Supplementary Online Content

Lu H-M, Li S, Black MH, et al. Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA Oncol*. Published online August 16, 2018. doi:10.1001/jamaoncol.2018.2956

This supplementary material has been provided by the authors to give readers additional information about their work.

eTable 1. Search Window of 625 Cancer-Associated Genes

Genes	Breast cancer	Ovarian cancer	Other cancers
	(*n* = 50)	(*n* = 30)	(*n* = 598)
ABCB1			X
ABCB11			X
ABCC2			X
ABCC4			X
ABL1			X
ABL2			X
ACO1			X
ACVR1B	X		X
ACVR2A			X
ACVR2B			X
ADNP			X
AJUBA			X
AKT1	X		X
AKT2			X
AKT3			X
ALK			X
ALKBH6			X
ALOX12B			X
ALPK2			X
AMER1			X
APC			X
APCDD1			X
APITD1			X
APOL2			X
AR			X
ARAF			X
ARFRP1			X
ARHGAP35			X
ARID1A	X		X
ARID1B			X
ARID2			X
ARID5B			X
ARL11			X
ASXL1			X
ASXL2			X
ASXL3			X
ATM	X		X
ATP5B			X
ATR	X		X
ATRIP			X
ATRX			X
AURKA			X
AURKB			X
AXIN1			X
AXIN2			X
AXL			X
AZGP1			X
B2M			X
B4GALT3			X

BACH1			X
BAK1			X
BAP1			X
BARD1	X		X
BCL2			X
BCL2L11			X
BCL2L2			X
BCL6			X
BCLAF1			X
BCOR			X
BCORL1			X
BCR			X
BLM			X
BMPR1A			X
BRAF			X
BRCA1	X	X	X
BRCA2	X	X	X
BRE			X
BRIP1		X	X
BRWD3			X
BTG1			X
BTK			X
BUB1B			X
C11orf30			X
CAP2			X
CARD11			X
CASP8			X
CBFB	X		
CBL			X
CBLB			X
CBLC			X
CCND1			X
CCND2			X
CCND3			X
CCNE1			X
CD1D			X
CD70			X
CD79A			X
CD79B			X
CDC27			X
CDC73			X
CDH1	X		X
CDH12			X
CDH18			X
CDK12		X	
CDK4			X
CDK6			X
CDK8			X
CDKN1A			X
CDKN1B	X		X
CDKN2A	X		X
CDKN2B		X	
CDKN2C			X

CEBPA			X
CENPL			X
CEP76			X
CERS2			X
CHD4			X
CHD8			X
CHEK1			X
CHEK2	X		X
CHUK			X
CIC			X
CNBD1			X
CNKSR1		X	
COL7A1			X
COMT			X
CRBN			X
CREBBP		X	
CRIPAK			X
CRKL			X
CRLF2	X		
CSF1R			X
CTCF	X		X
CTNNA1			X
CTNNB1			X
CUL4A			X
CUL4B	X		
CUX1			X
CYLD			X
CYP17A1			X
CYP1B1			X
CYP2C19			X
CYP2C8			X
CYP2D6			X
CYP3A4			X
CYP3A5			X
DAXX			X
DCAF6			X
DDB2			X
DDR1			X
DDR2			X
DDX11			X
DDX3X			X
DDX5			X
DIAPH1			X
DICER1			X
DIDO1			X
DIS3			X
DIS3L2			X
DKC1			X
DLC1			X
DNER			X
DNMT1			X
DNMT3A			X
DOCK8			X

DOT1L			X
DPYD			X
ECSCR			X
EGFR			X
EGR3			X
EIF2S2			X
EIF3A			X
EIF4A2			X
ELANE			X
ELF3			X
EME1			X
EME2			X
EML4			X
EP300	X		X
EPHA2			X
EPHA3			X
EPHA5			X
EPHB1			X
EPHB2			X
EPHB6			X
EPPK1		X	
ERBB2	X	X	X
ERBB3	X		X
ERBB4			X
ERCC1			X
ERCC2		X	
ERCC3			X
ERCC4			X
ERCC5			X
ERG			X
ESR1			X
ESR2			X
ETV1			X
ETV4			X
ETV5			X
ETV6			X
EWSR1			X
EXO1			X
EXT1			X
EXT2			X
EZH1			X
EZH2			X
EZR			X
FAH			X
FAM129B			X
FAM46C			X
FANCA			X
FANCB			X
FANCC			X
FANCD2			X
FANCE			X
FANCF			X
FANCG			X

FANCI			X
FANCL			X
FANCM	X		X
FAS			X
FAT1			X
FAT3			X
FBXW7			X
FCGR1A			X
FCGR2A			X
FCGR3A			X
FGF10			X
FGF12			X
FGF14			X
FGF19			X
FGF23			X
FGF3			X
FGF4			X
FGF6			X
FGF7			X
FGFBP1			X
FGFR1			X
FGFR2	X		X
FGFR3			X
FGFR4			X
FH			X
FLCN			X
FLT1			X
FLT3			X
FLT4			X
FOXA1	X		
FOXA2			X
FOXL2			X
FOXQ1			X
FUBP1			X
FZD1			X
GAB2			X
GATA1			X
GATA2			X
GATA3	X		
GBA			X
GID4			X
GJB2			X
GNA11			X
GNA13			X
GNAQ			X
GNAS			X
GNB1			X
GPC3			X
GPR124			X
GPS2			X
GRIN2A			X
GRM3			X
GSK3B			X

GSTP1			X
GUCY1A2			X
H3F3A			X
H3F3C			X
HAUS3			X
HDAC4			X
HES1			X
HFE			X
HGF			X
HIF1A			X
HIST1H1C			X
HIST1H1E			X
HIST1H2BD			X
HIST1H3B	X		
HIST1H4E			X
HLA-A			X
HLA-B			X
HLA-G			X
HMBS			X
HNF1A			X
HRAS			X
HSP90AB1			X
IDH1			X
IDH2			X
IGF1			X
IGF1R			X
IGF2			X
IKBKE			X
IKZF1			X
IL7R			X
ING1			X
INHA			X
INHBA			X
INPPL1			X
IPO7			X
IRF4			X
IRS2			X
ITGAV			X
ITK			X
ITPA			X
JAK1			X
JAK2			X
JAK3			X
JUN			X
KAT6A			X
KDM5A			X
KDM5C			X
KDM6A			X
KDR			X
KEAP1			X
KIF5B			X
KIT			X
KLF4			X

KLHL6			X
KMT2A			X
KMT2B			X
KMT2C			X
KMT2D			X
KRAS	X	X	X
LIFR			X
LMO1			X
LRP1B			X
LRP2			X
LRRK2			X
MALAT1	X		
MAN1B1			X
MAP2K1			X
MAP2K2			X
MAP2K4	X		X
MAP3K1	X		
MAP3K13			X
MAP3K15			X
MAP4K1			X
MAP4K3			X
MAPK1			X
MAPK8IP1			X
MAX			X
MBD1			X
MC1R			X
MCL1			X
MDM2			X
MDM4			X
MECOM			X
MED12			X
MED23	X		
MEF2A			X
MEF2B			X
MEN1			X
MET			X
MGA			X
MIR142			X
MITF			X
MLH1		X	X
MNDA			X
MORC4			X
MPL			X
MRE11A	X	X	X
MSH2		X	X
MSH6		X	X
MTAP			X
MTHFR			X
MTOR			X
MUTYH	X	X	X
MXRA5			X
MYB	X		
MYC			X

MYCL			X
MYCN			X
MYD88			X
MYLK			X
NAV3			X
NBN	X		X
NBPF1			X
NCOR1	X		
NEIL1			X
NF1	X		X
NF2			X
NFE2L2			X
NFKBIA			X
NKX2-1			X
NOTCH1			X
NOTCH2			X
NOTCH3			X
NOTCH4			X
NPM1			X
NQO1			X
NRAS		X	X
NRP2			X
NSD1			X
NTN4			X
NTRK1			X
NTRK2			X
NTRK3			X
NUP93			X
ODAM			X
OTUD7A			X
PAK3			X
PAK7			X
PALB2	X		X
PAPD5			X
PARP1			X
PARP2			X
PARP3			X
PARP4			X
PAX5			X
PBRM1			X
PCBP1			X
PCDH10			X
PDAP1			X
PDCD2L			X
PDGFRA			X
PDGFRB			X
PDK1			X
PDSS2			X
PHF6			X
PHOX2B			X
PIK3C2G		X	
PIK3C3			X
PIK3CA	X		X

PIK3CG			X
PIK3R1	X		X
PIK3R2			X
PLCG2			X
PML			X
PMS2		X	X
PMS2CL			X
PNRC1			X
POLD1			X
POLE			X
POLH			X
POLI			X
POLK			X
POLQ			X
PORCN			X
POU2AF1			X
POU2F2			X
PPM1D			X
PPP2R1A			X
PPP6C			X
PRDM1			X
PRKAR1A			X
PRKDC			X
PRLR			X
PRPF40B			X
PRSS1			X
PRSS8			X
PTCH1			X
PTEN	X	X	X
PTPN11			X
PTPRC			X
PTPRD			X
QKI			X
RAB40A	X		
RAC1			X
RAD21			X
RAD50	X	X	X
RAD51			X
RAD51B			X
RAD51C		X	X
RAD51D		X	X
RAD52			X
RAD54L			X
RAF1			X
RALY			X
RARA			X
RASA1			X
RB1	X	X	X
RBM10			X
RBMX			X
RECQL	X		
RECQL4			X
REL			X

RET			X
REV3L			X
RHBDF2			X
RHEB			X
RHOA			X
RICTOR			X
RIT1			X
RMI1			X
RMI2			X
RMRP			X
RNF43			X
ROS1			X
RPA1			X
RPA2			X
RPA4			X
RPL22			X
RPL5			X
RPS14			X
RPS15			X
RPS2			X
RPTOR			X
RUNX1	X		
RUNX1T1			X
RUNX3			X
RXRA			X
SBDS			X
SDHA		X	X
SDHAF2			X
SDHB			X
SDHC			X
SDHD			X
SERPINA1			X
SERPINB13			X
SETBP1			X
SETD2			X
SETDB1			X
SF1			X
SF3B1	X		
SGK1			X
SH2B3			X
SH2D1A			X
SIN3A			X
SIRPA			X
SIRT4			X
SLC19A1			X
SLC22A2			X
SLC25A13			X
SLCO1B3			X
SLX4			X
SMAD2			X
SMAD3			X
SMAD4			X
SMARCA4		X	X

