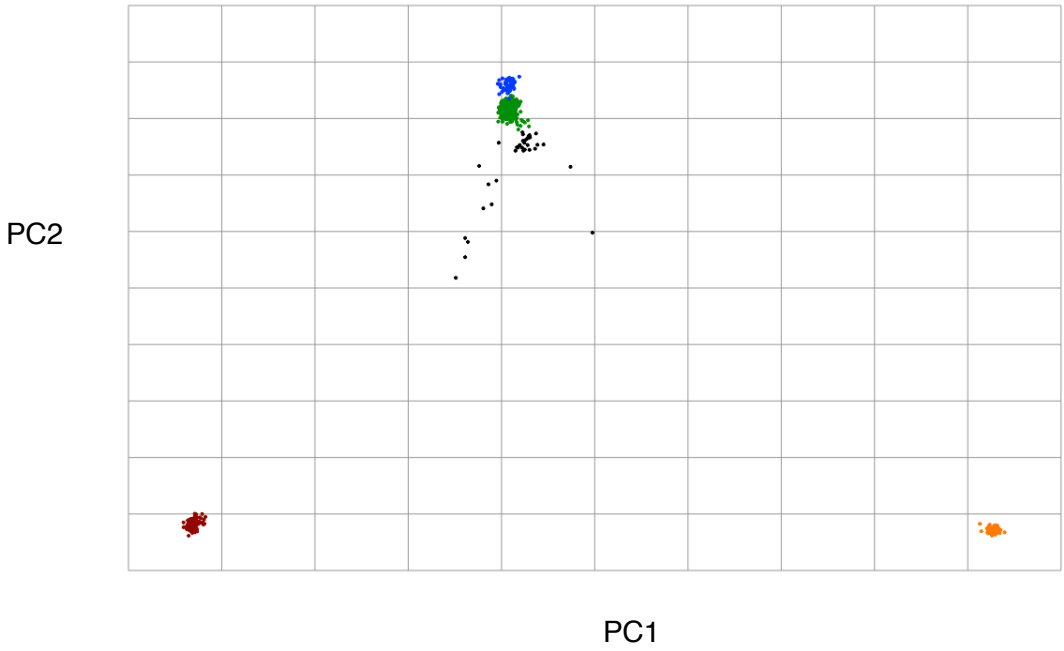
Genome-wide association study identifies a novel variant in *RAD51B* associated with male breast cancer risk

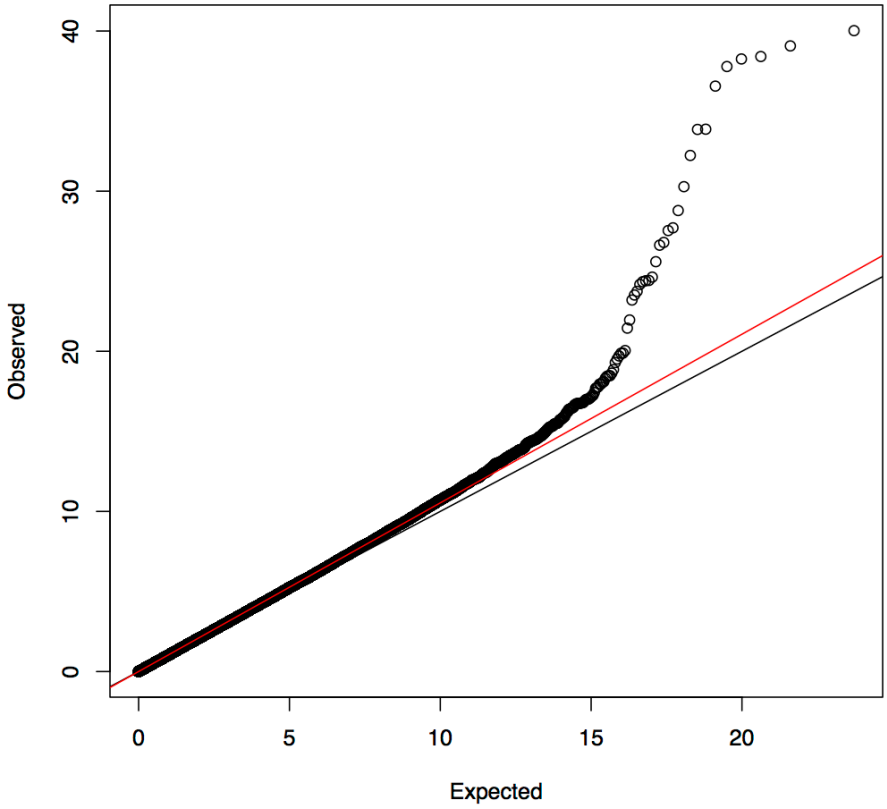
Nick Orr, Alina Lemnrau, Rosie Cooke, Olivia Fletcher, Katarzyna Tomczyk, Michael Jones, Nichola Johnson, Christopher J Lord, Costas Mitsopoulos, Marketa Zvelebil, Simon S McDade, Gemma Buck, Christine Blancher, KConFab Consortium, Alison H Trainer, Paul A James, Stig E Bojesen, Susanne Bokmand, Heli Nevanlinna, Johanna Mattson, Eitan Friedman, Yael Laitman, Domenico Palli, Giovanna Masala, Ines Zanna, Laura Ottini, Giuseppe Giannini, Antoinette Hollestelle, Ans M.W. van den Ouweland, Srdjan Novaković, Mateja Krajc, Manuela Gago-Dominguez, Jose Esteban Castelao, Håkan Olsson, Ingrid Hedenfalk, Douglas F Easton, Paul DP Pharoah, Alison M Dunning, D Timothy Bishop, Susan L Neuhausen, Linda Steele, Richard S Houlston, Montserrat Garcia-Closas, Alan Ashworth and Anthony J Swerdlow.

Supplementary Information

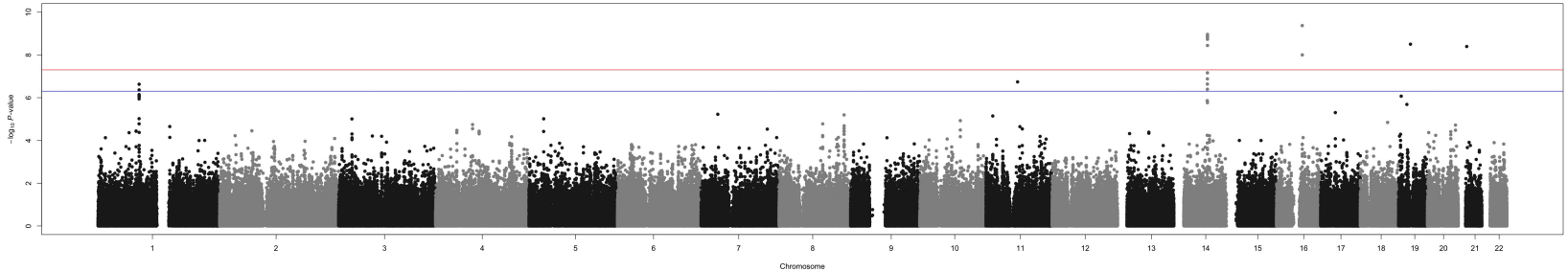
Supplementary Figure 1: Principal Components Analysis. The top two principal components were plotted from a PCA of HapMap data onto which the male breast cancer cases were projected. HapMap CEU, YRI and JPT+CHB samples are shown in blue, orange and red respectively; the MBC cases included in the final analyses are plotted in green; the PCA outliers excluded from the final analyses (n = 28) are plotted in black.



