SUPPLEMENTARY METHODS

Description of Study Populations and Genotyping

Our ascertainment criteria for cases included a living affected sister with disease willing to participate in the study. Among families that provided information, 96% reported Caucasian ancestry, 2% Ashkenazi, ~1% African American, 1.7% Native American, and <1% other. Our sib pairs consisted of 8 sets of self-reported monozygotic twins, (1.7%) and 6 pairs (1.3%) having non-shared paternity based upon allele sharing at the X-linked androgen receptor (*AR*) microsatellite (het=0.89) [1]. Non-shared paternity for a given sib pair was defined as an index case and her sibling sharing 0 alleles at *AR*. Index cases from these pairs were nonetheless retained for association testing. SNP assays were designed with assay Design 3.1 software into 3 separate assays (Supplementary Table S2). After a 6ml multiplex PCR amplification the resulting products were treated with Shrimp Alkaline Phosphatase (SAP) and single-base primer extension chemistry was conducted with mass modified dideoxyribonucleotides (iPlex Gold Chemistry). Extension products were processed with SpectroCLEAN resin, and spotted onto SpectroCHIPs and analyzed via MALDI-TOF mass spectrometry [2]. Automated genotype calls were made with SpectroTYPER v3.4 software. To reduce the likelihood of scoring errors due to genotyping platform disparities we genotyped both our AIM and association SNPs in a reference panel of HapMap samples and observed a concordance rate of 99.36% (2161/2175). Genotype data for all association SNPs were tested for deviations from Hardy-Weinberg proportions in cases and we observed no deviation for the 3 SNPs tested (p<0.001). Five AIMs showed deviations from Hardy-Weinberg proportions but were nonetheless retained for analysis with STRUCTURE 2.2 for population admixture.

Population Structure Analysis and Association Testing

We genotyped 59 AIMs utilizing Sequenom iPLEX mass spectrometry technology. We downloaded equivalent genotypes for HapMap self-identified reference samples and for controls

(n=1,142) and cases (1,145) from the CGEMS database. After genotyping our ECOG BrCa cases we retained 455 individuals for STRUCTURE analysis. Five AIMs (rs1040045, rs6451722, rs3907047, rs4746136 and rs798443) exhibited prior association to disease in the CGEMS dataset ($p<0.006$ to $p<0.037$) and were omitted from STRUCTURE analysis [3]. STRUCTURE analyses were run without any prior population assignment using 50,000 iterations with 10,000 burn-in cycles under the admixture model with initial □ =1.0 without specifying population membership. We utilized the infer □ option and estimated a separate a for each population under the F model (□ is defined as the Dirichlet parameter for degree of admixture). When we included 105 AMI (AmerInd) individuals as described by Kosoy, et. al [4] and increased the defined population cluster parameter to k=4 populations we observed no appreciable difference in the clustering of our ECOG cases or CGEMS controls to CEU HapMap reference samples (data not shown). More importantly we were also able to identify the most likely genetic ancestry of our cases and the CGEMS controls for which we lacked self-reported ethnicity. Association testing for imputed rs1434536 genotypes were performed by the method of Marchini [5]. Briefly, rs1434536 genotypes from CGEMS controls were imputed with IMPUTE from HapMap CEU SNPs from chromosome 4 region 96,289 – 96,296 kb, which surround rs1434536. IMPUTE uses a hidden Markov model and known HapMap haplotypes to impute missing data. Association testing with SNPTEST includes imputation uncertainty in the subsequent association test by modeling the observed data likelihood using the full data likelihood integrated over missing data.

