Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus

Melissa A. Buckley,[1,2] Nicholas T. Woods,[1,3] Jonathan P. Tyrer,[4] Gustavo Mendoza-Fandiño,[1] Kate Lawrenson,[5] Dennis J. Hazelett,[6,7,8] Hamed S. Najafabadi,[9,10] Anxhela Gjyshi,[1,2] Renato S. Carvalho,[1] Paulo C. Lyra, Jr.,[1] Simon G. Coetzee,[8] Howard C. Shen,[6] Ally W. Yang,[9] Madalene A. Earp,[11] Sean J. Yoder,[12] Harvey Risch,[13] Georgia Chenevix-Trench,[14] Susan J. Ramus,[15,16] Catherine M. Phelan,[1†] Gerhard A. Coetzee,[6,17] Houtan Noushmehr,[18] Timothy R. Hughes,[9,19,20] Thomas A. Sellers,[1] Ellen L. Goode,[12] Paul D. Pharoah,[4] Simon A. Gayther,[5,21] and Alvaro N.A. Monteiro,[1] on behalf of the Ovarian Cancer Association Consortium

[1]Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA
[2]University of South Florida Cancer Biology PhD Program, Tampa, FL 33612, USA
[3]Department of Oncological Sciences, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[4]The Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
[5]Women's Cancer Program at the Samuel Oschin Comprehensive, Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[6]Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA 90089, USA
[7]Department of Urology, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA 90089, USA
[8]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[9]Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
[10]Department of Human Genetics, McGill University, Montreal, QC H3A 0G1, Canada
[11]Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA
[12]Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA
[13]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA
[14]QIMR Berghofer Medical Research Institute, Brisbane, Australia
[15]School of Women's and Children's Health, University of New South Wales, Sydney, Australia
[16]The Kinghorn Cancer Center, Garvan Institute of Medical Research, Darlinghurst, Australia
[17]Van Andel Institute, Grand Rapids, MI 49503, USA
[18]Department of Neurosurgery, Henry Ford Health System, Detroit, MI 48202
[19]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada
[20]Canadian Institutes for Advanced Research, Toronto, ON M5G 1Z8, Canada
[21]Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[†]deceased

Running title: *Analysis of the 9p22 ovarian cancer risk locus*

Correspondence should be addressed to:
Alvaro N. Monteiro
H. Lee Moffitt Cancer Center & Research Institute
12902 Magnolia Drive, Tampa, FL 33612, USA.
Phone: 813-7456321Fax: 813-9036847
E-mail: alvaro.monteiro@moffitt.org

Keywords: Ovarian cancer, GWAS, SNPs, enhancer, susceptibility, scaffold/matrix attachment region

Conflict of interest statement: The authors state that they have no conflict of interest to disclose.

Abstract

Genome-wide association studies have identified 40 ovarian cancer risk loci. However, the mechanisms underlying these associations remain elusive. In this study, we conducted a two-pronged approach to identify candidate causal SNPs and assess underlying biological mechanisms at chromosome 9p22.2, the first and most statistically significant associated locus for ovarian cancer susceptibility. Three transcriptional regulatory elements with allele-specific effects and a scaffold/matrix attachment region were characterized and through physical DNA interactions *BNC2* was established as the most likely target gene. We determined the consensus binding sequence for BNC2 in vitro, verified its enrichment in BNC2 ChIP-Seq regions and validated a set of its downstream target genes. Fine-mapping by dense regional genotyping in over 15,000 ovarian cancer cases and 30,000 controls identified SNPs in the scaffold/matrix attachment region as among the most likely causal variants. This study reveals a comprehensive regulatory landscape at 9p22.2 and proposes a likely mechanism of susceptibility to ovarian cancer.

Significance: Mapping the 9p22.2 ovarian cancer risk locus identifies BNC2 as an ovarian cancer risk gene

INTRODUCTION

Epithelial ovarian cancer (EOC) is a poorly understood disease often diagnosed at late stages and with low 5-year survival rates. Although it used to be widely acknowledged that the ovarian surface epithelium (OSE) was the likely tissue of origin of EOC, recent evidence supports the notion that the epithelial lining of the fallopian tube and benign endometriosis contribute to the origin of invasive EOCs. Invasive EOCs may also originate from ectopic Müllerian tissue due to endosalpingiosis. The diverse cellular origins of EOC subtypes, in part, underlie the heterogeneity that characterizes ovarian cancer.

Less than half of all familial ovarian cancer cases and less than 15% of high grade serous EOC are due to highly penetrant pathogenic alleles of genes such as *BRCA1* and *BRCA2*. However, exhaustive family-based linkage studies have not identified additional highly penetrant EOC susceptibility genes (1). The excess familial risk of EOC may be explained, at least in part, by common variants with low to moderate penetrance. Genome-wide association studies (GWAS) have identified ~40 common variant loci associated with risk of EOC (2-13). Delineation of the mechanisms and likely causal variants at GWAS-identified loci may reveal novel chemoprevention and therapeutic strategies.

To evaluate the mechanisms by which single nucleotide polymorphisms (SNPs) may contribute to EOC, we conducted a functional dissection at the 9p22.2 locus, the first ovarian cancer risk locus identified through GWAS of European ancestry women (2). The SNP most significantly associated with high grade serous EOC risk was rs3814113, which is located 44 kb centromeric and 220 kb telomeric to the *BNC2* and *CNTLN* transcription start sites (TSS), respectively (2). The minor allele [C; MAF = 0.323] was associated with reduced risk of high grade serous EOC (combined data OR = 0.82; 95%CI = 0.79-0.86; P = 2.5×10^{-17}).

MATERIALS AND METHODS

Cell Lines

We used two immortalized normal OSE cell lines, iOSE4 and iOSE11 (14), and three immortalized normal fallopian tube surface epithelial cells (iFTSEC33, iFTSEC246, and iFTSEC283), a normal epithelial ovarian cell line, iOSE4^{CMYC}, immortalized with *hTERT* and transformed with *MYC* (15), and HEK293FT cells. Cell line aliquots were tested for mycoplasma (PCR-based method) and authenticated using STR analysis before being used for experiments, which were conducted before 20 passages after thawing.

FAIRE-Seq and ChIP-Seq for Histone Modifications

FAIRE-Seq (Formaldehyde Assisted Isolation of Regulatory Elements followed by sequencing) and ChIP-Seq (Chromatin immunoprecipitation followed by sequencing) for Histone H3 Lysine 27 Acetylation (H3K27Ac) and Histone H3 Lysine 4 Monomethylation (H3K4me1) were performed in iOSE4, iOSE11, iFTSEC33, iFTSEC246, iFTSEC283 (GSE68104) (16).

Enhancer Scanning

We used an optimized method to identify genomic regions with enhancer activity (17). Genomic tiles of ~2 kb were generated by PCR using bacterial artificial chromosome (BAC) Clone RPCI-11-185E1 (Empire Genomics) as the template and cloned in forward and reverse orientations upstream in the firefly luciferase reporter vector designed to test for enhancer activity (17). Primers can be found in Supplementary Table 1. Transfections included a plasmid expressing *Renilla sp.* luciferase as internal control and every tile was tested in two independent experiments. Tiles with significantly (two-tailed t-test; $p < 0.05$) higher luciferase counts than the control tile (TC) were tested for allele specific effects. For allele-specific luciferase assays, tiles

with the effect allele were considered significant if the luciferase counts were significantly higher or lower ($p < 0.05$) in at least one independent experiment than the tile with the reference allele.

Electrophoretic Mobility Shift Assays

Nuclear extracts were obtained from iOSE4^{CMYC} cells at 70-90% confluence and EMSAs were run as previously described (18).

Nuclear Scaffold Extraction

A lithium-based nuclear scaffold extraction was performed as previously described (19). Scaffold and genomic DNAs were quantified by qPCR using primers for Region 11, the ApoB S/MAR and the ApoB Neg regions (Supplementary Table 1). Samples were run using Sybr Green Spectrum on Applied Biosystems 7900 HT Real-Time PCR System. Enrichment was calculated by dividing the quantity of the scaffold DNA by the quantity of the digested genomic DNA. A Z-score for the region 11 and ApoB S/MAR was calculated as described previously (19) [Z score = (average of S/MAR – average of ApoB Neg)/std dev of ApoB Neg)]. A Z-Score > 8 indicates a site positive for scaffold binding (19). Each experiment includes three technical replicates.

Chromosome Conformation Capture (3C)

3C libraries were prepared as previously described (20). qPCR was performed by using Taq Polymerase PCR Kit (Qiagen) and Syto9 (Life Technologies). Samples were run using FAM Spectrum on an Applied Biosystems 7900 HT Fast Real-Time PCR System. EcoR1 digested BACs (RPCI-11-185E1 Empire Genomics, RPCI-11-179K24 Life Technologies, RPCI-11-106G11 Life Technologies) for the region were used for the standard curve. Interactions were calculated as a percentage of a restriction site directly adjacent to the bait restriction site. Sites

with a significantly higher frequency of interaction than the site adjacent to the anchor were considered significant ($p < 0.05$; two-tailed *t*-test). 3C was performed in two independent experiments and three technical replicates each.

Protein Binding Microarray

Fragments containing cDNAs of each of the zinc finger pairs were PCR amplified from a plasmid containing *BNC2* cDNA (a gift from Dr. Philippe Djian) using primers containing Gateway recombination sites (Supplementary Table 1). PCR products were cloned into pDONR221 using the BP recombination kit and transferred to pDEST15 as a fusion to Glutathione-S-transferase (GST) using LR recombination kit (Invitrogen). Purified GST-ZFs were eluted from beads with 50 mM reduced glutathione and 0.5 µg of each GST-ZF protein construct were applied individually to two differently designed arrays designated ME and HK as previously described (21, 22). ZFs typically bind to degenerate motifs and have the potential to have more than one recognition sequence (21). Each DNA probe sequence is given an E-score which is similar to the Area under the ROC curve statistical metric and an E-score above 0.45 was considered significant.

ChIP/ChIP-Seq for BNC2

Chromatin immunoprecipitations were performed as previously described (23) using a validated BNC2 antibody (Sigma Atlas) (see Supplementary Data). Real-time qPCR was performed using Sybr Green chemistry with primers at the -2184, -914, and -582 positions relative to the TSS (Supplementary Table 1) in an Applied Biosystems 7900HT Fast Real-Time PCR System. ChIP for each cell line was performed in four biological replicates. Overrepresentation test (release 20170413) was conducted with PANTHER version 11.1 released 2016-10-24 using all

genes in *Homo sapiens* database as a reference list and a Bonferroni correction for multiple testing. The uploaded list contained 965 genes of which 839 were mapped to GO-Slim.

For BNC2 ChIP-Seq four individual ChIP samples were pooled for each cell line (iOSE11 and iFTSEC283) in two biological replicates. Immunoprecipitated DNA was used to generate a sequencing library using the NuGEN Ovation Ultralow Library System with indexed adapters (NuGEN, Inc., San Carlos, CA). The library was PCR amplified and size-selected using AxyPrep Fragment Select beads (Corning Life Sciences – Axygen Inc., Union City, CA). Each enriched DNA library was then sequenced on an Illumina HiScan SQ sequencer to generate 20-30 million 100-base paired-end reads. The raw sequence data was de-multiplexed using the Illumina CASAVA 1.8.2 software (Illumina, Inc., San Diego, CA) and binding sites were identified using the MACS2 software (24) using input DNA as a control. See Supplementary Data for further details.

Nanostring

pNTAP-BNC2 (or the empty vector) was transfected with Fugene 6 into 293FT cells at 70% confluence. Cells were harvested after 24 h, RNA was isolated using Trizol RNA Isolation (Life Technologies), and cleaned using Qiagen RNeasy Mini Kit (Qiagen). The three biological replicates for HEK293FT cells with the empty vector or over-expressed BNC2 were applied to a Nanostring platform containing probes for 87 genes and 10 reference genes (Supplementary Table 2) used to normalize the data in the NanoString nSolver Analysis Software v 1.1. These genes had a %CV < 50. Genes were considered to be differentially expressed if $p < 0.05$ (two-tailed t-test).