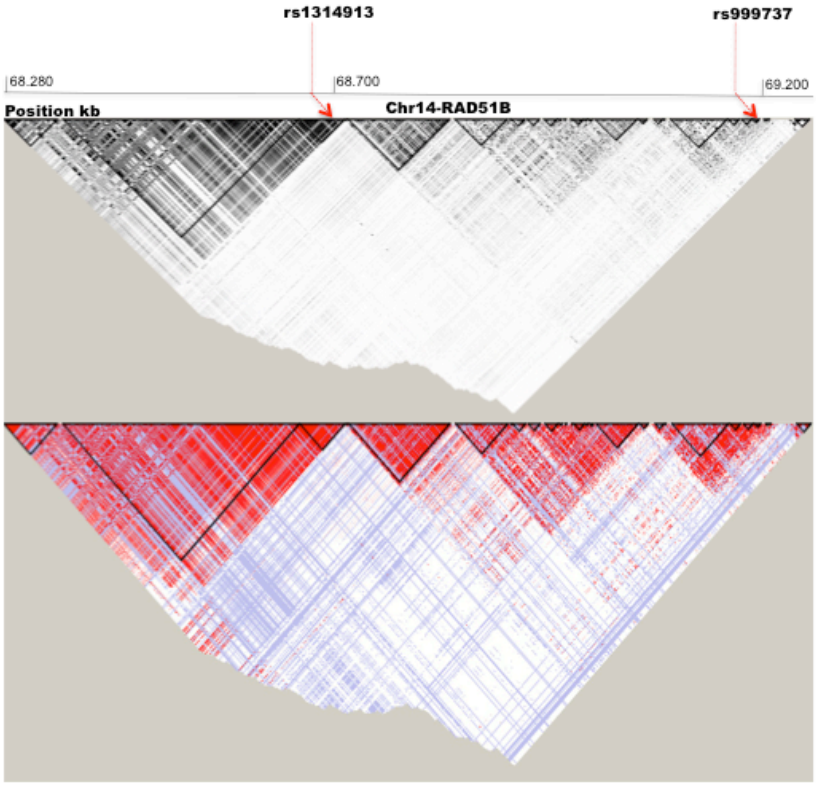
Supplementary Figure 2: QQ Plot of Association Statistics. The observed versus expected score test statistics are shown for 447,730 autosomal loci for which all poorly clustering loci (n = 30) with *P*-values ≤ 1 x 10^{-6} for association with male breast cancer were removed prior to estimating λ. The straight black line indicates distribution expected under the null hypothesis of no association and the red line indicates the inflation of the test statistics (λ=1.05).



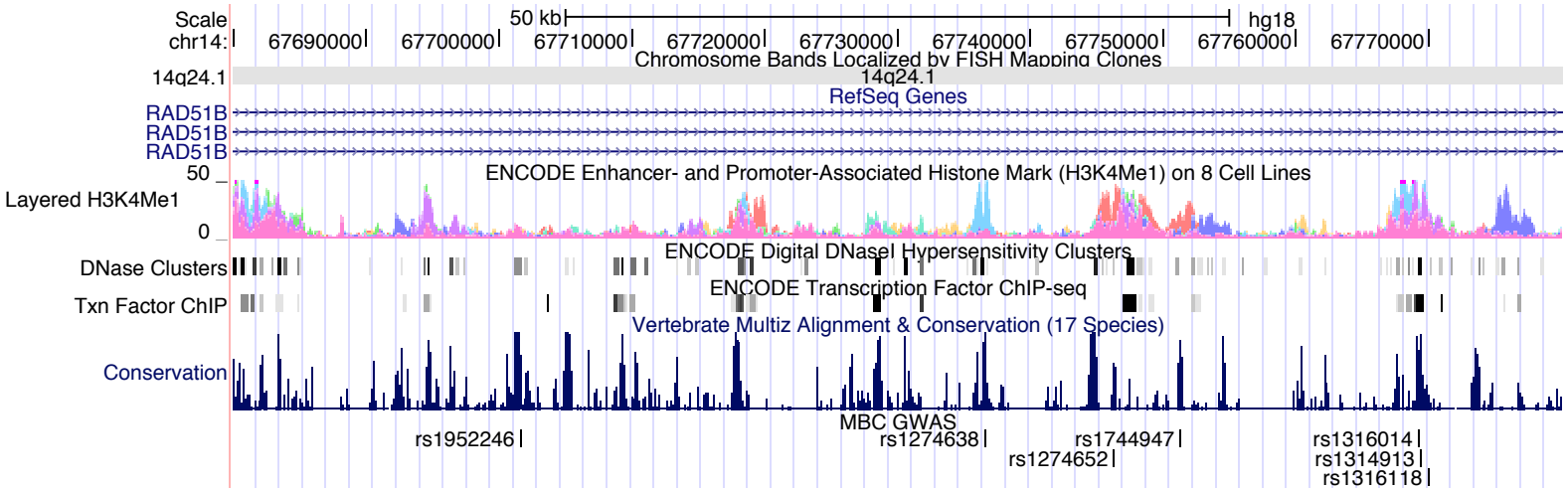
Supplementary Figure 3: Manhattan Plot of GWAS *P*-values. SNP association -$\log_{10}$ *P*-values are ordered firstly by chromosome number and then by position within each chromosome. The horizontal blue line indicates a threshold of suggestive significance at $P \leq 5 \times 10^{-7}$ and the red line indicates a genome-wide significance threshold of $P \leq 5 \times 10^{-8}$. Six SNPs mapping to regions with *P*-values below the suggestive significance threshold of $P \leq 5 \times 10^{-7}$ were selected for replication analysis.



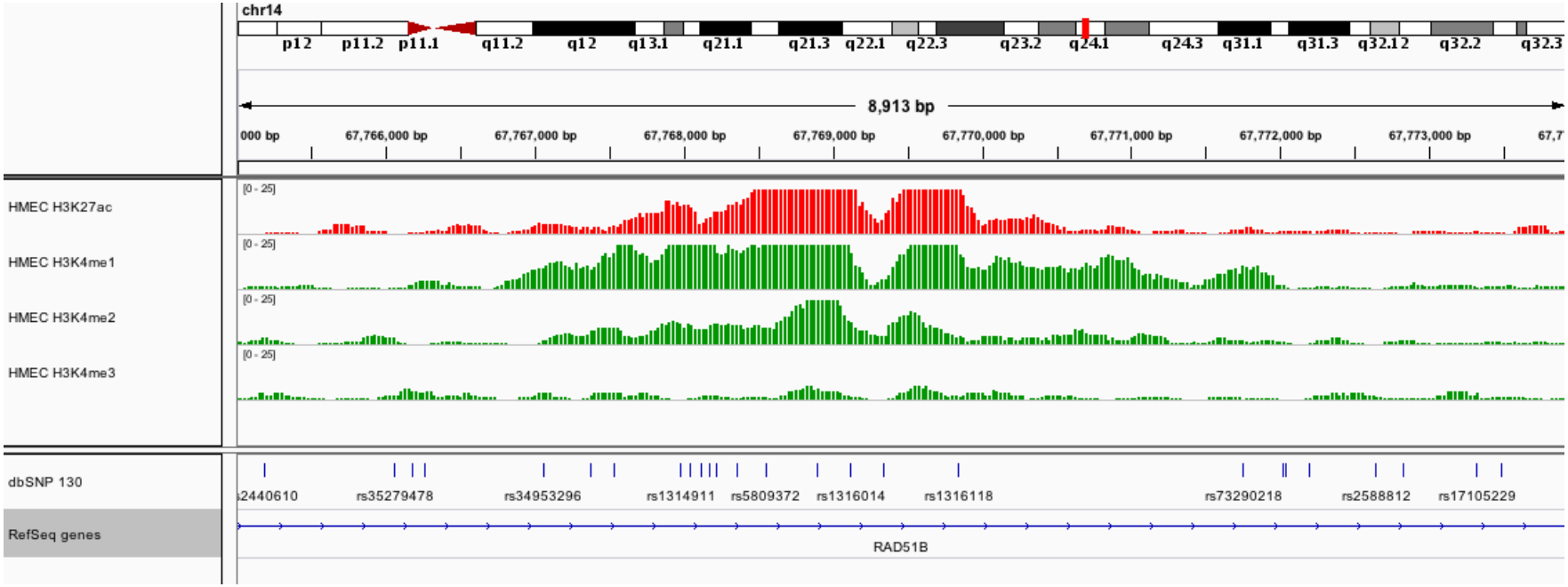
Supplementary Figure 4: Linkage Disequilibrium (LD) Structure at the 14q21.1 / *RAD51B* Locus. Pairwise LD plots generated using data from the HapMap CEU population and estimated by r^2 (top panel) and D' (bottom panel). The positions of r1314913 and rs999737 are indicated and are clearly separated by a region of strong recombination. Genomic coordinates are from NCBI build 36.