Cloning, Luciferase Assays, and qRT-PCR of *BMPR1B*

MB-MDA-231 and MCF-7 cell lines were obtained through American Type Culture Collection (Manassas, VA). All tissue culture reagents were purchased from Invitrogen (Carlsbad, CA) and Sigma (St. Louis, MO). PCR primers for *BMPR1B* were (forward primer: 5'-CCGCTCGAG GTCCCAGGACATTAAACTCTG-3', Reverse primer: 5'-TTTTCCTTTTGCGGCCGCGCATCA

TATCTTGAACAAGTT-3') containing *Xho I* and *Not I* restriction sites respectively for directional cloning into the MCS site Psi-CHECK-2.2. Twenty-five nanograms of genomic DNA were PCR amplified (95°C, 5min, 95°C 30sec, 55°C, 30sec, 72°C, 40sec, for 35 cycles, 72°C 3 min final extension) with 1□l of Taq (5U/□l Roche), 1X Taq Buffer, 1□M primers, and 200□M dNTPs. PCR products (~0.28kb) were restricted with the aforementioned enzymes, purified via gel electrophoresis and cloned into PsiCheck-2.2. Genotypes for MCF-7 and MDA-MB-231 cell lines at rs1434536 were determined by sequencing PCR products derived from 25ng genomic DNA isolated from cells grown in culture and the aforementioned primers.

MDA-MB-231 and MCF-7 cells seeded one day before, were transfected with plasmids bearing the T or C alleles in triplicates in 24-well plates at 80% confluency with a Lipofectamine 2000 (Invitrogen) complexed with a mixture of 25 ng psiCheck reporter plasmid and 75 ng stuffer DNA (pBlueScript) per well. miR125-b target site cleavage results in degradation of reporter mRNA, with a concomitant decrease in translated product, which can be detected by a luminescence-based assay system. Firefly luciferase expressed from psiCheck2.2 served as an internal normalization control.

Transfection of miR-125b Duplexes and qRT-PCR of *BMPR1B*

Sequences for siRNA duplexes: siR 5'- GGACUAUAGCUAAGCAGAUUCAGat-3' and 3'-UUCCUGAUAUCGAUUCGUCUAAGUCUA-5' (RNA nucleotides are shown in uppercase and DNA nucleotides shown in lower case) and has-miR-125b:hsa-miR-125b-1* (targets C allele of rs1434536) duplex: 5'-UCCCUGAGACCCUAACUUCUGA-3' and 5'-ACGGGUUAGGCUCUUGGGAGCU-3'. These duplexes target positions chr4:96,270,043 and chr4:96,294,738 respectively in *BMPR1B*. Cells were purified with RNA STAT-60 (IsoTex Diagnostics, Inc.) according to manufacturers directions. *BMPR1B* specific primers 5'-CAACAAAATTCTTCCCAGGAACT-3' and 5'-TGGTTCACAGAGTGCAACAATA-3' were used to amplify cDNAs. Samples were treated with DNase I (Ambion, Turbo DNA-free) and control reactions omitting M-MLV were also included to rule out

genomic DNA contamination. SYBR green technology was utilized for transcript quantitation. *GAPDH* intron spanning primers ('5-CATTGACCTCAACTACATG-3' and 5'-TCTCCATGGTGGTGAAGAC-3') were utilized as normalization controls. PCR conditions were: 95^{o}C for 5 min, followed by 40 cycles of 95^{o}C, 15 sec, 55^{o}C for 30sec, 72^{o}C for 30sec.

References

1. Haiman, C.A., et al., *The androgen receptor CAG repeat polymorphism and risk of breast cancer in the Nurses' Health Study.* Cancer Res, 2002. **62**(4): p. 1045-9.
2. Ragoussis, J., et al., *Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research.* PLoS Genet, 2006. **2**(7): p. e100.
3. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-59.
4. Kosoy, R., et al., *Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America.* Hum Mutat, 2009. **30**(1): p. 69-78.
5. Marchini, J., et al., *A new multipoint method for genome-wide association studies by imputation of genotypes.* Nat Genet, 2007. **39**(7): p. 906-13.

