Supplementary Table 1a. Participating Sites and Samples from Breast Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Cases** | **Control** | **White** | **African** | **Asian** | **Other** |
| **Geno centre** | **Acronym** | **Principal Investigator** | **Country** | **Design** | **Female** | **Female** | **Cases + Control** | **Cases + Control** | **Cases + Control** | **Cases + Control** |
| CIDR | 2SISTER | Jack Taylor | USA | Family-based study | 1130 | 0 | 1082 | 48 | 0 | 0 |
| Genome Quebec | ABCFS | John Hopper | Australia | Population-based case-control study | 1120 | 189 | 1309 | 0 | 0 | 0 |
| Cambridge | ABCS | Marjanka Schmidt | Netherlands | Hospital-based consecutive cases; population-based controls | 270 | 187 | 459 | 0 | 0 | 0 |
| Genome Quebec | ABCTB | Jane Carpenter | Australia | Hospital based multi site newly diagnosed breast cancer case | 687 | 189 | 1062 | 0 | 0 | 0 |
| Cambridge | ACP | Kenneth Muir | Thailand | Hospital based case-control study | 753 | 375 | 0 | 0 | 1395 | 0 |
| Genome Quebec | BBCC | Peter Fasching | Germany | Hospital based cases; population based controls | 442 | 642 | 695 | 0 | 0 | 0 |
| Cambridge | BBCS | Julian Peto | UK | Cancer registry and National Cancer Research network (NCRN) based cases; population based controls | 122 | 253 | 564 | 0 | 0 | 0 |
| Cambridge | BCEES | Jennifer Stone | Australia | Population-based case-control study | 785 | 442 | 1620 | 0 | 0 | 0 |
| Genome Quebec | BCFR-NY | Mary-Beth Terry | USA | Clinic-based family study | 465 | 835 | 493 | 0 | 0 | 0 |
| Genome Quebec | BCFR-PA | Mary Daly | USA | Clinic-based family study | 141 | 28 | 141 | 0 | 0 | 0 |
| Genome Quebec | BCFR-Utah | David Goldgar | USA | Clinic-based family study | 103 | 0 | 103 | 0 | 0 | 0 |
| Genome Quebec | BCINIS | Gadi Rennert | Israel | Population-based case-control | 1439 | 0 | 2163 | 0 | 0 | 0 |
| Cambridge | BREOGAN | Manuela Gago | Spain | Population-based case-control | 1379 | 724 | 2108 | 0 | 0 | 0 |
| Genome Quebec | BSUCH | Barbara Burwinkel | Germany | Hospital based cases;healthy blood donator controls | 269 | 729 | 437 | 0 | 0 | 0 |
| Genome Quebec | CAMA | Elad Ziv | Mexico | Population-based case-control study | 709 | 168 | 0 | 0 | 0 | 1366 |
| Genome Quebec | CBCS | Kristan Aranson | Canada | Population-based case-control study | 1025 | 657 | 1514 | 0 | 506 | 0 |
| Cambridge | CCGP | Manolis Saloustros | Greece | Hospital-based case-control study | 683 | 995 | 1015 | 0 | 0 | 0 |
| Genome Quebec | CECILE | Pascal Guenel | France | Population-based case-control study | 306 | 332 | 465 | 0 | 0 | 0 |
| Copenhagen | CGPS | Stig Bojesen | Denmark | Population-based case-control study | 1411 | 159 | 2127 | 0 | 0 | 0 |
| Genome Quebec | COLBCCC | Diana Torres | Colombia | Hospital-based case-control study | 633 | 716 | 0 | 0 | 0 | 1194 |
| CIDR | CPSII | Mia Gaudet | USA | Nested case-control study | 3055 | 561 | 6087 | 0 | 0 | 0 |
| Genome Quebec | CTS | Hoda Anton-Culver | USA | Nested case-control study | 1075 | 3032 | 1616 | 0 | 0 | 0 |
| Cambridge | DIETCOMPLYF | Miriam Dwek | UK | Multi-centre prospective case-only study | 711 | 541 | 711 | 0 | 0 | 0 |
| CIDR | EPIC | Elio Riboli | MULTIPLE | Nested case-control study | 3850 | 0 | 7497 | 0 | 0 | 0 |
| Genome Quebec | ESTHER | Hermann Brenner | Germany | Population-based case-control study | 296 | 3647 | 483 | 0 | 0 | 0 |
| Cambridge | GC-HBOC | Alfons Meindl | Germany | Clinic-based family study | 3647 | 187 | 5240 | 0 | 0 | 0 |
| Genome Quebec | GENICA | Hiltrud Brauch | Germany | Population-based case-control study | 460 | 1593 | 744 | 0 | 0 | 0 |
| Genome Quebec | GeparSixto | Peter Fasching | Germany | Clincal Trial | 390 | 284 | 390 | 0 | 0 | 0 |
| Cambridge | GESBC | Jenny Chang-Claude | Germany | Population-based case-control study | 360 | 0 | 541 | 0 | 0 | 0 |
| Cambridge | HaBCS | Thilo Doerk | Germany | Hospital-based case-control study | 934 | 181 | 1800 | 0 | 0 | 0 |
| Genome Quebec | HCSC | Atocha Romero | Spain | Hospital-based case-control study | 427 | 866 | 427 | 0 | 0 | 0 |
| Genome Quebec | HEBCS | Heli Nevanlinna | Finland | Hospital-based case-control study + additional familial cases | 281 | 0 | 458 | 0 | 0 | 0 |
| Genome Quebec | HERPACC | Keitaro Matsuo | Japan | Hospital-based case-control study | 282 | 177 | 0 | 0 | 565 | 0 |
| Genome Quebec | HKBCS | Ava Kwong | Hong Kong | Hospital-based case-control study | 564 | 283 | 0 | 0 | 1018 | 0 |
| Cambridge | HMBCS | Thilo Doerk | Belarus | Hospital based cases; population based controls | 212 | 454 | 463 | 0 | 0 | 0 |
| Cambridge | HUBCS | Thilo Doerk | Russia | Hospital based cases; population based controls | 312 | 251 | 538 | 0 | 0 | 0 |
| Genome Quebec | KARBAC | Annika Lindblom | Sweden | Population and hospital-based cases; geographically matched controls | 506 | 226 | 506 | 0 | 0 | 0 |
| Cambridge | KBCP | Arto Mannermaa | Finland | Population-based case-control study | 557 | 0 | 802 | 0 | 0 | 0 |
| Genome Quebec | KOHBRA | Sue Park | Korea | Hospital-based case-control study | 1464 | 245 | 0 | 0 | 2129 | 0 |
| Genome Quebec | LMBC | Diether Lambrechts | Belgium | Hospital-based case-control study | 806 | 665 | 2075 | 0 | 0 | 0 |
| Cambridge | MaBCS | Thilo Doerk | Macedonia | Hospital-based case-control study | 90 | 1269 | 183 | 0 | 0 | 0 |
| Genome Quebec | MARIE | Jenny Chang-Claude | Germany | Population-based case-control study | 512 | 93 | 801 | 0 | 0 | 0 |
| Genome Quebec | MBCSG | Paolo Radice | Italy | Clinic-based recruitment of familial cases; population-based controls | 788 | 289 | 1154 | 0 | 0 | 0 |
| Genome Quebec | MCBCS | Fergus Couch | USA | Hospital-based case-control study | 370 | 366 | 551 | 0 | 0 | 0 |
| Mayo | MCBCS (Mayo) | Fergus Couch | USA | Hospital-based case-control study | 558 | 181 | 598 | 0 | 0 | 0 |
| Cambridge | MCCS | Graham Giles | Australia | Nested case-control study | 862 | 40 | 1677 | 0 | 0 | 0 |
| CIDR | MCCS (CIDR) | Graham Giles | Australia | Nested case-control study | 190 | 815 | 355 | 0 | 0 | 0 |
| CIDR | MEC | Chris Haiman | USA | Nested case-control study | 674 | 725 | 1399 | 0 | 0 | 0 |
| Cambridge | MISS | Hakan Olsson | Sweden | Nested case-control study | 703 | 731 | 2248 | 0 | 0 | 0 |
| Genome Quebec | MMHS | Celine Vachon | USA | Nested case-control study | 306 | 1545 | 541 | 0 | 0 | 0 |
| Mayo | MMHS (Mayo) | Celine Vachon | USA | Nested case-control study | 78 | 235 | 1481 | 0 | 0 | 0 |
| Genome Quebec | MTLGEBCS | Mark Goldberg | Canada | Hospital-based case-control study | 343 | 1403 | 513 | 0 | 0 | 0 |
| Genome Quebec | MYBRCA | Soo Hwang-Teo | Malaysia | Hospital-based case-control study | 845 | 170 | 0 | 0 | 2103 | 0 |
| Genome Quebec | NBCS | Vessela Kristensen | Norway | Hospital-based case-control study | 1285 | 1258 | 1286 | 0 | 0 | 0 |
| CIDR | NBHS | Wei Zheng | USA | Population-based case-control study | 887 | 1 | 1329 | 354 | 0 | 0 |
| Genome Quebec | NC-BCFR | Esther John | USA | Population-based familial case-control study | 1264 | 796 | 960 | 254 | 503 | 0 |
| Genome Quebec | NGOBCS | Motoki Iwasaki | Japan | Hospital-based case-control study | 369 | 199 | 0 | 0 | 735 | 0 |
| CIDR | NHS | Peter Kraft | USA | Nested case-control study | 1594 | 366 | 3402 | 0 | 0 | 0 |
| CIDR | NHS2 | Peter Kraft | USA | Nested case-control study | 1609 | 1808 | 3517 | 0 | 0 | 0 |
| Genome Quebec | OFBCR | Irene Andrulis | Canada | Population-based familial case-control study | 1669 | 1908 | 2045 | 0 | 0 | 0 |
| Cambridge | ORIGO | Peter Devilee | Netherlands | Hospital-based case-control study | 1059 | 376 | 1721 | 0 | 0 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NCI | PBCS | Montse Garcia-Closas | Poland | Population-based case-control study | 1931 | 662 | 3976 | 0 | 0 | 0 |
| Cambridge | pKARMA | Kamila Czene | Sweden | Population-based case-control study | 2993 | 2045 | 9080 | 0 | 0 | 0 |
| NCI | PLCO | Robert Hoover | USA | Nested case-control study | 869 | 6087 | 1727 | 0 | 0 | 0 |
| Cambridge | POSH | Diana Eccles | UK | Clinic-based case-only study | 1091 | 858 | 1091 | 0 | 0 | 0 |
| Genome Quebec | PreFace | Peter Fasching | Germany | Clinical Trial | 991 | 0 | 991 | 0 | 0 | 0 |
| Cambridge | RBCS | Maartje Hooning | Netherlands | Hospital-based case-control study | 475 | 0 | 717 | 0 | 0 | 0 |
| Genome Quebec | SBCGS | Wei Zheng | China | Population-based case-control study | 840 | 242 | 0 | 0 | 1775 | 0 |
| Cambridge | SEARCH | Paul Pharoah | UK | Population-based case-control study | 4062 | 1828 | 6746 | 0 | 0 | 0 |
| Genome Quebec | SEBCS | Daehee Kang | Korea | Hospital-based case-control study | 1103 | 1791 | 0 | 0 | 2210 | 0 |
| Genome Quebec | SGBCC | Mikael Hartman | Singapore | Hospital-based cases, population based controls | 927 | 1107 | 0 | 0 | 1665 | 0 |
| CIDR | SISTER | Jack Taylor | USA | Population-based family study | 2187 | 738 | 3609 | 325 | 0 | 0 |
| Genome Quebec | SKKDKFZS | Uta Hamann | Germany | Hospital-based case-only study | 1097 | 1747 | 1097 | 0 | 0 | 0 |
| Cambridge | SMC | Alicja Wolk | Sweden | Nested case-control study | 1504 | 0 | 2213 | 0 | 0 | 0 |
| Genome Quebec | SuccessB | Peter Fasching | Germany | Clinical Trial | 440 | 709 | 440 | 0 | 0 | 0 |
| Genome Quebec | SuccessC | Peter Fasching | Germany | Clinical Trial | 1343 | 0 | 1343 | 0 | 0 | 0 |
| Genome Quebec | SZBCS | Jan Lubinski | Poland | Hospital-based case-control study | 387 | 0 | 561 | 0 | 0 | 0 |
| Mayo | TNBCC | Fergus Couch | MULTIPLE | Case series from multiple countries | 1439 | 69 | 1508 | 0 | 0 | 0 |
| Genome Quebec | TWBCS | Chen-Yang Shen | Taiwan | Hospital-based case-control study | 551 | 0 | 0 | 0 | 807 | 0 |
| Genome Quebec | UCIBCS | Hoda Anton-Culver | USA | Population-based case-control study | 507 | 256 | 767 | 0 | 0 | 0 |
| Cambridge | UKBGS | Anthony Swerdlow | UK | Nested case-control study | 1632 | 260 | 2337 | 0 | 0 | 0 |
| UKOPS | UKOPS | Usha Menon | UK | Population-based cohort |  | 705 | 976 | 0 | 0 | 0 |
| CIDR | WAABCS | Fumni Olopade | MULTIPLE | Hospital-based case-control study | 315 | 976 | 0 | 626 | 0 | 0 |
| CIDR | WHI | Ross Prentice | USA | Nested case-control study | 4937 | 311 | 9555 | 0 | 0 | 0 |