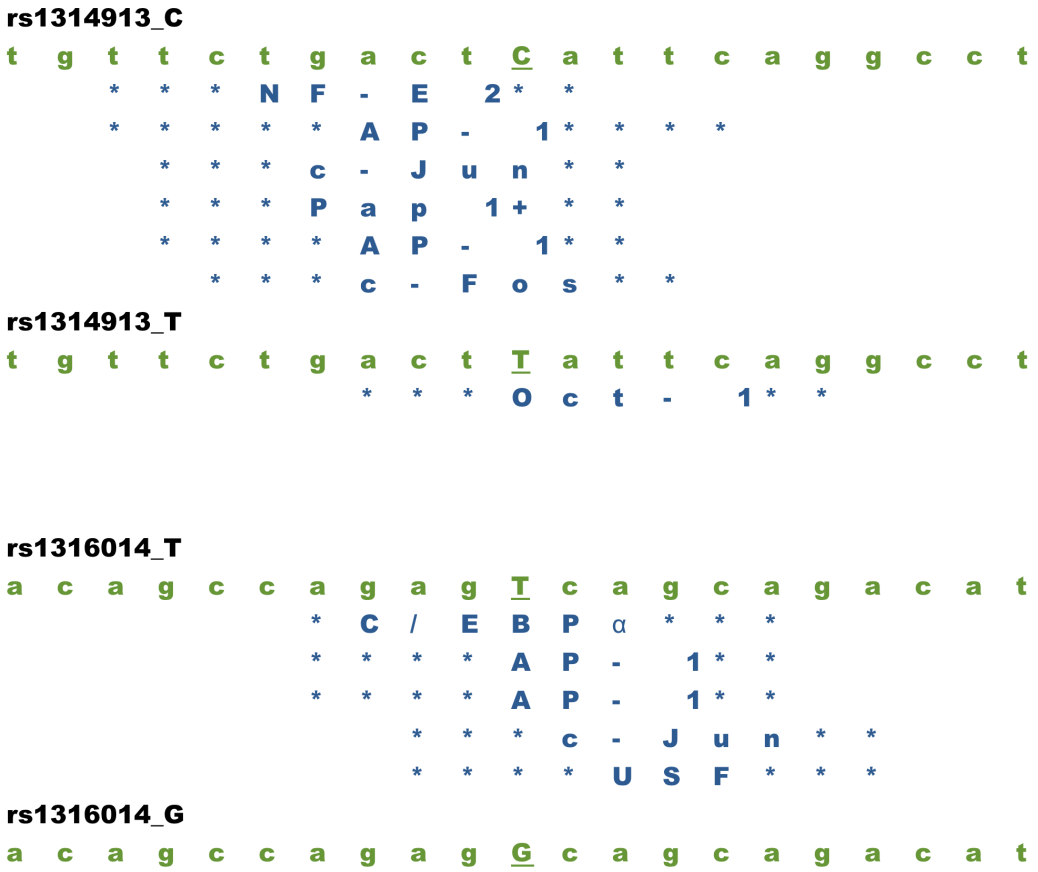
Supplementary Figure 5. Evolutionary conservation and transcription factor binding for seven *RAD51B* intron seven SNPs. Conservation scores for strongly associated SNPs at the *RAD51B* locus were computed using the phastCons algorithm with data from 17 vertebrate genomes[11] and used to filter SNPs that were not strongly conserved. The remaining seven SNPs were mapped on to ENCODE project data pertaining to gene regulation. Specifically, transcription factor binding sites, DNA hypersensitivity, and histone protein modification data were obtained from the ENCODE Regulation track on the UCSC genome browser. Only rs1316014 and rs1314913 map to regions with high conservation that show evidence of transcription factor binding by ChIP-seq, DNAse hypersensitivity and histone modification, all of which are indicative of an enhancer element.



Supplementary Figure 6. Histone mark data for human mammary epithelial cells. Broad histone modification data for human mammary epithelial cells indicates that rs1314913 is located in a region enriched for H3K4 mono and dimethylation and low trimethylation in both flanking regions, indicative of an enhancer region.



Supplementary Figure 7. Predicted transcription binding factor sites for the DNA sequence surrounding rs1314913 and rs1316014. The AliBaba2 transcription factor binding site prediction algorithm[14] was used in conjunction with the TRANSFAC binding motif database[15] to evaluate the effect of the minor allele of each SNP on transcription factor consensus binding sequences. The minor alleles of both SNPs abrogate predicted binding of AP-1 family member transcription factors *in silico*.



Supplementary Table 1: Descriptive details of individual samples sets for discovery and replication phases.

Phase	Study	Samples	Number	Source	Age (years)[§] (mean; range)	Years of diagnosis of cases	Percentage of cases histologically confirmed
Discovery	ICR, UK	Case	636	Population of England and Wales	65 (26-87)	2005-2011	100%
	Leeds Institute of Molecular Medicine, UK	Case	31	Population of Trent and Yorkshire, UK	62 (41-80)	1983-1990	100%
	University of Cambridge, UK	Case	138	Population of West Midlands, Trent, and Eastern Cancer Registries	61 (29-87)	1991-2010	unknown
	City of Hope, US	Case	115	Population of USA	61 (29-93)	1963-2001	96%
Replication	KConFab, Australia	Case	67	Population of Australia and New Zealand	61 (31-87)	1977-2011	100%
		Control	67	Population of Australia and New Zealand	61 (37-92)		
	Peter MacCallum Cancer Centre, Australia	Case	11	Hospital (PMCC familial cancer clinic)	66 (51-84)	1996-2010	100%
		Control	-	-	-		
	Copenhagen University Hospital, Denmark	Case	31	Population of Denmark	72 (53-88)	2003-2011	100%
		Control	30	Population of Denmark	75 (53-95)	2009	
	The Finnish Male Breast Cancer Study, Finland	Case	45	Hospital-based series (Helsinki, Oulu, Vassa and Kuopio)	61 (30-80)	1985-2011	82%
		Control	50	Adult blood donors from Finland	unknown		
	Sheba Medical Centre, Israel	Case	31	Hospital-based series (Tel-Aviv)	65 (39-79)	2006-2011	100%
		Control	27	Hospital-based (unrelated healthy visitors; Tel-Aviv)	66 (39-82)		
	ISPO, Florence, Italy	Case	101	Population of Tuscany	65 (35-87)	1990-2011	100%
		Control	106	Population of Tuscany	58 (43-65)		
	Sapienza University of Rome, Italy	Case	30	Hospital-based series (Rome)	61 (22-79)	1986-2010	100%
		Control	30	Hospital-based series (Rome)	60 (55-64)		
	Erasmus MC, The Netherlands	Case	38	Breast cancer families from the South West of the Netherlands	57 (28-86)	1964-2009	87%
		Control	38	Males from cystic fibrosis families who were spouses of at risk individuals	57 (28-86)		
	Institute of Oncology Ljubljana, Slovenia	Case	28	Population of Slovenia	56 (17-86)	1970-2010	100%
		Control	25	Population of Slovenia	35 (18-60)		
	Santiago (CHUS) and Vigo (CHUVI) University Hospitals, Galicia, Spain	Case	24	Population of Galicia, Spain	68 (49-88)	2000-2010	100%
		Control	19	Population of Galicia, Spain	68 (49-88)		
	Lund University, Sweden	Case	21	Hospital (Lund University Hospital)	63 (26-82)	1978-2010	100%
		Control	21	Spouses of cancer patients who themselves had no history of cancer	63 (44-77)		
	ICR, UK	Case	55	Population of England and Wales	64 (37-78)	2009-2011	100%
		Control	79	Population of England and Wales	58 (32-86)		

[§]Age at diagnosis of cases; age at blood draw of controls.