Supplementary Table S1 Sequenom MassARRAY assay designs of association and AIM SNPs

WELL*	SNP_ID†	2nd-PCRP‡	1st-PCRP	UP_SEQ§	UEP_DIR	EXT1_CALL	EXT2_CALL
Assay_1	rs9522149	AGAAAGGAGAGGAAACACCG	TCAGCAACTTCTAGTCCTCG	GGTCCTTGCAGCTCC	F	C	T
Assay_1	rs11652805	CCCTCAAAGTTTGGTGCATC	CTCTTCCTGGTCCTGAGATG	CTTTCTCTCTCCCAGC	F	C	T
Assay_1	rs9530435	ATCAGGCACATGGTAAGCAC	CTCCATCTGGTACATATGTG	gAAGCACTCAGCGAAG	R	T	C
Assay_1	rs2416791	TATAGCATCTACCATCAGCC	ATACTGCCCCATAAGGAGTG	aACCATCAGCCCAATTC	F	A	G
Assay_1	rs9855638[2]	GGTTAGTTTTGGTGAAGTCC	GACCTTGGCTTTTACCATAG	TTGTTGCTCATGCATTT	R	G	C
Assay_1	rs10108270	AACAGCATCTGAGACGCTTC	AGTGACCCTGGACACAATTC	TCAGGTGAGGACTTAGC	R	C	A
Assay_1	rs4666200	CCCTATCCTTGGTGATTTGG	CAGTCACAATTGGCAAGCAC	tACTTCAGAGCTATTGGC	R	G	A
Assay_1	rs9319336	ATGCAAGGTAATGCACCCTC	TCTACCTGCAGGTAAGTGTC	ACCCTCTCCCTGCTTCTAT	F	C	T
Assay_1	rs3907047	CAGAATCGGACATGATACCC	GAAAGTCCAGGAAGTTCAGG	ctTCAGCTCTCTGATCTCC	F	C	T
Assay_1	rs4908343	CCAAACCCCACAAGCTTAAC	AGGGAGAGAAGGTCAGTTAC	AACCCCTGGGCTATGACAA	F	A	G
Assay_1	rs1513181	CAGATTTCCCATAGCCTCTC	AGGTGAGACAGTTGGACAAG	GTTGAGCTTGAAAAATTCCC	F	C	T
Assay_1	rs3737576	GGTCCTGGTTCTTGTCAAAG	AGGAGGAAGAGCATAGTGAG	AGATTGTGAAAGACTGAAAT	F	A	G
Assay_1	rs1040045	GAGAGAAAGGGACACAGAAC	CCTCACCCCATCTACTCTTG	tagtATGGGGATTGGGGTAA	F	C	T
Assay_1	rs7803075	AATCCACATGAACTGCGCTC	ATCTATCACGTGGACCTTTG	cctCGCTCCTGGATCTTTTAC	F	A	G
Assay_1	rs731257	GTTGAGGTATCAGTGGAATC	TCCATCCTAATTGGAACTGC	TGGAATCACAAATTGTATCTC	R	G	A
Assay_1	rs7997709	TAACTGTGTTTCCCTCAGTG	ATGTGGATGGATTGCTCAAC	agaTTTCCCTCAGTGGTTAGC	R	T	C
Assay_1	rs4918842	ATTGCTCAGAAATGCTGTGG	TTAATCGGATGCTGAGCCTG	AAATGCTGTGGATATTGACTTA	F	C	T
Assay_1	rs10496971	AATGTCACCTTTAGGCAGAG	GGGAACATTGAGTCCTCAAG	ggggaTTTAGGCAGAGGCATTT	F	G	T
Assay_1	rs12629908	CCAATCTCTTACATCTCCTG	TCCATCATCCAGGAGCTTAG	CTTACATCTCCTGAAAAGAAATT	R	G	A
Assay_1	rs10007810	TCTTCTCTTGTAGACAGGGC	CGTGGCACATGCCATGTTTT	acttgAGACAGGGCCCTCTATCT	F	A	G