Supplementary Table 1b. Participating Sites and Samples from BRCA1 and BRCA2 Carrier Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Females** |  |  | **Males** |  |  | **Female and Male**  **Case + Control** | | |  |
| **Genotyping** | **Acronym** | **PI** | **Country** | **Case + Control** | **BRCA1 \*** | **BRCA2** | **Case + Control** | **BRCA1** | **BRCA2** | **White ^** | **Black** | **Asian** | **Other** |
| MAYO | BCFR-AU | Melissa Southey | AUSTRALIA | 81 | 40 | 41 | 1 | 0 | 1 | 82 | 0 | 0 | 0 |
| CIDR | BCFR-NC | Esther John | USA | 18 | 7 | 11 | 8 | 3 | 5 | 24 | 0 | 0 | 2 |
| CIDR | BCFR-NY | Mary Beth Terry | USA | 124 | 67 | 57 | 0 | 0 | 0 | 124 | 0 | 0 | 0 |
| GQ | BCFR-ON | Irene Andrulis | CANADA | 219 | 126 | 93 | 18 | 7 | 11 | 237 | 0 | 0 | 0 |
| CIDR | BCFR-PA | Mary Daly | USA | 52 | 45 | 7 | 0 | 0 | 0 | 52 | 0 | 0 | 0 |
| CIDR | BCFR-UT | David Goldgar | USA | 253 | 130 | 123 | 47 | 17 | 30 | 298 | 0 | 0 | 2 |
| CIDR | BFBOCC | Ramunas Janavicius/  Liene Nikitinia-Zake | LITHUANIA/  LATVIA | 267 | 249 | 18 | 4 | 4 | 0 | 271 | 0 | 0 | 0 |
| CIDR | BIDMC | Nadine Tung | USA | 140 | 86 | 54 | 0 | 0 | 0 | 140 | 0 | 0 | 0 |
| MAYO | BMBSA | Lizette Jansen van Rensburg | SOUTH AFRICA | 203 | 59 | 144 | 0 | 0 | 0 | 203 | 0 | 0 | 0 |
| GQ | BRICOH | Susan Neuhausen | USA | 320 | 181 | 139 | 82 | 19 | 63 | 400 | 0 | 0 | 2 |
| MAYO | CBCS | Thomas van Overeem Hansen | DENMARK | 343 | 203 | 140 | 1 | 1 | 0 | 344 | 0 | 0 | 0 |
| CIDR | CNIO | Javier Benitez/  Ana Osorio | SPAIN | 128 | 66 | 62 | 5 | 1 | 4 | 133 | 0 | 0 | 0 |
| CIDR | COH | Jeffrey Weitzel | USA | 652 | 431 | 221 | 5 | 3 | 2 | 388 | 0 | 0 | 269 |
| CAM | CONSIT TEAM | Paolo Radice | ITALY | 870 | 550 | 320 | 183 | 83 | 100 | 1053 | 0 | 0 | 0 |
| CIDR | DEMOKRITOS | Koulis Yannoukakos | GREECE | 271 | 235 | 36 | 7 | 4 | 3 | 278 | 0 | 0 | 0 |
| CIDR | DFCI | Judy Garber | USA | 283 | 150 | 133 | 0 | 0 | 0 | 283 | 0 | 0 | 0 |
| CAM | DKFZ | Ute Hamann | GERMANY | 85 | 60 | 25 | 7 | 5 | 2 | 92 | 0 | 0 | 0 |
| CAM | EMBRACE | Douglas Easton | UK/IRELAND | 3441 | 1749 | 1692 | 305 | 78 | 227 | 3744 | 0 | 0 | 2 |
| CIDR | FCCC | Andrew Godwin | USA | 123 | 78 | 45 | 18 | 3 | 15 | 141 | 0 | 0 | 0 |
| CAM | FPGMX | Ana Vega | SPAIN | 190 | 112 | 78 | 0 | 0 | 0 | 189 | 1 | 0 | 0 |
| CAM | GC-HBOC | Rita Schmutzler | GERMANY | 3039 | 1928 | 1111 | 162 | 44 | 118 | 3201 | 0 | 0 | 0 |
| GQ | GEMO | Sylvie Mazoyer/  Dominique Stoppa-Lyonnet/  Fabienne Lesueur | FRANCE/USA | 2459 | 1501 | 958 | 69 | 10 | 59 | 2528 | 0 | 0 | 0 |
| CIDR | GEORGETOWN | Claudine Isaacs | USA | 15 | 15 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 |
| CAM | G-FAST | Kathleen Claes | BELGIUM | 360 | 195 | 165 | 31 | 0 | 31 | 391 | 0 | 0 | 0 |
| CIDR | HCSC | Trinidad Caldes | SPAIN | 305 | 146 | 159 | 37 | 0 | 37 | 342 | 0 | 0 | 0 |
| CAM | HEBCS | Heli Nevannlina | FINLAND | 259 | 126 | 133 | 33 | 8 | 25 | 292 | 0 | 0 | 0 |
| CAM | HEBON | Matti Rookus | NETHERLANDS | 1528 | 901 | 627 | 15 | 8 | 7 | 1543 | 0 | 0 | 0 |
| CIDR | HRBCP | Ava Kwong | HONG KONG | 120 | 51 | 69 | 0 | 0 | 0 | 0 | 0 | 120 | 0 |
| MAYO | HUNBOCS | Edith Olah | HUNGARY | 398 | 282 | 116 | 26 | 8 | 18 | 424 | 0 | 0 | 0 |
| MAYO | HVH | Orland Diez | SPAIN | 256 | 120 | 136 | 20 | 2 | 18 | 276 | 0 | 0 | 0 |
| CIDR | ICO | Conxi Lazaro | SPAIN | 648 | 288 | 360 | 73 | 8 | 65 | 721 | 0 | 0 | 0 |
| GQ | IHCC | Jan Lubinski/Ania Jakabowska | POLAND | 205 | 205 | 0 | 0 | 0 | 0 | 205 | 0 | 0 | 0 |
| CAM | ILUH | Rosa Barkardottir | ICELAND | 147 | 0 | 147 | 43 | 0 | 43 | 190 | 0 | 0 | 0 |
| GQ | INHERIT | Jacques Simard | CANADA (QUEBEC) | 183 | 96 | 87 | 0 | 0 | 0 | 183 | 0 | 0 | 0 |
| CAM | IOVHBOCS | Marco Montagna | ITALY | 374 | 206 | 168 | 21 | 1 | 20 | 395 | 0 | 0 | 0 |
| CAM | IPOBCS | Manuel Teixeira | PORTUGAL | 281 | 117 | 164 | 12 | 0 | 12 | 293 | 0 | 0 | 0 |
| CIDR/MAYO | KCONFAB | Georgia Chenevix-Trench | AUSTRALIA | 1594 | 892 | 702 | 272 | 68 | 204 | 1866 | 0 | 0 | 0 |
| GQ | KOHBRA | Sue Park | KOREA | 502 | 194 | 308 | 65 | 20 | 45 | 1 | 0 | 566 | 0 |
| CIDR | KUMC | Priyanka Sharma | USA | 44 | 29 | 15 | 0 | 0 | 0 | 44 | 0 | 0 | 0 |
| CIDR | MAYO | Fergus Couch | USA | 387 | 258 | 129 | 4 | 2 | 2 | 391 | 0 | 0 | 0 |
| GQ | MCGILL | Mark Tischkowitz | CANADA (QUEBEC) | 88 | 54 | 34 | 0 | 0 | 0 | 88 | 0 | 0 | 0 |
| CIDR | MSKCC | Ken Offit | USA | 772 | 396 | 376 | 52 | 14 | 38 | 824 | 0 | 0 | 0 |
| CIDR/MAYO | MUV | Christian Singer | AUSTRIA | 806 | 541 | 265 | 22 | 4 | 18 | 828 | 0 | 0 | 0 |
| CIDR | NAROD | Steven Narod | CANADA | 380 | 286 | 94 | 0 | 0 | 0 | 301 | 0 | 32 | 47 |
| CIDR | NCI | Mark Greene | USA | 236 | 153 | 83 | 18 | 10 | 8 | 254 | 0 | 0 | 0 |
| CIDR | NNPIO | Evgeny Imyanitov | RUSSIA | 75 | 73 | 2 | 0 | 0 | 0 | 75 | 0 | 0 | 0 |
| CIDR | NORTHSHORE | Peter Hulick | USA | 139 | 82 | 57 | 0 | 0 | 0 | 139 | 0 | 0 | 0 |
| CIDR | NRG\_ONCOLOGY | Mark Greene | USA/  AUSTRALIA | 628 | 332 | 296 | 0 | 0 | 0 | 628 | 0 | 0 | 0 |
| GQ | OCGN | Irene Andrulis | CANADA | 382 | 208 | 174 | 19 | 7 | 12 | 401 | 0 | 0 | 0 |
| CIDR | OSU CCG | Amanda Toland | USA | 197 | 93 | 104 | 14 | 4 | 10 | 211 | 0 | 0 | 0 |
| MAYO | OUH | Mads Thomassen | DENMARK | 1000 | 568 | 432 | 105 | 27 | 78 | 1105 | 0 | 0 | 0 |
| CIDR | PBCS | Maria Caligo | ITALY | 98 | 91 | 7 | 4 | 0 | 4 | 102 | 0 | 0 | 0 |
| CAM/GQ | SEABASS | Soo Hwang-Teo | MALAYSIA | 111 | 61 | 50 | 12 | 9 | 3 | 0 | 0 | 123 | 0 |
| CIDR | SMC | Eitan Friedman | ISRAEL | 254 | 171 | 83 | 0 | 0 | 0 | 254 | 0 | 0 | 0 |
| CIDR | SWE-BRCA | Ake Borg/Johanna Rantala | SWEDEN | 498 | 434 | 64 | 11 | 8 | 3 | 509 | 0 | 0 | 0 |
| CIDR | UCHICAGO | Funmi Olopade | USA | 156 | 98 | 58 | 23 | 8 | 15 | 179 | 0 | 0 | 0 |
| CIDR | UCSF | Robert Nussbaum | USA | 170 | 100 | 70 | 0 | 0 | 0 | 170 | 0 | 0 | 0 |
| CAM | UKGRFOCR | Susan Ramus | UK | 74 | 57 | 17 | 0 | 0 | 0 | 74 | 0 | 0 | 0 |
| CIDR | UPENN | Kate Nathanson | USA | 850 | 489 | 361 | 90 | 44 | 46 | 927 | 7 | 0 | 6 |
| CIDR | UPITT | Darcy Thull | USA | 265 | 158 | 107 | 13 | 1 | 12 | 278 | 0 | 0 | 0 |
| CIDR | UTMDACC | Banu Arun | USA | 122 | 48 | 74 | 0 | 0 | 0 | 122 | 0 | 0 | 0 |
| CIDR/MAYO | VFCTG | Gillian Mitchell | AUSTRALIA | 469 | 244 | 225 | 32 | 13 | 19 | 500 | 0 | 1 | 0 |
| CIDR | WCP | Beth Karlan | USA | 213 | 157 | 56 | 0 | 0 | 0 | 213 | 0 | 0 | 0 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | \* 16 individuals carried both a BRCA1 and a BRCA2 mutation | |  |  |  |  |  |  |  |  |  |  |  |
|  | ^ includes Ashkenazi Jewish carriers | |  |  |  |  |  |  |  |  |  |  |  |