Supplementary Table 2a: GWAS Sample QC.

	Cases	Controls
Pre QC Total	957	2,912
Completion < 95%	-9	-42
QC Duplicates	-37	0
Unexpected Duplicates / Related	-23	-19
Ethnicity	-65	-56
Total included in analysis	823	2,795

Supplementary Table 2b: GWAS Locus QC.

	Omni Express	1.2M Custom	Merged
Pre QC Total	730,524	1,124,620	550,907
Completion < 95%	-7,706	-7,612	
MAF < 5% / Monomorphic			-95,576
X / Y / MT SNPs			-6,251
Hardy-Weinberg $P_{control} < 1x\ 10^{-5}$			-1,320
Total included in analysis	722,818	1,117,008	447,760

Supplementary Table 3: Stage 1, replication and combined association statistics for SNPs selected for validation.

Locus		MAF Controls	MAF Cases	*P*-value	OR_{trend}	95% CI
rs903263	GWAS	0.42	0.49	2.26×10^{-07}	1.34	1.20-1.50
1p31.1	Replication	0.45	0.47	0.42	1.08	0.90-1.30
PRKACB	Combined	0.43	0.49	1.15×10^{-06}	1.27	1.10-1.34
rs10897529	GWAS	0.35	0.42	1.54×10^{-07}	1.35	1.20-1.50
11q13.1	Replication	0.38	0.36	0.30	0.90	0.75-1.10
CDC42BPG	Combined	0.35	0.40	6.42×10^{-05}	1.22	1.10-1.34
rs1314913	GWAS	0.14	0.21	4.09×10^{-10}	1.55	1.35-1.78
14q24.1	Replication	0.15	0.22	1.71×10^{-04}	1.61	1.25-2.07
RAD51B	Combined	0.15	0.21	3.02×10^{-13}	1.57	1.39-1.77
rs3803662	GWAS	0.26	0.34	2.51×10^{-10}	1.46	1.30-1.64
16q12.1	Replication	0.27	0.37	2.38×10^{-06}	1.62	1.32-1.99
TOX3	Combined	0.26	0.35	3.87×10^{-15}	1.50	1.35-1.66
rs2160749	GWAS	0.13	0.08	1.37×10^{-08}	0.57	0.47-0.70
19p11	Replication	0.13	0.16	0.05	1.31	1.00-1.71
	Combined	0.13	0.11	1.96×10^{-04}	0.75	0.65-0.87
rs420519	GWAS	0.40	0.31	5.90×10^{-09}	0.71	0.63-0.80
21q21.1	Replication	0.40	0.43	0.16	1.15	0.95-1.39
	Combined	0.40	0.35	1.56×10^{-05}	0.80	0.73-0.89

Supplementary Table 4: Replication study specific estimates for validated loci.

Study	SNP	Controls				Cases						
		MAF	AA	AB	BB	MAF	AA	AB	BB	P-value	OR	95%CI
KConFab / Peter MacCallum Cancer Centre	rs1314913	0.13	48	15	1	0.17	47	25	0	0.32	1.44	0.70-2.94
	rs3803662	0.21	40	23	2	0.39	28	32	12	0.001	2.40	1.38-4.20
Copenhagen University Hospital, Denmark	rs1314913	0.06	23	3	0	0.18	19	7	1	0.09	3.17	0.80-12.65
	rs3803662	0.28	14	14	1	0.30	14	10	3	0.81	1.11	0.48-2.58
Helsinki University Central Hospital, Finland	rs1314913	0.10	38	10	0	0.21	29	13	3	0.05	2.23	0.97-5.12
	rs3803662	0.30	23	21	4	0.37	17	23	5	0.33	1.37	0.72-2.59
Sheba Medical Centre, Israel	rs1314913	0.21	14	10	0	0.25	12	12	0	0.56	1.40	0.45-4.38
	rs3803662	0.24	14	10	1	0.33	12	11	3	0.33	1.55	0.64-3.77
ISPO, Florence, Italy	rs1314913	0.18	70	23	6	0.29	43	40	5	0.02	1.82	1.11-2.98
	rs3803662	0.38	50	40	14	0.44	28	45	17	0.03	1.57	1.05-2.36
Sapienza University of Rome, Italy	rs1314913	0.18	19	11	0	0.33	12	15	2	0.05	2.56	0.98-6.83
	rs3803662	0.38	11	16	3	0.31	14	12	3	0.51	0.76	0.34-1.69
Erasmus MC, The Netherlands	rs1314913	0.24	10	6	1	0.25	5	5	0	0.90	1.09	0.28-4.30
	rs3803662	0.27	10	5	2	0.45	2	7	1	0.16	2.35	0.68-8.10
Institute of Oncology Ljubljana, Slovenia	rs1314913	0.13	18	6	0	0.13	20	7	0	0.94	1.05	0.30-3.71
	rs3803662	0.25	13	10	1	0.35	11	13	3	0.25	1.70	0.69-4.20
Galecia, Spain	rs1314913	0.17	24	12	0	0.29	20	10	5	0.11	1.89	0.86-4.15
	rs3803662	0.24	20	16	1	0.33	16	15	4	0.24	1.57	0.73-3.36
Lund University, Sweden	rs1314913	0.19	15	4	2	0.05	19	2	0	---	---	---
	rs3803662	0.19	14	6	1	0.33	8	12	1	0.11	2.39	0.79-7.20
ICR, UK	rs1314913	0.14	54	17	2	0.19	32	19	0	0.35	1.40	0.69-2.84
	rs3803662	0.22	48	20	6	0.35	23	24	7	0.02	1.83	1.07-3.13

Supplementary Table 5. Association statistics for imputed SNPs with *P*-values < 4.09×10^{-10} at the 14q12.1 locus.