Fine-mapping Association Analyses

To refine the observed signal at rs3814113 (2), fine-mapping was conducted using a customized Illumina iSelect genotyping array (iCOGS). SNPs were selected based on data from 1000 Genomes Project (1000GP) (25) CEU (April 2010) and Hapmap III within a 1 Mb interval of rs3814113 (chr9: 16407967-17407967)(26). We included tagging SNPs ($r^2 > 0.1$) with a minor allele called at least twice in the 1000GP and additional SNPs tagging remaining variation in the interval ($r^2 > 0.9$), requiring Illumina Design score > 0.8.

The iCOGS array was used to genotype cases and controls from constituent studies of the Ovarian Cancer Association Consortium as previously described (6), supplementing with data from three independent ovarian cancer GWAS. In iCOGS, we excluded samples if they were not of European ancestry, had a genotyping call rate of <95%, showed low or high heterozygosity, were not female or had ambiguous sex, or were duplicates (cryptic or intended). SNPs were excluded if they were mono-morphic, had a call rate of < 95%, showed evidence of deviation from Hardy-Weinberg equilibrium in controls or had low concordance between duplicate pairs(6). For two of the GWAS (from Mayo Clinic and the UK), we also excluded rare SNPs (MAF < 1% or allele count < 5, respectively). The final data set comprised 11,069 cases and 21,722 controls from iCOGS ('OCAC-iCOGS'), 2,165 cases and 2,564 controls from a GWAS from North America (27), 1,762 cases and 6,118 controls from a United Kingdom-based GWAS (2), and 441 cases and 441 controls from the Mayo Clinic. All subjects included in this analysis provided written informed consent as well as data and blood samples in accordance to ethical guidelines under protocols approved by institutional review boards of their respective study sites. Overall, 43 studies from 11 countries provided data on 15,437 women diagnosed with invasive EOC, 9,627 of whom were diagnosed with serous EOC, and 30,845 controls from the general population.

We imputed variants separately for the OCAC-iCOGS and each GWAS from 1000 Genomes Project data using the v3 April 2012 release as the reference panel using the

IMPUTE2 software (28) without pre-phasing. The final data set comprised genotypes for 4,234 SNPs of which 2,418 had been directly genotyped.

We evaluated the association between genotype and disease using logistic regression by estimating the associations with each additional copy of the minor allele (log-additive models). The analysis was adjusted for study and population substructure by including the eigenvectors of the first five European-specific principal components as covariates in the model. We used the same approach to evaluate SNP associations with serous ovarian cancer after excluding all cases with any other or unknown tumor subtype. For imputed SNPs, we used expected dosages in the logistic regression model to estimate SNP effect sizes and P values. We carried out analyses separately for OCAC-iCOGS and each GWAS and pooled data thereafter using a fixed-effects meta-analysis; thus, all results are based on the combined data. We also performed analyses adjusted for rs3814113 to evaluate evidence of independent signals.

RESULTS

Overview of Study Design

Here, we utilized two independent approaches to identify a list of candidate causal SNPs (Fig. 1). First, we conducted a comprehensive analysis to identify functional SNPs in linkage disequilibrium (LD, $r^2 > 0.3$) with rs3814113 with no prior assumption about their individual association to risk. Since all resided in non-coding regions, we hypothesized that SNP alleles determine the activity of regulatory elements in enhancers and promoter regions active in OSE and fallopian tube surface epithelial cells (FTSEC) (29). Second, we performed fine-mapping association analyses by densely genotyping over 15,000 ovarian cancer cases and 30,000 controls to identify a credible set of causal SNPs guide by association data. These parallel approaches identified the SNP most likely to be causal to ovarian cancer risk at the 9p22 locus.

Candidate Causal SNP Set for Functional Analysis

A total of 134 SNPs were chosen for functional analysis, based on their LD ($r^2 \geq 0.3$) with rs3814113 in European 1000 Genomes Project data (v3 April 2012 release). They are distributed over an 82 kb region ranging from the first intron of *BNC2* to ~44 kb centromeric to its transcription start site (TSS) (Fig. 2a and Supplementary Table 3).

Since all SNPs in the candidate functional set are in non-coding regions, several independent assays were used to identify transcriptional regulatory elements. First we analyzed data from FAIRE-Seq, and ChIP-Seq for H3K27Ac and H3K4Me1. FAIRE-Seq reveals regions of open chromatin while H3K27Ac or H3K4Me1 are markers for active chromatin and enhancers, respectively. The chromatin landscape profiles (Fig. 2a) were derived from iOSE and iFTSEC cells (16).

Analysis of FAIRE- and ChIP-Seq data identified twelve regions with evidence of enhancer activity in at least one cell line (Fig. 2a). Twenty-two candidate causal SNPs (Table 1)

are located within five regions containing FAIRE or ChIP-Seq features suggesting that these SNPs might have a functional impact (Fig. 2a). The relatively lenient threshold for LD and criteria to consider a region as a putative enhancer was designed to favor sensitivity at the initial stage of analysis (with high specificity being achieved by the integration of the two approaches).

Mapping SNPs to Regions of Enhancer Activity

To refine the analysis, we tested twelve genomic tiles (~2 kb each) (Fig. 2a), in both orientations, spanning the five candidate regions using a reporter assay to identify enhancer activity in iOSE4^{cMYC} ovarian cells (17). Although not present in a region with evidence of regulatory activity, we also tested one tile containing rs3814113 (Tile 12), the most significantly associated with high grade serous EOC in a previous study (2), and a control tile devoid of evidence for enhancer activity as judged by FAIRE and ChIP-Seq data (Fig. 2a, Tile C). Tiles in regions 6 (T6), 7 (T7.2, T7.3, T7.6), and 8 (T8) contained nine candidate causal SNPs and showed significant activity (two tailed *t*-test $p<0.05$ compared to the control tile C; two replicates) in at least one orientation (Fig. 2b).

Causal SNPs are hypothesized to display allele-specific effects. Therefore, we used site-directed mutagenesis in tiles T6, T7.2, T7.3, T7.6, and T8 to change each of the nine candidate causal SNPs from the reference to the effect allele and compared their activity. For tiles with multiple SNPs, the reference tile was the most common haplotype (Supplementary Fig. 1) (All populations; 1000 Genomes Project). Individual SNPs were mutated to determine the contribution of each SNP, with other SNPs in the haplotype retaining the reference SNP allele. Seven SNPs in T6 (rs62541878), T7.2 (rs62541920, rs12379183), T7.3 (rs1092647), and T8 (rs77507622, rs10810657, rs12350739) demonstrated significantly different transcription activity ($p<0.05$) between the reference and effect allele in at least one replicate (Fig. 2c-d). These seven SNPs were retained for analysis.

Allele Specific Activities in Electrophoretic Mobility Shift Assays

We conducted EMSAs using probes with both alleles for each of the seven SNPs in regions 6, 7 and 8 (Fig. 2e). Tiles that did not show activity in Fig. 2b were not tested. Tile 11 had significant transcription activity in only one reporter experiment but two SNPs within the region (rs113780397 and rs181552334) are correlated with the original SNP (r^2 of 0.818 and 0.5, respectively), and so four additional probes were tested. We also examined rs3814113, the most significant original GWAS SNP. EMSAs revealed allele specific nuclear extract binding for rs12379183, rs62541920 (Region 7), rs12350739, rs77507622 (Region 8) and rs181552334 (Region 11) (Fig. 2e) indicating these SNPs were strong causal candidates.

Region 11 Attaches to the Nuclear Scaffold

Region 11 overlapped with an open chromatin region, according to FAIRE-Seq data obtained in ovarian cells, and one SNP showed allele-specific binding in EMSA experiments (rs181552334). However, this region lacked H3K4Me1 and H3K27Ac marks and luciferase assays showed weak evidence for enhancer activity in ovarian cells (Fig. 2b). Interestingly, the region is A/T rich (> 60%), a feature in regions that anchor the cell's DNA to the nuclear scaffold/matrix (19). Moreover, Region 11 was predicted by MAR-Wiz to attach to the nuclear scaffold/matrix compared to the rest of the locus (Supplementary Fig. 2).

To determine whether Region 11 was attached to the nuclear scaffold in ovarian cells, we performed a nuclear scaffold extraction in iOSE11 cells (19), using HeLa cells as a control. Region 11 had significantly higher enrichment in the scaffold fraction of iOSE11 and HeLa cells than a previously defined negative control (ApoB Neg) (19) (Fig. 3a-c). A region previously defined as a S/MAR (ApoB S/MAR) (19) in HeLa cells did not have significantly higher enrichment in the scaffold fraction of iOSE11 cells than ApoB Neg (Fig. 3a-c) but had

significantly higher enrichment in the scaffold fraction of HeLa cells than ApoB Neg. These results indicate that Region 11 acts as S/MAR in ovarian cells. Visual inspection of HiC (High dimensional chromosome conformation capture) data from seven cell lines suggested the presence of a 1.8 Mb (chr9:15,750,000-17,550,000) topologically associating domain (TAD) in which the S/MAR (Region 11, rs181552334) is situated close to one of its borders (Fig. 3d-e). This TAD includes TSS for *BNC2, C9orf92*, and *CNTLN.*

Candidate Target Genes *BNC2 and CNTLN*

Two functional SNPs in Region 7 were located in an approximately 7kb region that includes the TSSs for two *BNC2* transcripts (Fig. 2a) denoted by FAIRE-Seq and H3K4me1 ChIP-Seq data in ovarian cells, and ENCODE layered H3K4me3 (promoters) ChIP-Seq data (Fig. 3f). This region is the major *BNC2* promoter, raising the hypothesis that *BNC2* may act as the mediator of risk at the 9p22.2 locus.

Region 8, containing two SNPs with allele specific activity in luciferase assays and EMSA, overlapped with FAIRE-Seq and ChIP-Seq data in ovarian cells with features indicative of an enhancer (Fig. 2a). To determine potential interacting promoters with the enhancer at region 8, we examined all genes (*c9orf92, BNC2, CNTLN* and *SH3GL2*) within a stretch of 1 MB at either side of the region containing the candidate SNPs (Fig. 3f). First, guided by H3K4me3 marks in seven non-ovarian cell lines from ENCODE, we identified their promoters close to TSSs (Fig. 3f). Next, we inferred whether the gene was expressed in ovarian cell lines using H3K27ac as a marker of active promoters combined with analysis of transcript levels from RNA sequencing (RNA-Seq) data for ovarian and fallopian tube epithelial cells (Fig. 3g). This analysis indicated that *BNC2* and *CNTLN* were expressed in ovarian cells, but *c9orf92* or *SH3GL2* were not (Fig. 3g).

Region 8 is physically close to the TSS of *BNC2* in Ovarian Cells

Next, we used Chromatin Conformation Capture (3C) to determine which promoters physically interacted with Region 8. In iOSE11 cells, Region 8, when compared to an adjacent site displayed two regions of frequent (Fig. 3h; blue arches) interactions. The interaction peak closer to the anchor is located upstream of the TSS but does not overlap with any known chromatin marks. The second interaction peak corresponds to region 7 considered to be the core promoter of *BNC2*. No significant interaction was detected between Region 8 and the *CNTLN* TSS (Fig. 3h). As expected, no interaction was detected between the S/MAR in Region 11 and promoters in the region (Fig. 3i). The modules in Regions 7 and 8 appear to affect the major promoter of *BNC2* and are a distal regulatory enhancer that physically interacts with the *BNC2* promoter, respectively.