Assay_1	rs772262	CACTTTTGACTTAAGACGGG	TTCAATACCTCTGGCCTCTC	cGACGGGTTTTTATCAGGACATA	F	A	G
Assay_1	rs4746136	GGTATGTCCTAGATGACAAG	AGCACATACCTGCAAGCACG	TGGACAGATAAACTTATTTGTGTA	R	G	A
Assay_1	rs2125345	AGTGTATGGTTTCTTTGTGG	TAACGTGAGTCAGACTGTAG	ggtagGGTTTCTTTGTGGGATTCT	R	G	A
Assay_1	rs6451722	CTCTGTAAGCAGCTATTGCC	TCTGCTCCTAAGGAAGATGC	tCAGCTATTGCCATTTTTTTCTCAT	F	A	G
Assay_1	rs7554936	AACCAGGGACTGCATACAAC	CATCCTAGTGAATGCCATCC	ccctTAAAGTCATAGGTGAACCTTC	R	T	C
Assay_1	rs7657799	ACAAGGCCCAATTGCTGAAG	AGCCAACTTGATTCTCTTTC	cccTGATCTACCTTGCAGGTATAATG	F	G	T
Assay_1	rs260690	CTCATAGTTGCTATGAACAG	TCTGTGGCCAACGTAAAAGG	ggcGTTGCTATGAACAGTTTAACAGT	R	C	A
Assay_1	rs4891825	GTGTAACAATCTCAATCCCC	CTAGGGTTGGTAAAGGATGG	atcgCAATCCCCCTTAATGTTTTCATC	F	A	G
Assay_1	rs6104567	ACAAGGCCCAGTATGATTG	GCTTGGCTTTAATATGGAGG	CAGTATGATTGATACATATCTAATTAA	F	G	T
Assay_1	rs1471939	TACCACCCATCTTAAACAGC	TGTTAACTCCAGAACAAGTG	cctCATCTTAAACAGCTATAGATATAGT	R	T	C
Assay_2	rs1407434	CCCATATCATCTCCACTCAG	TGAACCTAAAAAGCAAAGGG	GCCCTCAGTCCCTCT	R	T	C
Assay_2	rs2504853	CATCCTGAAGGTGATGGAAG	GAAATTCACAGGCTCCAGAC	ATGGAAGCCTTGCAT	F	C	T
Assay_2	rs870347	ACCTTTTTCAGCCTGACTCC	ATCATGCGACATCCAGGTAG	TGCTAAGTCCCTCACT	F	G	T
Assay_2	rs4821004	CTTGCAAGTGTGAAGAGCAG	CAAGGGCCGATGATATTTGC	GGGGAGGGAGCAAGCC	F	C	T
Assay_2	rs9845457	TTGCACTAGATCCGGGAAGC	CTTACTCCATCCCAGTACAG	ggCCGGGAAGCCGCTGC	R	G	A
Assay_2	rs2946788	TATCTACTCTGGCCAAACTC	CATTCCAATGAGCTTAAGCC	CCAAACTCAATAGCCACA	R	G	T
Assay_2	rs8113143	TGTGGGTTCTTGCTGTGTTG	AAGTGAGAGGATGAGAGGAG	GTTGGATAACACATCCCC	R	C	A
Assay_2	rs2030763	CTTCCTTTTCTTACCAACTGC	ATCCATGCGGATGGCTTAAC	ATGAATAAGCTGAGCTTCT	R	G	A
Assay_2	rs9809104	AAAACCACAGGACAGGACAG	TGACGTGGAGTGATTTGGAG	CAGGACAGTTATTCAGGAA	F	C	T
Assay_2	rs798443	GGTATTGCTAACATCTCCAG	CTCAGTGCAGATGGGAAATG	cAATTTCCACTAACAACGCA	F	A	G
Assay_2	rs2397060	AAAACATGTTTAGGGTTTG	CCTTCATTACAACCCAGGTA	ATGTTTAGGGTTTGAAGAAT	F	C	T