SNP	Location	BP from rs1314913	r^2 with rs1314313	*P*-Value	OR	95% CI
rs2588827	67,684,194	85,153	0.64	2.40×10^{-10}	0.64	0.56-0.74
rs2588819	67,698,433	70,914	0.64	2.85×10^{-10}	0.64	0.56-0.74
rs2588818	67,700,331	69,016	0.64	3.97×10^{-10}	0.65	0.57-0.74
rs1952246	67,701,565	67,782	0.64	2.47×10^{-10}	0.64	0.56-0.74
rs2243905	67,703,166	66,181	0.64	3.81×10^{-10}	0.65	0.57-0.74
rs2767384	67,706,464	62,883	0.61	3.68×10^{-10}	0.65	0.57-0.74
rs1958115	67,707,418	61,929	0.58	2.46×10^{-10}	0.64	0.56-0.74
rs1958113	67,707,659	61,688	0.54	4.08×10^{-10}	0.65	0.57-0.74
rs2588814	67,713,551	55,796	0.64	2.40×10^{-10}	0.64	0.56-0.74
rs2767382	67,713,799	55,548	0.64	1.75×10^{-10}	0.64	0.56-0.73
rs2038979	67,719,529	49,818	0.64	2.34×10^{-10}	0.64	0.56-0.74
rs2588808	67,729,934	39,413	0.64	3.11×10^{-10}	0.65	0.57-0.74
rs2588809	67,730,181	39,166	0.82	3.56×10^{-10}	0.65	0.57-0.74
rs2767378	67,733,019	36,328	0.87	2.45×10^{-10}	0.64	0.55-0.73
rs2255767	67,733,190	36,157	0.87	2.68×10^{-10}	0.64	0.56-0.73
rs1274638	67,736,475	32,872	0.87	2.70×10^{-10}	1.57	1.36-1.80
rs1274639	67,738,252	31,095	0.82	1.48×10^{-10}	1.56	1.36-1.79
rs1274640	67,738,281	31,066	0.94	3.75×10^{-10}	1.56	1.36-1.79
rs1028842	67,741,476	27,871	0.82	1.47×10^{-10}	1.56	1.36-1.79
rs1274642	67,742,456	26,891	0.77	1.66×10^{-10}	1.56	1.36-1.79
rs1274643	67,742,832	26,515	0.94	3.13×10^{-10}	1.56	1.36-1.80
rs1274644	67,742,870	26,477	0.82	1.66×10^{-10}	1.56	1.36-1.79
rs1274645	67,743,079	26,268	0.75	1.66×10^{-10}	1.56	1.36-1.79
rs1274646	67,743,142	26,205	0.82	1.66×10^{-10}	1.56	1.36-1.79
rs1274647	67,743,597	25,750	0.94	1.70×10^{-10}	1.56	1.36-1.79
rs1274648	67,743,638	25,709	0.94	3.14×10^{-10}	1.56	1.36-1.80
rs1274649	67,744,053	25,294	0.82	1.66×10^{-10}	1.56	1.36-1.79
rs725453	67,745,611	23,736	0.94	3.14×10^{-10}	1.56	1.36-1.80
rs1274650	67,745,926	23,421	0.77	1.64×10^{-10}	1.56	1.36-1.79
rs1274651	67,746,171	23,176	0.82	1.64×10^{-10}	1.56	1.36-1.79
rs1274652	67,746,273	23,074	0.94	3.24×10^{-10}	1.56	1.36-1.80
rs1298340	67,747,245	22,102	0.94	3.18×10^{-10}	1.57	1.36-1.80
rs1274655	67,747,731	21,616	0.94	3.18×10^{-10}	1.57	1.36-1.80
rs1295780	67,748,385	20,962	0.82	1.63×10^{-10}	1.56	1.36-1.79
rs1274656	67,749,234	20,113	0.94	3.18×10^{-10}	1.57	1.36-1.80
rs1274657	67,749,468	19,879	0.82	1.63×10^{-10}	1.56	1.36-1.79
rs1274658	67,749,760	19,587	0.82	1.63×10^{-10}	1.56	1.36-1.79
rs1274659	67,750,402	18,945	0.82	1.63×10^{-10}	1.56	1.36-1.79
rs1744947	67,751,229	18,118	0.77	1.63×10^{-10}	1.56	1.36-1.79
rs1296527	67,755,213	14,134	0.75	3.18×10^{-10}	1.57	1.36-1.80
rs1744949	67,756,147	13,200	0.77	1.59×10^{-10}	1.56	1.36-1.79
rs117351774	67,758,392	10,955	0.77	1.60×10^{-10}	1.56	1.36-1.79
rs116264287	67,758,455	10,892	0.77	1.60×10^{-10}	1.56	1.36-1.79
rs1274661	67,758,972	10,375	0.77	1.60×10^{-10}	1.56	1.36-1.79
rs1290939	67,759,566	9,781	0.70	1.96×10^{-10}	1.55	1.35-1.77
rs1295782	67,760,195	9,152	0.77	1.60×10^{-10}	1.56	1.36-1.79
rs1751382	67,762,403	6,944	0.77	2.14×10^{-10}	1.56	1.36-1.79
rs1314912	67,768,125	1,222	0.77	1.83×10^{-10}	1.56	1.36-1.79
rs1316170	67,768,225	1,122	0.87	3.58×10^{-10}	1.57	1.36-1.80
rs1316014	67,769,123	224	0.77	1.84×10^{-10}	1.57	1.37-1.79
rs1316118	67,769,850	503	0.87	4.00×10^{-10}	1.57	1.36-1.80
rs1314914	67,772,829	3,482	0.94	3.90×10^{-10}	1.57	1.37-1.81

Supplementary Table 6. Highly evolutionarily conserved SNPs that are strongly associated with risk of male breast cancer.

SNP	Location	*P*-Value	Conservation Score
rs1952246	67,701,565	2.47×10^{-10}	0.99
rs1274638	67,736,475	2.70×10^{-10}	1.00
rs1274652	67,746,273	3.24×10^{-10}	0.97
rs1744947	67,751,229	1.63×10^{-10}	0.96
rs1316014	67,769,123	1.84×10^{-10}	0.94
rs1314913	67,769,347	4.09×10^{-10}	1.00
rs1316118	67,769,850	4.00×10^{-10}	0.98

Supplementary Table 7: Summary statistics for known female breast cancer risk loci.

Locus	Proxy (r^2)*	Cytoband	Gene	*P*-Value	OR	95% CI
rs11249433		1p11.2		0.91	0.99	0.89-1.11
rs1045485	rs17468277 (1)	2q33.1	*CASP8*	0.32	0.92	0.78-1.08
rs13387042		2q35		0.05	0.90	0.80-1.00
rs4973768		3p24.1	*SLC4A7*	0.10	1.10	0.98-1.22
rs10941679±		5p12		N/A	N/A	N/A
rs10069690		5p15	*TERT*	0.76	1.02	0.90-1.15
rs889312		5q11.2		0.96	1.00	0.88-1.13
rs2046210	rs6900157 (0.94)	6q25.1	C6orf97	0.03	1.13	1.01-1.26
rs13281615‡		8q24.21		0.61	0.96	0.81-1.14
rs1011970		9p21.3	*CDKN2BAS*	0.06	1.14	0.99-1.32
rs865686‡		9q31.2		0.78	1.19	0.93-1.51
rs2380205‡		10p15.1	*ANKRD1*	0.88	1.09	0.92-1.31
rs10995190	rs10995189 (1)	10q21.2	*ZNF365*	0.03	0.83	0.71-0.98
rs704010‡		10q22.3	*ZMIZ1*	0.38	1.02	0.81-1.28
rs2981582		10q26.13	*FGFR2*	0.21	1.07	0.96-1.20
rs3817198		11p15.5	*LSP1*	0.71	1.02	0.91-1.15
rs614367		11q13.3		0.01	1.23	1.06-1.43
rs10771399‡		12p11.22		0.003	0.76	0.62-0.92
rs1292011		12q24.21		0.19	0.93	0.83-1.04
rs999737		14q24.1	*RAD51B*	0.08	0.89	0.78-1.01
rs3803662		16q12.1	*TOX3*	2.51×10^{-10}	1.46	1.30-1.64
rs6504950‡		17q22	*TOM1L1*	0.40	1.07	0.91-1.26
rs8170		19p13.1	*BABAM1*	0.39	0.94	0.81-1.08
rs2823093		21q21.1		0.64	0.97	0.85-1.10

*** Surrogate marker for the published female SNP if it was not genotyped in the current study and coefficient of correlation with published SNP.**

± No LD proxy and insufficient data to impute this SNP.

‡ *P*-value and estimates based on imputed data as the published female SNP was not genotyped in the current study.

Supplementary Methods

Subjects and samples

GWAS

920 cases with male breast cancer were ascertained from three UK and one US male breast cancer studies (Supplementary Table 1). The UK series comprised 636 cases from the Breakthrough Breast Cancer male breast cancer study collected between 2007 and 2011; 138 cases from the University of Cambridge; and 31 cases from the Leeds Institute of Molecular Medicine. The US samples comprised 115 IRB-consented prevalent cases, ascertained in the United States through intermountain state Cancer Registries (primarily from Utah; n=83), an on-line support group (all US; n=18), and referrals from physicians (n=4). The majority of men from the US study were enrolled between 1997-2001, though the years of diagnosis were wider (1963-2001).

For controls we used publicly accessible data generated by the UK Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 2,912 from the 1958 British Birth Cohort (58C; also known as the National Child Development Study)[1].