Fine mapping

Next, as part of our two-pronged approach, we conducted fine mapping of the 9p22 locus in 15,437 women diagnosed with invasive EOC and 30,845 controls (Fig. 4a). We evaluated the association between genotype and disease using logistic regression by estimating the associations with each additional copy of the minor allele (log-additive models) for 4,234 SNPs of which 2,418 were directly genotyped (Supplementary Table 4). SNP rs3814113 remained the most statistically significant association ($P = 2.10 \times 10^{-34}$) (Fig. 4a) with the minor allele [C] being protective. Next, we calculated the likelihood ratio of each SNP relative to the most significant SNP (rs3814113) of being the functional variant underlying the signal. For any given set of correlated associated SNPs, the strength of evidence was estimated by the log likelihood statistic from the logistic regression; thus difference in the log likelihood between the SNP with the strongest association and any other SNP provides a measure of the log odds in favor of the

most significant SNP being the SNP that is truly driving the observed association. There were 40 SNPs with odds of 1:1000 or better and were considered to be credible candidates for mediating the observed association. They were all in strong LD ($r^2 > 0.89$).

While 35 out of the 40 SNPs were part of the set of 134 SNPs assessed during functional analysis, five SNPs (rs34131140, rs112442786, rs113198237, rs199782476, and c9_pos16900214/rs62543587) were not (Supplementary Fig. 3). SNPs rs199782476 and rs62543587 did not overlap with any biofeatures (FAIRE-seq, H3K4me1, H3K27Ac) suggesting they were not functionally relevant. The remaining three SNPs were part of Tile 11 (rs112442786), Tile 12 (rs34131140) or the control (TC) Tile (rs113198237). Tiles 12 and TC did not display significant activity in enhancer scanning (Fig. 2b) suggesting that rs112442786, which resided in Tile 11 and mapped to the S/MAR region, may be functionally relevant (Supplementary Fig. 3).

Among the 40 SNPs, six SNPs with odds ranging from better than 1:4 (rs112442786) to better than 1:200 (rs181552334) mapped to the S/MAR (Region 11) (Fig. 4a). We repeated the association analyses adjusting for rs3814113 to identify additional independent signals in the region (Fig. 4b). Nine SNPs were significant at $P < 10^{-5}$ (Supplementary Table 4; Conditional tab) of which two (rs7848057, rs80039758) mapped to the S/MAR (Fig. 4b). In this group of 9 SNPs, rs10756825 and c9_pos16889285 were the next most significant associations ($p = 2.3 \times 10^{-5}$; $p = 6.23 \times 10^{-5}$, respectively) and mapped close to the enhancer in Region 8 (Fig. 4b). Finally, several attempts to remove the S/MAR using CRISPR-based genome editing techniques were not successful, suggesting that deletion of this region may impact the viability of ovarian and fallopian tube cells.

To identify eQTL associations for *c9orf92, BNC2, CNTLN* and *SH3GL2* we searched the GTEx dataset for single gene eQTLs in all tissues (GTEx Analysis Release V7; dbGaP Accession phs000424.v7.p2; fallopian tube not included due to small sample size). Although all

four genes displayed eQTL associations (*CNTLN* = 11,039; *C9ORF92* = 1; *SH3GL2* = 361; *BNC2* = 94) (Supplementary Table 5) only *BNC2* displayed eQTL associations with SNPs (rs10962662, rs10756823, and rs10124837; whole blood) present in our set of 40 credible candidate SNPs. Next, we searched for single SNP eQTL associations in all tissues for 40 credible candidate SNPs. The only three eQTL associations found were for SNPs rs10962662, rs10756823, and rs10124837 with *BNC2.*

The data from the functional analysis and fine mapping data provide evidence that the candidate causal SNPs at the locus exert their effects in a 1.8Mb TAD with *BNC2* as the most likely target gene at the locus.

In vitro Recognition of Specific DNA Sequences by BNC2 Zinc Fingers

BNC2 has three pairs of C2H2 zinc fingers (ZF) raising the possibility that it recognizes specific DNA sequences and is involved in transcription regulation (Fig. 5a) (30). To identify DNA sequences recognized by BNC2, GST-tagged constructs of each ZF pair (Supplementary Fig. 4) were expressed in bacteria and applied to a protein binding microarray (PBM) with overlapping, rationally randomized nucleotides, representing every possible motif up to 10 bp (21, 22). When aligned the top ten scoring sequences for each ZF pair generated a sequence logo using position weight matrix scoring (Fig. 5a). The motifs for ZF1,2 and 5,6 were consistent with the predicted C2H2 “recognition code” (31). Binding for ZF3,4, which yielded lower-confidence data, did not match the recognition code predictions (Fig. 5a) (32). The 3’ end of the ZF1,2 and ZF5,6 binding motifs had the same nucleotides at the exact same position and weight, consistent with the similarity in amino acid residue positions between ZF2 and ZF6 (Fig. 5a). Notably, the *BNC2* promoter region contains two BNC2 ZF5,6 PBM binding sequences (Supplementary Fig. 4).

We validated *BNC2* binding sequences identified with the PBM by conducting ChIP in iOSE11 and iFTSEC283 cells for endogenous BNC2 (Supplementary Fig. 4) at the PBM sites (-582 and -914 bp upstream of the TSS) at the *BNC2* locus. A significantly larger amount of DNA was immunoprecipitated with the BNC2 antibody than with the IgG control at the -582 (iOSE11 $p = 2.6 \times 10^{-3}$, iFTSEC283 $p = 8.3 \times 10^{-3}$) and -914 (iOSE11 $p = 1.8 \times 10^{-4}$, iFTSEC283 $p = 2.0 \times 10^{-6}$) bp sites, but not at the -2184 bp site (negative control; Supplementary Fig. 4). These data provides evidence that the sites identified in the PBM experiment are recognized by endogenous BNC2.

BNC2 Genome-wide Target Sites

To identify genomic sites bound by BNC2 in ovarian cells, we used ChIP-Seq in iOSE11 and iFTSEC283 cells (see Extended Data). MEME, a motif analysis tool, defined a motif centrally enriched in the ChIP-Seq peaks in both cell types (Fig. 5b-c). The motif identified by MEME appears to be a concatenation of the reverse complement motif for ZF1,2 and the motif for ZF5,6 with a 75% homology (Fig. 5b). The concatamer motif was significantly enriched in ChIP-Seq peak summits in iFTSEC283 and iOSE11 cells (Fig. 5c). ChIP-Seq data replicated BNC2 binding in the iOSE11 cells (chr9:16871799-16872039) at the -914 position tested in ChIP-qPCR (Supplementary Fig. 4).

Identification and Validation of BNC2 Target Genes

To identify putative target genes regulated by regulatory elements containing *BNC2* recognition sites, we generated a list of 995 genes/transcripts with TSS within 30 kb of the BNC2 ChIP-Seq peak centers found in both iOSE11 and iFTSEC283 cells (Supplementary Table 2). Next, we used PANTHER (33) and found that several functional classes were statistically overrepresented in our set including system development (GO:0048731), anatomical structure

development (GO:0048856), single-multicellular organism process (GO:0044707), multicellular organism development (GO:0007275), and tissue development (GO:0009888) (Supplementary Table 2).

From the above set, we selected a set of 87 genes that were: a) implicated in ovarian cancer; b) ovarian development; c) were part of KEGG pathways related to cancer; d) in which BNC2 ChIP-Seq peaks were found in their core promoter (within 1kb from the TSS) (Supplementary Table 2) and tested the extent to which their expression (measured by Nanostring) was modulated by overexpression of *BNC2* in HEK 293T.. Multiple unsuccessful attempts were performed to manipulate expression - silencing or ectopic overexpression - levels of BNC2 in ovarian cells, suggesting that BNC2 levels are tightly controlled. Several genes mapping to KEGG Focal Adhesion, ECM-receptor interaction or TGF-β Signaling Pathways and implicated in ovarian cancer or ovarian development showed significant changes in expression upon *BNC2* overexpression (Table 2; Supplementary Table 2). Although most genes showed a positive correlation with BNC2 overexpression, *FEM1A* and *IGTB5* showed an inverse correlation suggesting that BNC2 modulation of expression is likely to be context dependent (Table 2). Taken together, these experiments validate the BNC2 binding site in vivo and reveal putative downstream targets of BNC2 activity in ovarian cells.

DISCUSSION

Here, we started from early findings from GWAS for EOC risk and delineated a mechanistic hypothesis for susceptibility at the 9p22.2 locus. Using a two-pronged approach combining functional analysis and fine mapping we identified three genomic features (and enhancer to *BNC2*, the *BNC2* promoter, and a Substrate/Matrix Attachment Region) harboring twelve potentially functional SNPs (Fig. 1). Based on the likelihood statistics, the most likely causal SNPs were in a Substrate/Matrix Attachment Region (S/MAR) located in a 1.8 Mb topologically associating domain (TAD). Also, this TAD includes associated SNPs revealed in conditional analysis (adjusting for rs3814113) that locate to an enhancer region that interacts with the *BNC2* promoter in ovarian cells. Taken together our data implicate multiple candidate causal SNPs at the locus that converge to regulate *BNC2* in ovarian cancer susceptibility.

Our functional analysis revealed two SNPs in the *BNC2* promoter, two SNPs in an enhancer that physically interacts with the *BNC2* promoter, and a functional SNP in a S/MAR with allele-specific effects. Of these five SNPs, the strongest genetic evidence for causality is for rs181552334 in the S/MAR. An additional SNP, rs112442786 (r^2 to rs181552334 = 0.9556), located in the same region, which emerged in our fine-mapping approach was not directly tested and may also contribute to risk. S/MARs are thought to help maintain the local 3D chromatin structure by contributing to looping and modulate gene expression (34). Polymorphisms in S/MARs can regulate, in an allelic-specific manner, attachment to the nuclear scaffold/matrix (35). Interestingly, our EMSA experiments suggest allelic specific binding to nuclear proteins. Several attempts to remove the S/MAR using CRISPR-based genome editing techniques were not successful suggesting that ovarian cells may not be viable without this S/MAR. CRISPR-based deletion of a region including the S/MAR in 293T HEK cells led to a two-fold reduction of *BNC2* expression (36), with no changes in *CNTLN* expression, implicating the S/MAR in *BNC2* regulation. Notably, all three cis-eQTL associations detected for the 40 credible SNPs from fine

mapping were with *BNC2*; and all cis-eQTL associations for the four genes at the locus only *BNC2* showed eQTL with the set of credible SNPs. High dimensional chromatin interaction (HiC) data from seven cell lines indicates the presence of a 1.8 Mb topologically associating domain (TAD) in which the S/MAR is situated close to one of its borders. Our data suggest a role for several regulatory interactions, defined by a TAD containing multiple non-coding elements which target *BNC2* (Fig. 1).

Although future research will further delineate the relationship between *BNC2* and ovarian biology, recent reports support our findings. Hnisz et al. (37) identified a super enhancer in ovarian cells near *BNC2*, consistent with *BNC2* representing a cell identity gene or master regulator in ovarian cells. The *bonaparte* zebrafish (*Bnc2)* mutants display skin pigmentation defects (no body stripes), stunted growth, and dysmorphic ovaries coupled to infertility (38). In mice, *Bnc2* is expressed in ovarian theca cells, and female mice nullizygous for *Bnc2* display an excessive number of stromal cells combined with a reduced number of oocytes (39). Interestingly, rs12379183, in Region 7, is associated with sonographically detectable abnormalities in the ovaries (40). Moreover, a network-based integration of GWAS and gene expression in ovarian cancer focusing on transcription factors identified *BNC2* using a combination of coexpression and enrichment analysis as a gene contributing to a *HOX*-centric network associated with serous ovarian cancer risk (41). Finally, a recent analysis of genetic interactions between germline polymorphisms and tumor formation in specific tissues revealed a significant association between rs3814113 and ovarian cancer (42).