Supplementary Table 1c. Participating Sites and Samples from Colon Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | Cases | | Controls | | White | | African American | | Asian | | Other |  | Ethnicity |
| Genotyping | Acronym | PI | Country | Design | Males | Females | Males | Females | Case No. | Control No. | Case No. | Control No. | Case No. | Control No. | Case No. | Control No. | NA |
| CIDR | MECC | Gad Rennert,  Steve Gruber | Israel | Case-Control | 1735 | 1532 | 1192 | 1055 | 3267 | 2247 |  |  |  |  |  |  |  |
| USC | MECC | Gad Rennert,  Steve Gruber | Israel | Case-Control | 480 | 431 | 326 | 329 | 911 | 655 |  |  |  |  |  |  |  |
| CIDR | MSKCC | Kenneth Offit | USA | Case-Control | 53 | 73 | 0 | 0 | 126 | 0 |  |  |  |  |  |  |  |
| CIDR | SEARCH | Paul Pharoah | United Kingdom | Case-Control | 2516 | 1907 | 137 | 144 | 4423 | 281 |  |  |  |  |  |  |  |
| USC | SEARCH | Paul Pharoah | United Kingdom | Case-Control | 166 | 117 | 0 | 0 | 283 | 0 |  |  |  |  |  |  |  |
| CIDR | SPAIN | Victor Moreno | Spain | Case-Control | 628 | 353 | 499 | 442 | 981 | 941 |  |  |  |  |  |  |  |
| USC | SPAIN | Victor Moreno | Spain | Case-Control | 0 | 0 | 45 | 88 |  | 117 |  |  |  |  |  | 12 | 4 |
| USC | FIRE3 | Heinz-Josef Lenz | Germany | Clinical Trial Phase 3 | 184 | 72 | 0 | 0 | 262 | 0 |  |  |  |  |  |  | 6 |
| USC | TRIBE | Heinz-Josef Lenz | Italy | Clinical Trial Phase 3 | 270 | 177 | 0 | 0 | 447 | 0 |  |  |  |  |  |  |  |
| CIDR | MEC | Loic Le Marchand | USA | Cohort | 210 | 192 | 218 | 206 |  |  |  |  |  |  | 402 | 424 |  |
| CIDR | CFR-Hawaii | Loic LeMarchand | USA | Case-Control | 30 | 19 | 0 | 0 | 49 | 0 |  |  |  |  |  |  |  |
| CIDR | CFR-Australia | John Hopper /  Mark Jenkins | Australia | Case-Control | 139 | 120 | 10 | 15 | 259 | 25 |  |  |  |  |  |  |  |
| CIDR | CFR-Seattle | Polly Newcomb | USA | Case-Control | 529 | 484 | 300 | 298 | 1013 | 598 |  |  |  |  |  |  |  |
| CIDR | CRF-Mayo | Noralane Lindor | USA | Case-Control | 149 | 146 | 0 | 0 | 295 | 0 |  |  |  |  |  |  |  |
| CIDR | CFR-Ontario | Steve Gallinger | Canada | Case-Control | 201 | 174 | 129 | 86 | 375 | 215 |  |  |  |  |  |  |  |
| CIDR | CFR-USC | Graham Casey | USA | Case-Control | 85 | 107 | 0 | 0 | 192 | 0 |  |  |  |  |  |  |  |
| CIDR | Esther/  Verdi | Hermann Brenner | Germany | Case-Control | 300 | 166 | 313 | 166 | 466 | 479 |  |  |  |  |  |  |  |
| CIDR | Kiel | Jochen Hampe / Clemens Schafmeyer | Germany | Case-Control | 635 | 499 | 0 | 0 | 1134 | 0 |  |  |  |  |  |  |  |
| CIDR | ColoCare Germany + Seattle | Cornelia Ulrich / Bill Grady | Germany and USA | Case-Series | 292 | 179 | 18 | 32 | 471 | 50 |  |  |  |  |  |  |  |
| USC | ColoCare | Erin Siegel | USA | Case-Series | 230 | 197 | 0 | 0 | 356 |  | 8 |  | 3 |  | 5 |  | 55 |
| CIDR | NHS2 | Andrew Chan | USA | Cohort | 0 | 127 | 0 | 123 | 127 | 123 |  |  |  |  |  |  |  |
| CIDR | MCCS | Graham Giles | Australia | Cohort | 119 | 109 | 116 | 103 | 228 | 219 |  |  |  |  |  |  |  |
| CIDR | Korea | Wei Zheng | Korea | Case-Control | 2241 | 1292 | 1516 | 2076 |  |  |  |  | 3533 | 3592 |  |  |  |
| CIDR | Shanghai-Mens | Wei Zheng | China | Case-Control | 125 |  | 125 |  |  |  |  |  | 125 | 125 |  |  |  |
| CIDR | Shanghai - Womens | Wei Zheng | China | Case-Control |  | 125 |  | 125 |  |  |  |  | 125 | 125 |  |  |  |
| CIDR | Sweden - Lindblom | Annika Lindblom | Sweden | Cohort | 1768 | 1477 | 1485 | 1371 | 3245 | 2856 |  |  |  |  |  |  |  |

Supplementary Table 1d. Participating Sites and Samples from Endometrial Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Cases** |  | **Controls** |  | **White** | | **African** | | **Asian** | | **Other** | |
| Geno centre | Acronym | Principal Investigator | Country | Design | Female | Male | Female | Male | Cases | Controls | Cases | Controls | Cases | Controls | Cases | Controls |
| CIDR | RENDOCAS | Emma Tham | Sweden | Hospital based study | 574 |  |  |  | 574 |  |  |  |  |  |  |  |
| CIDR | ANECS | Amanda Spurdle | Australia | Population based case-control study | 584 |  |  |  | 584 |  |  |  |  |  |  |  |
| CIDR | BECS | Peter Fasching | Germany | Hospital based study | 223 |  |  |  | 223 |  |  |  |  |  |  |  |
| CIDR | LES | Diether Lambrechts | Belgium | Hospital based case-control study | 542 |  |  |  | 542 |  |  |  |  |  |  |  |
| CIDR | SEARCH | Douglas Easton | UK | Population based case-control study | 957 |  |  |  | 957 |  |  |  |  |  |  |  |
| CIDR | NSECG | Ian Tomlinson | UK | Population based case-control study | 782 |  |  |  | 782 |  |  |  |  |  |  |  |
| CIDR | CAHRES | Per Hall | Sweden | Population based case-control study | 593 |  |  |  | 593 |  |  |  |  |  |  |  |
| CIDR | MOMATEC |  | Norway | Population based case-control study | 850 |  |  |  | 850 |  |  |  |  |  |  |  |
| CIDR | MECS | Ellen Goode | USA | Hospital based case-control study | 271 |  |  |  | 271 |  |  |  |  |  |  |  |

Supplementary Table 1e. Participating Sites and Samples from Lung Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Cases** | | **Controls** | | **White** | | **African American** | | **Asian** | | **Other** |  |
| **Genotyping** | **Acronym** | **PI** | **Country** | **Design** | **Males** | **Females** | **Males** | **Females** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | **Case No.** | **Control No.** |
| CIDR | Norway | Aage Haugen | Norway | Hosp CC | 239 | 100 | 293 | 134 | 339 | 427 |  |  |  |  |  |  |
| CIDR | MDACC | Xifeng Wu | US | Hosp CC | 518 | 507 | 515 | 502 | 1005 | 990 | 4 | 3 | 1 | 1 | 15 | 23 |
| CIDR | HSPH | David Christiani | US | Hosp CC | 1461 | 1632 | 331 | 464 | 3020 | 745 | 42 | 5 | 10 | 7 | 22 | 38 |
| CIDR | Liverpool\_2008 | John Field | UK | nested CC | 62 | 46 | 70 | 48 | 108 | 118 |  |  |  |  |  |  |
| CIDR | Liverpool\_2013 | John Field | UK | nested CC | 193 | 157 | 225 | 177 | 342 | 390 | 1 | 3 |  | 2 | 5 | 5 |
| CIDR | CARET | Chu Chen,  Jen Doherty | US | nested CC | 421 | 191 | 421 | 192 | 578 | 579 | 22 | 22 | 5 | 5 | 7 | 7 |
| CIDR | NELCS | Angeline Andrew | US | Pop CC | 86 | 104 | 80 | 104 | 176 | 179 |  |  | 1 |  | 13 | 5 |
| CIDR | Tampa | Philip Lazarus | US | Hosp CC | 234 | 174 | 233 | 171 | 390 | 365 | 8 | 34 |  |  | 10 | 5 |
| CIDR | Resolucent | Penella J Woll ,  Dawn Teare | UK | family,  Pop CC | 343 | 344 | 173 | 270 | 591 | 390 | 2 |  | 3 |  | 93 | 53 |
| CIDR | ISRAEL | Gad Rennert | Israel | Pop CC | 467 | 264 | 349 | 209 | 731 | 557 |  | 1 |  |  |  |  |
| CIDR | Nijmegen | Lambertus A. Kiemeney | The Netherlands | Pop CC | 266 | 173 | 278 | 179 | 387 | 457 |  |  | 2 |  | 3 |  |
| CIDR | EAGLE | Maria Theresa Landi | Italy | Hosp CC | 1465 | 380 | 1423 | 441 | 1845 | 1864 |  |  |  |  |  |  |
| CIDR | CAPUA | Adonina Tardon | Spain | Hosp CC | 713 | 89 | 678 | 104 | 800 | 780 |  |  |  |  | 2 |  |
| CIDR | EPIC | Mattias Johansson | Europe | nested CC | 761 | 453 | 765 | 470 | 1214 | 1235 |  |  |  |  |  |  |
| CIDR | MEC | Loic Le Marchand | US | nested CC | 551 | 380 | 567 | 402 | 231 | 240 | 147 | 144 | 304 | 316 | 249 | 269 |
| CIDR | MSH-PMH | Rayjean Hung,  Geoffrey Liu | Canada | Clinic CC | 729 | 723 | 501 | 507 | 1446 | 1006 |  |  |  |  | 5 |  |
| CIDR | PLCO | Neil Caporaso | US | nested CC | 1040 | 660 | 917 | 654 | 1550 | 1114 | 86 | 395 | 31 | 33 | 33 | 29 |
| CIDR | MLD | Paul Brennan | Russia | Hosp CC | 833 | 308 | 712 | 426 | 1025 | 1084 |  |  |  |  | 116 | 55 |
| CIDR | Seoul | Yun-Chul Hong | Korea | Hosp CC | 206 | 96 | 199 | 291 |  |  |  |  | 302 | 490 |  |  |
| CIDR | ATBC | Demetrius Albanes | US | nested CC | 1040 |  | 721 |  | 1040 | 721 |  |  |  |  |  |  |
| CIDR | LCRI-DOD | Susanne Arnold | US | Pop CC | 50 | 50 | 65 | 72 | 98 | 133 | 1 | 1 |  |  | 1 | 3 |
| CIDR | MDCS | Jonas Manjer | Sweden | nested CC | 70 | 95 | 79 | 96 | 165 | 175 |  |  |  |  |  |  |
| CIDR | TLC | Matthew B. Schabath | US | case only | 212 | 247 |  |  | 432 |  | 11 |  |  |  | 16 |  |
| CIDR | Vanderbilt2 | Melinda Aldirch | US | Hosp CC | 428 | 370 | 429 | 370 | 740 | 735 | 58 | 56 |  | 1 |  | 7 |
| CIDR | SCHC | Jian-Min Yuan | Singapore | nested CC | 292 | 126 | 291 | 127 |  |  |  |  | 418 | 418 |  |  |
| CIDR | SCS | Jian-Min Yuan | China | nested CC | 178 |  | 325 |  |  |  |  |  | 178 | 325 |  |  |
| CIDR | Canadian screening | Stephen Lam,  Ming-Sound Tsao, Geoff Liu | Canada | nested CC | 117 | 152 | 202 | 267 | 263 | 455 | 2 | 4 | 4 | 9 |  |  |
| CIDR | EPIC | Mattias Johansson | Sweden | nested CC | 123 | 121 | 136 | 133 | 244 | 269 |  |  |  |  |  |  |
| Heidelberg | GLC | Angela Risch | Germany | Hosp CC | 686 | 343 | 514 | 171 | 1029 | 685 |  |  |  |  |  |  |
| Beijing Genome Institute | Nanjing Study | Hongbing Shen | China | Hosp/CC | 665 | 343 |  |  |  |  |  |  | 665 | 343 |  |  |
| Copenhagen | Denmark Population | Stig Bojeson | Denmark | Pop CC | 324 | 333 | 131 | 115 | 657 | 246 |  |  |  |  |  |  |

