SNP Selection for the iCOGS Array

General Strategy

SNPs were selected for the iCOGS separately by each consortium. Each consortium was given a share of the array: nominally 50,000 SNPs each for BCAC, PRACTICAL and OCAC; 35,000 for CIMBA; and 15,000 for COMMON. In practice the allocations were larger as a result of overlaps. For both BCAC and PRACTICAL, the allocation was divided into three categories: GWAS replication, fine mapping and candidates.

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 previous 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 which 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 $r^2=1$ (not $r^2>0.9$ as in the GWAS).

To merge the three categories, we first included all the selected fine-mapping SNPs (see below), and then included SNPs from the merged GWAS and candidate lists in priority order, in the ratio 49:1.

Common SNPs 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 (67,988 BCAC; 57,033 OCAC; 68,638 PRACTICAL; 51,758 CIMBA; 23,211 COMMON).

Selection of GWAS SNPs

BCAC

The main selection was based on a systematic analysis of nine breast cancer GWAS. Each of the GWAS was imputed based on Hapmap phase II to permit combined analysis. We then defined ranked lists of SNPs based on six analyses: overall 1df test, 1df weighted by family history, a 2df test, cases diagnosed <40, cases diagnosed <50, and ER negative cases. We also defined lists for each of the nine component studies.

Additionally, we included ranked lists from three additional scans: BPC3 GWAS of ER-negative breast cancer; a scan of triple negative breast cancer (Fergus Couch); and a scan of African-American breast cancer cases (Chris Haiman). We also included a ranked list of SNPs associated with survival, from a combined analysis of GWAS (partially overlapping the set used for the risk analysis).

PRACTICAL

The main selection was based on a systematic analysis of four prostate cancer GWAS (UKGPCS, CGEMS, BPC3 and CAPS). All GWAS were imputed based on Hapmap phase II to permit combined analysis. We then defined ranked lists of SNPs based on six analyses: overall 1df test, overall 2df test, cases diagnosed <55, aggressive disease cases (Gleason 8+), and Gleason score as a continuous trait. We also included lists for low PSA vs. normal controls (based on ProtecT controls using in the UKGPCS scan) and for PSA as a continuous trait (based on CAPS).

OCAC

We combined the results from two ovarian cancer GWAS from North America and the UK. Imputation to HapMap2 was performed using 60 CEU founders as reference. Data on 2,508,744 genotyped SNPs or SNPs imputed with r^2>0.3 were available for analysis. The North American and UK studies were analysed separately and the results combined using fixed effects meta-analysis. The 2.5 million SNPs were ranked according to the *P*-values for each of four analyses performed: North America study only (all invasive and serous histology) and combined GWAS meta-analysis (all invasive and serous histology). The minimal ranking for each SNP was obtained across the four sets of results. SNPs with minor allele frequency less than 3 percent or SNPs that were already genotyped or in perfect LD with UK GWAS phase 2 SNPs were excluded. In addition we included a set of ranked SNPs identified through an analysis of survival time using case data from the two GWAS.

CIMBA

BRCA1 list: SNPs were selected for inclusion on the iCOGS array based on 8 separate using data from the *BRCA1* GWAS. Markers were evaluated for associations with (1) breast cancer (2) ovarian cancer (3) breast cancer restricted to Class 1 mutations (4) breast cancer restricted to Class 2 mutations (5) breast cancer by tumour ER-status (6) breast cancer restricted to *BRCA1* 185delAG mutation carriers (7) breast cancer restricted to *BRCA1* 5382insC mutation carriers and (8) breast cancer by contrasting the genotype distributions in *BRCA1* mutation carriers, against the distribution in population-based controls. Analyses (1) and (2) were based on both imputed and observed genotypes, whereas the rest were based on only the observed genotypes. All

analyses were performed separately, and SNPs were ranked according to the 1 d.f. score-test for trend P-value. All lists were then combined, to produce a global ranked list of SNPs such that at any cutoff, it contained the top SNPs from each for the analyses in the nominal proportions: 61.5%, 20%, 2.5%, 2.5%, 2.5%, 0.5%, 0.5% and 10.0% for analyses (1) to (8) respectively.

BRCA2 list: SNPs were selected on the basis of the strength of their associations with breast cancer risk from the discovery stage of the *BRCA2* GWAS , using imputed genotype data for 1.4M SNPs identified through CEU+TSI samples on HapMap3 release 2. A ranked list of SNPs was produced on the basis of a 1-df trend test statistic, after excluding highly correlated SNPs (r^2>0.4)

A global ranked CIMBA SNP list which included SNPs with the following nominal proportions: 55.5% from the *BRCA1* GWAS and fine mapping, 41.6% from the *BRCA2* GWAS, 2.9% for CIMBA candidate SNPs.