Genome-wide and candidate gene association studies suggest that this locus may also be pleiotropic in humans with effects on ovary, skin, and skeletal biology. SNPs in the 9p22.2 locus have been associated with skin pigmentation in Europeans (43) and Asians (44), with freckling (45), and height (46).Functional analysis revealed rs12350739 as the likely causal variant contributing to saturation of skin color (47). SNP rs12350739 was identified in the

present study mapping to a Region 8 (a candidate *BNC2* enhancer). An introgressed region of Neanderthal DNA (Chr9: 16,720,121-16,786,930) proposed to confer adaptive advantage to colder climates through changes in skin pigmentation is also present at this locus (48, 49). Finally, this locus has also been shown to modify ovarian cancer risk in carriers of *BRCA1* and *BRCA2* pathogenic variants (50).

We acknowledge limitations of this work including the that regulatory networks may be significantly altered during development (51), the incomplete knowledge of the regulatory landscape in ovarian cells (*e.g.* lack of data on CTCF repressor marks and of information on other non-coding RNA elements) and the possibility of missing rare alleles that contribute to the phenotype that could be revealed using the larger Haplotype Reference Consortium data for imputation. Despite these limitations, our data identify plausible and likely biological mechanisms operating to modulate ovarian cancer risk. In summary, we confirmed the region as a highly associated susceptibility locus and propose that the mechanism of ovarian cancer susceptibility at the 9p22.2 locus is likely mediated by changes in a transcriptional regulatory network involving several regulatory elements (enhancers and S/MAR) acting on *BNC2.*

ACKNOWLEDGEMENTS

We thank Alexandra Valle, Jiqiang Yao and Xueli Li for technical assistance and Anindya Dutta for helpful discussions. This work was supported by the following awards: NIH Genetic Association and Mechanisms in Oncology (GAME-ON) through NCI U19 (CA148112) to T.A. Sellers, K99/R00 (CA184415) to K. Lawrenson and R01 (CA136924) to G.A. Coetzee, a grant from the Ovarian Cancer Research Foundation (258807) to S.A. Gayther, and in part by the Cancer Informatics, Proteomics and the Molecular Genomics Core Facilities at the Moffitt Cancer Center through its NCI CCSG grant (P30-CA76292; T.A. Sellers). M. A. Buckley is an ARCS (Achievement Rewards for College Scientists) fellow and a recipient of the Ruth L. Kirschstein National Research Service Award (F31 CA165528). Details of funding for individual investigators and studies are provided at the end of the manuscript.

Additional funding for individual studies

Funding of the constituent studies was provided by the American Cancer Society (CRTG-00-196-01-CCE); the California Cancer Research Program (00-01389V-20170, N01-CN25403, 2II0200); the Canadian Institutes for Health Research (MOP-86727); Cancer Council Victoria; Cancer Council Queensland; Cancer Council New South Wales; Cancer Council South Australia; Cancer Council Tasmania; Cancer Foundation of Western Australia; the Cancer Institute of New Jersey; Cancer Research UK (C490/A16561, C490/A6187, C490/A10119, C490/A10124, C536/A13086, C536/A6689); the Celma Mastry Ovarian Cancer Foundation; the Danish Cancer Society (94-222-52); the ELAN Program of the University of Erlangen-Nuremberg; the Eve Appeal; the Helsinki University Hospital Research Fund; Helse Vest; Imperial Experimental Cancer Research Centre (C1312/A15589); the Norwegian Cancer Society; the Norwegian Research Council; the Ovarian Cancer Research Fund; Nationaal Kankerplan of Belgium; Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for

Cancer Control from the Ministry of Health Labour and Welfare of Japan; Grants-in-Aid for Scientific Research on Priority Areas and Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Science, Sports, Culture and Technology of Japan; Takeda Science Foundation; the L & S Milken Foundation; the Polish Ministry of Science and Higher Education (4 PO5C 028 14, 2 PO5A 068 27); Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Initiatives Foundation; the Roswell Park Cancer Institute Alliance Foundation; the US National Cancer Institute (K07-CA095666, K07-CA143047, K22-CA138563, N01-CN55424, N01-PC067010, N01-PC035137, P01-CA017054, P01-CA087696, P01-CA87969, P30-CA15083, P50-CA105009, P50- CA136393, P50-CA159981, R01-CA014089, R01-CA016056, R01-CA017054, R01-CA049449, R01-CA050385, R01-CA054419, R01- CA058598, R01-CA058860, R01-CA061107, R01-CA061132, R01-CA063678, R01-CA063682, R01-CA064277, R01-CA067262, R01- CA071766, R01-CA074850, R01-CA076016, R01-CA080742, R01-CA080978, R01-CA083918, R01-CA087538, R01- CA092044, R01-095023, R01-CA106414, R01-CA122443, R01-CA112523, R01-CA114343, R01-CA126841, R01- CA136924, R01-CA149429, R03-CA113148, R03-CA115195, R37-CA070867, R37-CA70867, U01-CA069417, U01- CA071966, UM1 CA186107, UM1 CA176726 and Intramural research funds); the US Army Medical Research and Material Command (DAMD17-98-1- 8659, DAMD17-01-1-0729, DAMD17-02-1-0666, DAMD17-02-1-0669, W81XWH-10-1-0280, W81XWH-07-0449, W81XWH-10-1-02802); the National Health and Medical Research Council of Australia (199600 and 400281); the German Federal Ministry of Education and Research of Germany Programme of Clinical Biomedical Research (01 GB 9401); the German Cancer Research Center (DKFZ); the Minnesota Ovarian Cancer Alliance; the Mayo Foundation; the Fred C. and Katherine B. Andersen Foundation; the Lon V. Smith Foundation (LVS-39420); the Oak Foundation; the OHSU Foundation; the Mermaid I project; the Rudolf-Bartling Foundation; the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge, Imperial College London, University College Hospital

“Womens Health Theme”, the Royal Marsden Hospital; WorkSafeBC, and OvCaRe: British Columbia’s Ovarian Cancer Research Team. This study made use of data generated by the Wellcome Trust Case Control consortium; funding for the project was provided by the Wellcome Trust under award 076113.

Authors' contributions

MAB, NTW, ELG, SAG, PDP, and ANAM conceived the project and designed the experiments. MAB, NTW, GMF, KL, PCL, DH, HCS, RK, AG, SY, AY, and MC performed the experiments. MAB, NTW, DH, HCS, HSN, YAC, SC, GB, MCL, BLF, TRH, HN, GAC, PDP, and ANAM performed the analysis and interpreted the results. MAB, HCS, KL, AG, RSC, and SC contributed to resource building that underpins the work. MAB, NTW, KL, HCS, JF, GAC, CP, HSN, ELG, PDP, TAS, SAG, PDP, and ANAM contributed to the discussion and overall data interpretation. MAB, NTW, and ANAM wrote the paper. Other authors and collaborators (below) provided intellectual input and contributed to sample/clinical information collection for individual studies for association analyses. Contributions were recognized according to OCAC guidelines. All authors approved the final manuscript.

OCAC Collaborators

Y. Ann Chen,[1] Katja KH. Aben,[2,3] Hoda Anton-Culver,[4] Fiona Bruinsma,[5] Elisa V. Bandera,[6] Yukie T. Bean,[7,8] Matthias W. Beckmann,[9] Line Bjorge,[10,11] Natalia Bogdanova,[12] Louise A. Brinton,[13] Angela Brooks-Wilson,[14,15] Clareann H. Bunker,[16] Ralf Butzow,[17,18] Ian G. Campbell,[19,20,21] Karen Carty,[22,23] Jenny Chang-Claude,[24,25] Linda S. Cook,[26] Daniel W. Cramer,[27,28] Julie M. Cunningham,[29] Cezary Cybulski,[30] Agnieszka Dansonka-Mieszkowska,[31] Andreas du Bois,[32,33] Ed Dicks,[34] Jennifer A. Doherty,[35,36] Thilo Dörk,[12] Matthias Dürst,[37] Douglas F. Easton,[38,39] Diana M. Eccles,[40] Robert P. Edwards,[41] Arif B. Ekici,[42] Peter A. Fasching,[10,43] Brooke L. Fridley,[1] Yu-Tang Gao,[44] Aleksandra Gentry-Maharaj,[45] Graham G. Giles,[6,46] Rosalind Glasspool,[23] Marc T. Goodman[47,48] Jacek Gronwald,[30] Philipp Harter,[32,33] Hanis N. Hasmad,[49] Alexander Hein,[9] Michelle A.T. Hildebrandt,[50] Peter Hillemanns,[12] Estrid Hogdall,[51,52] Edwin S. Iversen,[53] Anna Jakubowska,[30] Allan Jensen,[51] Beth Y. Karlan,[54] Joseph L. Kelley,[55] Lambertus A. Kiemeney,[2] Liv C. V. Thomsen,[10,11] Susanne K. Kjaer,[51,56] Jolanta Kupryjanczyk,[31] Diether Lambrechts,[57,58] Sandrina Lambrechts,[59] Alice W. Lee,[60] Jenny Lester,[54] Douglas A. Levine,[61] Dong Liang,[62] Jolanta Lissowska,[63] Karen Lu,[64] Jan Lubinski,[30] Lene Lundvall,[65]

Leon F.A.G. Massuger,[66] Keitaro Matsuo,[67] Valerie McGuire,[68] John R. McLaughlin,[69] Ian McNeish,[23] Usha Menon,[70] Roger L. Milne,[5,45] Francesmary Modugno,[21,71,72] Kirsten B. Moysich,[73] Roberta B. Ness,[74] Heli Nevanlinna,[17] Kunle Odunsi,[75] Sara H. Olson,[76] Irene Orlow,[77] Sandra Orsulic,[54,77] James Paul,[23] Celeste L. Pearce,[78,79] Tanja Pejovic,[7,8] Liisa M. Pelttari,[17] Jennifer B. Permuth,[80] Malcolm C. Pike,[75,78] Elizabeth M. Poole,[81,82] Barry Rosen,[83] Mary Anne Rossing,[36] Anja Rudolph,[24] Iwona K. Rzepecka,[31] Reidun K. Kopperud,[10,11] Xiao-Ou Shu,[84] Yurii B. Shvetsov,[85] Nadeem Siddiqui,[23] Weiva Sieh,[86] Honglin Song,[34] Melissa C. Southey,[20] Beata Spiewankiewicz,[87] Lara Sucheston-Campbell,[88] Soo-Hwang Teo,[49,89] Kathryn L. Terry,[27,82] Pamela J. Thompson,[47,48] Lotte Thomsen,[90] Katharina Bischof,[10,11] Ya-Yu Tsai,[80] Shelley S. Tworoger,[80,81,82] Anne M. van Altena,[66] Ignace Vergote,[59] Christine S. Walsh,[54] Shan Wang-Gohrke,[91] Nicolas Wentzensen,[13] Alice S. Whittemore,[92] Lynne R. Wilkens,[85] Anna H. Wu,[78] Xifeng Wu,[50] Yin-Ling Woo,[93] Hannah Yang,[13] Wei Zheng,[94] Argyrios Ziogas,[4] Andrew Berchuck,[95] Georgia Chenevix-Trench,[96] on behalf of the AOCS management group,[96,97] Joellen M. Schildkraut,[98] Linda E. Kelemen,[99] Matthew L. Freedman,[100] .