Supplementary Table 1f. Participating Sites and Samples from Ovarian Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Cases** | **Controls** | | **White** | | **African American** | | **Asian** | | **Other** | |
| **Genotyping** | **Acronym** | **PI** | **Country** | **Design** | **Females** | **Males** | **Females** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | **Case No.** | **Control No.** |
| CIDR | AAS | Joellen Schildkraut,  Patricia Moorman | USA | Case-control | 296 | 0 | 475 | 1 | 0 | 295 | 475 | 0 | 0 | 0 | 0 |
| MAYO | AOCS/  ACS | Georgia Chenevix-Trench, Penelope Webb | Australia | Case-control | 1,504 | 0 | 1,206 | 1420 | 1167 | 4 | 1 | 56 | 18 | 24 | 20 |
| MAYO | AUS | Georgia Chenevix-Trench, Penelope Webb | Australia | Case-control | 112 | 0 | 0 | 106 | 0 | 1 | 0 | 5 | 0 | 0 | 0 |
| MAYO | BAV | Peter Fasching | Germany | Case-control | 293 | 0 | 287 | 292 | 284 | 0 | 1 | 1 | 1 | 0 | 1 |
| MAYO/  CIDR | BEL | Diether Lambrechts | Belgium | Case-control | 799 | 0 | 1,306 | 789 | 1297 | 4 | 5 | 2 | 2 | 4 | 2 |
| CAM | BGS | Anthony Swerdlow | UK | Cohort | 226 | 0 | 0 | 226 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MAYO | BVU | Digna Velez Edwards | USA | Case-control | 149 | 0 | 496 | 135 | 391 | 9 | 102 | 1 | 1 | 4 | 2 |
| CAM | CAM | James Brenton | UK | Case-only | 231 | 0 | 0 | 228 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| SHANGHAI | CHA | Kexin Chen, Fengju Song | China | Case-control | 1,244 | 0 | 2,072 | 1 | 0 | 0 | 0 | 1243 | 2072 | 0 | 0 |
| SHANGHAI | CHN | Li Yan, Kang Shan | China | Case-only | 390 | 0 | 0 | 0 | 0 | 0 | 0 | 390 | 0 | 0 | 0 |
| MAYO | CNI | Javier Benítez,  María J. García,  Cristina Rodriguez-Antona | Spain | Case-control | 83 | 0 | 179 | 81 | 176 | 0 | 0 | 2 | 2 | 0 | 1 |
| CIDR | DKE | Joellen Schildkraut,  Andrew Berchuck | USA | Case-only | 93 | 0 | 0 | 80 | 0 | 10 | 0 | 2 | 0 | 1 | 0 |
| CIDR | DOV | Mary Anne Rossing | USA | Case-control | 1,346 | 0 | 1,568 | 1246 | 1460 | 12 | 36 | 62 | 45 | 26 | 27 |
| CIDR | EPC | Charlotte Onland-Moret, Elio Riboli | Europe | Nested case-control | 437 | 0 | 876 | 431 | 872 | 0 | 2 | 3 | 1 | 3 | 1 |
| CIDR | GER | Jenny Chang-Claude | Germany | Case-control | 205 | 0 | 376 | 203 | 376 | 0 | 0 | 2 | 0 | 0 | 0 |
| MAYO | GRC | Drakoulis Yannoukakos | Greece | Case-only | 327 | 0 | 0 | 325 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| CAM | GRR | Kirsten Moysich | USA | Case-only | 22 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | HAW | Marc Goodman | USA | Case-control | 397 | 0 | 626 | 105 | 172 | 6 | 9 | 275 | 412 | 11 | 33 |
| CAM | HJO | Thilo Doerk-Bousset, Matthias Duerst | Germany | Case-control | 244 | 0 | 0 | 242 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| CAM | HMO | Thilo Doerk-Bousset, Natalia Bogdanova | Germany | Case-control | 66 | 0 | 287 | 65 | 283 | 0 | 0 | 1 | 1 | 0 | 3 |
| CAM | HOC | Ralf Butzow | Finland | Case-control | 265 | 0 | 280 | 264 | 280 | 0 | 0 | 1 | 0 | 0 | 0 |
| CIDR | HOP | Francesmary Modugno, Kirsten Moysich,  Roberta Ness | USA | Case-control | 549 | 0 | 1,217 | 524 | 1189 | 21 | 23 | 2 | 1 | 2 | 4 |
| MAYO | HSK | Florian Heitz | Germany | Case-only | 123 | 0 | 0 | 122 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| CAM | HUO | Thilo Doerk-Bousset | Germany | Case-control | 73 | 0 | 235 | 49 | 126 | 0 | 0 | 2 | 16 | 22 | 93 |
| MAYO | ICN | Florian Heitz | UK | Case-only | 415 | 0 | 0 | 390 | 0 | 4 | 0 | 9 | 0 | 12 | 0 |
| CIDR | JPN | Keitaro Matsuo | Japan | Case-control | 150 | 0 | 232 | 0 | 1 | 0 | 0 | 150 | 231 | 0 | 0 |
| MAYO | KRA | Sue Park | Korea | Case-control | 310 | 0 | 688 | 0 | 0 | 0 | 0 | 310 | 688 | 0 | 0 |
| CIDR | LAX | Beth Karlan | USA | Case-only | 476 | 0 | 0 | 384 | 0 | 27 | 0 | 34 | 0 | 31 | 0 |
| CAM | LUN | Håkan Olsson | Sweden | Case-control | 41 | 0 | 1,577 | 41 | 1576 | 0 | 0 | 0 | 0 | 0 | 1 |
| MAYO | MAC | Ellen Goode | USA | Case-only | 213 | 0 | 0 | 205 | 0 | 2 | 0 | 2 | 0 | 4 | 0 |
| CIDR | MAL | Susanne Kruger Kjaer | Denmark | Case-control | 384 | 0 | 649 | 384 | 649 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | MAS | Soo-Hwang Teo,  Yin Ling Woo | Malaysia | Case-control | 179 | 0 | 181 | 0 | 0 | 0 | 0 | 152 | 158 | 27 | 23 |
| MAYO | MAY | Ellen Goode | USA | Case-control | 1,170 | 0 | 1,146 | 1145 | 1135 | 6 | 5 | 7 | 4 | 12 | 2 |
| MAYO | MCC | Graham Giles,  Laura Baglietto,  Gianluca Severi | Australia | Nested case-control | 136 | 0 | 141 | 135 | 141 | 0 | 0 | 1 | 0 | 0 | 0 |
| MAYO | MDA | Karen Lu,  Michelle Hildebrandt | USA | Case-control | 313 | 0 | 298 | 307 | 297 | 1 | 0 | 1 | 0 | 4 | 1 |
| MAYO | MEC | Wendy Setiawan | USA | Case-control | 67 | 0 | 79 | 6 | 6 | 14 | 15 | 19 | 28 | 28 | 30 |
| CIDR | MOF | Thomas Sellers,  Jennifer Permuth Wey, Catherine Phelan,  Alvaro Monteiro | USA | Case-control | 414 | 0 | 459 | 371 | 412 | 19 | 22 | 9 | 14 | 15 | 11 |
| CIDR | MSK | Douglas Levine | USA | Case-control | 238 | 0 | 245 | 201 | 205 | 13 | 26 | 15 | 6 | 9 | 8 |
| CIDR | NCO | Joellen Schildkraut | USA | Case-control | 994 | 0 | 925 | 837 | 734 | 142 | 179 | 12 | 3 | 3 | 9 |
| CIDR | NEC | Daniel Cramer,  Kathryn Terry | USA | Case-control | 532 | 0 | 586 | 502 | 569 | 13 | 7 | 12 | 6 | 5 | 4 |
| CIDR | NHS | Shelley Tworoger,  Meir Stampfer,  Walter Willett | USA | Nested case-control | 342 | 0 | 316 | 337 | 314 | 3 | 0 | 2 | 2 | 0 | 0 |
| CIDR | NOR | Helga B Salvesen | Norway | Case-control | 186 | 0 | 342 | 184 | 342 | 1 | 0 | 1 | 0 | 0 | 0 |
| CIDR | NTH | Lambartus Kiemeney,  Leon Massuger | Netherlands | Case-control | 263 | 0 | 588 | 255 | 588 | 1 | 0 | 2 | 0 | 5 | 0 |
| MAYO | OPL | Penelope Webb | Australia | Case-only | 510 | 0 | 0 | 484 | 0 | 0 | 0 | 13 | 0 | 13 | 0 |
| MAYO | ORE | Tanja Pejovic | USA | Case-only | 92 | 0 | 0 | 84 | 0 | 2 | 0 | 5 | 0 | 1 | 0 |
| CIDR | OVA | Linda Cook, Nhu Le | Canada | Case-control | 756 | 0 | 797 | 669 | 734 | 2 | 0 | 63 | 46 | 22 | 17 |
| NCI | PLC | Nicolas Wentzensen | USA | Cohort | 277 | 0 | 1,257 | 263 | 1119 | 7 | 94 | 5 | 43 | 2 | 1 |
| CIDR | POC | Jacek Gronwald | Poland | Case-control | 183 | 0 | 0 | 183 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| NCI | POL | Nicolas Wentzensen | Poland | Case-control | 272 | 0 | 0 | 272 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | PVD | Estrid Hogdall,  Claus Hogdall | Denmark | Case-only | 197 | 0 | 0 | 193 | 0 | 1 | 0 | 1 | 0 | 2 | 0 |
| CAM | RBH | Georgia Chenevix-Trench | Australia | Case-only | 141 | 0 | 0 | 139 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| CIDR | RMH | Paul Pharoah | UK | Case-only | 182 | 0 | 0 | 174 | 0 | 2 | 0 | 1 | 0 | 5 | 0 |
| CAM | RPC | Kirsten Moysich | USA | Case-only | 106 | 0 | 0 | 99 | 0 | 5 | 0 | 1 | 0 | 1 | 0 |
| MAYO/  CIDR | SEA | Paul Pharoah | UK | Case-control | 2,180 | 0 | 1,869 | 2154 | 1844 | 7 | 4 | 4 | 7 | 15 | 14 |
| CIDR | SIS | Dale Sandler | USA | Cohort | 131 | 0 | 1,507 | 118 | 1306 | 9 | 150 | 2 | 15 | 2 | 36 |
| MAYO | SMC | Alicja Wolk | Sweden | Cohort | 83 | 0 | 93 | 83 | 93 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | SOC | Ian Campbell, Diana Eccles | UK | Case-only | 301 | 0 | 0 | 297 | 0 | 1 | 0 | 0 | 0 | 3 | 0 |
| MAYO | SRO | Jim Paul, Nadeem Siddiqui, Ros Glasspool,  Iain McNeish,  Susana Banerjee | UK | Case-only | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | STA | Alice Whittemore,  Weiva Sieh | USA | Case-control | 424 | 0 | 464 | 282 | 307 | 16 | 51 | 77 | 60 | 49 | 46 |
| CAM | SWH | Wei Zheng | China | Case-control | 135 | 0 | 135 | 0 | 0 | 0 | 0 | 135 | 135 | 0 | 0 |
| CIDR | SZB | Jacek Gronwald | Poland | Controls | 0 | 0 | 181 | 0 | 180 | 0 | 0 | 0 | 0 | 0 | 1 |
| MAYO | TBO | Rebecca Sutphen, Catherine Phelan | USA | Case-control | 176 | 0 | 138 | 176 | 138 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | TOR | Catherine Phelan,  Steven Narod, Harvey Risch | Canada | Case-control | 474 | 0 | 486 | 445 | 477 | 1 | 1 | 17 | 3 | 11 | 5 |
| CIDR | UCI | Hoda Anton-Culver | USA | Case-control | 311 | 0 | 348 | 258 | 295 | 2 | 3 | 22 | 21 | 29 | 29 |
| MAYO | UHN | Marcus Bernardini | Canada | Case-only | 211 | 0 | 0 | 177 | 0 | 7 | 0 | 13 | 0 | 14 | 0 |
| CIDR | UKO | Usha Menon,  Simon Gayther | UK | Case-control | 755 | 0 | 998 | 737 | 979 | 6 | 11 | 4 | 2 | 8 | 6 |
| CIDR | UKR | Paul Pharoah | UK | Case-only | 49 | 0 | 0 | 48 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| CIDR | USC | Leigh Pearce,  Anna Wu | USA | Case-control | 926 | 0 | 1,026 | 607 | 787 | 40 | 34 | 113 | 93 | 166 | 112 |
| CAM | VAN | David Huntsman | Canada | Case-only | 221 | 0 | 0 | 171 | 0 | 1 | 0 | 34 | 0 | 15 | 0 |
| MAYO | WMH | Anna deFazio | Australia | Case-only | 176 | 0 | 0 | 145 | 0 | 1 | 0 | 16 | 0 | 14 | 0 |
| MAYO | WOC | Jolanta Kupryjanczyk | Poland | Case-control | 200 | 0 | 207 | 200 | 207 | 0 | 0 | 0 | 0 | 0 | 0 |