Replication Series

Patients and controls for the replication series comprised 482 cases and 492 controls ascertained from: KConFab, Melbourne, Australia (67 cases, 67 controls); Peter MacCallum Cancer Centre, Melbourne, Australia (11 cases); Copenhagen University Hospital, Denmark (31 cases, 30 controls); Finnish Male Breast Cancer Study, Finland (45 cases, 50 controls); Sheba Medical Centre, Israel (31 cases, 27 controls); ISPO Cancer Research and Prevention Unit, Florence, Italy (101 cases, 106 controls); Sapienza University of Rome, Italy (30 cases, 30 controls); Erasmus MC, Netherlands (38 cases, 38 controls); Institute of Oncology Ljubljana, Slovenia (28 cases, 25 controls); Santiago (CHUS) and Vigo (CHUVI) University Hospitals, Galicia, Spain (24 cases, 19 controls); Lund University, Sweden (21 cases, 21 controls) and ICR, UK (55 cases, 79 controls). Detailed information regarding the replication series samples is provided in Supplementary Table 1.

Ethics

Collection of blood samples and clinical data from subjects was performed in accordance with local guidelines and regulations. Ethical review board approval was in accordance with the Declaration of Helsinki.

Genotyping

For the Breakthrough Breast Cancer cases, DNA was extracted from peripheral blood using standard methods and was quantified using PicoGreen (Invitrogen, Carlsbad CA) according to the manufacturer's instructions. Samples were plated and normalised to a concentration of 50 ng / μl using a Hamilton MicroLab Star liquid handling workstation (Hamilton Robotics, Bonaduz, Switzerland). For the remainder of the case samples, DNA

extracted from peripheral blood was supplied at a concentrations of ≥ 50 ng / μl by the contributing studies. These concentrations were confirmed in-house as above and were adjusted as necessary. Genotyping was performed by the High-Throughput Genomics Core at the Wellcome Trust Centre for Human Genetics (Oxford, UK) using OmniExpress bead chips (Illumina, San Diego CA) according to standard protocols.

Replication genotyping was conducted using Sequenom Mass Array iPlex Gold chemistry (Sequenom, San Diego, CA) and a Sequenom Mass Array 4 MALDI-TOF mass spectrometer and by KASP competitive allele specific PCR (Kbiosciences Ltd, Hertfordshire UK) performed on an ABI 7900HT Fast RT PCR system (Applied Biosystems, Carlsbad CA). Details of specific primers, probes, reagents and conditions are available upon request.

GWAS Quality Control

The Illumina OmniExpress bead chip contains assays for 730,524 markers. Genotype clustering was performed using the GenTrain 2.0 algorithm in GenomeStudio 2011.1 (Illumina, San Diego, CA). We excluded loci with call rates < 95% (n = 7,706) and excluded all samples for which < 95% of loci were successfully genotyped (n = 9). We included 37 pairs (approximately 4%) of duplicate samples to assess assay reproducibility. Average genotyping concordance between members of each duplicate pair was 99.95%. From each duplicate pair we excluded the sample with the lowest completion rate from all subsequent analyses. To detect unexpected duplicates and to detect samples, particularly from the three UK studies, that were present in more than one study, we computed pairwise identity-by-state (IBS) measures for every possible pair of samples. We detected 23 unexpected duplicate cases either within (n = 9) or between (n = 14) studies from the UK collections. We excluded 37 cases with self-reported non-Caucasian ancestry. To further address confounding due to ethnicity related substructure we applied k-means cluster analysis to the first two principal components from a principal components analysis of the HapMap CEU, YRI and JPT+CHB samples onto which the male breast cases were projected and eliminated 28 outliers from the main Caucasian cluster (Supplementary Figure 3). For the control samples from the 1958 British Birth Cohort samples genotyped on Illumina 1.2M Duo Custom arrays we obtained raw data from the European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) and clustered loci using the GenTrain 2.0 algorithm. We excluded 7,612 loci with completion rates < 95% and from a total of 2,912 samples we excluded: 42 individuals with completion rates < 95%; one from each of 19 pairs of related individuals; 56 individuals with non-Caucasian ancestry. We merged the case data and control data for 550,907 loci that were common to both the Illumina OmniExpress and 1.2M Duo Custom arrays and then excluded: 95,576 loci that had a MAF < 5%, or that were monomorphic; 6,251 SNPs that were not autosomal; 1,320 that showed significant deviation (P-value < 1 x 10^{-5}) from Hardy-Weinberg proportions in controls. Overall genotype completion rate for the merged case and control data set was 99.81%. The analytic dataset comprised 823 cases, 2,795 controls and 447,760 autosomal loci. GWAS

genotyping QC measures are summarised in Supplementary Tables 2a and 2b.

Replication Genotyping Quality Control
We excluded samples for which there was insufficient DNA for genotyping (n = 5) and for which > 2 loci failed to genotype (n = 62). The call rate for all replication SNPs was ≥ 95% and the average call rate for all six loci was 99.14%. The analytic dataset for replication genotyping comprised 438 cases, 474 controls and six SNPs.

Statistical and Bioinformatics Analysis
Locus and sample genotyping completion rates, IBS estimates, LD estimates and exact tests of deviation from Hardy-Weinberg equilibrium were computed using the Genotype Library and Utilities (GLU) package (http://code.google.com/p/glu-genetics). Principal components analysis (PCA) was conducted using EigenStrat[2] and k-means cluster analysis was performed using the kmeans routine in R (http://cran.r-project.org). Analysis of association between each SNP and risk of male breast cancer was performed using unconditional logistic regression assuming a log-additive genetic model and *P*-values were determined from a one degree-of-freedom score test using GLU. We examined the distribution of test statistics for evidence of systematic inflation by estimating the genomic control measure of Devlin *et al* [3] and by inspecting QQ plots generated using the estlambda function of GenABEL[4]. Prior to calculating the inflation factor, we manually inspected the cluster plots of all loci with $P < 1 \times 10^{-6}$ and excluded any SNP with poor clustering. We also performed an analysis in which we adjusted the estimates using the top ten principal components from a PCA analysis of the genotype data. Since this adjustment altered neither the inflation factor nor the effect estimates and did not substantially alter the ranking of the top hits, we report only the unadjusted analysis. For analysis of replication data we estimated the effect at each locus using unconditional logistic regression adjusted for study. Replication study specific estimates are shown in Supplementary Table 4. Imputation of unobserved loci was conducted by pre-phasing the combined case-control data, using SNPs common to both the Illumina 1.2M DuoCustom and OmniExpress arrays that had satisfied the QC measures outlined in Supplementary Table 2b, using SHAPEIT[5] and imputing genotypes using Impute v2.2[6] with reference haplotypes from phase one of the 1,000 Genomes Project. Imputation accuracy was assessed was assessed using the internal "leave-one-out" validation implemented in IMPUTE2. There was 98.3% concordance between known and imputed genotypes with call probabilities ≥ 90% using this method. SNPTEST[7] was used to test association between imputed SNPs with call probabilities ≥ 90% and male breast cancer accounting for genotype uncertainty. We generated regional association plots using SNAP[8] and LD plots using Haploview[9] along with data from the HapMap CEU population[10]. On the predication that a functional variant will be evolutionarily conserved, we filtered SNPs by a conservation score derived from pairwise alignments of genomic sequence from 17 vertebrates, including mammalian, amphibian, bird, and fish species, obtained

from the phastCons17[11] track of the UCSC Genome Browser (http://genome.ucsc.edu/) and we used data from the ENCODE project, also accessed from the UCSC Genome Browser, to inspect ChIP-seq and H3K4Me1 histone modification data. We visualised additional H3K4Me1 data pertaining to human mammary epithelial cells using the Broad Institute Integrative Genomics Viewer[12,13]. We used the AliBaba2 transcription factor binding site prediction algorithm[14] along with the TRANSFAC binding motif database[15] to make inferences regarding the effect of individual SNPs on transcription factor binding site consensus sequences.