[1]Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA
[2]Department for Health Evidence, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands
[3]Netherlands Comprehensive Cancer Center Organization, Utrecht, The Netherlands
[4]Genetic Epidemiology Research Institute, UCI Center for Cancer Genetics Research and Prevention, School of Medicine, Department of Epidemiology, University of California Irvine, Irvine, CA, USA
[5]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia
[6]Cancer Prevention and Control, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA
[7]Department of Obstetrics & Gynecology, Oregon Health & Science University, Portland, OR, USA
[8]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA
[9]Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander-University, Erlangen Nuremberg Comprehensive Cancer Center, Erlanegen EMN, Germany.
[10]Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway
[11]Centre for Cancer Biomarkers, Department of Clinical Medicine, University of Bergen, Bergen, Norway
[12]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
[13]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
[14]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada.
[15]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC Canada.
[16]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA
[17]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, HUS, Finland
[18]Department of Pathology, Helsinki University Central Hospital, Helsinki, HUS, Finland.
[19]Cancer Genetics Laboratory, Research Division, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Australia
[20]Department of Pathology, University of Melbourne, Parkville, Victoria, Australia
[21]Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria, Australia
[22]Department of Gynaecological Oncology, Glasgow Royal Infirmary, Glasgow, G31 2ER, UK
[23]CRUK Clinical Trials Unit, The Beatson West of Scotland Cancer Centre, 1053 Great Western Road, Glasgow G12 0YN, UK

[24]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[24]University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg Germany
[26]Division of Epidemiology and Biostatistics, Department of Internal Medicine, University of New Mexico, Albuquerque, NM, USA
[27]Obstetrics and Gynecology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[28]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA
[29]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
[30]International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.
[31]Department of Pathology, The Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland
[32]Department of Gynaecology and Gynaecologic Oncology, Kliniken Essen-Mitte, Essen, Germany
[33]Department of Gynaecology and Gynaecologic Oncology, Dr. Horst Schmidt Kliniken Wiesbaden, Wiesbaden, Germany
[34]Department of Oncology, The Centre for Cancer Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK
[35]Department of Community and Family Medicine, Section of Biostatistics & Epidemiology, Dartmouth Medical School, Hanover, NH, USA
[36]Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA
[37]Department of Gynecology, Jena University Hospital, Friedrich Schiller University Jena, Jena, Germany.
[38]The Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
[39]Department of Oncology, The Centre for Cancer Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK
[40]Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK
[41]Department of Obstetrics Gynecology/RS, Division of Gynecological Oncology, University of Pittsburgh, Pittsburgh, PA, USA
[42]Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.
[43]University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, Los Angeles, CA, USA
[44]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China
[45]Women's Cancer, UCL EGA Institute for Women's Health, London, UK
[46]Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[47]Cancer Prevention and Control, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[48]Community and Population Health Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[49]Cancer Research Initiatives Foundation, Sime Darby Medical Center, Subang Jaya, Malaysia
[50]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[51]Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark
[52]Molecular Unit, Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark
[53]Department of Statistics, Duke University, Durham, NC, USA
[54]Women's Cancer Program at the Samuel Oschin Comprehensive, Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[55]Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[57]Vesalius Research Center, VIB, University of Leuven, Leuven, Belgium
[58]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Belgium.

[59]Division of Gynecologic Oncology; Leuven Cancer Institute, University Hospitals Leuven, KU Leuven, Leuven, Belgium
[60]Department of Health Science, California State University, Fullerton, Fullerton, CA 92831, USA
[61]NYU Langone Medical Center, New York University, New York, NY, USA
[62]College of Pharmacy and Health Sciences, Texas Southern University, Houston, TX, USA
[63]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland
[64]Department of Gynecologic Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[65]Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
[66]Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands
[67]Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan
[68]Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA, USA
[69]Public Health Ontario, Toronto, ON, Canada
[70]MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, University College London, UK
[71]Women's Cancer Research Program, Magee-Women's Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA
[72]Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[73]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA
[74]The University of Texas School of Public Health, Houston, TX, USA
[75]Department of Gynecologic Oncology, Roswell Park Cancer Institute, Buffalo, NY.
[76]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
[77]Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA
[78]Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA 90089, USA
[79]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA
[80]Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA
[81]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[82]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[83]Department of Gynecology-Oncology, Princess Margaret Hospital, and Department of Obstetrics and Gynecology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[84]Epidemiology Center and Vanderbilt, Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
[85]Cancer Epidemiology Program, University of Hawaii Cancer Center, Hawaii, USA
[86]Department of Population Health Science and Policy, and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[87]Department of Gynecologic Oncology, Institute of Oncology, Warsaw, Poland
[88]College of Pharmacy, The Ohio State University, Columbus, OH, USA
[89]University Malaya Medical Centre, University of Malaya, Kuala Lumpur, Maylaysia.
[90]Department of Pathology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
[91]Molecular Biology Laboratory, Department of Obstetrics and Gynecology, University of Ulm, Ulm, Germany
[92]Department of Health Research and Policy- Epidemiology, Stanford University School of Medicine, Stanford, CA, USA
[93]Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia
[94]Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA
[95]Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA
[96]QIMR Berghofer Medical Research Institute, Brisbane, Australia

[97]Peter MacCallum Cancer Centre, East Melbourne, Australia
[98]Department of Public Health Sciences, University of Virginia, Charlottesville, VA USA
[99]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA
[100]Department of Medical Oncology, The Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA

REFERENCES

1. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nature reviews Cancer. 2004;4:850-60.

2. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nature genetics. 2009;41:996-1000.

3. Shen H, Fridley BL, Song H, Lawrenson K, Cunningham JM, Ramus SJ, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nature communications. 2013;4:1628.

4. Permuth-Wey J, Lawrenson K, Shen HC, Velkova A, Tyrer JP, Chen Z, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nature communications. 2013;4:1627.

5. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics. 2013;45:371-84.

6. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nature genetics. 2013;45:362-70.

7. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nature genetics. 2010;42:880-4.

8. Chen K, Ma H, Li L, Zang R, Wang C, Song F, et al. Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nature communications. 2014;5:4682.

9. Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nature genetics. 2015;47:164-71.

10. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature genetics. 2010;42:874-9.

11. Kelemen LE, Lawrenson K, Tyrer J, Li Q, Lee JM, Seo JH, et al. Genome-wide significant risk associations for mucinous ovarian carcinoma. Nature genetics. 2015;47:888-97.

12. Kar SP, Beesley J, Amin Al Olama A, Michailidou K, Tyrer J, Kote-Jarai Z, et al. Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types. Cancer discovery. 2016;6:1052-67.

13. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics. 2017;49:680-91.

14. Lawrenson K, Benjamin E, Turmaine M, Jacobs I, Gayther S, Dafou D. In vitro three-dimensional modelling of human ovarian surface epithelial cells. Cell Proliferation. 2009.

15. Lawrenson K, Grun B, Benjamin E, Jacobs IJ, Dafou D, Gayther SA. Senescent fibroblasts promote neoplastic transformation of partially transformed ovarian epithelial cells in a three-dimensional model of early stage ovarian cancer. Neoplasia. 2010;12:317-25.

16. Coetzee SG, Shen HC, Hazelett DJ, Lawrenson K, Kuchenbaecker K, Tyrer J, et al. Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. Human molecular genetics. 2015;24:3595-607.

17. Buckley M, Gjyshi A, Mendoza-Fandino G, Baskin R, Carvalho RS, Carvalho MA, et al. Enhancer scanning to locate regulatory regions in genomic loci. Nature protocols. 2016;11:46-60.

18. Baskin R, Woods NT, Mendoza-Fandino G, Forsyth P, Egan KM, Monteiro AN. Functional analysis of the 11q23.3 glioma susceptibility locus implicates PHLDB1 and DDX6 in glioma susceptibility. Sci Rep. 2015;5:17367.

19. Keaton MA, Taylor CM, Layer RM, Dutta A. Nuclear scaffold attachment sites within ENCODE regions associate with actively transcribed genes. PloS one. 2011;6:e17912.

20. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics. 2009;41:882-4.

21. Lam KN, van Bakel H, Cote AG, van der Ven A, Hughes TR. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. Nucleic acids research. 2011;39:4680-90.

22. Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nature protocols. 2009;4:393-411.

23. Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. Genes & development. 2006;20:601-12.

24. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008;9:R137.

25. 1000 Genomes Project Consortium; Abecasis GRA, D.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Gibbs, R.A.; Hurles, M.E.; McVean, G.A. . A map of human genome variation from population-scale sequencing. Nature. 2010;467:1061-73.

26. The International HapMap P. The International HapMap Project. Nature. 2003;426:789-96.

27. Permuth-Wey J, Kim D, Tsai YY, Lin HY, Chen YA, Barnholtz-Sloan J, et al. LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer research. 2011;71:3896-903.

28. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009;5:e1000529.

29. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nature genetics. 2011;43:513-8.

30. Vanhoutteghem A, Djian P. Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. Proceedings of the National Academy of Sciences of the United States of America. 2004;101:3468-73.

31. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nature biotechnology. 2015;33:555-62.

32. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol. 2006;24:1429-35.

33. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nature protocols. 2013;8:1551-66.

34. Linnemann AK, Platts AE, Krawetz SA. Differential nuclear scaffold/matrix attachment marks expressed genes. Human molecular genetics. 2009;18:645-54.

35. Kisseljova NP, Dmitriev P, Katargin A, Kim E, Ezerina D, Markozashvili D, et al. DNA polymorphism and epigenetic marks modulate the affinity of a scaffold/matrix attachment region to the nuclear matrix. European journal of human genetics : EJHG. 2014;22:1117-23.

36. Cesaratto L, Grisard E, Coan M, Zandona L, De Mattia E, Poletto E, et al. BNC2 is a putative tumor suppressor gene in high-grade serous ovarian carcinoma and impacts cell survival after oxidative stress. Cell death & disease. 2016;7:e2374.

37. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155:934-47.

38. Lang MR, Patterson LB, Gordon TN, Johnson SL, Parichy DM. Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. PLoS genetics. 2009;5:e1000744.

39. Vanhoutteghem A, Delhomme B, Herve F, Nondier I, Petit JM, Araki M, et al. The importance of basonuclin 2 in adult mice and its relation to basonuclin 1. Mechanisms of development. 2016;140:53-73.

40. Wentzensen N, Black A, Jacobs K, Yang HP, Berg CD, Caporaso N, et al. Genetic variation on 9p22 is associated with abnormal ovarian ultrasound results in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. PloS one. 2011;6:e21731.

41. Kar SP, Tyrer JP, Li Q, Lawrenson K, Aben KK, Anton-Culver H, et al. Network-Based Integration of GWAS and Gene Expression Identifies a HOX-Centric Network Associated with Serous Ovarian Cancer Risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2015;24:1574-84.

42. Carter H, Marty R, Hofree M, Gross AM, Jensen J, Fisch KM, et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. Cancer discovery. 2017;7:410-23.

43. Jacobs LC, Wollstein A, Lao O, Hofman A, Klaver CC, Uitterlinden AG, et al. Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. Human genetics. 2013;132:147-58.

44. Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM, et al. Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. BMC evolutionary biology. 2013;13:150.

45. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS genetics. 2010;6:e1000993.

46. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics. 2014;46:1173-86.

47. Visser M, Palstra RJ, Kayser M. Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. Human molecular genetics. 2014.

48. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science. 2014;343:1017-21.

49. Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507:354-7.

50. Vigorito E, Kuchenbaecker KB, Beesley J, Adlard J, Agnarsson BA, Andrulis IL, et al. Fine-Scale Mapping at 9p22.2 Identifies Candidate Causal Variants That Modify Ovarian Cancer Risk in BRCA1 and BRCA2 Mutation Carriers. PloS one. 2016;11:e0158801.

51. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444:499-502.

Table 1: Twenty-two SNPs correlated with rs3814113 overlap with areas of regulatory activity.