SMARCB1		X	
SMARCD1			X
SMARCE1			X
SMC1A			X
SMC3			X
SMO			X
SNX25			X
SOCS1			X
SOD2			X
SOS1			X
SOX10			X
SOX17			X
SOX2			X
SOX9			X
SPEN	X		
SPOP			X
SPRY4			X
SRC			X
SRSF2			X
SRY			X
STAG2	X		
STAT3			X
STAT4			X
STK11		X	X
STK19			X
STK38			X
STX2			X
SUFU			X
SULT1A1			X
SUZ12			X
SYK			X
SZRD1			X
TAF1			X
TBC1D12			X
TBL1XR1	X		X
TBX3	X		
TCEB1			X
TCF7L2			X
TELO2			X
TERT			X
TET2			X
TFG			X
TGFBR1			X
TGFBR2			X
TIMM17A			X
TIPARP			X
TLK2			X
TLR4			X
TMEM127			X
TMPRSS2			X
TNF			X
TNFAIP3			X
TNFRSF14			X

TOP1			X
TOP3A			X
TOP3B			X
TP53	X	X	X
TP53BP1			X
TPMT			X
TPX2			X
TRAF3			X
TRAF7			X
TRIM37			X
TRRAP			X
TSC1			X
TSC2			X
TSHR			X
TSHZ2			X
TSHZ3			X
TYMS			X
TYR			X
U2AF1			X
U2AF2			X
UGT1A1			X
UMPS			X
UROD			X
USP1			X
USP9X			X
VANGL2			X
VEZF1	X		
VHL			X
WAC			X
WAS			X
WASF3			X
WDR48			X
WISP3			X
WNK1			X
WRN			X
WT1			X
XPA			X
XPC			X
XPO1			X
XRCC2			X
XRCC3			X
ZFHX3			X
ZNF217			X
ZNF703			X
ZRANB3			X
ZRSR2			X

eTable 2. Clinical Characteristics of BC Patients With or Without Pathologic Information

	BC patients with hormone receptor status N (%)	BC patients without hormone receptor status N (%)
Number of patients	4,447	5,192
Race/Ethnicity (%)[a]		
Caucasian	3,244 (72.9)	3,964 (76.3)
African American	501 (11.3)	484 (9.3)
Hispanic	433 (9.7)	441 (8.5)
Asian	269 (6.0)	303 (5.8)
Age at diagnosis, years		
Mean ± SD	48.9 ± 11.5	48.6 ± 11.5
<=45	1,782 (40.1)	2,178 (41.9)
46-60	1,830 (41.2)	2,046 (39.4)
>60	735 (16.5)	783 (15.1)
Not provided	100 (2.2)	185 (3.6)
Gender (%)		
male	51 (1.1)	72 (1.4)
female	4,396 (98.9)	5,120 (98.6)
Breast cancer histology (%)		
IDC	3,476 (78.2)	3,353 (64.6)
ILC	284 (6.4)	209 (4.0)
IDC and ILC	87 (2.0)	54 (1.0)
Other	600 (13.5)	1,576 (30.4)
Bilateral disease (%)		
Yes	506 (11.4)	815 (15.7)
No	3,941 (88.6)	4,377 (84.3)
First-/second-degree relative with any cancer (%)		
Yes	3,726 (83.8)	4,426 (85.2)
No	139 (3.1)	143 (2.8)
Not Provided	582	623
First-/second-degree relative with BC (%)		
Yes	2374(53.4)	3030(58.4)
No	1491(33.5)	1539(29.6)
Not provided	582(13.1)	623(12.0)
First-/second-degree relative with OV (%)		
Yes	425(9.6)	556(10.7)
No	3440(77.4)	4013(77.3)
Not provided	582(13.1)	623(12.0)
Carriers of pathogenic variants in HBOC genes (%)		
ATM	28 (0.6)	51 (1.0)
BRCA1	56 (1.3)	72 (1.4)
BRCA2	101 (2.3)	133 (2.6)
BRIP1	7 (0.2)	5 (0.1)

CDH1	7 (0.2)	5 (0.1)
CHEK2	47 (1.1)	63 (1.2)
PALB2	28 (0.6)	33 (0.6)
PTEN	4 (0.1)	4 (0.1)
RAD51C	4 (0.1)	3 (0.1)
RAD51D	2 (0.0)	1 (0.0)
TP53	11 (0.2)	21 (0.4)
BARD1	7 (0.2)	5 (0.1)
NF1	4 (0.1)	8 (0.2)
PTEN	4 (0.1)	4 (0.1)
CDH1	7 (0.2)	5 (0.1)
MLH1	14 (0.3)	10 (0.2)
MSH2	5 (0.1)	4 (0.1)
MSH6	21 (0.5)	44 (0.8)
PMS2	59 (1.3)	45 (0.9)
MRE11A	4 (0.1)	6 (0.1)
RAD50	5 (0.1)	6 (0.1)
NBN	1 (0.0)	6 (0.1)
BRIP1	7 (0.2)	5 (0.1)
RAD51C	4 (0.1)	3 (0.1)
RAD51D	2 (0.0)	1 (0.0)
CDKN2A	6 (0.1)	11 (0.2)

[a]Inferred ethnicity using principal component analysis (PCA) based on exome sequencing data (eMethods)

eTable 3. BC and OV Associations Identified From Comparisons of Cases With In-Lab Controls and Reference Population

Genes	In-Lab Control				
	Cases[†]	Controls[†]	OR (95% CI)[a]	*p*-value	FDR-corrected
BC					
*ATM**	79 (11512)	14 (6029)	2.97 (1.67-5.68)	4.02×10^{-5}	1.05×10^{-3}
*CHEK2**	110 (13553)	24 (6442)	2.19 (1.40-3.56)	2.66×10^{-4}	5.54×10^{-3}
MSH6	65 (17362)	11 (7586)	2.59 (1.35-5.44)	1.69×10^{-3}	.03
*PALB2**	61 (15532)	5 (7020)	5.53 (2.24-17.65)	6.50×10^{-6}	2.25×10^{-4}
*TP53**	32 (17275)	3 (7400)	4.58 (1.43-23.36)	4.73×10^{-3}	.06
OV					
ATM	16 (2475)	14 (6156)	2.85 (1.30-6.32)	4.36×10^{-3}	.04
*MSH6**	23 (3944)	11 (7819)	4.16 (1.95-9.47)	6.55×10^{-5}	1.48×10^{-3}
*RAD51C**	6 (1588)	0 (4846)	-	2.24×10^{-4}	3.82×10^{-3}
*TP53**	9 (3680)	1 (7550)	18.50 (2.56-808.10)	3.05×10^{-4}	4.15×10^{-3}
Genes	**gnomAD[b]**				
	Cases[†]	**Controls[†]**	**OR (95% CI)[a]**	***p*-value**	**FDR-corrected**
BC					
*ATM**	78 (11512)	482 (245934)	3.47 (2.70-4.42)	1.29×10^{-18}	3.62×10^{-17}
*CHEK2**	110 (13553)	840 (245838)	2.39 (1.94-2.92)	1.41×10^{-14}	3.25×10^{-13}
MSH6	59 (17362)	146 (245790)	5.74 (4.16-7.82)	1.96×10^{-22}	7.09×10^{-21}
*PALB2**	61 (15532)	215 (246188)	4.51 (3.34-6.02)	4.12×10^{-19}	1.30×10^{-17}
*TP53**	24 (17275)	54 (246118)	6.34 (3.75-10.44)	9.48×10^{-11}	1.71×10^{-9}
OV					
ATM	15 (2466)	482 (245959)	3.12 (1.73-5.20)	1.75×10^{-4}	2.77×10^{-3}
*MSH6**	21 (3944)	146 (245790)	9.01 (5.41-14.30)	3.87×10^{-13}	1.05×10^{-11}
*RAD51C**	5 (1588)	130 (246142)	5.98 (1.91-14.33)	1.88×10^{-3}	.02
*TP53**	7 (3680)	54 (246130)	8.69 (3.33-19.16)	3.26×10^{-5}	6.88×10^{-4}

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)
[b]The differences of case mutation samples between comparisons with In-Lab control and gnomAD are due to calibration of varying site coverage in gnomAD (Methods)

eTable 4. BC and OV Associations Among Caucasians

Genes	In-Lab Control				
	Cases[†]	Controls[†]	OR (95% CI)[a]	*p*-value	FDR-corrected
BC					
*ATM**	72 (8828)	8 (4006)	4.11 (1.98-9.89)	$9.38x10^{-6}$	$8.44x10^{-4}$
*CHEK2**	107 (10192)	22 (4311)	2.07 (1.30-3.44)	$1.30x10^{-3}$	.03
MSH6	50 (13698)	7 (5374)	2.81 (1.27-7.35)	$7.25x10^{-3}$	.11
*PALB2**	45 (11144)	5 (4640)	3.76 (1.50-12.14)	$1.61x10^{-3}$	.03
OV					
ATM	14 (2124)	8 (4162)	3.44 (1.35-9.49)	$5.35x10^{-3}$	.04
*MSH6**	19 (3302)	7 (5412)	4.47 (1.80-12.59)	$3.53x10^{-4}$	$5.77x10^{-3}$
*RAD51C**	5 (1390)	0 (3078)	-	$2.90x10^{-3}$	.03
*TP53**	9 (3088)	1 (5180)	15.13 (2.10-661.56)	$9.33x10^{-4}$	.01
Genes	**gnomAD[b]**				
	Cases[†]	**Controls[†]**	**OR (95% CI)[a]**	***p*-value**	**FDR-corrected**
BC					
*ATM**	71 (8828)	265 (111480)	3.40 (2.58-4.44)	$5.66x10^{-16}$	$1.33x10^{-14}$
*CHEK2**	107 (10192)	475 (111460)	2.48 (1.99-3.07)	$1.53x10^{-14}$	$3.26x10^{-13}$
MSH6	45 (13698)	56 (111290)	6.55 (4.32-9.87)	$1.15x10^{-17}$	$3.01x10^{-16}$
*PALB2**	45 (11144)	98 (111668)	4.62 (3.17-6.64)	$4.65x10^{-14}$	$9.11x10^{-13}$
OV					
ATM	14 (2124)	265 (111517)	2.79 (1.50-4.77)	$8.90x10^{-4}$	.01
*MSH6**	17 (3302)	56 (111290)	10.28 (5.59-17.97)	$2.38x10^{-11}$	$5.66x10^{-10}$
*RAD51C**	4 (1390)	66 (111652)	4.88 (1.29-13.12)	.01	.13
*TP53**	7 (3088)	31 (111626)	8.18 (3.04-18.95)	$6.17x10^{-5}$	$1.14x10^{-3}$