Supplementary Table 1g. Participating Sites and Samples from Prostate Cancer Consortia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Cases** | **White** | | **African American** | | **Asian** | | **Other** |  | |
| **Genotyping Center** | **Acronym** | **PIs** | **Country** | **Design** | **Males** | **Males** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | **Case No.** | **Control No.** | | **Case No.** | **Control No.** |
| Cambridge | Aarhus | Karina Dalsgaard Sorensen, Torben Falck Orntoft | Denmark | Hospital-based, Retrospective, Observational | 1140 | 570 | 1130 | 558 | 1 | 0 | 4 | 7 | | 5 | 5 |
| NCI | AHS | Laura E. Beane-Freeman, Michael Alavanja,  Stella Koutros | USA | Nested case-control study within prospective cohort | 514 | 1314 | 514 | 1314 | 0 | 0 | 0 | 0 | | 0 | 0 |
| NCI | ATBC | Demetrius Albanes | USA | Prospective, nested case-control | 1474 | 2205 | 1449 | 2189 | 0 | 0 | 5 | 4 | | 20 | 12 |
| USC | BioVu | Melinda Aldrich,  Dana C. Crawford | USA | Cases identified in a biobank linked to electronic health records | 213 | 0 | 0 | 0 | 213 | 0 | 0 | 0 | | 0 | 0 |
| CIDR | CCI | Matthew Parliament, Nawaid Usmani | Canada | Case series, Hospital-based | 285 | 0 | 275 | 0 | 4 | 0 | 4 | 0 | | 2 | 0 |
| NCI | ProGene (CeRePP) | Olivier Cussenot,  Géraldine Cancel-Tassin | France | Case-Control, Prospective, Observational, Hospital-based | 1064 | 881 | 952 | 771 | 109 | 107 | 1 | 2 | | 2 | 1 |
| BGI | CHIPGECS | Yong-Jie Lu, Guangwen Cao, Hong-Wei Zhang ,  Ninghan Feng, Xin Guo, Guomin Wang, Zan Sun | China | Case-control | 533 | 666 | 0 | 0 | 0 | 0 | 532 | 666 | | 1 | 0 |
| CIDR | COH | Susan L. Neuhausen | USA | hospital-based cases and controls from outside | 263 | 269 | 259 | 269 | 0 | 0 | 3 | 0 | | 1 | 0 |
| CIDR | COSM | Alicja Wolk | Sweden | Population-based cohort | 2406 | 1204 | 2389 | 1193 | 0 | 0 | 11 | 6 | | 6 | 5 |
| Copenhagen | CPCS1 | Børge G. Nordestgaard | Denmark | Case-control - Denmark | 552 | 269 | 551 | 269 | 0 | 0 | 1 | 0 | | 0 | 0 |
| Copenhagen | CPCS2 |  | Denmark |  | 461 | 238 | 461 | 238 | 0 | 0 | 0 | 0 | | 0 | 0 |
| USC | CPDR | Shiv Srivastava,  Jennifer C. Cullen,  George Petrovics | USA | Retrospective cohort | 145 | 44 | 0 | 0 | 145 | 44 | 0 | 0 | | 0 | 0 |
| NCI | CPS-II | Susan M. Gapstur,  Victoria L. Stevens | USA | Nested case-control derived from a prospective cohort study | 4743 | 4508 | 4688 | 4453 | 11 | 14 | 33 | 28 | | 11 | 13 |
| CIDR | EPIC | Tim J. Key, Ruth C. Travis, Elio Riboli | Multi Center in EU | Case-control - Germany, Greece, Italy, Netherlands, Spain, Sweden, UK | 697 | 739 | 681 | 723 | 4 | 0 | 3 | 8 | | 9 | 8 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Centre National de Genotypage (CNG) | EPICAP | Florence Menegaux | France | Case-control, Population-based, ages less than 75 years at diagonosis, Hérault, France | 64 | 63 | 0 | 0 | 64 | 63 | 0 | 0 | 0 | 0 |
| Cambridge | ERSPC | Christopher Bangma, Monique J. Roobol |  | Population-based randomised trial | 75 | 75 | 73 | 73 | 2 | 0 | 0 | 1 | 0 | 1 |
| Cambridge | ESTHER | Hermann Brenner | Germany | Case-control, Prospective, Observational, Population-based | 341 | 333 | 339 | 333 | 0 | 0 | 1 | 0 | 1 | 0 |
| CIDR | FHCRC | Janet L. Stanford | USA | Population-based, case-control, ages 35-74 years at diagnosis, King County, WA, USA | 434 | 421 | 418 | 403 | 2 | 1 | 6 | 12 | 8 | 5 |
| CIDR | Gene-PARE | Barry Rosenstein,  Harry Ostrer |  | Hospital-based | 1330 | 0 | 274 | 0 | 48 | 0 | 996 | 0 | 12 | 0 |
| CIDR | Hamburg-Zagreb | Marija Gamulin,  Davor Lessel | Croatia | Hospital-based, Prospective | 154 | 154 | 154 | 154 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | HPFS | Sara Lindstrom,  Edward Giovannucci, Kathryn L. Penney,  Lorelei Mucci | USA | Nested case-control | 1233 | 1095 | 1212 | 1081 | 14 | 4 | 4 | 8 | 3 | 2 |
| Cambridge | IMPACT | Rosalind A. Eeles | UK | Observational | 60 | 993 | 58 | 975 | 0 | 3 | 2 | 10 | 0 | 5 |
| Cambridge | IPO-Porto | Manuel R. Teixeira | Portugal | Hospital-based | 386 | 190 | 384 | 190 | 0 | 0 | 0 | 0 | 2 | 0 |
| USC | Karuprostate | Laurent Brureau, Luc Multigner, Pascal Blanchet | W. Indies | Case-control, Retrospective, Population-based | 384 | 411 | 0 | 0 | 384 | 411 | 0 | 0 | 0 | 0 |
| Genome Quebec | KULEUVEN | Frank Claessens,  Thomas Van den Broeck, Steven Joniau | Belgium | Hospital-based, Prospective, Observational | 175 | 103 | 174 | 103 | 0 | 0 | 1 | 0 | 0 | 0 |
| USC | LAAPC | Sue Ann Ingles | USA | Population-based, Case-control | 789 | 621 | 456 | 283 | 2 | 0 | 7 | 6 | 324 | 332 |
| USC | Malaysia | Azad Razack, Jasmine Lim, Soo-Hwang Teo, Meng H. Tan, Aik T. Ong | Malaysia | Case-control | 210 | 210 | 1 | 0 | 0 | 0 | 208 | 209 | 1 | 1 |
| CIDR | MCC-Spain | Manolis Kogevinas,  Gemma Castaño-Vinyals, Javier Llorca Diaz | Spain | Case-control | 542 | 443 | 534 | 425 | 1 | 1 | 4 | 11 | 3 | 6 |
| CIDR | MCCS | Graham G. Giles,  Melissa C. Southey | Australia | Nested case-control, Melbourne, Victoria | 780 | 334 | 776 | 334 | 0 | 0 | 3 | 0 | 1 | 0 |
| USC | MD Anderson | Sara S. Strom | USA |  | 1139 | 316 | 532 | 0 | 47 | 0 | 39 | 0 | 521 | 316 |
| CIDR | MDACC\_AS | Christopher J. Logothetis, Jeri Kim | USA | A prospective cohort study | 633 | 0 | 532 | 0 | 47 | 0 | 39 | 0 | 15 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| USC | MEC | Christopher A. Haiman, Brian E. Henderson,  Fredrick Schumacher | USA | Population-based | 1310 | 1396 | 625 | 664 | 490 | 530 | 32 | 30 | 163 | 172 |
| USC | WFPCS | Jennifer J. Hu | USA |  | 59 | 66 | 0 | 0 | 59 | 66 | 0 | 0 | 0 | 0 |
| USC | MOFFITT | Jong Y. Park | USA | Hospital-based | 602 | 346 | 429 | 226 | 129 | 101 | 7 | 3 | 37 | 16 |
| USC | NMHS | Jay Fowke | USA | Case-control, clinic based, Nashville TN | 188 | 201 | 0 | 0 | 188 | 201 | 0 | 0 | 0 | 0 |
| Cambridge | CONOR (Oslo) | Lovise Maehle,  Eli Marie Grindedal, Johanna Schleutker,  Fredrik Wiklund | Norway | Population-based, Retrospective, Observational | 1513 | 0 | 1487 | 0 | 2 | 0 | 11 | 0 | 13 | 0 |
| CIDR | Canary PASS | Daniel W. Lin | USA | Prospective, Multi-site, Observational Acive Surveillance Study | 380 | 0 | 369 | 0 | 0 | 0 | 9 | 0 | 2 | 0 |
| USC | PCaP | Jeannette T. Bensen,  James Mohler,  Elizabeth T.H. Fontham, Gary J. Smith | USA | Population-based , case only | 1022 | 0 | 0 | 0 | 1022 | 0 | 0 | 0 | 0 | 0 |
| Cambridge | PCMUS | Radka Kaneva, Vanio Mitev, Chavdar Slavov | Bulgaria | Case-control - Sofia, Bulgaria | 195 | 90 | 195 | 90 | 0 | 0 | 0 | 0 | 0 | 0 |
| CIDR | PHS | Meir Stampfer,  Sara Lindstrom, Peter Kraft, Kathryn L. Penney | USA | Nested case-control | 664 | 286 | 642 | 271 | 4 | 2 | 11 | 11 | 7 | 2 |
| NCI | PLCO | Sonja I. Berndt,  Stephen Chanock,  Gerald Andriole | USA | Nested case-control | 1010 | 1275 | 999 | 1187 | 0 | 58 | 1 | 1 | 10 | 29 |
| CIDR | Poland | Cezary Cybulski | Poland | Case-control | 510 | 345 | 509 | 344 | 0 | 0 | 1 | 1 | 0 | 0 |
| USC? | PRAGGA | Manuela Gago Dominguez, Jose Esteban Castelao | Spain | Case-control | 133 | 104 | 132 | 102 | 0 | 0 | 1 | 2 | 0 | 0 |
| CIDR | PROCAP | Henrik Gronberg,  Fredrik Wiklund | Sweden | Population-based, Retrospective, Observational | 677 | 339 | 675 | 332 | 0 | 1 | 1 | 6 | 1 | 0 |
| Cambridge | PROFILE | Rosalind A. Eeles | UK | Hospital-based, Prospective, Observational | 32 | 88 | 30 | 85 | 1 | 2 | 1 | 1 | 0 | 0 |
| CIDR | PROGReSS | Ana Vega | Spain | Hospital-based, Prospective, Observational | 696 | 349 | 692 | 348 | 0 | 0 | 2 | 1 | 2 | 0 |
| Cambridge | ProMPT | David E. Neal | UK | A study to collect samples and data from subjects with and without prostate cancer. Retrospective, Experimental | 1002 | 12 | 975 | 12 | 11 | 0 | 7 | 0 | 9 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cambridge | ProtecT | Jenny L. Donovan,  Freddie C. Hamdy,  David E. Neal,  Richard Martin | UK | Trial of treatment. Samples taken from subjects invited for PSA testing from the community at nine centres across United Kingdom | 4 | 1448 | 4 | 1429 | 0 | 2 | 0 | 11 | 0 | 6 |
| USC | PROtEuS | Marie-Elise Parent | Canada | Case-control, population-based | 72 | 58 | 0 | 0 | 72 | 58 | 0 | 0 | 0 | 0 |
| CIDR | QLD | Jyotsna Batra,  Suzanne Chambers,  Amanda Spurdle | Australia | Case-control | 3489 | 1356 | 3425 | 1336 | 3 | 0 | 52 | 15 | 9 | 5 |
| Cambridge | RAPPER | Alison Dunning,  Catharine West,  Neil Burnet |  | Multi-centre, hosptial based blood sample collection study in patients enrolled in clinical trials with prospect-ive collection of radiotherapy toxicity data | 2350 | 0 | 2255 | 0 | 52 | 0 | 29 | 0 | 14 | 0 |
| USC | SABOR | Ian M. Thompson Jr. | USA | Prostate Cancer Screening Cohort | 366 | 366 | 0 | 0 | 106 | 106 | 0 | 0 | 260 | 260 |
| USC | SCCS | William J. Blot, Wei Zheng | USA | Case-control in cohort, South-eastern USA. Prospective, Observational, Population-based | 257 | 1601 | 0 | 0 | 257 | 1601 | 0 | 0 | 0 | 0 |
| USC | SCPCS | Maureen Sanderson | USA | Population-based , Retrospective, Observational | 64 | 39 | 0 | 0 | 64 | 39 | 0 | 0 | 0 | 0 |
| Cambridge/CIDR | SEARCH | Paul Pharoah,  Nora Pashayan,  Alison Dunning | UK | Case-control - East Anglia, UK | 2932 | 1520 | 2852 | 1504 | 30 | 3 | 30 | 9 | 20 | 4 |
| USC | SFPCS | Esther M. John | USA | Population-based case-control study, Retrospective, Observational | 378 | 249 | 290 | 212 | 86 | 37 | 1 | 0 | 1 | 0 |
| Cambridge | SNP\_Prostate\_Ghent | Kim De Ruyck, Piet Ost | Belgium | Hospital-based, Retrospective, Observational | 334 | 141 | 325 | 141 | 1 | 0 | 3 | 0 | 5 | 0 |
| Cambridge | SPAG | Claire Aukim-Hastie, Samantha Larkin,  Paul A. Townsend | UK | Hospital-based, Retrospective, Observational | 47 | 192 | 47 | 189 | 0 | 0 | 0 | 1 | 0 | 2 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CIDR | STHM2 | Henrik Gronberg,  Fredrik Wiklund | Sweden | Population-based, Retrospective, Observational | 3148 | 1576 | 0 | 3104 | 1557 | 12 | 0 | 18 | 11 | 14 | 8 |
| CIDR | SWOG-PCPT | Catherine M. Tangen,  Ian M. Thompson |  | Case-control from a randomized clinical trial | 1211 | 1424 | 0 | 1097 | 1080 | 88 | 239 | 22 | 71 | 4 | 34 |
| CIDR | SWOG-SELECT | Catherine M. Tangen,  Ian M. Thompson |  | Case-cohort from a randomized clinical trial | 1847 | 3122 | 0 | 1507 | 2215 | 263 | 697 | 41 | 85 | 36 | 125 |
| Cambridge | TAMPERE | Johanna Schleutker | Finland | Case-control - Finland, Retrospective, Observational, Population-based | 2544 | 1226 | 0 | 2534 | 1215 | 0 | 0 | 4 | 8 | 6 | 3 |
| Ontario Cancer Institute Genomics Center | Toronto | Robert J. Hamilton,  Neil E. Fleshner,  Antonio Finelli |  | Prospective hospital-based biopsy cohort | 821 | 599 | 0 | 677 | 464 | 60 | 28 | 65 | 89 | 19 | 18 |
| USC | UGANDA | Stephen Watya | Uganda |  | 567 | 489 | 0 | 0 | 0 | 567 | 489 | 0 | 0 | 0 | 0 |
| Cambridge, USC, CIDR | UKGPCS | Kenneth Muir,  Rosalind A. Eeles,  ZSofia Kote-Jarai | UK | ICR, UK | 14107 | 7601 | 0 | 13168 | 7494 | 708 | 6 | 145 | 49 | 86 | 52 |
| Cambridge | ULM | Christiane Maier | Germany | Case-control - Germany | 475 | 190 | 0 | 471 | 188 | 1 | 0 | 1 | 1 | 2 | 1 |
| CIDR / USC | WUGS | Bettina Drake,  Adam S. Kibel,  Aleksandra Klim,  Graeme Colditz | USA | Cases Series, USA | 930 | 153 | 0 | 704 | 0 | 211 | 153 | 7 | 0 | 8 | 0 |