Region	chr9 Coordinates Tile	SNP Name	Effect Allele	Reference Allele	R^2	MAF	P value in Song et al.
1	16,837,392-16,838,723						
2	16,848,158-16,848,790						
3	16,850,432-16,851,014						
4	16,852,717-16,853,479						
5	16,857,377-16,857,907 T5	B	A	C	0.719	A=0.3904	
6	16,860,790-16,861,348 T6	rs62541878	T	A	0.3	T=0.0513	
7	16,863,768-16,874,127 T7.1	rs11792249	G	T	0.3	G=0.0513	
		rs2153271	T	C	0.539	T=0.2879	$4.66x10^{-10}$
	T7.2	**rs62541920**	**A**	**G**	**0.3**	**A=0.0511**	
		rs12379183	**G**	**A**	**0.445**	**G=0.2462**	**$1.36x10^{-10}$**
	T7.3	rs10962647	G	T	0.3	G=0.0515	
	T7.4 & T7.5	rs10962648	C	G	0.3	C=0.0515	
	T7.6	rs62541922	C	T	0.317	C=0.0487	
		rs62541923	A	C	0.3	A=0.0507	
	T7.7	rs11789875	A	G	0.3	A=0.0489	
	T7.8	rs10962649	T	C	0.3	T=0.0489	
		rs10810650	T	C	0.589	T=0.2963	
8	16,883,570-16,885,692 T8	rs10810657	A	T	0.528	A=0.2915	
		rs12350739	**A**	**G**	**0.508**	**A=0.1875**	
		rs77507622	**G**	**A**	**0.3**	**G=0.0493**	
9	16,899,790-16,900,338						
10	16,901,238-16,902,039						
11	16,907,559-16,908,180 T11	rs113780397	A	G	0.818	A=0.4395	
		rs9697099	A	T	0.301	T=0.4814	
		<u>rs181552334</u>	**<u>G</u>**	**<u>A</u>**	**<u>0.527</u>**	**<u>G=0.4395</u>**	
		rs76718132	T	C	0.379	NA	
		rs117224476	G	T	0.44	NA	
		rs77795022	G	T	0.442	NA	
12	16,915,387-16,915,739						

LD (r2 ≥ 0.3) to rs3814113 based on 1000GP data v3.

MAF: minor allele frequency; rs2153271 are reported in dbSNP as the reverse orientation to the genome; NA, not available. SNPs in bold represent the final five SNPs remaining at the end of the functional analysis; SNP shown in bold and underlined indicates the only SNP that is common to the two analytical approaches.

Table 2: Validation of ChIP-seq Data by Nanostring

Gene	TSS to Peak Center	P-Value	Expression Correlation to BNC2	PBM	Notes
FAM49B	21553	0.000147821	-	1,2;5,6	Ovarian cancer
ITGB5	8074	0.000463581	-	1,2;5,6	Focal Adhesion
JUN	20445	0.00626649	+	1,2; 5,6	Focal adhesion, WNT and MAPK signaling
TGFBR3	5215|-14508	0.0390608	+	5,6	TGF-beta Signaling Pathway
CCND3	-25565	0.00106617	+	1,2;5,6	Focal adhesion, WNT and MAPK signaling
CEP55	-24627|24647	0.002645	+	1,2;5,6	Ovarian Development
FEM1A	-65	0.0206431	-	1,2; 5,6	Promoter with Peak

Expression correlation to BNC2: indicates whether the expression of the target gene is positively (upregulated) (+) or negatively (downregulated) (-) correlated with the overexpression of *BNC2*.

FIGURE LEGENDS

Figure 1: Rules guiding SNP selection and prioritization of causal SNPs at the 9q22 locus using a two-pronged strategy. Functional dissection guided by Linkage Disequilibrium with the most significantly associated risk SNP (left flowchart) identified five SNPs in three regulatory elements, an enhancer to BNC2, the BNC2 promoter, and a Substrate/Matrix attachment region (S/MAR). Analysis guided by fine mapping data (right flowchart) points to the S/MAR as the region with the SNPs most highly associated with risk (in bold red) which was also identified in the functional analysis. Conditional analysis (adjusted for rs3814113; green font) revealed independent signals at the locus.

Figure 2: Candidate functional SNPs overlapping with regions of regulatory activity in ovarian cells. *A*. Within the region of the 9p22 locus containing linked SNPs, twelve regions contain FAIRE peaks (gray bars), H3K27Ac peaks (orange bars), and/or H3K4Me1 peaks (maroon bars) in iOSE and iFTSEC cells. Some regulatory regions do not overlap with candidate SNPs (yellow highlight). Regions highlighted in red overlap with candidate functional SNPs (thin blue bars). Numbered blue bars represent the location of 2 kb tiles cloned into luciferase reporter vectors. ***B*.** Box and whisker plots showing the luciferase activity from duplicate experiments with eight biological replicates of each tile in both orientations. Asterisks denote tiles exhibiting significant transcription activity compared to a control tile (C) located in a genomic region inactive in ovarian cells as judged by features in the figure. Tiles moving forward in the functional assays are colored red. ***C** and **D.*** Luciferase assays reveal significant allele-specific differences in transcription activation for rs62541878, rs62541920, rs12379183, rs1092647, rs10810657, rs12350739, and rs77507622, as indicated by red boxes and asterisks in forward (***C***) or reverse orientation (***D***). Reference and effect allele tiles are shown in blue and red fonts, respectively. ***E*.** EMSA showing allele-specific differences in mobility between the reference and effect alleles. SNPs in Regions 7 (rs12379183 and rs6251920), and 8 (rs12350739 and rs77507622) display differences in complex formation between the reference and effect alleles. SNPs with allele-specific differences are indicated by red text.

Figure 3: Region 11 is attached to the nuclear scaffold in ovarian cells. *A-C.* Genomic DNA (total or attached to the nuclear scaffold) was extracted from ovarian iOSE11 and HeLa cells. For each region the ratio of scaffold-attached to total DNA is depicted. Significance was defined by a Z score ≥21 (Z score = (average of scaffold attached DNA – average of negative control)/standard deviation of negative control). R11, Region 11; ApoB, ApoliproteinB gene, used as positive control in HeLa cells. ***D.*** HiC (High dimensional chromosome conformation capture) interaction frequency data from cell lines obtained from the Yue lab HiC browser (http://promoter.bx.psu.edu/hi-c/view.php). Dashed line indicates the position of the S/MAR. ***E.*** Depiction of the location of the topologically associating domain (TAD) inferred from the interaction data in ***D*** and the relative positions of the S/MAR and the other two genes located in the TAD. ***F.*** A snapshot from the genome browser displays UCSC genes as well as FAIRE peaks (gray), H3K27Ac peaks (orange), and/or H3K4Me1 peaks (maroon) in iOSE, iFTSE, and ovarian cancer cells generated in the laboratory. The four genes within the region considered as potential target genes for ovarian cancer susceptibility include *c9orf92*, *BNC2*, *CNTLN*, and

SH3GL2. ENCODE H3K4me3 peaks (purple), used to identify the promoters of these four genes (highlighted in yellow). H3K27ac tracks (orange) inform the extent to which these promoters are active and show that *BNC2* and *CNTLN* promoters are active in ovarian cells while *c9orf92* and *SH3GL2* are not. ***G.*** RNA-Seq for these four genes indicates the presence of transcripts for *BNC2* and *CNTLN* but not for *SH3GL2* and *c9orf92*. ***H-I.*** Chromosome conformation capture (3C) analysis indicates that Region 8 interacts with the *BNC2* promoter (***H)*** while region 11 (right) does not show a significant interaction compared to the adjacent site (***I***). Anchor regions for 3C are highlighted in red. Red dashed line indicates the interaction to adjacent probes. Each graph is aligned with chromatin mark and transcript information from the genome browser. Regions containing SNPs are indicated by blue boxes. Blue arches depict the interactions.

Figure 4: Fine mapping of the interval at 9p22 (chr9: 16407967-17407967) locus in 15,437 women diagnosed with invasive EOC and 30,845 controls. Plotted using LocusZoom (http://locuszoom.sph.umich.edu/locuszoom). **A.** rs3814113 (the most significant SNP in the original analysis (Song et al.) is shown as a purple diamond as remains as the most significant association in fine mapping analysis for serous ovarian cancer. SNPs are colored according to LD to rs3814113. **B.** Conditional analysis adjusting for rs3814113.

Figure 5: BNC2 recognizes specific nucleotide sequence. *A*. BNC2 is characterized as a C2H2 zinc finger protein with three pairs of ZFs (called 1,2; 3,4; 5,6). BNC2 Zinc Finger binding sites were identified in vitro by applying recombinant proteins of each ZF pair to a protein binding microarray. Position weight matrices of all potential binding sites with significant scores for each BNC2 ZF pair are shown as logos. Motifs predicted based on the protein sequence of the ZF domains aligned with ZF1,2 and ZF5,6. The 3' end of the sequences recognized by ZF1,2 and ZF5,6 reveal the same nucleotides. Inspection of the amino acid sequences for ZF2 and ZF6 show that amino acid residues at position -1, 2, 3, 6, and 10 within the alpha helix that specifically interact with DNA nucleotides (in red) are the same. **B.** The ChIP-Seq motif identified by MEME seems to be a concatenation of the predicted motif for ZF1,2 and the predicted reverse complement motif for ZF 5,6. **C.** Enrichment of motif relative to ChIP-Seq peak summits.

Figure 1

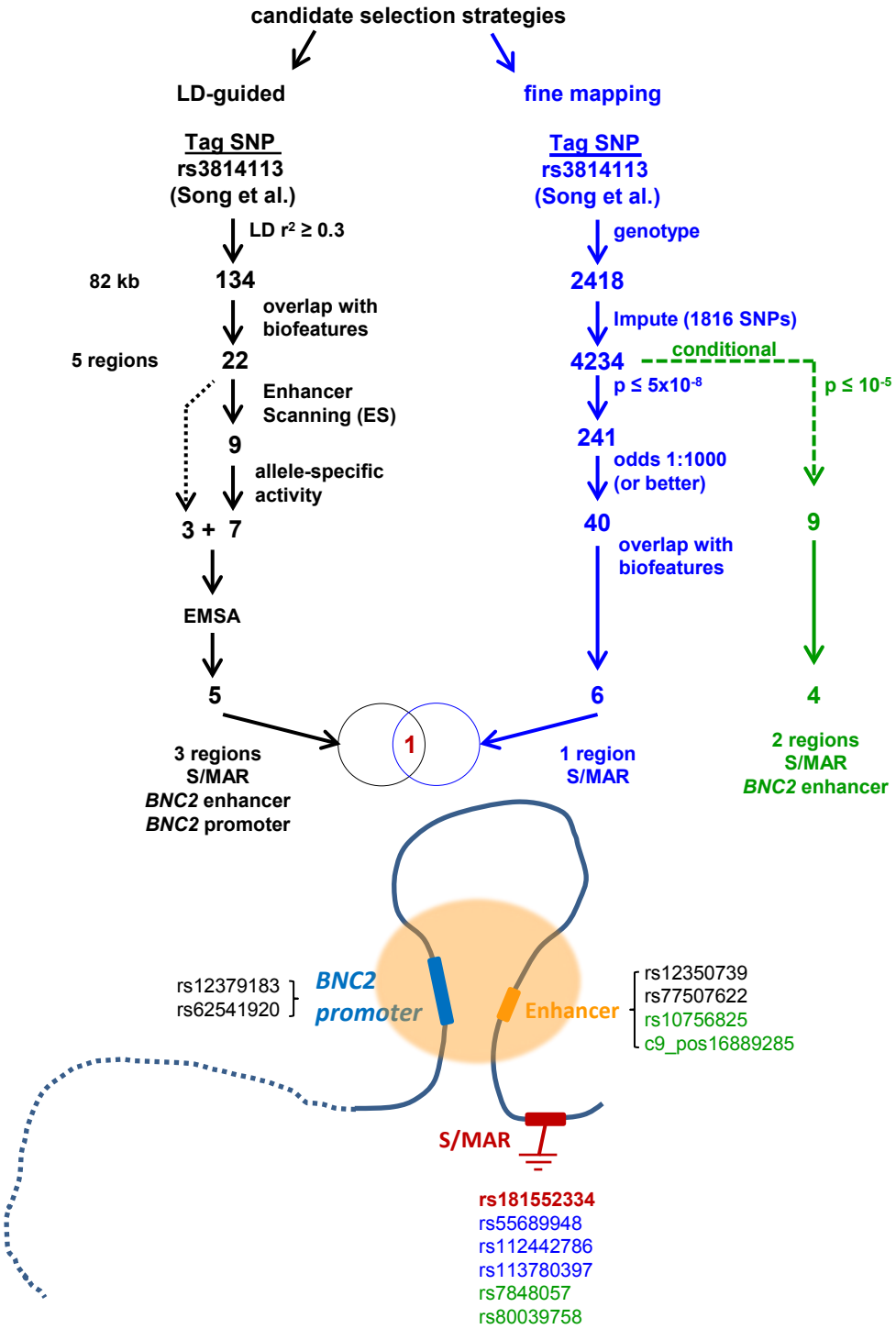

candidate selection strategies
LD-guided
fine mapping
Tag SNP
rs3814113
(Song et al.)
LD r² ≥ 0.3
82 kb
134
overlap with
biofeatures
5 regions
22
Enhancer
Scanning (ES)
9
allele-specific
activity
3 + 7
EMSA
5
3 regions
S/MAR
BNC2 enhancer
BNC2 promoter
Tag SNP
rs3814113
(Song et al.)
genotype
2418
Impute (1816 SNPs)
4234
conditional
p ≤ 5x10-8
p ≤ 10-5
241
odds 1:1000
(or better)
40
overlap with
biofeatures
9
6
4
1
1 region
S/MAR
2 regions
S/MAR
BNC2 enhancer
rs12379183
rs62541920
BNC2
promoter
Enhancer
rs12350739
rs77507622
rs10756825
c9_pos16889285
S/MAR
rs181552334
rs55689948
rs112442786
rs113780397
rs7848057
rs80039758