Supplementary Note

kConFab Consortia Membership

Morteza Aghmesheh, Oncology Research Centre, Prince of Wales Hospital, Randwick, NSW 2031; David Amor, Genetic Health Services, Victoria Royal Children's Hospital, Melbourne VIC 3050; Lesley Andrews, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick, NSW 2031; Yoland Antill, Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, VIC 3002; Shane Armitage, Molecular Genetics Lab Royal Brisbane and Women's Hospital; Leanne Arnold, Molecular Genetics Lab, Royal Brisbane and Women's Hospital; Rosemary Balleine, Department of Translational Oncology, C/- Department of Medical Oncology, Westmead Hospital, Westmead, NSW 2145; Agnes Bankier, PO Box 5444 HEIDELBERG WEST, C/o The Austin Hospital, 3081 Australia; Patti Bastick, St George Hospital Medical Oncology Dept, Gray Street, Kogarah NSW 2000, Australia; John Beilby, Pathology Centre, Queen Elizabeth Medical Centre, Nedlands WA 6009; Ian Bennett, Silverton Place, 101 Wickham Terrace, Brisbane QLD 4000; Barbara Bennett, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick, NSW 2031; Geoffrey Berry, Dept of Public Health and Community Medicine, University of Sydney, Sydney, NSW 2006; Anneke Blackburn, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601; Michael Bogwitz, Familial Cancer Centre, The Royal Melbourne Hospital, Grattan Street, Parkville Victoria 3050, Australia; Meagan Brennan, NSW Breast Cancer Institute, PO Box 143, Westmead NSW 2145; Melissa Brown, Department of Biochemistry, University of Queensland, St. Lucia, QLD 4072; Michael Buckley, Molecular and Cytogenetics Unit, Prince of Wales Hospital, Randwick NSW 2031; Matthew Burgess, Clinical Genetics Service, Austin Health, Victoria 3084, Australia; Jo Burke, Royal Hobart Hospital, GPO Box 1061L, Hobart TAS 7001; Phyllis Butow, Medical Psychology Unit, Royal Prince Alfred Hospital, Camperdown, NSW 2204; Keith Byron, Australian Genome Research Facility, Walter & Eliza Hall Medical Research Institute, Royal Melbourne Hospital, Parkville, VIC 3050; David Callen, Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide/Hanson Institute, P.O. Box 14, Rundle Mall, SA 5000; Ian Campbell, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, VIC 3002; Deepa Chauhan, School of Psychology, University of Sydney; Georgia Chenevix-Trench, Queensland Institute of Medical Research, Royal Brisbane Hospital,

Herston QLD 4029; Alice Christian, Genetics Department, Central Region Genetics Service, Wellington Hospital, New Zealand; Christine Clarke, Westmead Institute for Cancer Research, University of Sydney, Westmead Hospital, Westmead NSW 2145; Alison Colley, Department of Clinical Genetics, Liverpool Health Service, PO Box 103, Liverpool NSW 2170; Dick Cotton, Mutation Research Centre, St Vincent's Hospital, Victoria Parade, Fitzroy VIC 3065; James Cui, Epidemiology and Preventive Medicine, Monash University, Prahan Vic 3004, Australia; Bronwyn Culling, Molecular and Clinical Genetics, Level 1 Building 65, Royal Prince Alfred Hospital, Camperdown NSW 2050; Margaret Cummings, Department of Pathology, University of Queensland Medical School, Herston NSW 4006; Sarah-Jane Dawson, Molecular Genetics Department, Cambridge University, England; Anna DeFazio, Dept. Gynaecological Oncology, Westmead Institute for Cancer Research, Westmead Hospital, Westmead NSW 2145; Martin Delatycki, Heidelberg Repatriation Hospital, PO Box 5444, Heidelberg West, Vic 3081, Australia; Rebecca Dickson, Level 2, Block 51, Royal North Shore Hospital, North Shore NSW 2408; Joanne Dixon, Central Regional Genetic Services, Wellington Hospital, Private bag 7902, Wellington, New Zealand; Alexander Dobrovic, Department of Pathology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Tracy Dudding, Hunter Genetics, Hunter Area Health Service, PO Box 84, Waratah, 2298 NSW; Ted Edkins, Clinical Chemistry, Princess Margret Hospital for Children, Box D184, Perth WA 6001; Stacey Edwards, Department of Biochemistry and Molecular Biology, University of Queensland, St Lucia Qld 4072, Australia; Maurice Eisenbruch, Department of Multicultural Health, University of Sydney, NSW 2052; Gelareh Farshid, Tissue Pathology, IMVS, Adelaide SA 5000; Susan Fawcett, Family Cancer Clinic, Monash Medical Centre, Clayton VIC 3168; Georgina Fenton, South West Family Cancer Clinic, Liverpool Hospital, Liverpool, BC NSW 1871; Michael Field, Royal North Shore Hospital, Level 2, Vindin House, St Leonards NSW 2065; Frank Firgaira, GTG, 60 - 66 Hanover Street, Fitzroy, 3065; James Flanagan, Epigenetics Unit, Department of Surgery and Oncology, Imperial College London, London W12 0NN, England; Jean Fleming, Eskitis Institute of Cell & Molecular Therapies, School of Biomolecular and Biomedical Sciences, Griffith University, Nathan QLD 4111; Peter Fong, Medical Oncology Department, Regional Cancer and Blood Services, Level 1 Building 7, Auckland City Hospital, 2 Park Rd. Grafton, Auckland 1023, New Zealand; John Forbes, University of Newcastle, Newcastle Mater Hospital, Waratah NSW 2298; Stephen Fox, Pathology Department, Level 1, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne Vic 3002; Juliet French, School of Molecular and Microbial Sciences, University of Queensland, St Lucia Qld 4072; Michael Friedlander, Department of Medical Oncology, Prince of Wales Hospital, Randwick NSW 2031; Clara Gaff, Victorian Clinical Genetics Service, Royal Melbourne Hospital, Parkville VIC 3052; Mac Gardner, Genetic Health Services Victoria, 10th Floor The Murdoch Institute, Royal Children's Hospital, Parkville VIC 3052; Mike Gattas, Queensland Clinical Genetic Service, Royal Children's Hospital, Bramston Terrace, Herston QLD 4020; Peter George, Clinical Biochemistry Unit, Canterbury Health Labs, PO Box 151, Christchurch, New

Zealand; Graham Giles, Cancer Epidemiology Centre, Anti Cancer Council of Victoria, 1 Rathdowne Street, Carlton South VIC 3052; Grantley Gill, Department of Surgery, Royal Adelaide Hospital, Adelaide SA 5000; Jack Goldblatt, Genetic Services Of WA, King Edward Memorial Hospital, 374 Bagot Road, Subiaco WA 6008; Sian Greening, Illawarra Cancer Centre, Wollongong Hospital, Private Mail Bag 8808, South Coast Mail Centre, NSW 2521; Scott Grist, Department of Haematology and Genetic Pathology, Flinders Medical Centre, SA; Eric Haan, Department of Medical Genetics, Women's and Children's Hospital, North Adelaide SA 5006; Kate Hardie, Room 430 Bldg 76, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia QLD 4072; Marion Harris, Familial Cancer Clinic, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne VIC 3002; Stewart Hart, Breast and Ovarian Cancer Genetics, Monash Medical Centre, 871 Centre Road, Bentleigh East VIC 3165; Nick Hayward, Queensland Institute for Medical Research, Royal Brisbane Hospital, Post Office, Herston QLD 4029; Sue Healey, Queensland Institute of Medical Research (QIMR), 300 Herston Road, Herston Qld Q4006; Louise Heiniger, Medical Psychology Research Unit, The University of Sydney, Sydney NSW 2006; John Hopper, Centre for M.E.G.A. Epidemiology, University of Melbourne, Level 1, 723 Swanston Street, Carlton VIC 3010; Evelyn Humphrey, Royal Hobart Hospital, GPO Box 1061L, Hobart TAS 7001; Clare Hunt, Southern Health Familial Cancer Centre, Monash Medical Centre, Special Medicine Building, 246 Clayton Rd, Clayton Victoria 3168, Australia; Paul James, Genetic Health Sevices, Monash Medical Centre, Clayton Vic; Mark Jenkins, Centre for M.E.G.A. Epidemiology, The University of Melbourne, 723 Swanston Stree, Carlton VIC 3053; Alison Jones, Molecular Genetics Lab, Royal Brisbane and Women's Hospital, QLD; Rick Kefford, Westmead Hospital, Westmead NSW 2145; Alexa Kidd, Clinical Genetics Departments, Central Regional Genetics Service, Wellington Hospital, New Zealand; Belinda Kiely, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown Sydney NSW 1450; Judy Kirk, Familial Cancer Service, Department of Medicine, Westmead Hospital, Westmead NSW 2145; Jessica Koehler, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031; James Kollias, Breast Endocrine and Surgical Unit, Royal Adelaide Hospital, North Terrace SA 5000; Serguei Kovalenko, Genetic Technologies Limited, 60-66 Hanover Street, Fitzroy Vic 3065; Sunil Lakhani, UQ Centre for Clinical Research, Level 6 Building 71/918, University of Queensland, The Royal Brisbane & Women's Hospital Herston, 4029; Amanda Leaming, Wesley Breast Clinic, Chasely Street, Auchenflower, Brisbane Qld 4066; Jennifer Leary, Familial Cancer Laboratory, Westmead Hospital, Westmead NSW 2145; Jacqueline Lim, Dept of Psychological Medicine, Royal North Shore Hospital, St Leonards NSW 2065; Geoff Lindeman, Breast Cancer Laboratory, Walter and Eliza Hall Institute, PO Royal Melbourne Hospital, Parkville VIC 3050; Lara Lipton, Medical Oncology and Clinical Haematology Unit, Western Hospital, Footscray VIC; Liz Lobb, Medical Psychology Research Unit, Brennan MacCallum Building, The University of Sydney, Camperdown, 2006; Graham Mann, Westmead Institute for Cancer Research, Westmead Millennium Institute, Westmead NSW 2145;