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)
[b]The differences of case mutation samples between comparisons with In-Lab control and gnomAD are due to calibration of varying site coverage in gnomAD (Methods)

eTable 5. BC and OV Associations Tested With the Combined Multivariate Collapsing Method

Genes	OR[a]	p-value	FDR-corrected
BC			
*ATM**	2.82	6.45×10^{-4}	.01
*CHEK2**	1.94	6.83×10^{-5}	3.04×10^{-3}
MSH6	2.68	1.28×10^{-3}	.02
*PALB2**	3.58	1.01×10^{-3}	.02
OV			
ATM	3.26	2.46×10^{-3}	.03
*MSH6**	3.26	2.48×10^{-3}	.03
*RAD51C**	-	6.62×10^{-3}	.06
*TP53**	3.25	.06	.26

*Known breast or ovarian cancer gene
[a]Odds Ratio estimated using CMC (Combined Multivariate and Collapsing), adjusted for the first three principle components of ancestry (eMethods)

eTable 6. BC and OV Associations Tested Among Females

Genes	In-Lab Control			
	Cases[†]	Controls[†]	OR (95% CI)[a]	*p*-value
BC				
*ATM**	81 (11368)	12 (3751)	2.24 (1.21-4.51)	7.57×10^{-3}
*CHEK2**	109 (12152)	14 (3826)	2.46 (1.41-4.66)	6.20×10^{-4}
MSH6	65 (17950)	5 (5170)	3.75 (1.53-11.95)	1.31×10^{-3}
*PALB2**	63 (16095)	3 (4772)	6.25 (2.04-31.12)	9.16×10^{-5}
OV				
ATM	16 (2466)	12 (3902)	2.12 (0.94-4.91)	.05
*MSH6**	23 (3944)	5 (5212)	6.11 (2.27-20.58)	2.91×10^{-5}
*RAD51C**	6 (1588)	0 (2906)	-	1.94×10^{-3}
*TP53**	9 (3680)	1 (5020)	12.30 (1.70-538.09)	2.67×10^{-3}
Genes	**gnomAD[b]**			
	Cases[†]	**Controls[†]**	**OR (95% CI)[a]**	***p*-value**
BC				
*ATM**	78 (11512)	221 (111194)	3.43 (2.61-4.46)	5.37×10^{-17}
*CHEK2**	110 (13553)	371 (111191)	2.44 (1.96-3.03)	5.72×10^{-14}
MSH6	59 (17362)	57 (111138)	6.64 (4.54-9.74)	8.73×10^{-22}
*PALB2**	61 (15532)	116 (111322)	3.78 (2.72-5.20)	1.58×10^{-14}
OV				
ATM	15 (2466)	221 (113218)	3.07 (1.69-5.19)	2.32×10^{-4}
*MSH6**	21 (3944)	57 (111133)	10.43 (6.00-17.49)	1.35×10^{-13}
*RAD51C**	5 (1588)	44 (111311)	7.99 (2.47-20.13)	6.24×10^{-4}
*TP53**	7 (3680)	19 (111267)	11.16 (3.96-27.74)	1.32×10^{-5}

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)
[b]The differences of case mutation samples between comparisons with In-Lab control and gnomAD are due to calibration of varying site coverage in gnomAD (Methods)

eTable 7. BC Associations After Excluding Patients With First- and Second-Degree Relatives With OV

Genes		In-Lab Control			
	Cases[†]	Controls[†]	OR (95% CI)[a]	*p*-value	FDR-corrected
BC					
*ATM**	75 (10270)	14 (6038)	3.17 (1.77-6.07)	1.30×10^{-5}	3.42×10^{-4}
*CHEK2**	98 (12148)	24 (6442)	2.17 (1.38-3.56)	3.71×10^{-4}	7.80×10^{-3}
MSH6	60 (16451)	11 (7785)	2.59 (1.35-5.46)	2.06×10^{-3}	.04
*PALB2**	58 (14274)	5 (7132)	5.82 (2.35-18.59)	2.98×10^{-6}	1.04×10^{-4}

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)

eTable 8. BC and OV Associations After Excluding Patients With Other Cancers

Genes	In-Lab Control			
	Cases[†]	Controls[†]	OR (95% CI)[a]	*p*-value
BC				
*ATM**	77 (10806)	14 (5994)	3.07 (1.72-5.87)	2.32×10^{-5}
*CHEK2**	101 (12672)	24 (6442)	2.15 (1.36-3.51)	4.08×10^{-4}
MSH6	55 (17086)	11 (7784)	2.28 (1.18-4.84)	.01
*PALB2**	64 (15230)	5 (7275)	6.14 (2.50-19.54)	9.34×10^{-7}
OV				
ATM	14 (2232)	14 (6156)	2.77 (1.22-6.28)	8.74×10^{-3}
*MSH6**	22 (3544)	11 (7826)	4.44 (2.06-10.15)	3.28×10^{-5}
*RAD51C**	5 (1422)	0 (4888)	-	5.78×10^{-4}
*TP53**	7 (3320)	1 (7550)	15.95 (2.05-716.36)	1.45×10^{-3}

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)

eTable 9. BC and OV Associations Stratified by Loss of Function, Missense and Other Types of Variants

Genes	Loss of function variants			
	Case[†]	InlabCtrl[†]	OR (95% CI)[a]	*p*-value
BC				
*ATM**	45 (11510)	8 (6152)	3.01 (1.41-7.41)	2.12×10^{-3}
*CHEK2**	96 (11434)	18 (5818)	2.73 (1.64-4.80)	2.36×10^{-5}
MSH6	60 (16414)	10 (7326)	2.68 (1.36-5.88)	1.76×10^{-3}
*PALB2**	58 (15532)	5 (7020)	5.26 (2.13-16.81)	1.53×10^{-5}
OV				
ATM	11 (2748)	8 (6230)	3.13 (1.14-8.96)	.02
*MSH6**	23 (3944)	10 (7819)	4.58 (2.09-10.79)	2.45×10^{-5}
RAD51C	4 (1588)	0 (4846)	-	3.70×10^{-3}
*TP53**	-	-	-	-
Genes	**Missense variants**			
	Case[†]	**InlabCtrl[†]**	**OR (95% CI)[a]**	***p*-value**
BC				
*ATM**	21 (11514)	3 (5950)	3.62 (1.08-18.97)	.03
*CHEK2**	14 (14802)	6 (6830)	1.08 (0.39-3.42)	1.00
MSH6	5 (18233)	1 (7779)	2.13 (0.24-100.88)	.68
*PALB2**	-	-	-	-
OV				
ATM	3 (2110)	3 (5679)	2.69 (0.36-20.13)	.35
*MSH6**	-	-	-	-
RAD51C	-	-	-	-
*TP53**	9 (3680)	1 (7550)	18.50 (2.56-808.10)	3.05×10^{-4}
Genes	**Other types of variants**			
	Case[†]	**InlabCtrl[†]**	**OR (95% CI)[a]**	***p*-value**
BC				
*ATM**	13 (11793)	3 (5957)	2.19 (0.60-11.99)	.29
*CHEK2**	-	-	-	-
MSH6	-	-	-	-
*PALB2**	-	-	-	-
OV				
ATM	2 (2816)	3 (6346)	1.50 (0.13-13.12)	.65
*MSH6**	-	-	-	-
RAD51C	2 (2078)	0 (5383)	-	-
*TP53**	-	-	-	-

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)

eTable 10. Case-Case Analysis (ER^+PR^+ vs ER^-PR^{-b}) Among BC Patients

Genes	In-Lab Control			
	ER^+PR^+ †	ER^-PR^- †	OR (95% CI)[a]	*p*-value
BC				
*ATM**	27 (3690)	3 (1726)	4.23 (1.30-21.83)	$9.40x10^{-3}$
*CHEK2**	42 (3866)	5 (1794)	3.93 (1.55-12.75)	$1.33x10^{-3}$
MSH6	14 (5375)	8 (2519)	0.82 (0.32-2.26)	.65
*PALB2**	19 (4907)	12 (2295)	0.74 (0.34-1.67)	.44

*Known breast or ovarian cancer gene
†Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)
[b]ER+PR+ represents all HR+ and TPBC; ER-PR- represents all HR- and TNBC

eTable 11. Case-Case Analysis (Onset Age <45 vs ≥45 yr) Among BC and OV Patients

Gene	In-Lab Control			
	onset age <45yr[†]	**onset age ≥45yr[†]**	**OR (95% CI)[a]**	***p*-value**
BC				
*ATM**	33 (4930)	46 (6764)	0.98 (0.61-1.58)	1.00
*CHEK2**	57 (5630)	53 (7923)	1.52 (1.02-2.25)	.03
MSH6	29 (7208)	35 (10154)	1.17 (0.69-1.97)	.53
*PALB2**	22 (6783)	42 (9517)	0.73 (0.42-1.26)	.26
OV				
*MSH6**	8 (854)	13 (3082)	2.23 (0.80-5.83)	.10
*RAD51C**	1 (344)	4 (1264)	0.92 (0.02-9.32)	1.00
*TP53**	3 (807)	6 (2913)	1.81 (0.29-8.49)	.42

