M13 REV (**Supplementary table 6**) were used to PCR amplify the *attP* recombination sites and CDDB gene from the pDonr221 vector (Invitrogen). The PCR product was gel purified using Qiaquick Gel Extraction Kit (Qiagen). This region was then cloned into pCR-Script via KpnI and SacI restriction sites. The *attP* recombination sites and CDDB gene were then sub-cloned into pGL3-Promoter vector via KpnI and SacI creating the plasmid further referred to as pGL3-BP. To create a positive control for our studies, the SV40 enhancer from pGL3-Control vector (Promega) was amplified with the following primers, 2KB TEST GC enh FWD and 2KB TEST GC enh REV (**Supplementary table 6**), which include the gateway cloning *attB* sites. The SV40 Enhancer was then cloned using the gateway technology (Invitrogen). Positive clones were then tested by restriction digest with *Mlu*I and *Xho*I. The same procedure was used for all PCR products obtained from the BAC containing the 8q21 locus.

Cell Culture and Transfection

We used a normal epithelial ovarian cell lines immortalized with TERT and *MYC* (4C2). 4C2 cells were plated at 5×10^3 cells in 50 μ Ls of NOSE media per well in 96 well plates. Twenty four hours later 100 ngs of plasmids containing the SV40 promoter and potential enhancer regions and 0.5 ng of the internal control, pRL-CMV (Promega; a vector expressing *Renilla* luciferase driven by a constitutive promoter) were transfected using Fugene HD (Promega). Each construct was transfected in eight replicates. pGL3-BP (empty vector) and pGL3-BP-Enhancer were used as negative and positive controls, respectively. Cells were harvested 24 hours post transfection and luciferase expression was measured using a dual-glo luciferase assay (Promega).

Analysis

Firefly luciferase levels were normalized using the internal control (*Renilla* luciferase) for each individual well to determine the relative luciferase levels. The relative luciferase levels of each construct were divided by the average relative luciferase levels of pGL3-BP to determine the relative luciferase levels fold change over the empty vector. An unpaired two tailed t-test between each construct and pGL3-BP was then used to determine which constructs had a significant change in luciferase activity compared to the empty vector.

Co-expression networks (Supplementary fig. 11)

Gene co-expression networks were generating using Genemania (<http://www.genemania.org/>)⁸. Networks were generating enabling all co-expression networks (n=154) and using a query-dependent weighting assigned based on the query genes (the weights are chosen automatically using linear regression). Networks were generating with only the query genes or with the query genes plus the 20 most connected additional genes in the datasets (most data is from Gene Expression Omnibus). Individual files with details about which specific dataset showing each interaction and weighted scores are available upon request.

Two genes appear as linked when their expression levels are similar across conditions in a study. A link represents a weighted interaction network where each pair of genes is assigned an association weight (zero indicating no interaction, or a positive value that reflects the strength of interaction or the reliability the observation).

Acknowledgements

Supplementary note

Funding

D.F.E. is a Principal Research Fellow of Cancer Research UK. G.C.-T. and P.M.W. are supported by the National Health and Medical Research Council. B.K. holds an American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN). L.E.K. is supported by a Canadian Institutes of Health Research Investigator award (MSH-87734).