Figure 2

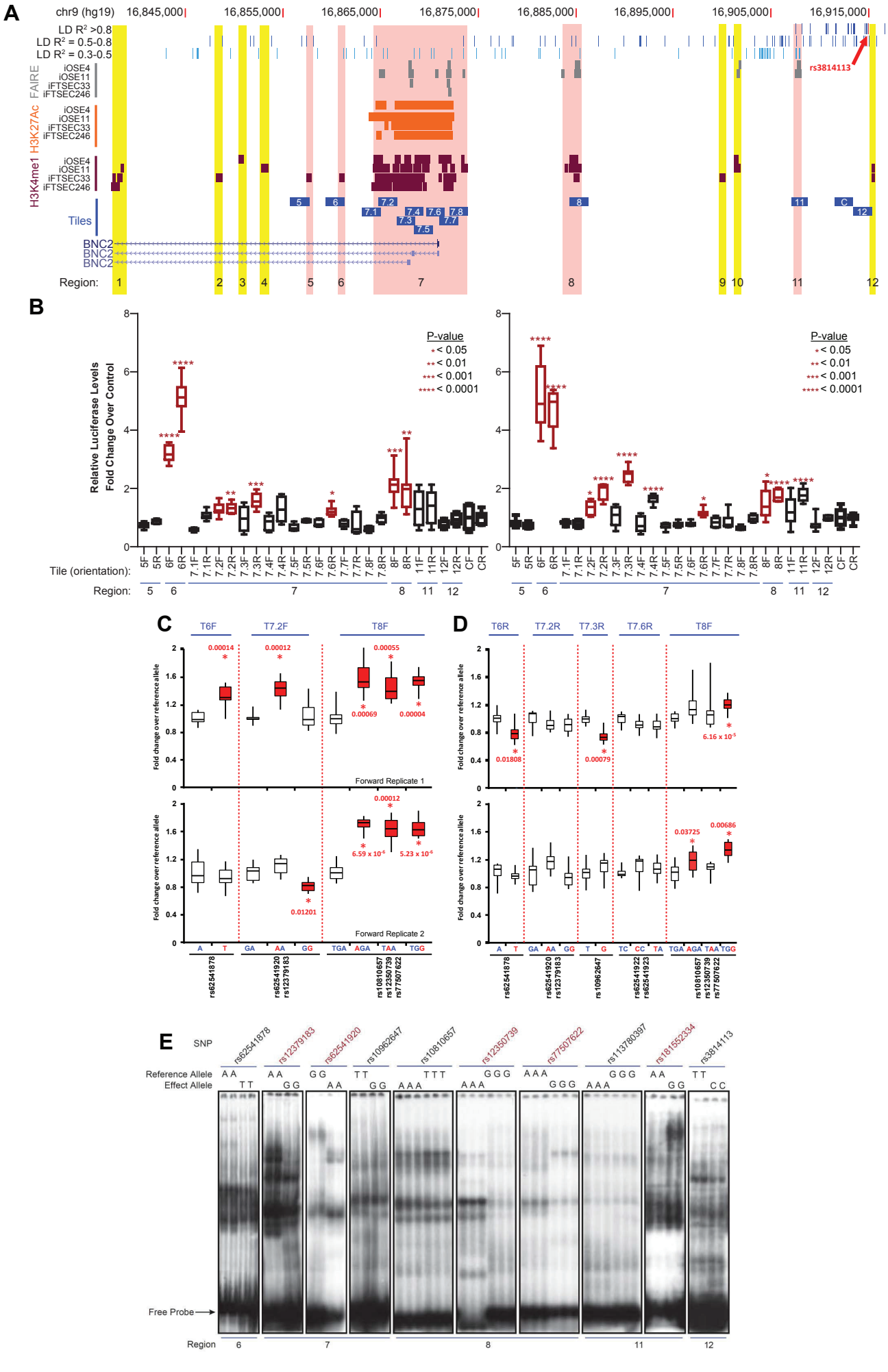

A
chr9 (hg19)
16,845,000
16,855,000
16,865,000
16,875,000
16,885,000
16,895,000
16,905,000
16,915,000
LD R² >0.8
LD R² = 0.5-0.8
LD R² = 0.3-0.5
rs3814113
FAIRE
iOSE4
iOSE11
iFTSEC33
iFTSEC246
H3K27Ac
H3K4me1
Tiles
BNC2
Region:
B
Relative Luciferase Levels
Fold Change Over Control
P-value
* < 0.05
** < 0.01
*** < 0.001
**** < 0.0001
Tile (orientation):
5F 5R 6F 6R 7.1F 7.1R 7.2F 7.2R 7.3F 7.3R 7.4F 7.4R 7.5F 7.5R 7.6F 7.6R 7.7F 7.7R 7.8F 7.8R 8F 8R 11F 11R 12F 12R CF CR
Region: 5 6 7 8 11 12
C
T6F
T7.2F
T8F
Fold change over reference allele
0.00014
0.00012
0.00055
0.00069
0.00004
Forward Replicate 1
0.00012
6.59 x 10⁻⁶
5.23 x 10⁻⁶
0.01201
Forward Replicate 2
A T GA AA GG TGA AGA TAA TGG
rs62541878
rs62541920
rs12379183
rs10810657
rs12350739
rs77507622
D
T6R
T7.2R
T7.3R
T7.6R
T8F
0.01808
0.00079
6.16 x 10⁻⁵
0.03725
0.00686
A T GA AA GG T G TC CC TA TGA AGA TAA TGG
rs62541878
rs62541920
rs12379183
rs10962647
rs62541922
rs62541923
rs10810657
rs12350739
rs77507622
E
SNP
rs62541878
rs12379183
rs62541920
rs10962647
rs10810657
rs12350739
rs77507622
rs113780397
rs181552334
rs3814113
Reference Allele
Effect Allele
A A T T
A A G G
G G A A
T T G G
T T T A A A
G G G A A A
A A A G G G
G G G A A A
A A G G
T T C C
Free Probe
Region
6 7 8 11 12

Figure 3

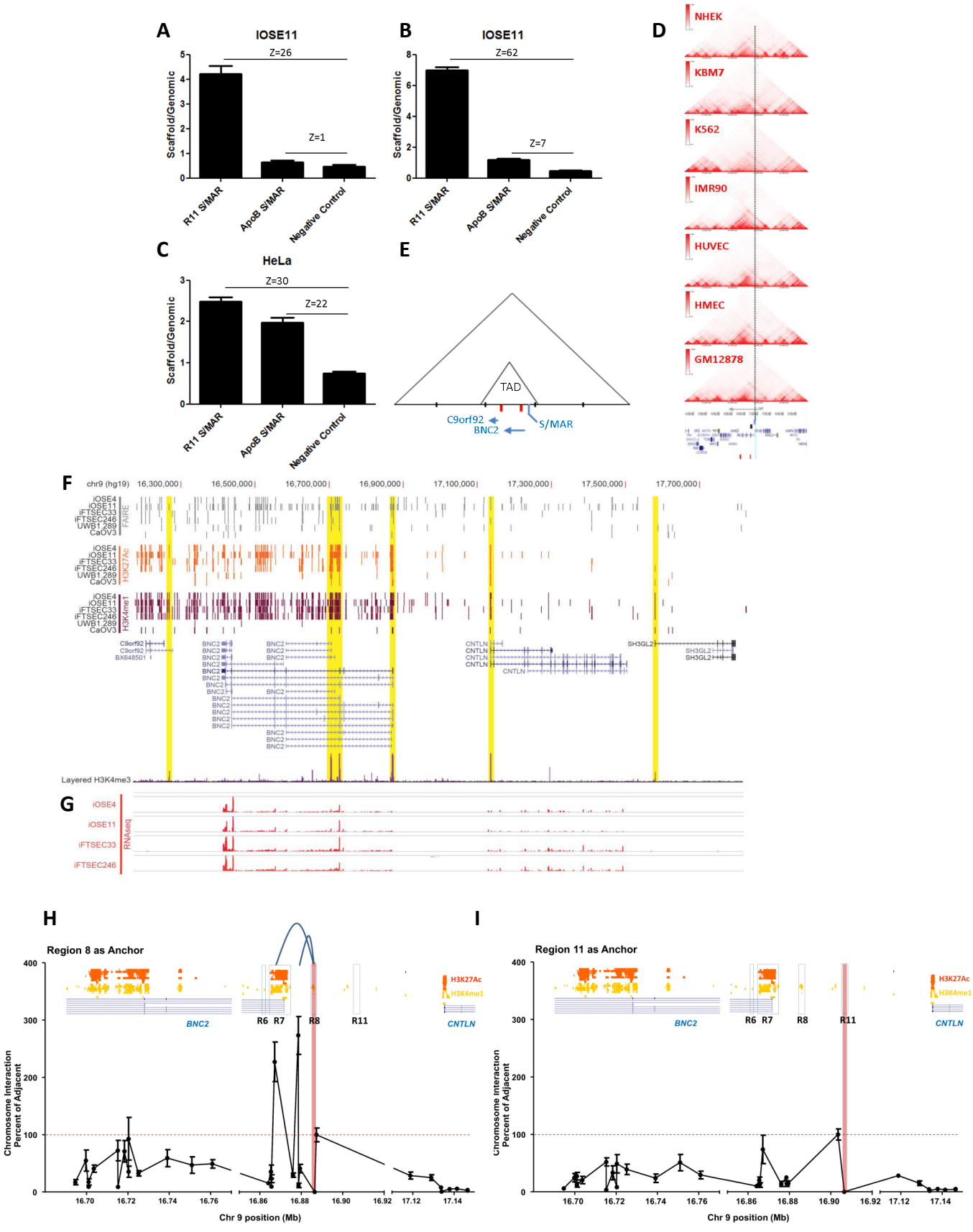

A
IOSE11
Z=26
Z=1
Scaffold/Genomic
R11 S/MAR
ApoB S/MAR
Negative Control
B
IOSE11
Z=62
Z=7
C
HeLa
Z=30
Z=22
D
NHEK
KBM7
K562
IMR90
HUVEC
HMEC
GM12878
E
TAD
C9orf92
BNC2
S/MAR
F
chr9 (hg19)
16,300,000
16,500,000
16,700,000
16,900,000
17,100,000
17,300,000
17,500,000
17,700,000
FAIRE
H3K27Ac
H3K4me1
Layered H3K4me3
G
RNAseq
H
Region 8 as Anchor
BNC2
R6 R7 R8 R11
CNTLN
Chromosome Interaction Percent of Adjacent
Chr 9 position (Mb)
I
Region 11 as Anchor