*Known breast or ovarian cancer gene
[†]Mutated allele count (total allele number)
[a]Odds Ratio and 95% Confidence Interval, p-values derived from Fisher exact test (Methods)

eTable 12. Pathogenic Variants in BC Patients

Chromosome	Variant	#Mutation
11	ATM NM_000051 c.8584+2T>C	1
11	ATM NM_000051 c.9005delT p.F3002Sfs*4	1
11	ATM NM_000051 c.7189C>T p.Q2397*	1
11	ATM NM_000051 c.2295delT p.N765Kfs*12	1
11	ATM NM_000051 c.1249delA p.T417Pfs*3	1
11	ATM NM_000051 c.5549delT p.L1850Yfs*67	1
11	ATM NM_000051 c.5908C>T p.Q1970*	1
11	ATM NM_000051 c.1027_1030delGAAA p.E343Ifs*2	1
11	ATM NM_000051 c.8476_8477dupAA p.N2826Kfs*32	1
11	ATM NM_000051 c.7913G>A p.W2638*	1
11	ATM NM_000051 c.3511C>T p.Q1171*	1
11	ATM NM_000051 c.564delT p.R189Efs*2	1
11	ATM NM_000051 c.2T>C p.M1?	1
11	ATM NM_000051 c.6200C>A p.A2067D	1
11	ATM NM_000051 c.1396C>T p.Q466*	1
11	ATM NM_000051 c.8737G>T p.D2913Y	1
11	ATM NM_000051 c.2284_2285delCT p.L762Vfs*2	1
11	ATM NM_000051 c.2849T>G p.L950R	1
11	ATM NM_000051 c.3894dupT p.A1299Cfs*3	1
11	ATM NM_000051 c.8395_8404del10 p.F2799Kfs*4	1
11	ATM NM_000051 c.4804_4805delGT p.V1602Lfs*2	1
11	ATM NM_000051 c.6154G>A p.E2052K	1
11	ATM NM_000051 c.977_978delTA p.I326Rfs*3	1
11	ATM NM_000051 c.7630-2A>C	1
11	ATM NM_000051 c.8786+1G>C	1
11	ATM NM_000051 c.8977C>T p.R2993*	1
11	ATM NM_000051 c.9166delG p.V3056Cfs*19	1
11	ATM NM_000051 c.205C>T p.Q69*	1
11	ATM NM_000051 c.1402_1403delAA p.K468Efs*18	1
11	ATM NM_000051 c.7638_7646del9 p.R2547_S2549del	1
11	ATM NM_000051 c.901+1G>T	1
11	ATM NM_000051 c.2921+1G>A	1
11	ATM NM_000051 c.742C>T p.R248*	1
11	ATM NM_000051 c.829G>T p.E277*	1
11	ATM NM_000051 c.7705_7706delGA p.D2569*	1
11	ATM NM_000051 c.8050C>T p.Q2684*	1
11	ATM NM_000051 c.2250G>A p.K750K	2
11	ATM NM_000051 c.7463G>A p.C2488Y	2
11	ATM NM_000051 c.4507C>T p.Q1503*	2

11	ATM NM_000051 c.4394T>C p.L1465P	2
11	ATM NM_000051 c.1369C>T p.R457*	2
11	ATM NM_000051 c.7788G>A p.E2596E	2
11	ATM NM_000051 c.1564_1565delGA p.E522Ifs*43	2
11	ATM NM_000051 c.6100C>T p.R2034*	2
11	ATM NM_000051 c.8147T>C p.V2716A	2
11	ATM NM_000051 c.9022C>T p.R3008C	2
11	ATM NM_000051 c.6679C>T p.R2227C	2
11	ATM NM_000051 c.3848T>C p.L1283P	3
11	ATM NM_000051 c.901+1G>A	3
11	ATM NM_000051 c.3802delG p.V1268*	3
11	ATM NM_000051 c.170G>A p.W57*	4
11	ATM NM_000051 c.7271T>G p.V2424G	4
11	ATM NM_000051 c.8418+5_8418+8delGTGA	4
22	CHEK2 NM_007194 c.920dupG p.E308Rfs*4	1
22	CHEK2 NM_007194 c.1462-2A>G	1
22	CHEK2 NM_007194 c.279G>A p.W93*	1
22	CHEK2 NM_007194 c.499G>A p.G167R	1
22	CHEK2 NM_007194 c.1368dupA p.E457Rfs*33	1
22	CHEK2 NM_007194 c.1486C>T p.Q496*	1
22	CHEK2 NM_007194 c.409C>T p.R137*	1
22	CHEK2 NM_007194 c.247delC p.Q83Kfs*27	1
22	CHEK2 NM_007194 c.591delA p.V198Ffs*7	1
22	CHEK2 NM_007194 c.283C>T p.R95*	2
22	CHEK2 NM_007194 c.277delT p.W93Gfs*17	2
22	CHEK2 NM_007194 c.1263delT p.S422Vfs*15	2
22	CHEK2 NM_007194 c.917G>C p.G306A	2
22	CHEK2 NM_007194 c.349A>G p.R117G	5
22	CHEK2 NM_007194 c.1427C>T p.T476M	6
22	CHEK2 NM_007194 c.216T>G p.Y72*	7
22	CHEK2 NM_007194 c.444+1G>A	8
22	CHEK2 NM_007194 c.1100delC p.T367Mfs*15	67
2	MSH6 NM_000179 c.3155_3156delAG p.E1052Vfs*13	1
2	MSH6 NM_000179 c.742delC p.R248Efs*31	1
2	MSH6 NM_000179 c.2062_2063delGT p.V688Lfs*9	1
2	MSH6 NM_000179 c.2314C>T p.R772W	1
2	MSH6 NM_000179 c.3261delC p.F1088Sfs*2	1
2	MSH6 NM_000179 c.3725G>A p.R1242H	1
2	MSH6 NM_000179 c.3173-1G>C	1
2	MSH6 NM_000179 c.10C>T p.Q4*	1
2	MSH6 NM_000179 c.3172+1G>T	1

2	MSH6 NM_000179 c.3439-2A>G	2
2	MSH6 NM_000179 c.3261dupC p.F1088Lfs*5	2
2	MSH6 NM_000179 c.3959_3962delCAAG p.A1320Efs*6	2
2	MSH6 NM_000179 c.260+1G>C	2
2	MSH6 NM_000179 c.3226C>T p.R1076C	3
2	MSH6 NM_000179 c.2945delC p.P982Lfs*15	45
16	PALB2 NM_024675 c.1327A>T p.K443*	1
16	PALB2 NM_024675 c.3362delG p.G1121Vfs*3	1
16	PALB2 NM_024675 c.49-2A>T	1
16	PALB2 NM_024675 c.424A>T p.K142*	1
16	PALB2 NM_024675 c.658delA p.S220Vfs*3	1
16	PALB2 NM_024675 c.2006delA p.E669Gfs*3	1
16	PALB2 NM_024675 c.3426_3429delACTT p.L1142Ffs*20	1
16	PALB2 NM_024675 c.1059delA p.K353Nfs*3	1
16	PALB2 NM_024675 c.3048delT p.F1016Lfs*17	1
16	PALB2 NM_024675 c.93dupA p.L32Tfs*11	1
16	PALB2 NM_024675 c.2142_2143insTAA p.D714_D715ins*	1
16	PALB2 NM_024675 c.2748+1G>T	1
16	PALB2 NM_024675 c.3004_3007delGAAA p.E1002Tfs*4	1
16	PALB2 NM_024675 c.35delA p.E12Gfs*6	1
16	PALB2 NM_024675 c.3201+1G>C	1
16	PALB2 NM_024675 c.2968G>T p.E990*	1
16	PALB2 NM_024675 c.1914_1929delTGAGTCAAAAATGTTT p.F638Lfs*17	1
16	PALB2 NM_024675 c.1571C>G p.S524*	1
16	PALB2 NM_024675 c.1085_1086delTT p.L362Rfs*5	1
16	PALB2 NM_024675 c.745_749delCCTTT p.P249Tfs*6	1
16	PALB2 NM_024675 c.451C>T p.Q151*	1
16	PALB2 NM_024675 c.2052delC p.R686Gfs*23	1
16	PALB2 NM_024675 c.3507_3508delTC p.H1170Ffs*19	1
16	PALB2 NM_024675 c.1479delC p.T494Lfs*67	1
16	PALB2 NM_024675 c.2470dupT p.C824Lfs*2	2
16	PALB2 NM_024675 c.2488delG p.E830Sfs*21	2
16	PALB2 NM_024675 c.3350+4A>G	2
16	PALB2 NM_024675 c.3549C>G p.Y1183*	2
16	PALB2 NM_024675 c.3285_3286insGTTAATG p.N1096Vfs*5	2
16	PALB2 NM_024675 c.1317delG p.F440Lfs*12	2
16	PALB2 NM_024675 c.3549C>A p.Y1183*	2
16	PALB2 NM_024675 c.1037_1041delAAGAA p.K346Tfs*13	2
16	PALB2 NM_024675 c.3323delA p.Y1108Sfs*16	3
16	PALB2 NM_024675 c.2167_2168delAT p.M723Vfs*21	4

16	PALB2 NM_024675 c.172_175delTTGT p.Q60Rfs*7	4
16	PALB2 NM_024675 c.1240C>T p.R414*	5
16	PALB2 NM_024675 c.509_510delGA p.R170Ifs*14	5
17	TP53 NM_000546 c.455C>T p.P152L	1
17	TP53 NM_000546 c.467G>A p.R156H	1
17	TP53 NM_000546 c.638G>C p.R213P	1
17	TP53 NM_000546 c.542G>A p.R181H	1
17	TP53 NM_000546 c.844C>T p.R282W	1
17	TP53 NM_000546 c.659A>G p.Y220C	2
17	TP53 NM_000546 c.1024C>T p.R342*	2
17	TP53 NM_000546 c.374C>T p.T125M	2
17	TP53 NM_000546 c.818G>A p.R273H	3
17	TP53 NM_000546 c.641A>G p.H214R	3
17	TP53 NM_000546 c.524G>A p.R175H	5
17	TP53 NM_000546 c.848G>A p.R283H	5
17	TP53 NM_000546 c.743G>A p.R248Q	5

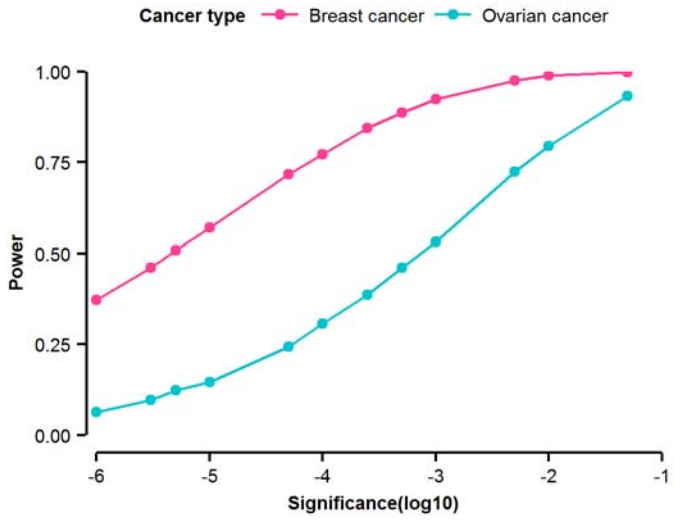
eTable 13. Pathogenic Variants in OV Patients