Assay_2	rs4984913	GGAAGTGGTCCTCTTCTTAC	ACCCGGAACTTTCGTGGTGT	aagaaCAGGAAGTGGGCACA	F	A	G
Assay_2	rs2627037	AGCGCCGAACTTCAATTATC	GTGCCTTCCTTTTCGGAATC	TTGTCTGAATCTCCAGTTTAC	R	G	A
Assay_2	rs3943253	TGTGGCTTAGGAGTGACATC	ATCCAGTGTAGAAAGAGCCG	TGACATCGTAATACCACTTGG	R	G	A
Assay_2	rs13400937	CTTACCACCCGTGAAATAAC	CCAAAGTTTGTTCCAAATCTG	aaACATTTCAGGAAGTTGAATT	F	G	T
Assay_2	rs734873	ACTGTCCTGTGTCAAGAACC	GATGTCTTGATGATTCCTCC	gggaCCTAGGGCAAGAGAGTAA	R	T	C
Assay_2	rs3745099	CAGTTACTTTTCTCCCCTGC	AAGTAGAAGGTGAGTGAGGG	ggccGCTATTTTCTCGGCACCTT	R	G	A
Assay_2	rs1040404	AACTCAAGTGTCTCCTGAGC	CAGCTGAGCATTTTGTAGTG	tgGTGATACTATTTTCTACCACA	F	C	T
Assay_2	rs10236187	AGAAGGAACGGCAGACAAAG	CCTAGGTGGGAGTAAAAGTG	ggagtCAGACAAAGCCTCACATTA	F	C	A
Assay_2	rs1325502	TCTGGATAAACATTCTGGCG	CATCACCCAGAATGGCAAAC	gagtgCTGGCGTTGCTGCATGTTT	R	G	A
Assay_2	rs10513300	TACCTCTGCAATGCCCTATC	AAGAGCACATACTCCATACC	CCCTATCTTATTATCATATGAGTTC	F	C	T
Assay_2	rs12130799	GTGTTACTCAATGGAGCTCT	GGTTCTGGGATATTGTTGGG	gTTCTCTATTGTATCTCCAATGTCT	F	A	G
Assay_2	rs6422347	TGAAGGCCGACTTCACGGA	ATGTTGACCTCCCTCTCCC	ggtcCGGAGCTGGTGACATTTTAAC	F	C	T
Assay_2	rs1408801	GTGATAGTTTTACAGTTTCC	ACATGCATGTGTATTGCAGG	ggcaTTTTACAGTTTCCTAAACCATG	R	G	A
Assay_2	rs4463276	TCGGCTTGTTTCCTTTTTTG	ACAACAAGGAAAATGAGCCC	tttgGTGGGTACACAGTAAGTGTATA	R	G	A
Assay_2	rs4717865	GTTCTAGATTCAGACCCTGC	CATCGGAGAGGCAAATTGAC	ggggaCCTGCTGCTGCTACCCAGCCTC	R	G	A
Assay_2	rs1760921	ATACGCAAAACCACTGCCAC	TACTGGCCATATTCTCTCTC	ccacaACTGCCACATCCGTCCCATACCT	F	A	G
Assay_2	rs6556352	CAATGCATATGTACTGCTTCC	AGCATTCTATAAACCGACAG	ccTCCATAAAAATGAAATATCATTTAAC	R	T	C
Assay_2	rs7421394	AGTTTAAGAGGTTTGACAGG	TTTTCCACGTGAACATACCC	cctctGTTTGACAGGATAATTTCTGAGA	R	G	A
Assay_3	rs1219648	GACAAGCCATGGCCATCCTT	TCTTCCATGGTACCGGTTTC	GGCCATCCTTGAAGAG	R	G	A
Assay_3	rs1970801	CAGTAGGCCATAAATGTGGG	CAAATTGCTTTATGGGGAAG	GACACCCATTTCTTACCT	R	C	A
Assay_3	rs6831418	GGACTTCCTTACTAGAGCAC	CCTCACAGAATTAAGAGTGC	TGTTTCCTTTCCTCTCC	R	T	C