Deborah Marsh, Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards NSW 2065; Sue Anne McLachlan, Department of Oncology, St Vincent's Hospital, 41 Victoria Parade, Fitzroy VIC 3065; Bettina Meiser, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031; Cliff Meldrum, Molecular Pathology Dept, 1st Floor, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne Vic 3002; Roger Milne, Centro Nacional de Investigaciones Oncologicas, C/ Melchor Fernández Almagro, 3E-28029 Madrid, Spain; Gillian Mitchell, Family Cancer Clinic, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Beth Newman, School of Public Health Road, Queensland University of Technology, Victoria Park, Kelvin Grove QLD 4059; Shona O'Connell, Southern Health Familial Cancer Centre, Special Medicine Building, 246 Clayton Road, Clayton Vic 3168; Imelda O'Loughlin, St Vincent's Breast Clinic, PO Box 4751, Toowoomba QLD 4350; Richard Osborne, Dept of Public Health and Community Medicine, 200 Berkeley Street, Carlton VIC 3053; Nick Pachter, Familial Cancer and Clinical Genetics, Royal Melbourne Hospital, Grattan Street, Parkville VIC 3050, Australia; Briony Patterson, Tas Clinical Genetics Service, Royal Hobart Hospital, GPO Box 1061, Hobart Tasmania 7001, Australia; Lester Peters, Radiation Oncology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Kelly Phillips, Department of Medical Oncology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Melanie Price, Medical Psychology, University of Sydney, Sydney, 2006; Lynne Purser, The Centre for Genetics Education NSW Health, PO Box 317, St Leonards NSW 1590, Australia; Tony Reeve, Cancer Genetics Laboratory, University of Otago, PO Box 56, Dunedin, New Zealand; Jeanne Reeve, Northern Regional Genetic Service, Auckland Hospital, New Zealand; Robert Richards, Dept of Cytogenetics and Molecular Genetics, Women and Children's Hospital, Adelaide SA 5006; Edwina Rickard, Familial Cancer Centre, Westmead Hospital, Westmead NSW 2145; Bridget Robinson, Oncology Service, Christchurch Hospital, Private Bag 4710, Christchurch, New Zealand; Barney Rudzk, Molecular Pathology Department, The University of Melbourne, Parkville Vic 3050; Mona Saleh, Centre for Genetic Education, Prince of Wales Hospital, Randwick NSW 2031; Elizabeth Salisbury, UNSW Prince of Wales Hospital, Randwick, 2031 NSW; Joe Sambrook, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Christobel Saunders, School of Surgery and Pathology, QE11 Medical Centre, M block 2nd Floor, Nedlands WA 6907; Robyn Sayer, Gynaecological Cancer Centre, Royal Hospital for Women, Randwick NSW 2011; Elizabeth Scott, South View Clinic, Suite 13, Level 3 South Street, Kogarah NSW 2217; Clare Scott, Research Department, WEHI, C/o Royal Melbourne Hospital, Parkville, 3050; Rodney Scott, Hunter Area Pathology Service, John Hunter Hospital, Locked Bag 1 Regional Mail Centre, NSW 2310; Ram Seshadri, Department of Haematology, Flinders Medical Centre, Bedford Park SA 5042; Adrienne Sexton, Familial Cancer Centre, Royal Melbourne Hospital, Grattan Street, Parkville Vic 3050; Raghwa Sharma, Dept of Tissue Pathology, Westmead Hospital, Westmead NSW 2145; Andrew Shelling, Obstetrics and Gynaecology, University of Auckland, New Zealand; Melissa Southey, Genetic

Epidemiology Laboratory, Department of Pathology, University of Melbourne, VIC 3010; Amanda Spurdle, Cancer Unit, Queensland Institute of Medical Research, Herston QLD 4029; Graeme Suthers, South Australian Clinical Genetics Service, Centre for Medical Genetics, Women and Children's Hospital, North Adelaide SA 5006; Pamela Sykes, Molecular Pathology, Flinders Medical Centre, Flinders Drive, Bedford Park, 5042; Jessica Taylor, Familial Cancer and Genetics Medicine, Royal Melbourne Hospital, 2nd Floor Grattan Street, Parkville Vic 3050, Australia; Donna Taylor, Department of Radiology, Royal Perth Hospital, Perth WA; Ella Thompson, Cancer Genetics, Research Department, 3rd level, Peter MaCallum Cancer Centre, East Melbourne Vic 3002; Heather Thorne, Research Department, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002; Sharron Townshend, Genetic Services of Western Australia, 3rd Floor Agnes Walsh House, 374 Bagot Rd, Subiaco WA 6008; Alison Trainer, University of NSW, Prince of Wales Hospital, Barker Street, Randwick NSW 2031; Lan Tran, Medical Psychology Unit, University of Sydney, NSW 2006; Kathy Tucker, Heredity Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031; Janet Tyler, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031; Jane Visvader, The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Parkville VIC 3050; Logan Walker, Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, P.O. Royal Brisbane Hospital, Herston Qld 4027, Australia; Ian Walpole, Genetic Services of WA, King Edward Memorial Hospital, Subiaco WA 6008; Robin Ward, Department of Medical Oncology, Prince of Wales Hospital, Randwick NSW 2031; Paul Waring, Department of Pathology, University of WA, 35 Stirling Highway, CRAWLEY WA 6009 Perth; Bev Warner, Cabrini Hospital, 183 Wattletree Rd, Malvern VIC 3144; Graham Warren, St Vincent's Breast Clinic, PO Box 4751, Toowoomba QLD 4350; Rachael Williams, Family Cancer Clinic, St Vincent's Hospital, Darlinghurst NSW 2010; Judy Wilson, Block 4, Level 5, Royal North Shore Hospital, St Leonards NSW 2065; Ingrid Winship, Department of Genetics, Royal Melbourne Hospital, Parkville, 3050; Mary Ann Young, Familial Cancer Clinic, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002.

References

1. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* **35**, 34-41 (2006).
2. Reich, D., Price, A.L. & Patterson, N. Principal component analysis of genetic data. *Nat Genet* **40**, 491-2 (2008).
3. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
4. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
5. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).

6. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
7. 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).
8. Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-9 (2008).
9. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
10. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
11. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-50 (2005).
12. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
13. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
14. Grabe, N. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* **2**, S1-15 (2002).
15. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31**, 374-8 (2003).