Chromosome	Variant	#Mutation
2	MSH6 NM_000179 c.900dupG p.K301Efs*11	1
2	MSH6 NM_000179 c.3311_3312delTT p.F1104Wfs*3	1
2	MSH6 NM_000179 c.2906_2907delAT p.Y969Lfs*5	1
2	MSH6 NM_000179 c.1059dupT p.G354Wfs*4	1
2	MSH6 NM_000179 c.3261delC p.F1088Sfs*2	1
2	MSH6 NM_000179 c.3103C>T p.R1035*	1
2	MSH6 NM_000179 c.3261dupC p.F1088Lfs*5	1
2	MSH6 NM_000179 c.10C>T p.Q4*	2
2	MSH6 NM_000179 c.3439-2A>G	2
2	MSH6 NM_000179 c.2945delC p.P982Lfs*15	12
11	ATM NM_000051 c.7886_7890delTATTA p.I2629Sfs*25	1
11	ATM NM_000051 c.2921+1G>A	1
11	ATM NM_000051 c.6100C>T p.R2034*	1
11	ATM NM_000051 c.4507C>T p.Q1503*	1
11	ATM NM_000051 c.1369C>T p.R457*	1
11	ATM NM_000051 c.1439_1448del10 p.L480Sfs*13	1
11	ATM NM_000051 c.7788G>A p.E2596E	1
11	ATM NM_000051 c.7271T>G p.V2424G	1
11	ATM NM_000051 c.7913G>A p.W2638*	1
11	ATM NM_000051 c.7463G>A p.C2488Y	1
11	ATM NM_000051 c.1689delG p.M563Ifs*4	1
11	ATM NM_000051 c.7638_7646del9 p.R2547_S2549del	1
11	ATM NM_000051 c.6200C>A p.A2067D	1
11	ATM NM_000051 c.1564_1565delGA p.E522Ifs*43	3
17	RAD51C NM_058216 c.904+5G>T	1
17	RAD51C NM_058216 c.97C>T p.Q33*	1
17	RAD51C NM_058216 c.630T>G p.Y210*	1
17	RAD51C NM_058216 c.653_654delAG p.E218Vfs*33	1
17	RAD51C NM_058216 c.224dupA p.Y75*	1
17	RAD51C NM_058216 c.837+1G>T	1
17	TP53 NM_000546 c.818G>A p.R273H	1
17	TP53 NM_000546 c.467G>A p.R156H	1
17	TP53 NM_000546 c.472C>T p.R158C	1
17	TP53 NM_000546 c.638G>C p.R213P	1
17	TP53 NM_000546 c.542G>A p.R181H	1
17	TP53 NM_000546 c.659A>G p.Y220C	1
17	TP53 NM_000546 c.524G>A p.R175H	1
17	TP53 NM_000546 c.848G>A p.R283H	2

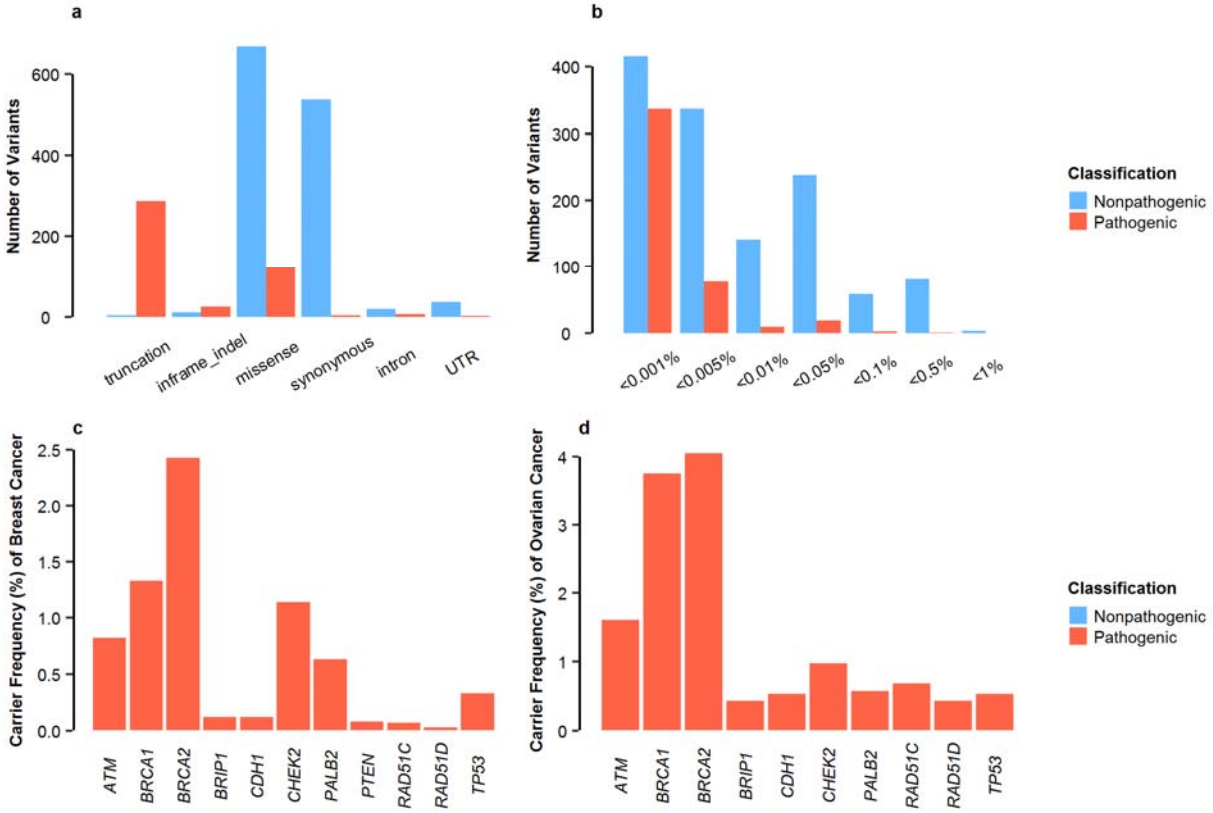
eTable 14. Pathogenic Variants in In-Lab Controls

Chromosome	Variant	#Mutation
2	MSH6 NM_000179 c.718C>T p.R240*	1
2	MSH6 NM_000179 c.2300C>T p.T767I	1
2	MSH6 NM_000179 c.2230dupG p.E744Gfs*12	1
2	MSH6 NM_000179 c.979delA p.T327Lfs*11	1
2	MSH6 NM_000179 c.2945delC p.P982Lfs*15	7
11	ATM NM_000051 c.4198A>T p.K1400*	1
11	ATM NM_000051 c.8565T>G p.S2855R	1
11	ATM NM_000051 c.6347+1G>A	1
11	ATM NM_000051 c.7926A>C p.R2642S	1
11	ATM NM_000051 c.3510dupA p.Q1171Tfs*8	1
11	ATM NM_000051 c.2921+1G>A	1
11	ATM NM_000051 c.742C>T p.R248*	1
11	ATM NM_000051 c.829G>T p.E277*	1
11	ATM NM_000051 c.7705_7706delGA p.D2569*	1
11	ATM NM_000051 c.8050C>T p.Q2684*	2
11	ATM NM_000051 c.6679C>T p.R2227C	1
11	ATM NM_000051 c.3802delG p.V1268*	1
11	ATM NM_000051 c.8418+5_8418+8delGTGA	1
16	PALB2 NM_024675 c.2920_2921delAA p.K974Efs*5	1
16	PALB2 NM_024675 c.3256C>T p.R1086*	1
16	PALB2 NM_024675 c.1479delC p.T494Lfs*67	1
16	PALB2 NM_024675 c.1037_1041delAAGAA p.K346Tfs*13	1
16	PALB2 NM_024675 c.509_510delGA p.R170Ifs*14	1
17	TP53 NM_000546 c.374C>T p.T125M	1
17	TP53 NM_000546 c.743G>A p.R248Q	2
22	CHEK2 NM_007194 c.1096-1G>A	1
22	CHEK2 NM_007194 c.433C>T p.R145W	1
22	CHEK2 NM_007194 c.247delC p.Q83Kfs*27	1
22	CHEK2 NM_007194 c.591delA p.V198Ffs*7	2
22	CHEK2 NM_007194 c.917G>C p.G306A	1
22	CHEK2 NM_007194 c.1427C>T p.T476M	4
22	CHEK2 NM_007194 c.216T>G p.Y72*	3
22	CHEK2 NM_007194 c.444+1G>A	1
22	CHEK2 NM_007194 c.1100delC p.T367Mfs*15	10

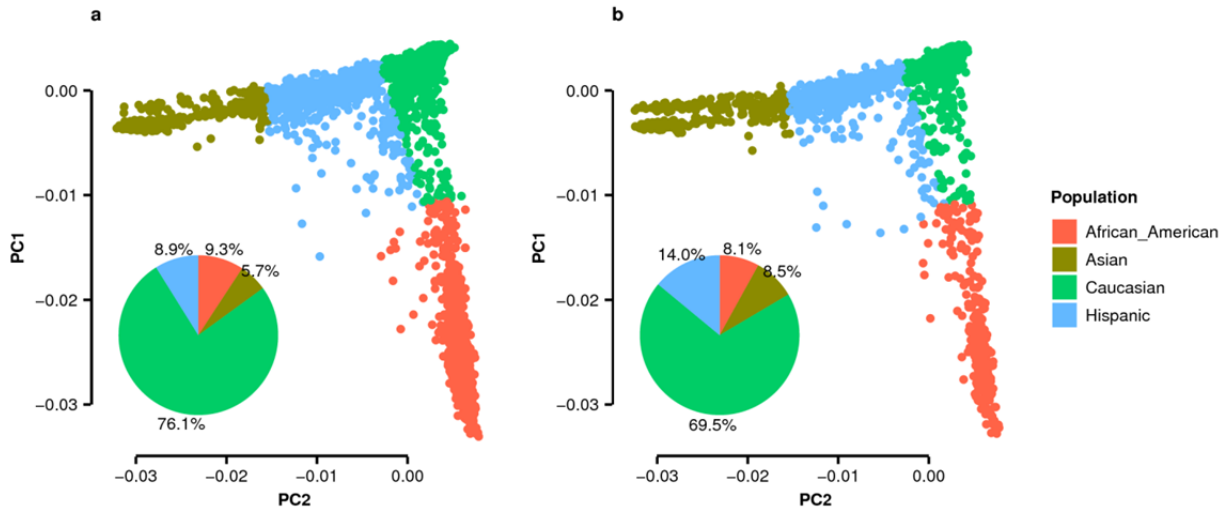
eFigure 1. Power (1-β) Versus Significance Level (α) Given 10000 BC, 2000 OV Cases and 4000 Controls. Performed burden test to compare total frequencies of mutations after collapsing rare (MAF ≤0.05%) variants within genes under additive inheritance model with effect size = 2, replication = 10000 times.