Supplementary Table S1 Sequenom MassARRAY assay designs of association and AIM SNPs

Assay_3	rs1434536	CTTGAACATCGTCCTGCTTC	TGGGAGCTTCTCTGTCTTTG	GGTTCAGACCTCACCT	R	G	A
Assay_3	rs11097457	TGCTCTTGTGTTGTAAGAGG	AGATACAAGCCATCTGCACG	gggACATGTCAACAAAGATAGG	F	A	G

Notes

* Assays 1 and 2 dervied from In4 markers of Kosoy, et al, Assay 3 SNP for association testing

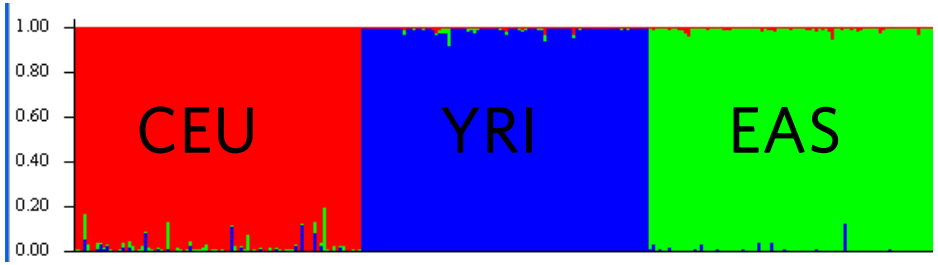
† rs9855638 replaced In_4 SNP rs6548616 with r^2 >0.9 in 3 HapMap populations

‡ ACGTTGGATG mass tags were appended to each PCR primer

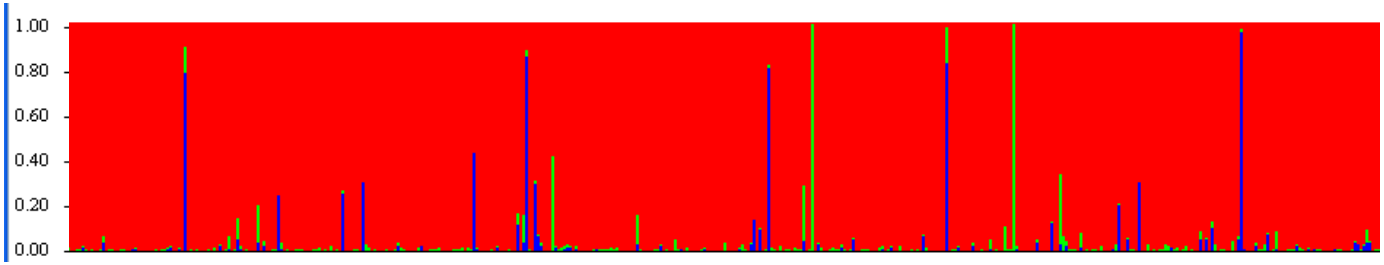
§ lower case nucleotides in extension primer are non-complementary to amplified regior

Supplemental Figure 1

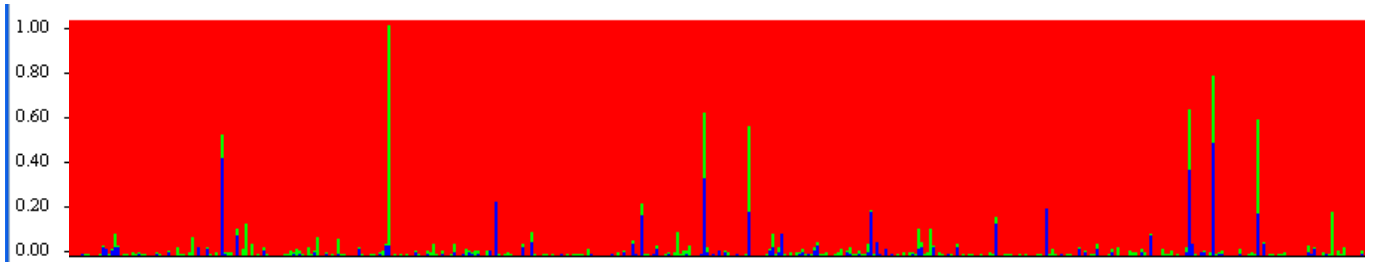
HapMap Samples (n=270)



ECOG BrCa Cases (n=455)



CGEMS Controls (~450 shown)



Patient Sample Set	Number Patients	>80% YRI membership	>80% EAS membership	>90% CEU membership
CGEMS Controls	1142	0	9	1064 (0.932)
ECOG BrCa Cases	455	4	2	428 (0.941)