Selection of Fine-mapping SNPs

For the BCAC regions, we identified 21 loci which were associated with breast cancer risk at genome-wide significance (either published or through unpublished breast GWAS and BCAC replication data), as at 31st March 2010. 18 of these regions were included in the BCAC allocation and 3 (8q24, *ESR1* and *CDKN2A*) on the common allocation. For PRACTICAL, we included 27 regions (in addition to those in the common allocation). For OCAC we included 6 regions. CIMBA included fine mapping SNPs for the *RAD51* and 19p13 regions. For the common allocation, we included four regions (8q24, *ESR1*, *CDKN2A* and *TERT*) and aimed to cover the loci associated with any cancer type.

For each locus, we defined intervals around the most strongly associated SNP (or SNPs, if there was evidence of more than one signal). We defined intervals based on Hapmap (CEU), such as to include all SNPs with r^2>0.1 with the target SNPs. We then identified polymorphisms in each interval from the 1000 genomes dataset (CEU) available in April 2010, together with Hapmap phase III. We identified all variants for which the minor allele was called at least twice, and for which the Illumina Design score was >0.8. From the resulting set of typeable SNPs, we selected all SNPs with r^2>0.1 with any of the target SNPs, together with a set of SNPs that tagged all the remaining SNPs at r^2>0.9.

Selection of Common SNPs

Published GWAS hits were taken from the NHGRI GWAS catalogue (May 2010), including SNPs at $P<10^{-7}$ with any trait. The most strongly associated SNPs from GWAS for four other cancer types and other related phenotypes were also included (see table 2). The common SNPs also include a set of tagging SNPs covering 111 DNA repair genes, SNPs associated with allelic imbalance, ancestry informative markers (AIMs) and SNPs on MT and Y covering the major haplogroups.

Table 1. Summary of SNPs available in the BCAC allocation.

Category	List	Supplied by	SNPs on iCOGS
Fine-mapping	Fine-mapping	Maya Ghoussaini/Alison Dunning/Angie Cox	8,643
GWAS	Combined_1df_unweighted	Douglas Easton	27,668
GWAS	Combined_1df_weighted	Douglas Easton	28,588
GWAS	Combined_2df	Douglas Easton	7,278
GWAS	Combined_ERneg	Douglas Easton	2,910
GWAS	Combined_<40	Douglas Easton	4,283
GWAS	Combined_<50	Douglas Easton	5,826
GWAS	All Combined	Douglas Easton	36,767
GWAS	GxG conditional	Douglas Easton	13,210
GWAS	UK2	Douglas Easton	5,837
GWAS	BBCS	Julian Peto/Olivia Fletcher	1,308
GWAS	ABCFS	John Hopper	1,094
GWAS	CGEMS	David Hunter	1,020
GWAS	VUMC	Quentin Waisfisz	1,091
GWAS	GCHBOC	Alfons Meindl	1,038
GWAS	SASBAC	Per Hall	1,010
GWAS	HEBCS	Heli Nevanlinna	1,087
GWAS	MARIE	Jenny Chang-Claude	1,024
GWAS	GxHRT	Jenny Chang-Claude	1,391
GWAS	ABCFS_local	John Hopper	1,079
GWAS	SASBAC_local	Per Hall	514
GWAS	BPC3 GWAS	David Hunter	5,738
GWAS	Triple Negative GWAS	Fergus Couch	4,887
GWAS	African American GWAS	Chris Haiman	840
GWAS	Survival	Paul Pharoah/Heli Nevanlinna	9,618
GWAS	TOTAL		61,250
Candidate	Previous BCAC SNPs	All	237
Candidate	Candidate SNPs	*Various*	1,328
Candidate	ALL		1,537
TOTAL			**72,266***

* 68,007 were originally selected. The remaining 4,268 are from further down the lists and match SNPs from other consortia.