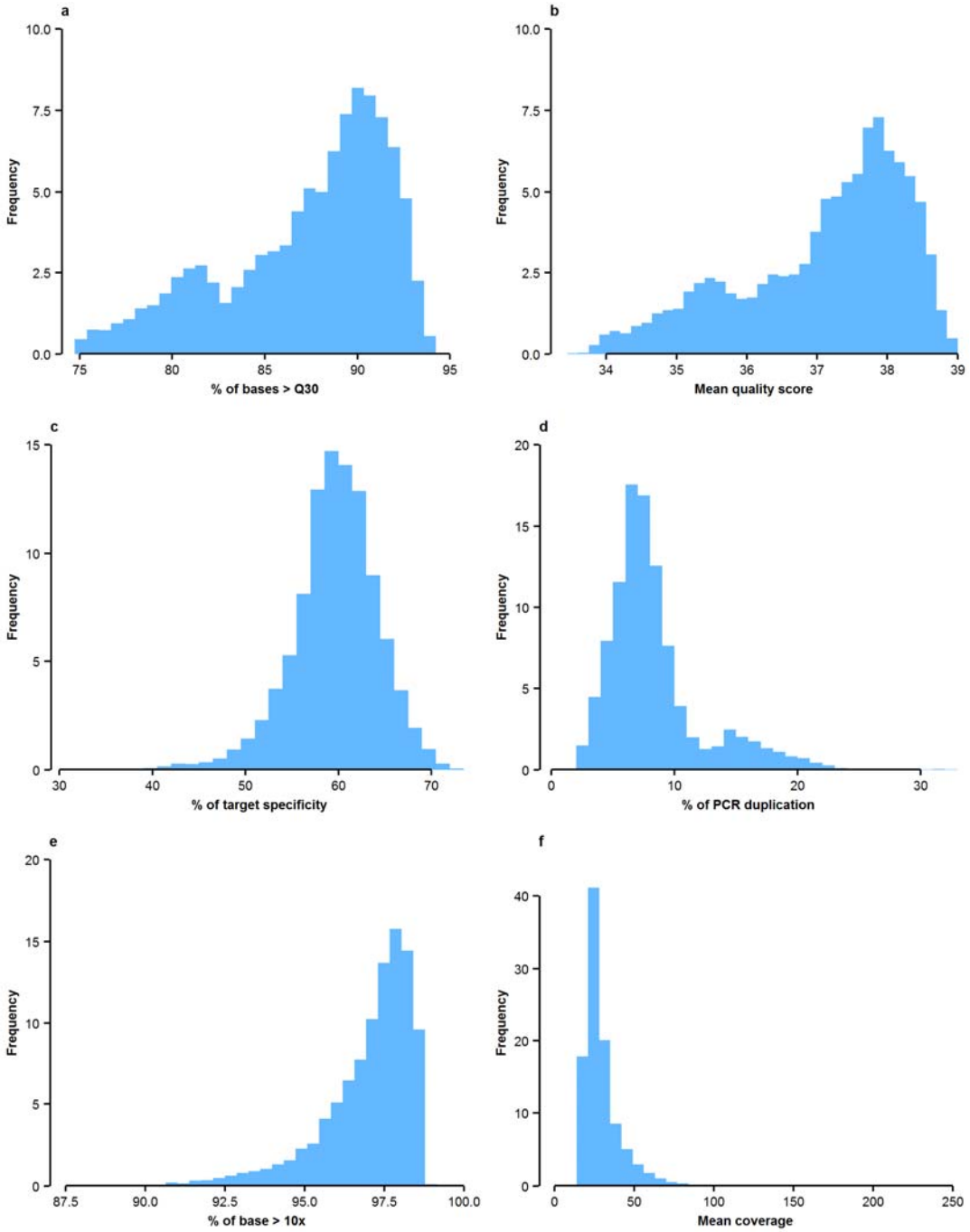
eFigure 2. Distribution of Variants in Characterized BC/OV Genes. (a) Distribution of pathogenic variants by variant types. (b) gnomAD allele frequency distribution for pathogenic and non-pathogenic variants**.** carrier frequency of pathogenic variants among (c) breast cancer and (d) ovarian cancer patients**.**



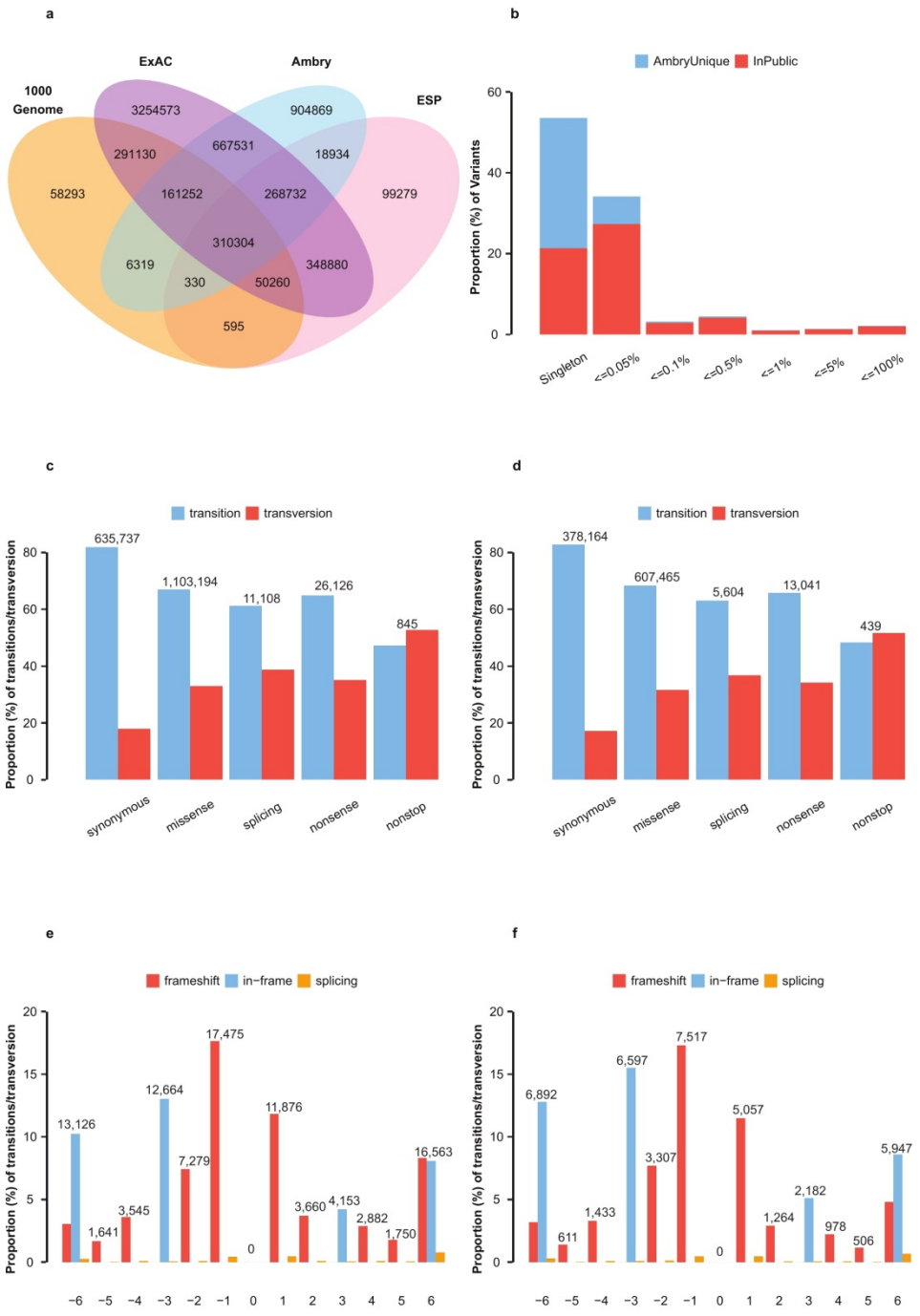
eFigure 3. Inferred Ethnicity of (a) 11,416 Cancer Patients and (b) 3,988 In-Lab Controls. Principal component analsis (PCA) based on exome sequencing data**.**