**Supplementary Table 2:** Transmitting institutions for organization of SNPs on the Oncoarray along with the proportion of the array allocated to specific cancers, areas of overlapping effects among cancers and for fine mapping among cancers.

|  |  |  |
| --- | --- | --- |
| Site | Submitting Center | Proportional allocation |
| Lung | Dartmouth | 13.5% |
| Ovary\* | Cambridge | 13.8% |
| Colorectal | USC | 13.1% |
| Breast\* | Cambridge | 25.0% |
| Prostate | USC/ICR | 24.5% |
| *BRCA1/2\** | Cambridge | 6.0% |
| Common (non fine-mapping) | Cambridge, Dartmouth, CGR, USC | 4.1% |
| Common fine-mapping | Cambridge | Included in cancer-specific loci |

*\**Breast and *BRCA1/2*, and ovary have shared lists based on meta-analyses. To simplify the final merging process, Cambridge assembled a single list from all three groups with a total allocation of 44.8% (split in the above proportions). The top 1800 SNPs identified by ECAC were included among those submitted as common non fine-mapping SNPs.

USC- University of Southern California; ICR=Institute for Cancer Research

**Supplementary Table 3. Description of SNPs that were manually reviewed for clustering.**

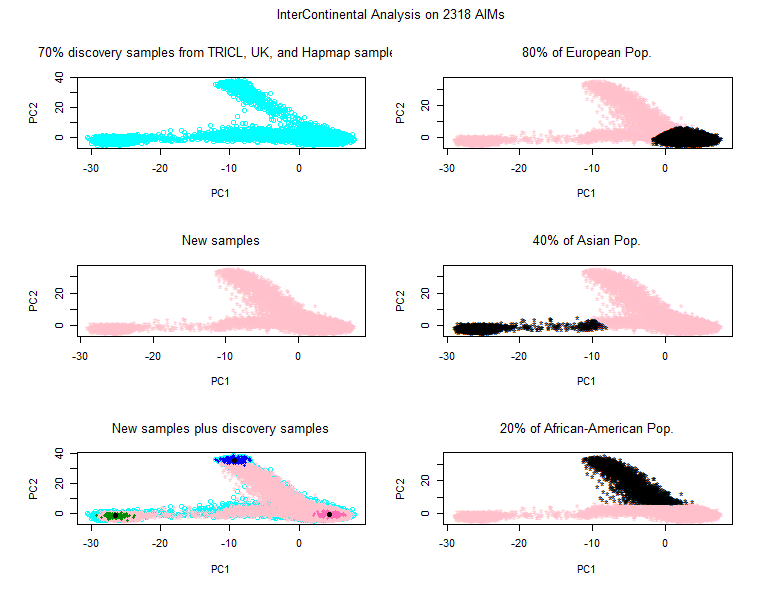
|  |  |  |
| --- | --- | --- |
| **Metric to select SNPs for Manual Inpection** | **Total SNPs in category** | **SNPs to be checked after excluding those in previous categories** |
| Very rare MAF < 0.01% (Call Rate > 99%) | 10107 | 10107 |
| SNP call rate 95% | 1737 | 1737 |
| SNP call rate 96% | 2492 | 2492 |
| SNP call rate 97% | 4081 | 4081 |
| SNP call rate 98% | 8256 | 6079 |
| SNP call rate 99.0% to 99.5% | 11467 | 6785 |
| Monomorphic with call rate < 99% | 1561 | 1016 |
| Low intensity AB R Mean <0.2 | 2997 | 2347 |
| Low intensity AB R Mean 0.2 - 0.3 | 3671 | 480 |
| AB T Mean <0.25 or >0.75 (Het. cluster close to hom. cluster ) | 6059 | 4576 |
| AB T Mean <0.2 or >0.8 (Het. cluster extreme ) | 2748 | 90 |
| AB T Deviation >0.09 (Wide het. cluster ) | 3011 | 1902 |
| AB Freq > 0.5 | 550 | 240 |
| Illumina Cluster Separation score <0.2 | 7756 | 241 |
| Illumina Cluster Separation score 0.2-0.25 | 7515 | 3591 |
| Illumina Cluster Separation 0.25 - 0.30 | 5240 | 1532 |
| Illumina Cluster Separation 0.30 - 0.39 | 14148 | 5504 |
| Rare in 1000 Genomes EUR (MAF< 1%) - Common in observed MAF > 5% | 67 | 40 |
| 1000G MAF 0.5% to 5% - observed MAF >10% or <0.005% | 2736 | 1393 |
| 1000G MAF > 5% - observed MAF <1% or at least 10% different | 4529 | 3440 |
| Candidate variants | 5768 | 5468 |
| **Total** | **100728** | **57673** |