Table 2. Summary of SNPs available in the PRACTICAL allocation

Category	List	Supplied by	SNPs on iCOGS
Fine-mapping	Fine-mapping	Maya Ghoussaini/ Alison Dunning/ Zsofia Kote-Jarai	13,739
GWAS	Combined_1df	Douglas Easton	31,310
GWAS	Combined_2df	Douglas Easton	5,577
GWAS	Combined_<55	Douglas Easton	7,209
GWAS	Combined_Advanced	Douglas Easton	19,558
GWAS	Combined_Gleason	Douglas Easton	12,332
GWAS	All Combined	Douglas Easton	47,938
GWAS	GxG_conditional	Douglas Easton	16,908
GWAS	lowPSA	Douglas Easton	5,290
GWAS	PSAlevel	Douglas Easton	6,190
GWAS	UK_Australia	Douglas Easton	1,656
GWAS	CGEMS	Douglas Easton	1,079
GWAS	CAPS	Frederik Wiklund	1,776
GWAS	BPC3	BPC3	4,520
GWAS	Survival	Frederik Wiklund	13,116
GWAS	ALL		74,001
Candidate	Candidate	*Various*	1,336
Candidate	Tophits	Douglas Easton	76
Candidate	ALL		1,398
TOTAL			**85,278**

* 68,638 were originally selected. The remaining 16,640 are from further down the lists and match SNPs from other consortia

Table 3. Summary of SNPs available in the OCAC allocation

Category	List	Supplied by	SNPs on iCOGS
GWAS	Susceptibility	Tom Sellers/Paul Pharoah	22,254
GWAS	Survival	Ellen Goode/Paul Pharoah	7,656
Fine mapping		Honglin Song	3,726
Candidate		Ellen Goode	2,295
Candidate		Marc Goodman	3,982
Candidate		Linda Kelemen	771
Candidate		Kirsten Moysich	2,842
Candidate		Catherine Phelan	3,417
Candidate		Leigh Pearce	474
Candidate		Joellen Schildkraut	3,241
Candidate		Tom Sellers	825
Candidate		Georgia Chenevix-Trench	1,993
TOTAL			**51,853**

Table 4. Summary of SNPs available in the CIMBA allocation

Category	List	Supplied by	SNPs on iCOGS
GWAS	*BRCA1*	Antonis Antoniou/Fergus Couch	32,529
GWAS	*BRCA2*	Ken Offit/Mia Gaudet	19,029
Fine Mapping	19p13	Antonis Antoniou/Fergus Couch	527
Fine Mapping	*RAD51*	Olga Sinilnikova	132
Candidates		Various	1,419
TOTAL			**51,901**

Table 5. Summary of SNPs available in the COMMON allocation

Category	List	Supplied by	SNPs on iCOGS
PopStructure	AIM	Jonathan Tyrer	740
PopStructure	Y	Jim Wilson	82
PopStructure	MT	Jim Wilson	50
Fine-mapping	Fine-mapping	Maya Ghoussaini/Alison Dunning	3608
GWAS	Published_cancer	Douglas Easton/Diether Lambrechts	141
GWAS	Published_other	Douglas Easton/Diether Lambrechts	422
GWAS	Lung cancer	Richard Houlston	690
GWAS	Endometrial cancer	Mandy Spurdle	1548
GWAS	Melanoma/naevi	Tim Bishop	1240
GWAS	Melanoma/naevi	Stuart MacGregor	502
GWAS	Testis cancer	Clare Turnbull	750
GWAS	Age at menarche	Cathy Elks	2338
GWAS	Age at menopause	Cathy Elks	2632
GWAS	Telomere length	Karen Pooley	1629
GWAS	Mammographic Breast Density	Rulla Tamini/MODE Consortium	1870
GWAS	Mammographic Breast Density	John Hopper	462
GWAS	Serum oestradiol level	Deborah Thompson	1267
GWAS	Serum testosterone level	Deborah Thompson	1299
GWAS	Serum SHBG level	Deborah Thompson	1271
GWAS	Other hormone level	Deborah Thompson	353
GWAS	Endometriosis	Georgia Chenevix-Trench	30
Candidate	DNA Repair	Zsofia Kote-Jarai	2346
Candidate	Allelic Imbalance	Jacques Simard	479
Candidate	Cancer SNP Panel		1168
Candidate	Rare Variants	Mandy Spurdle (ENIGMA), Melissa Southey, Zsofia Kote-Jarai	107
TOTAL			**25,153**

Table 6. Fine-mapping regions (BCAC)