Funding of the constituent studies was provided by the American Cancer Society (CRTG-00-196-01-CCE); the California Cancer Research Program (00-01389V-20170, N01-CN25403, 2II0200); the Canadian Institutes for Health Research (MOP-86727); Cancer Council Victoria; Cancer Council Queensland; Cancer Council New South Wales; Cancer Council South Australia; Cancer Council Tasmania; Cancer Foundation of Western Australia; the Cancer Institute of New Jersey; Cancer Research UK (C490/A6187, C490/A10119, C490/A10124, C536/A13086, C536/A6689); the Celma Mastry Ovarian Cancer Foundation ; the Danish Cancer Society (94-222-52); the ELAN Program of the University of Erlangen-Nuremberg; the Eve Appeal; the Helsinki University Central Hospital Research Fund; Helse Vest; Imperial Experimental Cancer Research Centre (C1312/A15589); the Norwegian Cancer Society; the Norwegian Research Council; the Ovarian Cancer Research Fund; Nationaal Kankerplan of Belgium; Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health Labour and Welfare of Japan; the L & S Milken Foundation; the Polish Ministry of Science and Higher Education (4 PO5C 028 14, 2 PO5A 068 27); Malaysian Ministry of Higher Education (UM.C/HIR/MOHE/06) and Cancer Research Initiatives Foundation; the Roswell Park Cancer Institute Alliance Foundation; the US National Cancer Institute (K07-CA095666, K07-CA143047, K22-CA138563, N01-CN55424, N01-PC067010, N01-PC035137, P01-CA017054, P01-CA087696, P30-CA15083, P50-CA105009, P50-CA136393, R01-CA014089, R01-CA016056, R01-CA017054, R01-CA049449, R01-CA050385, R01-CA054419, R01-CA058598, R01-CA058860, R01-CA061107, R01-CA061132, R01-CA063682, R01-CA064277, R01-CA067262, R01-CA071766, R01-CA074850, R01-CA076016, R01-CA080742, R01-CA080978, R01-CA083918, R01-CA087538, R01-CA092044, R01-095023, R01-CA106414, R01-CA122443, R01-CA112523, R01-CA114343, R01-CA126841, R01-CA136924, R01-CA149429, R03-CA113148, R03-CA115195, R37-CA070867, R37-CA70867, U01-CA069417, U01-CA071966 and Intramural research funds); the US Army Medical Research and Material Command (DAMD17-98-1-8659, DAMD17-01-1-0729, DAMD17-02-1-0666, DAMD17-02-1-0669, W81XWH-10-1-0280); the National Health and Medical Research Council of Australia (199600 and 400281); the German Federal Ministry of Education and Research of Germany Programme of Clinical Biomedical Research (01 GB 9401); the state of Baden-Württemberg through Medical Faculty of the University of Ulm (P.685); the Minnesota Ovarian Cancer Alliance; the Mayo Foundation; the Fred C. and Katherine B. Andersen Foundation; the Lon V. Smith Foundation (LVS-39420); the Oak Foundation; the OHSU Foundation; the Mermaid I project; the Rudolf-Bartling Foundation; the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge, Imperial College London, University College Hospital “Womens Health Theme” and the Royal Marsden Hospital; WorkSafeBC

This study would not have been possible without the contributions of the following: G. Peuteman, T. Van Brussel and D. Smeets (BEL); the staff of the genotyping unit, S LaBoissière and F Robidoux (Genome Quebec); U. Eilber and T. Koehler (GER); L. Gacucova (HMO); P. Schürmann, F. Kramer, W. Zheng, T.-W. Park-Simon, K. Beer-Grondke and D. Schmidt (HJO); G.S. Keeney, S. Windebank, C. Hilker and J. Vollenweider (MAY); the state cancer registries of AL, AZ, AR, CA, CO, CT, DE, FL, GA, HI, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WYL (NHS); L. Paddock, M. King, U. Chandran, A. Samoila, and Y. Bensman (NJO); L. Brinton, M. Sherman, A. Hutchinson, N. Szeszenia-Dabrowska, B. Peplonska, W. Zatonski, A. Soni, P. Chao and M. Stagner (POL); C. Luccarini, P. Harrington the SEARCH team and ECRIC (SEA); the Scottish Gynaecological Clinical Trials group and SCOTROC1 investigators (SRO); W-H. Chow, Y-T. Gao (SWH); I. Jacobs, M. Widschwendter, E. Wozniak, N. Balogun, A. Ryan and J. Ford (UKO); Carole Pye (UKR); A. Amin Al Olama, K. Michilaidou, K. Kuchenbäker (COGS).

Members of the Australian Cancer Study are listed below

Investigators : David C. Whiteman MBBS, PhD, Penelope M. Webb MA, D Phil, Adele C. Green MBBS, PhD, Nicholas K. Hayward PhD, Peter G. Parsons PhD, David M. Purdie PhD

Clinical collaborators : B. Mark Smithers FRACS, David Gotley FRACS PhD, Andrew Clouston FRACP PhD, Ian Brown FRACP

Project Manager: Suzanne Moore RN, MPH.

Database : Karen Harrap BIT, Troy Sadkowski BIT.

Research Nurses: Suzanne O'Brien RN MPH, Ellen Minehan RN, Deborah Roffe RN, Sue O'Keefe RN, Suzanne Lipshut RN, Gabby Connor RN, Hayley Berry RN, Frances Walker RN, Teresa Barnes RN, Janine Thomas RN, Linda Terry RN MPH, Michael Connard B Sc, Leanne Bowes B Sc, MaryRose Malt RN, Jo White RN.