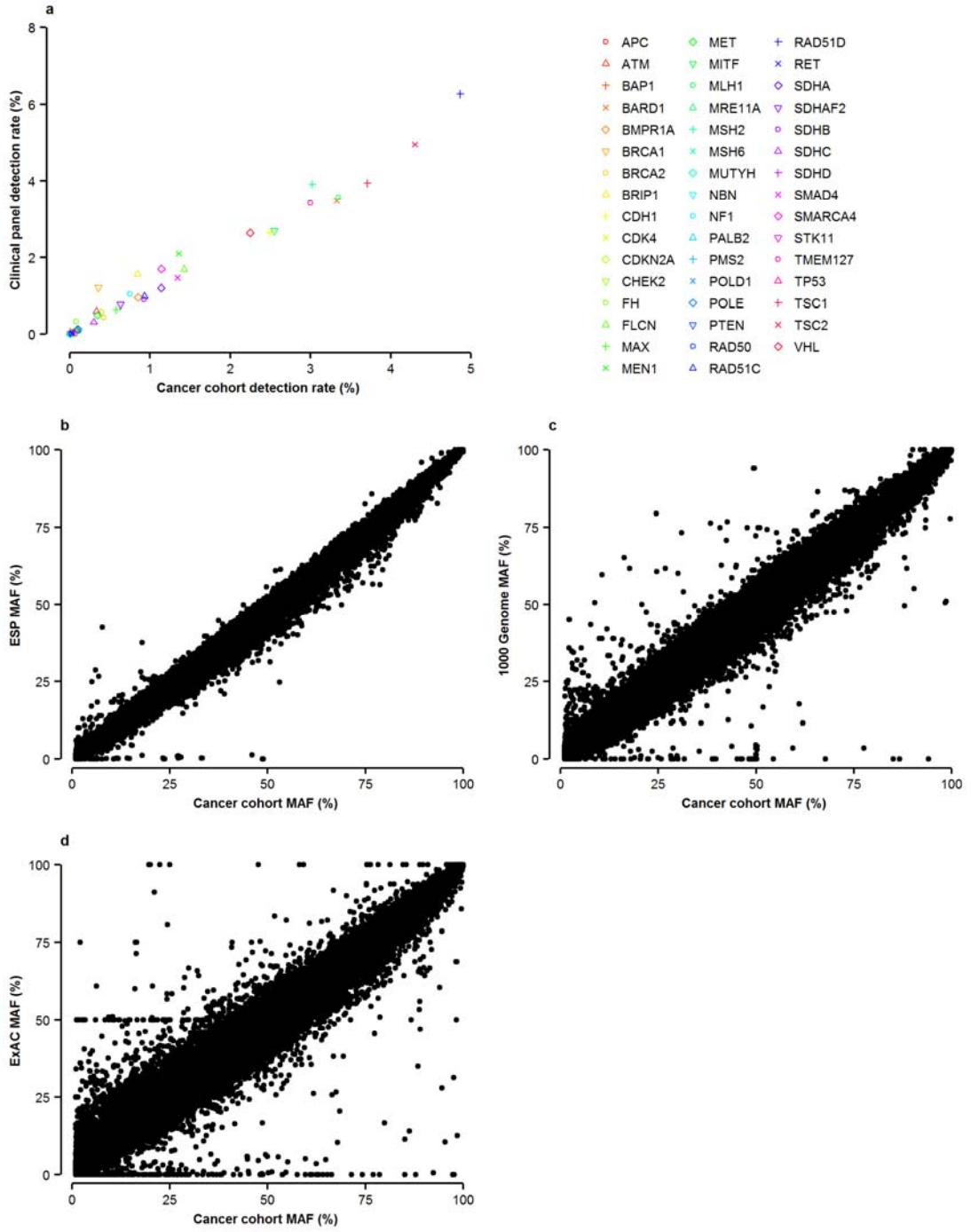
eFigure 4. Quality Metrics of Samples. (a) Distribution of percentage of bases > Q30. **b)** Distribution of mean base quality score. (**c**) Distribution of target specificity. (**d**) Distribution of percentage of PCR duplications. (**e**) Distribution of percentage of bases > 10x. (**f**) Distribution of mean coverage.



eFigure 5. Characteristics of Detected Variants. (**a**) Comparison of detected variants in cancer patients and public reference populations. (**b**) Allele frequency distribution of variants in all cases and In-lab controls. Distribution of mutaion context and functional category of SNVs in (**c**) cases and (**d**) In-lab controls. Distribution of size and functional category of indels in (**e**) cases and (**f**) In-lab controls.



eFigure 6. Evaluation of Variant Calling Accuracy. (**a**) Comparison of the detection rate of pathogenic, likely pathogenic and variants with uncertain significance between target sequenced cancer panel testing and exome sequencing data in cancer samples. Comparison of MAF between Caucasian cancer patients and public reference data: (**b**) ESP European American, (**c**) G1K Caucasian and (**d**) ExAC Caucasian.



eAppendix. Methods

DNA sequencing and quality

Samples were sent to a single laboratory (Ambry Genetics, Aliso Viejo, CA). At least 1000ng of genomic deoxyribonucleic acid (gDNA) was isolated using the QIAsymphony® DSP Midi DNA kit (Qiagen, Hilden, Germany) from peripheral blood samples submitted in collection tubes with either EDTA or ACD as the anticoagulants. Saliva samples submitted in Oragene kits were isolated using the prepIT DNA Extration Kit (DNA Genotek, Ottawa, Ontario, Canada). Isolated DNA was quantified using a NanoDrop UV spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) with quality metrics of A260/280 = 1.8-2.0 and A260/230 >1.6.

Exome target enrichment was performed using custom Integrated DNA Technologies xGen Lockdown probes (IDT, Coralville, IA). Briefly, one microgram of DNA was sheared into 150- to 300-bp fragments, which were then repaired, ligated to adapters, and purified for subsequent polymerase chain reaction amplification. Amplified products were then captured by biotinylated DNA library baits in solution following the manufacturer's instructions. Bound DNA was isolated with streptavidin-coated beads and re-amplified. The final isolated products were sequenced using the Illumina HiSeq 4000 sequencing system with 150-bp paired-end reads (Illumina, San Diego, CA). Initial data processing and base calling, including extraction of cluster intensities, is done using RTA 1.17.21.3 (Real Time Analysis, HiSeq Control Software version 2.0.10). Demultiplexing by barcode is done by the bcl2fastq Conversion Software v1.8 (Illumina Inc., San Diego, CA).

Raw sequencing reads were aligned to the reference human genome (GRCh37) by Novoalign V3.02.05 (Novocraft Technologies Sdn Bhd, Malaysia), and duplicated reads were removed using Picard 2.1.0 (Broad Institute, Cambridge, MA). Samples with >75% of Q30 bases, mean base quality > 30 and >80% of bases above 10x in target exon regions were retained for analysis (eFigure 4). Samples that did not meet >80% of bases >10x in target exon regions were re-sequenced and merged together to increase coverage. Samples achieved target specificity of 59.6% (±4.56%) and a PCR duplicate rate of 8.4% (±3.8%), resulting in 97.0% (±1.4%) of bases in target region >10x and mean coverage of 29.4x (±12.9x). Variant calls were generated using Genome Analysis Toolkit (GATK v3.2.2, Broad Institute, Cambridge, MA) following GATK best practices for variant calling on targeted exons (with ±2bp paddings) of 31 Mbs. Variants with coverage <10x were defined as low quality variants and therefore recalibrated as missing values. We also evaluated the coverage for each site across all samples.

Positions for which >20% of individuals had coverage <10x were filtered out as low coverage sites. We conducted 3 validation runs using HapMap sample NA18507 to evaluate reproducibility of our sequencing and bioinformatics analysis. High concordance rate (97% of variants) were observed for the same sample among 3 different sequencing runs, as well as between each sequencing run and genotype in HapMap Project[1].

We identified 297,044,999 high-quality variants with mean read coverage 46.5x (±53.7x), as well as average heterozygous ratio 48.9%±11.0% and 89.7%±29.9% for heterozygous and homozygous variants, respectively. The high-quality variants (2,307,597 unique) include 288,289,860 (2,189,444 unique) SNVs and 8,755,139 (118,153 unique) indels, which included 35.5% heterozygous deletions, 15.2% homozygous deletions, 34.9% heterozygous insertions and 14.4% homozygous insertions. The hit rates of clinical relevant variants were computed ($\frac{\#\text{Samples with} \geq 1 \text{ pathogenic or uncertain significance variants detected}}{\#\text{Samples analyzed}}$) and compared between exome sequencing and multi-gene panel testing using Pearson correlation test.

The landscape of variants identified in the present study among BC/OV cases and In-Lab controls is described below (eFigure 5-6). Of the 2,307,597 unique high-quality variants, 33.5% were synonymous SNVs, 59.3% missense SNVs, 0.6% splicing SNVs, 1.4% nonsense SNVs, 0.1% stop-loss SNVs, 1.8% in-frame indels, 3.2% frameshift indels and 0.1% splicing indels. Overall, 1.7% of detected variants have MAF ≥5%, 1.1% have 1%<MAF<5%, and 97.2% have MAF <1%. Most (93.9%) of the variants with MAF above 1% were also identified in publicly available reference populations (ExAC, ESP and G1K, eFigure 5). As most of our samples are Caucasian, the MAF distribution of unique Caucasian polymorphisms (MAF ≥1%) shared between our samples and reference controls were highly correlated (ESP, r=0.997; G1K, r=0.991; ExAC, r=0.986, eFigure 6b-d). We further compared the detection of clinically relevant variants between this study and targeted NGS panel based on multi-gene hereditary cancer panel testing from the same lab, and observed strong concordance (r=0.965) (eFigure 6a). Taken together, the consistent capture of both common polymorphisms and rare cancer-susceptibility variants are indicative of highly accurate variant calling.

The proportion of transversions (tv) for each type of variant increased from the least to most damaging variant type among cancer cases: synonymous (18%), missense (33%), nonsense (35%), splicing (39%) and stop-loss (53%) (eFigure 5c). In contrast, while we identified approximately half as many transitions and tvs among controls compared to cases, the tv proportion in controls was also slightly lower for each variant type: synonymous (17%), missense(32%), nonsense (34%), splicing (37%) and stop-loss (52%) (eFigure 5d), reflecting the effect of

selection against tvs in the control population. Frameshift variants, which are more likely to be pathogenic, were less common than in-frame indels (eFigure 5e). The number of frameshift variants decreased dramatically by indel size, while the decrease by size for in-frame indels was minimal. The distributions of indels in cancer (eFigure 5e) and control (eFigure 5f) samples were similar.

Public reference populations

ESP (NHLBI Exome Sequencing Project, http://evs.gs.washington.edu/EVS/)

The final release of ESP dataset (ESP6500) contains 6,823 individuals, including 4,419 EAs (European American), 2,336 AAs (African American) and 46 samples of other ethnicity based on self-reported race. Samples were sequenced at two centers: the University of Washington and the Broad Institute. At the University of Washington, exome capture was performed using Roche/Nimblegen capture, and at the Broad using Agilent reagents. 4,300 EAs and 2,203 AAs were kept for analysis after data curation and quality filtration. The genotype data of the 6,503 samples were released in VCF format. Both the genotype and the coverage files are downloaded from the ESP website.

G1K (1000 Genomes Project, http://www.1000genomes.org/)

The 1000 Genomes Project provides an extensive public catalog of human genetic variation across multiple populations. The sequence dataset were downloaded from the NCBI site (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/). G1K phase III represents individuals across 26 populations comprised of five continental groups: European (EUR), Eastern Asian (EAS), Southern Asian (SAS), American (AMR) and African (AFR) ancestry. This phase III release (September, 2015) contains genotype calls on 2,504 samples with an average genome coverage of 4x (http://www.1000genomes.org/about) in VCF format. We removed 37 possible related samples, identified based on the kinship coefficients estimated within each ethnic group using the KING robust relationship inference method[2], resulting in 2,467 unrelated controls.