**Supplementary Table 4:** Intercontinental ancestry analysis with OncoArray samples of 56,107 with 95% call rate or higher using FastPop.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ancestry** | **European** | **Asian** | **African** | **Other** |
| **Ancestry Definition from Fastpop** | **80%** | **40%** | **20%** |  |
| **Number of samples** | **47,554** | **4,625** | **2,525** | **1,403** |

**Supplementary Figure 1**. Comparison of the Oncoarray to several other Illumina arrays by imputing genotypes to the 1000 genomes release 3.3 or the Haplotype consortium for chromosome 22 using ShapeIt version 2 and Beagle, version 3.3.imputationResults.OncoArray.olivierBins.pdf

**Supplementary Figure 2.** Scores of discovery set in blue, the predicted scores from SNP weights in discovery set in pink. Three populations in Hapmap2 display CEU in hot pink, CHB in green, and YRI in blue. Three plots on the right side indicate 80% European, 40% Asian, and 20% African-American proportions of population memberships.



**Supplementary Information About the OncoArray Consortium**

The Consortium was formed to develop and genotype a new custom genotyping array (the “OncoArray”). The Oncoarray consortium brings together multiple disease-based consortia, including partnerships between the NCI-funded Genetic Associations and Mechanisms in Oncology (GAME-ON) initiative consortia (TRICL, FOCI, DRIVE, ELLIPSE and CORECT), the Breast Cancer Association Consortium (BCAC) and the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). The project has been funded through substantial grants from the NCI to the GAME-ON initiative and the Division of Cancer Epidemiology and Genetics (DCEG), Genome Canada/Genome Quebec/CIHR through the Personalized Risk Stratification for Prevention and Early Detection of Breast Cancer international project, Cancer Research UK (University of Cambridge) and a EU FP7 grant (“COGS”), together with many other grants.

The OncoArray Consortium has assembled more than 400,000 samples from existing studies and several biobanks. The OncoArray, which includes approximately 530K SNP markers, is a custom array that was manufactured by Illumina. Genotyping began in October 2013. The array includes a backbone of approximately 260,000 single nucleotide polymorphisms (SNPs) that provide genome-wide coverage of most common variants, together with markers of interest for each of the five diseases identified through genome-wide association studies (GWAS), fine-mapping of known susceptibility regions, sequencing studies, and other approaches. The array also includes loci of interest identified through studies of other cancer types, and other loci of interest to multiple cancer types (including loci associated with cancer related phenotypes, drug metabolism and radiation response). Additionally, SNPs relating to quantitative phenotypes such as BMI, height, and breast density that correlate with common cancer risks are also included.

OncoArray Steering Committee:

* + Transdisciplinary Research in Cancer of the Lung (TRICL)
* [Christopher Amos, Ph.D.](mailto:Christopher.I.Amos@Dartmouth.edu), Dartmouth College
* [Loic Le Marchand, M.D., M.P.H., Ph.D.](mailto:loic@crch.hawaii.edu), Cancer Research Center of Hawaii, University of Hawaii
  + Follow-up of Ovarian Cancer Genetic Association and Interaction Studies (FOCI)
* [Thomas Sellers, Ph.D., M.P.H.](mailto:thomas.sellers@moffitt.org), H. Lee Moffitt Cancer Center & Research Institute
  + - [Georgia Chenevix-Trench, Ph.D.](mailto:georgi.trench@qimr.edu.au), QIMR Berghofer
* [Paul Pharoah, Ph.D.](mailto:paul.pharoah@srl.cam.ac.uk), University of Cambridge
  + ColoRectal Transdisciplinary Study (CORECT)
    - * [Stephen Gruber, M.D., Ph.D., M.P.H.](mailto:sgruber@usc.edu), University of Southern California
  + Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE)
    - * [Stephen Chanock, M.D.](http://dceg.cancer.gov/about/staff-directory/biographies/O-Z/chanock-stephen), DCEG, NCI
      * [Alison Dunning, Ph.D.](mailto:amd24@medschl.cam.ac.uk), University of Cambridge
      * [Douglas Easton, Ph.D.](mailto:Douglas@srl.cam.ac.uk), University of Cambridge
      * [Rosalind Eeles, Ph.D., F.C.R.P., F.R.C.R.](mailto:rosalind.eeles@icr.ac.uk), The Institute of Cancer Research
* Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE)
  + - * [David Hunter, M.B.B.S., Sc.D.](mailto:david.hunter@channing.harvard.edu), Harvard University
      * [Douglas Easton, Ph.D.](mailto:Douglas@srl.cam.ac.uk), University of Cambridge
      * [Stephen Chanock, M.D.](http://dceg.cancer.gov/about/staff-directory/biographies/O-Z/chanock-stephen), DCEG, NCI
* Breast Cancer Association Consortium
* Genome Canada/Genome Quebec/CIHR funded Personalized Risk Stratification for Prevention and Early Detection of Breast Cancer international project
  + [Jacques Simard, Ph.D.](mailto:jacques.simard@crchul.ulaval.ca), Laval University
  + Douglas Easton, Ph.D., University of Cambridge
* Cancer Research UK
* Douglas Easton, Ph.D., University of Cambridge
* [Alison Dunning, Ph.D.](mailto:amd24@medschl.cam.ac.uk), University of Cambridge
* Paul Pharoah, Ph.D.
* [Georgia Chenevix-Trench, Ph.D.](mailto:georgi.trench@qimr.edu.au), QIMR Berghofer
* CIMBA
* Antonis Antoniou, Ph.D., University of Cambridge
* [Georgia Chenevix-Trench, Ph.D.](mailto:georgi.trench@qimr.edu.au), QIMR Berghofer
* [Jacques Simard, Ph.D.](mailto:jacques.simard@crchul.ulaval.ca), Laval University
* NCI EGRP
* Daniela Seminara, PhD., M.P.H. (liaison)
* NCI DCEG
* Stephen Chanock, M.D.

**SNP Selection for the Oncoarray**

General Principles

* SNP selection should be decided in collaboration between all the collaborating groups, i.e. all U19s plus any other groups providing resources (funding or datasets) for the initial project.
* The SNP content should made publicly available.
* The array will be made freely available for purchase by other groups.
* The Oncoarray will include 600,000 beadtypes (somewhat less than 600,000 SNPs, because ambiguous C/T or A/T SNPs require two beadtypes).
* The content should be divided between the disease groups. As an initial proposal, these should be divided as follows:
* Common content 60,000 (10% of the content)
* The remaining cost to be divided in proportion to the total samples/budget (both CIDR and non-CIDR). Additionally, we decided to allocate 260,000 SNPs for a GWAS backbone so the remaining allocations were made proportional to the proportion of samples that were genotyped.
* Each disease group decided how to select SNPs however a guideline was adopted to ensure a level of consistency.
* Broadly speaking, the disease-specific components should include:
  + Follow-up of combined GWAS/replication
  + Fine-mapping of known hits
  + Follow-up of rare variant/sequencing experiments
  + Ad-hoc candidates
* The relative contributions of each list were up to each disease group to decide.

Common Content

* SNP selection from meta-analysis across diseases (either overall or using mixture model)
* Lists for other cancers (say up to 1000 each, depending on availability)
* Fine-mapping of regions that are hits for more than one cancer type

(*TERT*, 8q24 (proximal and distal to *MYC*), *HNF1B, TET2, RAD51B*, 11q13)

* QTLs:
* Menarche
* Menopause
* Anthropometric (height, weight, BMI, WHR) *Try to include longer list this time from GIANT*
* Telomere length
* Confirmed GW significant hits for all cancers and cancer-related phenotypes (e.g. smoking)
* Nominations from cross-site pathway analyses.
* Y and MT

GWAS replication

* Generally, best to base on full available data, i.e. combined GWAS + replication, and imputed to 1KG.
* Remove highly correlated SNPs (r2>0.8), but include additional surrogates for the most strongly associated markers.
* Overall disease
* Subtypes
* Ethnicity specific analysis (lengths of these lists for Asian/African ancestry will depend on how many samples are likely to be genotyped, but we should try to include some).
* Survival (where there is available data)
* Attempt to include all SNPs, at least for overall disease, that appear to be measurably predictive of risk in a predictive risk score analysis (for prostate, initial analyses suggest at least 10,000)

Fine mapping

Define regions to map, based on both LD and relevant genomic features (e.g. to cover regulatory regions pertinent to genes of interest, if known)

Define complete catalogue of variants (from 1KG, augmented with other sequence data if available)

Attempt to include all variants correlated with best hit, plus dense tagging set of remaining variants

Parameters will depend on number of regions to map and size. For iCOGS, r2>0.1 was used.

Technical/organizational issues

* NCI DCEG was designated as responsible for the final list and its submission to Illumina. Each disease group was responsible for generating its own list.
* DCEG provided a shared space to exchange lists
* Admissible design score (0.8 was used, but a lower score was allowed for fine-mapping and candidates)

Merging process

Merging performed as a tree (scripts already available to implement this).

* For each disease, make separate lists for each category, ranked by importance
* Merge all the replication lists (choosing surrogates as necessary)
* Merge the replication, fine-mapping, rare variant and candidate lists, to make a final ranked list for each disease (*these lists can be 50-100% larger than the allocation, to allow for overlaps*).
* Final merging (across disease sites), in proportion to the SNP allocation (no surrogates chosen at this stage, only exact duplicates removed).
* GWAS framework included as an additional list, discarding SNPs selected for replication (not from the other lists) if an adequate GWAS SNP existed.