Clinical Contributors:

Australian Capital Territory: Charles Mosse FRACS, Noel Tait FRACS

New South Wales: Chris Bambach FRACS, Andrew Biankan FRACS, Roy Brancatisano FRACS, Max Coleman FRACS, Michael Cox FRACS, Stephen Deane FRACS, Gregory L. Falk FRACS, James Gallagher FRACS, Mike Hollands FRACS, Tom Hugh FRACS, David Hunt FRACS, John Jorgensen FRACS, Christopher Martin FRACS, Mark Richardson FRACS, Garrett Smith FRACS, Ross Smith FRACS, David Storey FRACS

Queensland: John Avramovic FRACS, John Croese FRACP, Justin D'Arcy FRACS, Stephen Fairley FRACP, John Hansen FRACS, John Masson FRACP, Les Nathanson FRACS, Barry O'Loughlin FRACS, Leigh Rutherford FRACS, Richard Turner FRACS, Morgan Windsor FRACS

South Australia: Justin Bessell FRACS, Peter Devitt FRACS, Glyn Jamieson FRACS, David Watson FRACS

Victoria: Stephen Blamey FRACS, Alex Boussioutas FRACP, Richard Cade FRACS, Gary Crosthwaite FRACS, Ian Faragher FRACS, John Gribbin FRACS, Geoff Hebbard FRACP, George Kiroff FRACS, Bruce Mann FRACS, Bob Millar FRACS, Paul O'Brien FRACS, Robert Thomas FRACS, Simon Wood FRACS

Western Australia: Steve Archer FRACS, Kingsley Faulkner FRACS, Jeff Hamdorf FRACS

Members of the Australian Ovarian Cancer Study are listed below

R Stuart-Harris; NSW- F Kirsten, J Rutovitz, P Clingan, A Glasgow, A Proietto, S Braye, G Otton, J Shannon, T Bonaventura, J Stewart, S Begbie, M Friedlander, D Bell, S Baron-Hay, A Ferrier (*dec.*), G Gard, D Nevell, N Pavlakis, S Valmadre, B Young, C Camaris, R Crouch, L Edwards, N Hacker, D Marsden, G Robertson, P Beale, J Beith, J Carter, C Dalrymple, R Houghton, P Russell, L Anderson, M Links, J Grygiel, J Hill, A Brand, K Byth, R Jaworski, P Harnett, R Sharma, G Wain; QLD- D Purdie, D Whiteman, B Ward, D Papadimos, A Crandon, M Cummings, K Horwood. A Obermair, L Perrin, D Wyld, J Nicklin; SA- M Davy, MK Oehler, C Hall, T Dodd, T Healy, K Pittman, D Henderson, J Miller, J Pierdes, A Achan; TAS- P Blomfield, D Challis, R McIntosh, A Parker; VIC- B Brown, R Rome, D Allen, P Grant, S Hyde, R Laurie M Robbie, D Healy, T Jobling, T Manolitsas, J McNealage, P Rogers, B Susil, E Sumithran, I Simpson, I Haviv, K Phillips, D Rischin, S Fox, D Johnson, S Lade, P Waring, M Loughrey, N O'Callaghan, B Murray, L Mileshkin, P Allan; V Billson, J Pyman, D Neesham, M Quinn, A Hamilton, C Underhill, R Bell, LF Ng, R Blum, V Ganju; WA- I Hammond, A McCartney (*dec.*), C Stewart, Y Leung, M Buck, N Zeps (WARTN)

REFERENCES

1. Permuth-Wey, J. et al. LIN28B Polymorphisms Influence Susceptibility to Epithelial Ovarian Cancer. *Cancer Res.* **71**, 3896-3903 (2011).
2. Song, H. et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat.Genet.* **41**, 996-1000 (2009).
3. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am.J.Hum.Genet.* **78**, 629-44 (2006).
4. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat.Genet.* **39**, 906-13 (2007).
5. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am.J.Hum.Genet.* **82**, 290-303 (2008).
6. Forbes, S.A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945-50 (2011).
7. Adzhubei, I.A. et al. A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-9 (2010).
8. Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* **9 Suppl 1**, S4 (2008).