Supplementary Table 2
SNPs Mapping to miRNA Sites

SNP_ID	miRNA recogniton site	chrom	SNP_Pos*	Motif_Chrom:position	genelist†
rs10157828	GCAGCCA	chr1	17267200	chr1:17267200-17267206	ER- overexpressed
rs4366378	GGCCAGT	chr1	208914363	chr1:208914359-208914365	ER+ overexpressed
rs6565	AACCATA	chr1	28085744	chr1:28085739-28085745	ER- overexpressed
rs6619	TTGGGAG	chr1	37803283	chr1:37803283-37803289	ER+ overexpressed
rs1058240	CCGTTGA	chr10	8156604	chr10:8156602-8156608	ER+ overexpressed
rs11191382	AATGGGT	chr10	104488596	chr10:104488593-104488599	ER+ overexpressed
rs1334891	CCAGGTT	chr10	99070861	chr10:99070857-99070863	ER+ overexpressed
rs7074516	GTGTGAG	chr10	98344914	chr10:98344912-98344918	ER- overexpressed
rs7922412	CAAGGGA	chr10	124805384	chr10:124805379-124805385	ER+ overexpressed
rs10279	GTATTAT	chr11	8925885	chr11:8925881-8925887	ER+ overexpressed
rs1056562	AAGGGAT	chr11	117630835	chr11:117630830-117630836	ER- overexpressed
rs10790248	ACACTAC	chr11	117630882	chr11:117630877-117630883	ER- overexpressed
rs12288903	AAGTCCA	chr11	45860170	chr11:45860164-45860170	ER+ overexpressed
rs3741325	GTGCCAT	chr11	117911199	chr11:117911194-117911200	ER+ overexpressed
rs3741360	TCCAGAG	chr11	66056924	chr11:66056919-66056925	ER+ overexpressed
rs8432	AGCTGCT	chr11	66056091	chr11:66056091-66056097	ER+ overexpressed
rs8995	CCACCCC	chr11	63351648	chr11:63351645-63351651	ER- overexpressed
rs2857672	CACCAGC	chr12	50994544	chr12:50994544-50994550	ER- overexpressed
rs859147	CTCTATG	chr12	25152535	chr12:25152535-25152541	ER+ overexpressed
rs1327179	ATACTGT	chr13	20626320	chr13:20626318-20626324	ER- overexpressed
rs403904	AAGGCAT	chr13	35144233	chr13:35144228-35144234	ER+ overexpressed
rs1565970	AGTCTTA	chr14	51967826	chr14:51967824-51967830	ER+ overexpressed
rs10468050	AGGCACT	chr15	69860993	chr15:69860990-69860996	ER+ overexpressed
rs16956198	CTGTTGA	chr15	69858995	chr15:69858992-69858998	ER+ overexpressed
rs17811309	AAAGGGA	chr15	69860243	chr15:69860239-69860245	ER+ overexpressed
rs2072692	GGGATGC	chr15	87816037	chr15:87816035-87816041	ER- overexpressed
rs30122	CATTAAC	chr16	14266734	chr16:14266731-14266737	ER+ overexpressed
rs30126	GAGACGG	chr16	14263266	chr16:14263262-14263268	ER+ overexpressed
rs1051443	TTAGCTC	chr17	6294757	chr17:6294751-6294757	ER- overexpressed
rs7687	TTCCCCC	chr17	41459142	chr17:41459141-41459147	ER+ overexpressed
rs1046294	ACAACCT	chr19	40352386	chr19:40352380-40352386	ER- overexpressed
rs12427	GCTGGAG	chr19	48962659	chr19:48962655-48962661	ER- overexpressed
rs12891	GAGCCAG	chr19	8233196	chr19:8233196-8233202	ER+ overexpressed
rs7257398	AAGCACA	chr19	59433680	chr19:59433674-59433680	ER- overexpressed
rs2287086	GTGCAAA	chr2	60539999	chr2:60539993-60539999	ER- overexpressed
rs6729137	AATGCAT	chr2	5757912	chr2:5757907-5757913	ER- overexpressed
rs6737419	GTGCAAA	chr2	3570576	chr2:3570570-3570576	ER- overexpressed
rs873033	TCTAGAG	chr2	85390883	chr2:85390879-85390885	ER- overexpressed
rs1048055	GTGCCAT	chr20	1558062	chr20:1558057-1558063	ER- overexpressed
rs2281807	GAGCCAG	chr20	1558201	chr20:1558195-1558201	ER- overexpressed
rs6043374	GTAAACC	chr20	1557952	chr20:1557946-1557952	ER- overexpressed
rs6091230	ACTGCAG	chr20	48926602	chr20:48926602-48926608	ER+ overexpressed
rs2834602	TTACTAG	chr21	35012300	chr21:35012294-35012300	ER+ overexpressed
rs12172608	AGGTGCA	chr22	45137237	chr22:45137236-45137242	ER+ overexpressed
rs6007891	CAGTTTT	chr22	45135497	chr22:45135493-45135499	ER+ overexpressed