**SNP Selection - prostate**

- Known index signals

- SNPs from COGS

- SNPs from meta-analysis of all cases in European Americans

- SNPs from meta-analysis of advanced cases in European Americans

- SNPs from meta-analysis of all cases in African Americans

- SNPs from meta-analysis of advanced cases in African Americans

- SNPs from meta-analysis of all cases in all groups

- SNPs from meta-analysis of advanced cases in all groups

- Fine-mapping of known regions in European-Americans or African-Americans

- Top SNPs from analysis of an Exome chip

- Rare variants from the International Consortium for Prostate Cancer Genetics (~1000)

- Candidates (~2000)

- Results from analysis of a PSA GWAS

**SNP Selection – breast**

Fine-mapping of known regions

Replication: combined analysis from GWAS + iCOGS (imputed to 1KG):

Overall disease (1df and 2df tests)

Disease <40

ER-negative

Grade

Breast density

Survival

Asian ancestry

African ancestry

Exome chip (~5,000SNPS)

Rare variants from COMPLEXO, other consortial nominations (allocate ~1,000)

Variants from whole genome sequencing

Candidates (allocate ~2,000)

**SNP Selection – Lung 43,206 variants were nominated**

GWAS and GWAS Meta-analyses replication

* + Meta-analysis of 16 individual GWAS
  + HapMap 2 based meta-analysis
  + 1000Genome based meta-analysis
  + GWAS in Asian and African-American

Tagging and Fine-mapping

* + confirmed loci (5p15; 6p21-11; 9p21.3; 15q15.1; 15q25; 12p13.33, 22q12.2)

Individual Group Variants

* + Candidate genes including IPF, asthma, COPD
  + Rare variants from sequencing projects – TCGA data on lung adenocarcinoma, squamous carcinoma and head and neck cancers
  + Lung eQTL variants
  + Inflammation variants
  + Histology pathway analysis
  + COPD variants
  + Tobacco metabolism and smoking phenotypes variants (placed in common area)

# Detailed Oncoarray QC Guidelines are available at the Oncoarray wiki:

<http://consortia.ccge.medschl.cam.ac.uk/oncoarray> )

## 1. Genotype Calling

Call all genotypes with the v2c cluster file. (Download from <http://consortia.ccge.medschl.cam.ac.uk/oncoarray/onco_v2c.zip>).

Export Illumina TOP alleles from Genome Studio.

## 2. Sample QC

### 2.1 Initial call rate filtering (by consortium)

Exclude samples with call rate <80%

Exclude SNPs with call rate <80%

Exclude samples with call rate <95%

Exclude SNPs with call rate <95%

### 2.2 Ancestry

Define set of uncorrelated markers (~3,000) including all AIMS.

Use this set of markers and PCAs to define individuals of European/Asian/African American ancestry, or other. The Dartmouth group has defined principal components for identifying Continental ancestry and the R-package FastPop is available at <http://sourceforge.net/projects/fastpop/> .

### 2.3 Heterozygosity

We excluded samples with heterozygosity<5% or > 40% and heterozygosity if p<10-6, (|Z|>4.892 We tested Asian and European populations separately.

### 2.4 Sex checks

We excluded unexpected genotypic males/females/males (using X and Y markers). We also excluded XO, XXY, and samples with low X heterozygosity (<5%). We used a list of 300 Y markers confirmed to identify males and to check non-autosomal cluster patterns (chr\_Y\_SNPs\_for\_sex\_checking.csv). We excluded from the test chromosome X, SNPs that showed a high level of heterozygous calls in males and/or that had autosomal cluster patterns in males. A set of markers were reclassified as pseudoautosomal from the description provided by Illumina (chr\_X\_SNPs\_with\_autosomal\_clusters.csv.)

### 2.5 Duplicate concordance

We identified duplicates within studies and within disease sites among studies.

We checked for expected duplicates and if consistent, we excluded the sample with the lower call rate.

We identified unexpected duplicates. We worked with study data-managers to attempt to resolve any discrepancies andremoved both if they were not resolved.

We checked for concordance with previous GWAS or iCOGS genotyping or sequencing.

We excluded individuals discordant with previous consortium genotyping *(if the study co-ordinator could not resolve)*.

### 2.6 Relatives

We identified likely relatives. Individuals with estimated 0.55>ibd>0.45 were evaluated as likely first degree relatives.

These may be excluded by some of the consortia. For case-control pairs of relatives, we excluded the control. Otherwise we excluded the lower call rate sample.

## 3. SNP QC performed within Consortia

### 3.1 Call rate

We excluded SNPs that were zeroed by the cluster file because they had no alternate alleles.

We exclude samples with call rate <80%

We then excluded SNPs with call rate <80%

We then excluded samples with call rate <95%

Finally, we excluded SNPs with call rate <95%

**3.2 Hardy-Weinberg**

We check Hardy-Weinberg equilibrium and excluded SNPs if *P*<10-7 in controls or *P*<10-12 in cases.

(In CIMBA, all subjects were treated as controls.) Tests were performed only within Europeans as a procedure for removing SNPs that had poor performance. 4. SNP QC Exclusions Combined Across Consortia

### 4.1 Combine list of failures

All consortia excluded SNPS that failed for low call rate or for departing from HWE in any other consortium.

Chromosome Y exclusions were taken only from Practical. Practical used chromosome X HWE exclusions from BCAC.

### 4.2 Duplicate calling concordance

If the genotypes for pairs of duplicates differed by greater than 2% for any SNP, then we excluded that SNP as unreliable.

Duplicate concordance figures were combined from up to 5,250 duplicates from BCAC, OCAC, Practical, CIMBA.

### 4.3 Duplicate probes

There are a number of variants on the Oncoarray with the same probe in the same position (or a few with the same alleles but the sequence from the opposite strand.)

A list (onco\_duplicate\_variants\_excluded.csv) of 765 was compiled of duplicate probes that should be excluded. The probe with the worse QC scores and call rate was chosen for exclusion.

### 4.4 Cluster Plot Checking

We excluded SNPs for which the cluster plot was evaluated as “Failed” by two independent reviews.

## 5. Additional Steps Before Imputation

### 5.1 Rare SNPs with poor call rate

We excluded SNPs with call rate below 98% and MAF <0.01 (Europeans) in any consortium from the imputation input files. (The genotyped calls for these SNPs can still be analyzed, but were not used for imputations.)

### 5.2 Non-ideal cluster plots

SNPs with cluster plots that were scored as Possible (P) or Subset interference (S) in the second round of checking were. These are either rare SNPs where there is no clear heterozygote cluster or SNPs with more than three clouds because of interference from other SNPs or possible copy number variation.

### 5.3 Variants unmatched to a 1000 Genomes variant

Strand information was obtained by blasting the Illumina TOP sequences against the 1000 genomes sequences. Some manifest positions identified by “rs” numbers were updated from dbSNP and the new positions confirmed by sequence matching.

The variants on the chip were then matched to the variants from Phase 3 variant set provided for the Impute software. (<https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html>)

Variants were matched by position and alleles. Genotypes for variants not matched to a 1000G variant were not used for imputation by Impute.

### 5.4 Frequency Comparison to 1000 Genomes variants

Allele frequencies for controls from BCAC, OCAC and Practical were combined into a single frequency for Europeans (from 108,000 samples) and Asians (11,000 samples). These were tested against the expected frequency from 1000G using a ..

A difference statistic is calculated by the formula:  
(|p1-p2|-0.01)+^2/((p1+p2)(2-p1-p2))

where p1 and p2 are the frequencies our dataset and in the 1000 genomes respectively.

A cutoff of 0.008 in Europeans and 0.012 in Asians was needed to pass. Very rare SNPs are less likely to be rejected.

SNPs where the frequency would match if the alleles were flipped were excluded.

A list of strands and matched 1000G variants is provided on the Oncoarray Wiki.

A list of SNP where the Illumina TOP alleles need flipping in order to match the 1000 Genomes alleles is provided.

## 6. Principal Components

*Define Oncoarray consortium PCs and validate against some consortium specific PC definitions.* We defined a set of PCs for the European, Asian, and African subsets, which serve as covariates for them plus a global set to use for those of mixed ethnicity.

**7. Access to Genotyping Results**

The populations that have been genotyped as a part of the OncoArray are presented in Supplementary Table 1. This table provides a description of the design of the studies that are participating in the OncoArray along with the reported ethnic background of the participating studies. Samples that were genotyped at the Center for Inherited Disease Research or with genotyping supported by NIH funding, will be available for analysis through the dbGAP portal. Data from other samples along with more detailed phenotyping data may be available through collaborative requests to the participating consortia. Websites that provide details about the process for obtaining genotyping information are available for lung cancer at the Transdisciplinary Research in Cancer of the Lung website ([www.u19tricl.org](http://www.u19tricl.org)), for prostate cancer through PRACTICAL (http://practical.ccge.medschl.cam.ac.uk/), for Breast Cancer at BCAC (http://apps.ccge.medschl.cam.ac.uk/consortia/bcac/), for Ovarian Cancer at OCAC (http://apps.ccge.medschl.cam.ac.uk/consortia/ocac//), for colon cancer at CORECT (http://epi.grants.cancer.gov/gameon/), for endometrial cancer at (http://apps.ccge.medschl.cam.ac.uk/consortia/ocac/ http://epi.grants.cancer.gov/eecc/), and for *BRCA1* and *BRCA2* mutation carriers at CIMBA (http://apps.ccge.medschl.cam.ac.uk/consortia/cimba/). In total after all quality control exclusions there are 494,763 SNPs that were retained for analysis.

As shown in Supplementary Figure 2 and Supplementary Table 3, the ancestry definitions of 80% (European), 40% (Asian), and 20% (African) in intercontinental ancestry analysis led to no samples being assigned to multiple continental origins. The ancestry analytical approach was made available to all members of the consortium for classifying individuals by ethnic background.

**OncoArray Imputation**

We used as reference Dataset the 1000 Genomes Project (GP) Phase 3 ([Haplotype release date October 2014](https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html)) for chromosomes 1 to 22. The 1000 Genomes Project Phase 1 ([Haplotype ChrX release date Aug 2012](https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html)) was used for chromosome X, since the phased data for Chr X from 1000GP Phase 3 is not available.

The OncoArray whole genome data were imputed in a two-stage procedure using SHAPEIT (shapeit.v2.r790.Ubuntu\_12.04.4.static) to derive phased genotypes, and IMPUTEv2 (impute\_v2.3.2\_x86\_64\_static) to perform imputation of the phased data.

We used the default parameters used to derive phased genotypes with SHAPEIT, increasing:

- the number of burn-in iterations used by the algorithm to reach a good starting point to 10 ("--burn 10"),

- the number of pruning iterations used by the algorithm to find a parsimonious graph for each individual to 10 ("--prune 10"),

- and the number of iterations used by the algorithm to compute transition probabilities in the haplotype graphs to 50 ("--main 50")

We performed imputation with IMPUTEv2 using ~5Mb non-overlapping intervals for the whole genome. The flag  "-use\_prephased\_g" was provided to indicate that pre-phased haplotypes were being used. In addition we excluded from imputation the 1000 GP variants whose minor allele frequency in Europeans and East Asians was lower than 0.001. The missing genotypes at typed SNPs were replaced with imputed genotypes using the option "-pgs\_miss". The number of reference haplotypes to use as templates when imputing missing genotypes was increased to 800 ("-k\_hap 800"), and the buffer region was increased to 500kb ("-buffer 500").

For the fine mapping regions we also imputed the non-genotyped data with IMPUTEv2 but without prephasing in SHAPEIT in order to improve imputation accuracy. For this we also increased:

- the default number of Markov chain Monte Carlo (MCMC) iterations (including burn-in) to 50 ("-iter 50"),

- the number of MCMC iterations to discard as burn-in to 15 ("-burnin 15"),

- and the number of haplotypes to use as templates when phasing observed genotypes to 100 ("-k 100").

**Duplicated position issues**

SHAPEIT cannot handle duplicated variants (same position, and same alleles). The program stops when these variants are detected.

Therefore, we included for imputation only one of the variants that match the same position.