Region	Chromosome	Interval	Target SNP(s)	SNPs on iCOGS
1p11	1	120,300,000-121,185,600	rs11249433	92
CASP8	2	201,274,373-202,274,373	rs1045485 rs10931936	540
2q35	2	217,440,000-217,651,000	rs13387042	280
SLC4A7	3	26,913,000-27,794,000	rs4973768 rs2307032	928
5p12	5	44,430,000-45,400,000	rs7716600	574
MAP3K1	5	56,019,414-56,324,567	rs889312	325
9q31.2	9	109,780,000-110,150,000	rs865686	428
10p15.1	10	5,500,000-6,300,000	rs2380205	970
ZNF365	10	63,875,000-64,440,800	rs10995190 rs10509168	441
ZMIZ1	10	80,012,900-80,600,000	rs704010	512
FGFR2	10	123,200,000-123,700,000	rs2981579	427
LSP1	11	1,820,000-1,998,000	rs3817198 rs909116	171
11q13	11	68,692,000-69,375,000	rs614367 rs624797	603
PTHLH	12	27,850,000-28,550,000	rs1975930	456
12q24.21	12	113,820,000-114,350,000	rs1292011	542
RAD51L1	14	67,800,000-68,200,000	rs999737 rs8009944	418
TOX3	16	51,076,400-51,608,550	rs3803662	396
COX11	17	50,171,000-50,640,000	rs6504950	539

Table 7. Fine-mapping regions (PRACTICAL)

Region	Chromosome	Interval	Target SNP(s)	SNPs on iCOGS
2p15	2	62,500,000-63,500,000	rs2710647 rs721048	659
2p21	2	43,200,000-43,800,000	rs1465618	381
ITGA6	2	172,900,000-173,500,000	rs12621278	595
3p12	3	86,800,000-87,700,000	rs2660753 rs6788616	786
3q21	3	129,000,000-129,650,000	rs2999081	542
3q26.2	3	171,200,000-171,700,000	rs10936632	287
PDLIM5	4	95,200,000-95,800,000	rs12500426 rs17021918	333
4q24	4	106,150,000-106,750,000	rs7679673	534
5q35	5	175,500,000-176,050,000	rs11134973	298
6q24	6	139,600,000-140,000,000	rs12528855	284
6q25	6	160,200,000-161,100,000	rs651164 rs9364554	687
7p21	7	11,900,000-12,550,000	rs1347436	666
JAZF1	7	27,400,000-28,100,000	rs10486567	544
LMTK2	7	97,450,000-98,000,000	rs6465657	441
NKX3.1	8	23,150,000-23,700,000	rs2928679 rs1512268	675
9q33	9	123,400,000-123,950,000	rs1571801	368
MSMB	10	51,150,000-51,250,000	rs10993994	91

CTBP2	10	126,250,000-127,000,000	rs4962416	602
11p15	11	1,700,000-2,300,000	rs7127900	514
11q13	11	68,200,000-69,250,000	rs11228565 rs7931342	759
12q13	12	51,280,000-51,700,000	rs4919743	368
HNF1B	17	33,100,000-33,250,000	rs11649743 rs4430796	175
17q25	17	66,250,000-67,075,000	rs1859962	800
19q13.1	19	43,150,000-43,700,000	rs8102476	382
KLK3	19	55,600,000-56,500,000	rs17632542	754
22q13	22	41,300,000-42,200,000	rs5759167	837
Xp11.2	X	51,000,000-51,800,000	rs5945619	377

Table 8. Fine mapping regions (OCAC)

Region	Chromosome	Interval	Target SNP	SNPs on iCOGS
2q31	2	176,918,757 – 176,258,582	rs2072590	725
3q25	3	157,385,185 – 158,268,891	rs2665390	952
8q24	8	129,184,399 – 129,693,249	rs10088218	715
9p22	9	16,831,349 – 17,053,703	rs3814113	450
17q21	17	43,276,905 – 44,260,055	rs9303542	1,262
19p13	19	16,991,401 – 17,321,999	rs8170/rs2363956	522

Table 9. Fine-mapping regions (CIMBA)

Region	Chromosome	Interval	Target SNP(s)	SNPs on iCOGS
RAD51	15	38,392,069-39,064,843	rs1801320	132
19p13	19	16,991,401 – 17,321,999	rs8170/rs2363956	527

Table 10. Fine-mapping regions (COMMON)

Region	Chromosome	Interval	Target SNP(s)	SNPs on iCOGS
TERT	5	1,280,000-1,415,000	rs401681 rs2736109 rs3816659 rs33964002 rs2853669 rs2736100	119
ESR1	6	151,600,000-152,650,000	rs2046210 rs3757318 rs3020314	933
8q24	8	127,630,906-129,184,370	rs13281615 rs13262406 rs1562430 rs12543663 rs10086908 rs1016343 rs13252298 rs6983561 rs620861 rs6983267 rs10090154 rs16901979 rs13254738 rs7000448	1870
CDKN2A	9	21,600,000-22,200,000	rs7023329 rs101970 rs11515 rs7023329 rs10757278 rs10757257 rs2218220 rs4636294	686