Supplementary Table 2
SNPs Mapping to miRNA Sites

rs495702	CACTTCA	chr3	173598334	chr3:173598328-173598334	ER- overexpressed
rs1046322	GAGTGAC	chr4	6355349	chr4:6355347-6355353	ER+ overexpressed
rs1434536	CTCAGGG	chr4	96294988	chr4:96294988-96294994	ER+ overexpressed
rs1141538	ATGCTGC	chr5	137303098	chr5:137303092-137303098	ER+ overexpressed
rs1438688	CCCCGCC	chr5	148619255	chr5:148619255-148619261	ER+ overexpressed
rs1062577	ATTCTTT	chr6	152465598	chr6:152465594-152465600	ER+ overexpressed
rs1225737	TGCCTTA	chr6	11090638	chr6:11090633-11090639	ER+ overexpressed
rs508477	GGTGTGT	chr6	13472323	chr6:13472320-13472326	ER- overexpressed
rs7756717	CTGAGCC	chr6	11090925	chr6:11090921-11090927	ER+ overexpressed
rs8523	ATTTCTC	chr6	11089039	chr6:11089037-11089043	ER+ overexpressed
rs9341070	ACACTCC	chr6	152461890	chr6:152461884-152461890	ER+ overexpressed
rs9341074	AATGGGT	chr6	152464148	chr6:152464142-152464148	ER+ overexpressed
rs10263074	TTTATCT	chr7	87375999	chr7:87375994-87376000	ER- overexpressed
rs6616	ATTTCTC	chr7	16790503	chr7:16790502-16790508	ER+ overexpressed
rs4986994	AACTGAC	chr8	18124767	chr8:18124761-18124767	ER+ overexpressed
rs12710570	GTGCAAT	chrX	115506264	chrX:115506260-115506266	ER- overexpressed
rs6567569	TGCAGAA	chrX	3534309	chrX:3534306-3534312	ER- overexpressed
rs741500	TGTTACT	chrX	10377020	chrX:10377016-10377022	ER- overexpressed

* UCSC Build 36 coordinates

† Defined in Smith, et al BMC Bioinformatics 9:63,2008

Supplementary Table S3-Original CGEMS Association Findings

		CGEMS cases (n=1145)			CGEMS controls (n=1142)			CGEMS			
SNP	GT	Count	Prop	HWE	Count	Prop	HWE	OR	95% C.I.	P-value*	Rank†
rs1219648	A/A	351	0.31	0.152	432	0.38	0.851	1.00		8.0E-06	2
Het	A/G	542	0.47		534	0.47		1.25	1.04-1.50		
Minor	G/G	249	0.22		169	0.15		1.81	1.42-2.04		
rs1970801	G/G	165	0.15	0.950	215	0.30	0.337	1.00		1.7E-04	79
Het	T/G	534	0.47		577	0.51		1.21	0.95-1.52		
Major	T/T	436	0.38		344	0.19		1.65	1.29-2.11		
rs6831418	C/C	250	0.22	0.005	374	0.33	0.028	1.0		1.2E-04	52
Het	C/T	515	0.45		588	0.52		0.88	0.73-1.06		
Minor	T/T	373	0.33		175	0.15		1.43	1.13-1.82		

*Unadjusted P-value from the score test with df=2 in Logistic regression from original CGEMS GWAS
†Rank of the SNPs from original CGEMS GWAS, for rs1970801, test was for minor allele G