ExAC (Exome Aggregation Consortium, http://exac.broadinstitute.org/)

Exome Aggregation Consortium provides an aggregated exome sequencing data from a total of 60,706 samples. These exomes were processed using the Genome Analysis Toolkit (GATK) HaplotypeCaller pipeline v3.1-144. The raw variants were filtered by GATK Variant Quality Recalibration (VQSR) to achieve SNV sensitivity 99.8% and Indel sensitivity 95.1%. The sequence data (Mar, 2016), obtained from multiple platforms (Illumina,

SOLiD and 454), was downloaded from ftp://ftp.broadinstitute.org/pub/ExAC_release/current/subsets/. Excluding samples from The Cancer Genome Atlas (TCGA) results in 53,105 individuals across seven super ethnic groups: Non-Finnish European (NFE), Finnish (FIN), Eastern Asian (EAS), Southern Asian (SAS), Latinos (AMR), African (AFR) and Other (OTH) ancestry.

gnomAD (Genome Aggregation Database, http://gnomad.broadinstitute.org/)

Genome Aggregation Database provides exome sequence data from 123,136 individuals (including samples in ExAC) and whole-genome sequencing data from 15,496 individuals from multiple disease-specific and population genetic studies. In addition to the seven ethnic groups in ExAC, gnomAD additionally included Ashkenazi Jewish, resulting in a total of 138,632 individuals from eight populations. The sequencing data was processed using the same GATK HaplotypeCaller pipeline. A newly developed random forest classifier and other hard filters were applied to variant QC, which perform better than the standard GATK VQSR for large samples. Known tumor samples and related individuals were removed. The sequence data and coverage files were downloaded from http://gnomad.broadinstitute.org/downloads. The coverage file was used for coverage check in the comparison of cancer cases versus gnomAD and only high-quality calls with PASS flag were included. Similar to our cohort, ExAC exome sequencing data has 80% of individuals with a minimum of 10× coverage, which composes half of the gnomAD data.

Data analysis

Relatedness and ancestry analysis

11,455 breast and ovarian cancer cases and 4,010 In-Lab controls were initially sequenced and considered for analysis. Prior to association testing, we applied the KING (Kinship-based Inference for Genome-wide association studies) method[2] to identify and exclude cases (n=39) and controls (n=22) based on potential cryptic relatedness. We also performed principal component analysis (PCA) using R package SNPRelate[3] to infer ancestry and confirm population membership. Using similar method in previous study[4], we computed the genetic covariance matrix based on 29,974 bi-allelic autosomal SNPs with MAF ≥1% and missing rate ≤10% in our BC/OV patients and In-Lab controls. We projected samples in the 2-dimensional space defined by the first two PCs and computed the Euclidean distance from the centroids of ethnic clusters, which allowed us to infer race/ethnicity based on minimizing the distance between the clusters and the sample projection[5] (Figure 1).

Power calculation

To test for significant association with BC and OV, we compiled a list of 625 genes previously shown to be associated with cancer[6-13] (eTable 1) and compared frequencies observed in cancer cases with those of In-Lab controls and gnomAD. Similar to methods utilized by others[6,14], we chose *a priori* to examine and report associations among these genes rather than every gene in the exome, because with ~10,000 cases and ~4,000 controls, we had only 37% power to detect an OR=2.0 at exome-wide significance ($\alpha=1x10^{-6}$ to correct for multiple testing) using gene-based Fisher's exact tests[15] (eFigure 1). With this sample size, we had 76% power to detect an OR=2.0 with the same test after Bonferroni correction for 625 genes. Power analyses were performed with SEQPower[16].

ClinVar

To assess the clinical significance of variants, we downloaded variant significance data from ClinVar[17] (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar), which archives interpretations of variants and clinical significance (pathogenic, likely pathogenic, uncertain significance, likely benign and benign). For tests of association, we included variants classified as pathogenic/likely pathogenic by at least two clinical genetics groups with assertion criteria (Ambry, SCRP, InVitae, GeneDx, Emory and/or InSiGHT)[18], and having more pathogenic evidence than uncertain or benign evidence if conflicting.

Filtering and inclusion criteria

For cancer cases and In-lab controls:

1) excluded common variants (MAF ≥1%) in cases, In-Lab controls or any public reference population (ESP, G1K, ExAC, gnomAD).

2) excluded Sanger identified or recorded NGS artifacts; excluded variants that were only detected under low coverage (≤20x) and low heterozygous ratio (≤30%); excluded possible artifacts identified by experienced bioinformatics scientists using IGV (Integrative Genomics Viewer); excluded variants with missing rate ≥30% in cases or controls; excluded variants in genes or regions with severe pseudogene issue (eg. *PMS2* variants in exons 9, 11-15).

3) excluded benign (including likely benign) variants classified by Ambry Genetics or ClinVar; excluded variants with gnomAD MAF ≥0.1%; protected pathogenic (including likely pathogenic) variants in 625 cancer genes

(eg. known common pathogenic variant *CHEK2* 1100delC, gnomAD MAF=0.2%), but excluded *CHEK2* p.I157T and p.S428F because they are low penetrance variants[18] with gnomAD MAFs >0.1%.

4) included protein truncating variants (stop-loss, nonsense, frameshift and splicing); removed truncating variants detected among In-Lab controls with MAF ≥0.1% but not present in the large reference population gnomAD (potential artifacts); removed truncating variants not influenced by nonsense mediated RNA decay (NMD) (variants located beyond the last 55bp of penultimate exon and not in functional domains[18]).

For cancer cases and gnomAD controls:

1) included the pathogenic variants identified in BC/OV cases as described above.
2) included known pathogenic variants and protein truncating variants in gnomAD as described above.
3) for all variants identified in cases and gnomAD, excluded sites with gnomAD median coverage ≤10x or missing rate ≥30% at 10x coverage, and excluded variants with gnomAD MAF ≥0.1%.

Pathogenic variants in characterized BC/OV genes

Among 11,416 BC and/or OV samples, we identified 1,723 variants, including 540 (31.4%) synonymous, 36 (2.1%) in-frame insertion/deletions, 790 (45.9%) missense, 292 (16.9%) truncating variants, and 65 (3.8%) intronic/UTR in 11 well-characterized BC and/or OV genes[19-22]. Among these 1,723 variants, 448 were pathogenic (including truncating variants) based on the inclusion criteria described above. Notably, 99.0% of truncations and 69.4% of in-frame insertion/deletions were pathogenic, while 99.4% of synonymous and 84.6% of missense variants were non-pathogenic (eFigure 2a). All pathogenic variants had minor allele frequency (MAF) <0.5% and 99.8% had MAF <0.1% in gnomAD (eFigure 2b). Only one pathogenic variant *CHEK2* c.1100delC had gnomAD MAF >0.1%, and we retained it for analysis, as described above.

Among 9,639 BC patients, 7.1% were carriers of pathogenic variants, which were observed across all 11 genes: 1.3% were carriers of pathogenic variants in *BRCA2*, 2.4% in *BRCA1* and 3.4% in *non-BRCA1/2* genes (eFigure 2c). Among 2,051 patients with OV, 13.6% were carriers of pathogenic variant, which were identified in 10 of the 11 genes (eFigure 2d). Specifically, 3.8% of OV patients had pathogenic variants in *BRCA1*, 4.0% in *BRCA2* and 5.8% in non-*BRCA1/2* genes.

Sensitivity analyses

To evaluate whether gene-phenotype associations varied by methods, or were altered after adjustment for ancestry, we also conducted Combined Multivariate and Collapsing (CMC)[15] test (adjusted for the first three principle components of ancestry) (eTable 5). To examine the potential impact of sex distribution among cases and controls, we performed sensitivity analyses among females (eTable 6). To determine whether the observed *MSH6* association with BC is mainly driven by patients' family history of ovarian cancer, we excluded BC cases with 1st or 2nd degree relatives with OV, and re-analyzed the *MSH6* association with BC (eTable 7). To determine whether our findings were impacted by inclusion/exclusion of cases with multiple cancer primaries, we estimated risks after exclusion of individuals with dual BC and OV (*n*=274; data not shown), and after exclusion of those with personal history of other cancers (*n*=818; eTable 8). To assess whether results were specific to certain pathogenic variant types, we examined associations stratified by variant type, defined as "loss of function", "missense" or "other", and comparisons yielded similar results for overall BC and OV associations (eTable 9). We further explored genetic association with clinical features of BC and OV by performing case-case analyses: ER^+PR^+ vs. ER^-PR^- for BC (eTable 10) and early onset (<45 years vs. ≥45 years of age) for BC and OV (eTable 11).

eReferences

1. International HapMap, C. The international HapMap project. *Nature* **426**, 789-96 (2003).
2. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
3. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-8 (2012).
4. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).
5. Shukla, S.A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* **33**, 1152-8 (2015).
6. Lu, C. *et al.* Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* **6**, 10086 (2015).
7. Cybulski, C. *et al.* Germline *RECQL* mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-6 (2015).
8. Sun, J. *et al.* Mutations in *RECQL* gene are associated with predisposition to breast cancer. *PLoS Genet* **11**, e1005228 (2015).
9. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-9 (2013).
10. Lawrence, M.S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501 (2014).
11. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-58 (2013).
12. Pritchard, C.C. *et al.* Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J Mol Diagn* **16**, 56-67 (2014).
13. Frampton, G.M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **31**, 1023-31 (2013).
14. Kanchi, K.L. *et al.* Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun* **5**, 3156 (2014).
15. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-21 (2008).
16. Wang, G.T., Li, B., Santos-Cortez, R.P., Peng, B. & Leal, S.M. Power analysis and sample size estimation for sequence-based association studies. *Bioinformatics* **30**, 2377-8 (2014).
17. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-8 (2016).
18. Couch, F.J. *et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol* **3**, 1190-1196 (2017).
19. Easton, D.F. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**, 2243-57 (2015).
20. Tung, N. *et al.* Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J Clin Oncol* (2016).
21. Loveday, C. *et al.* Germline mutations in *RAD51D* confer susceptibility to ovarian cancer. *Nat Genet* **43**, 879-82 (2011).
22. Ramus, S.J. *et al.* Germline mutations in the *BRIP1*, *BARD1*, *PALB2*, and *NBN* genes in women with ovarian cancer. *J Natl Cancer Inst* **107**(2015).