Figure 4

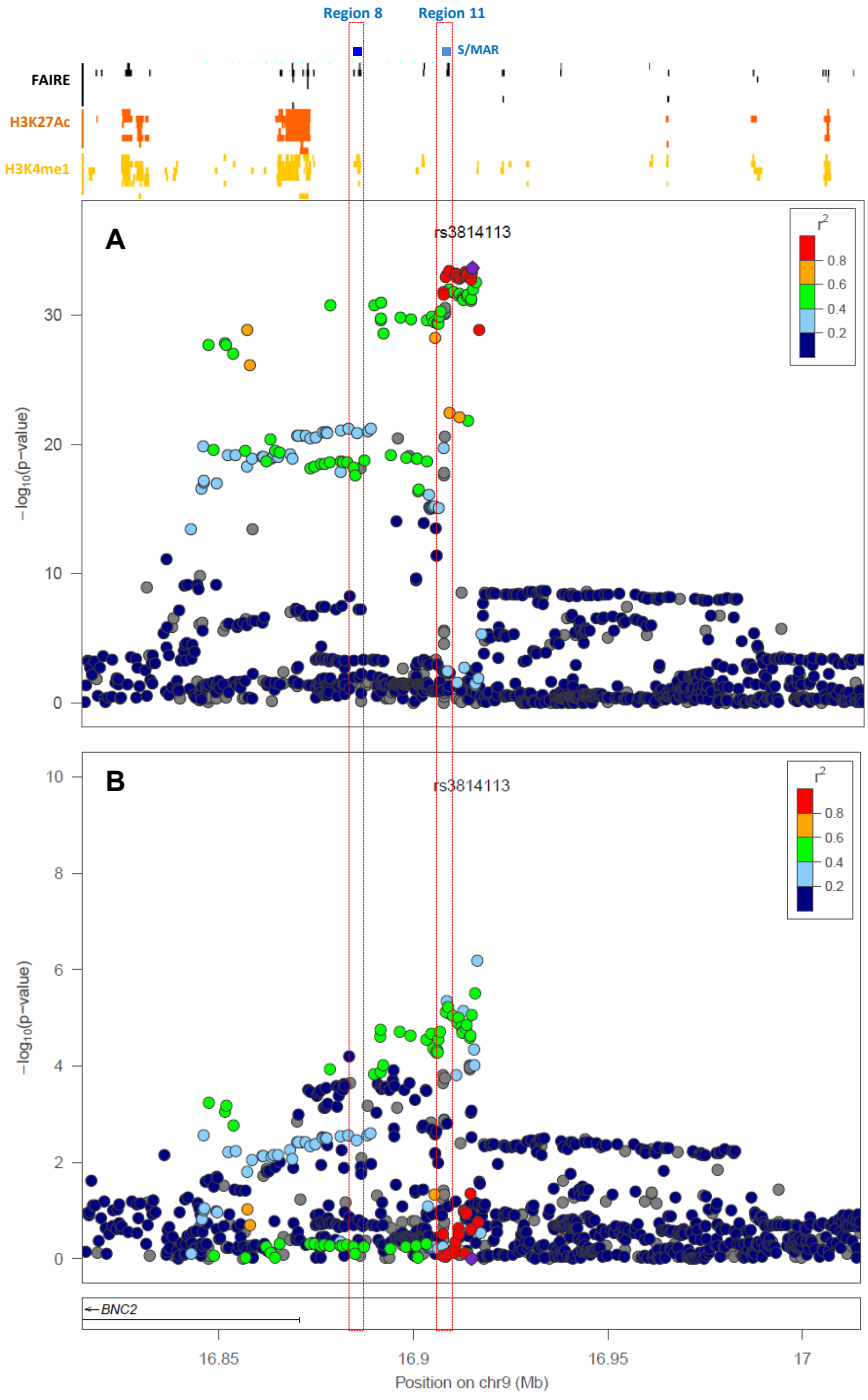
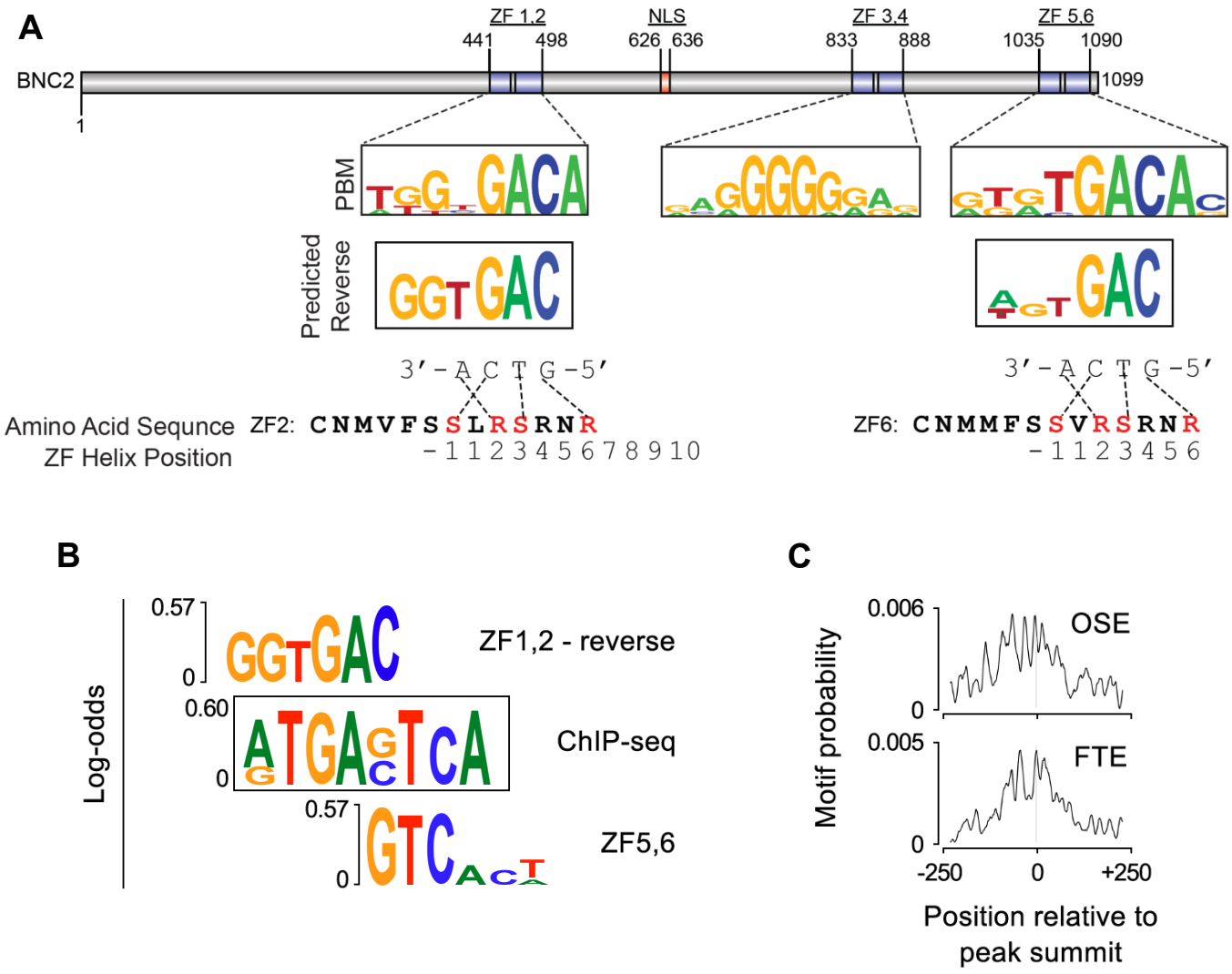

Region 8
Region 11
S/MAR
FAIRE
H3K27Ac
H3K4me1
A
rs3814113
r²
0.8
0.6
0.4
0.2
−log10(p−value)
30
20
10
0
B
rs3814113
r²
0.8
0.6
0.4
0.2
−log10(p−value)
10
8
6
4
2
0
←BNC2
16.85
16.9
16.95
17
Position on chr9 (Mb)

Figure 5


A
ZF 1,2
441
498
NLS
626
636
ZF 3,4
833
888
ZF 5,6
1035
1090
BNC2
1
1099
PBM
Predicted Reverse
3′-A C T G-5′
3′-A C T G-5′
Amino Acid Sequnce
ZF Helix Position
ZF2: C N M V F S S L R S R N R
-1 1 2 3 4 5 6 7 8 9 10
ZF6: C N M M F S S V R S R N R
-1 1 2 3 4 5 6
B
Log-odds
0.57
0
0.60
0
0.57
0
ZF1,2 - reverse
ChIP-seq
ZF5,6
C
Motif probability
0.006
0
OSE
0.005
0
FTE
-250
0
+250
Position relative to peak summit

Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus

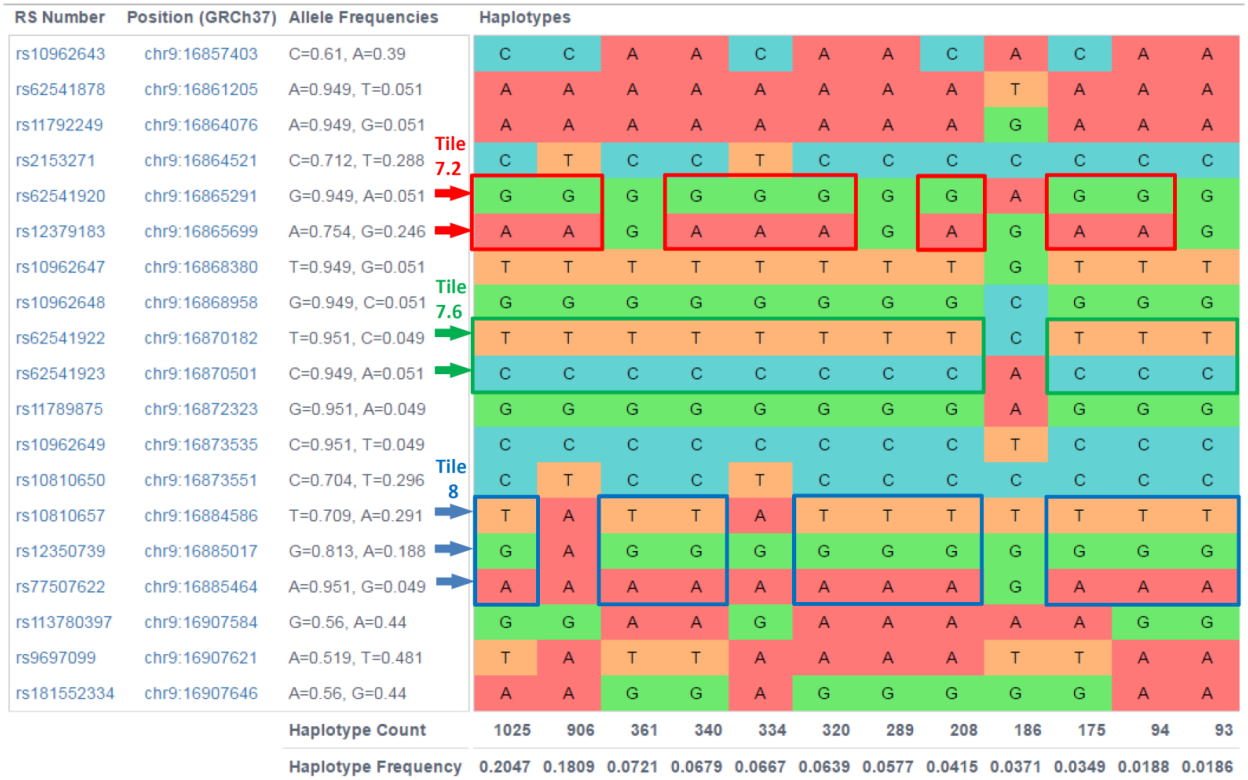
Melissa A. Buckley et al.

This file contains:

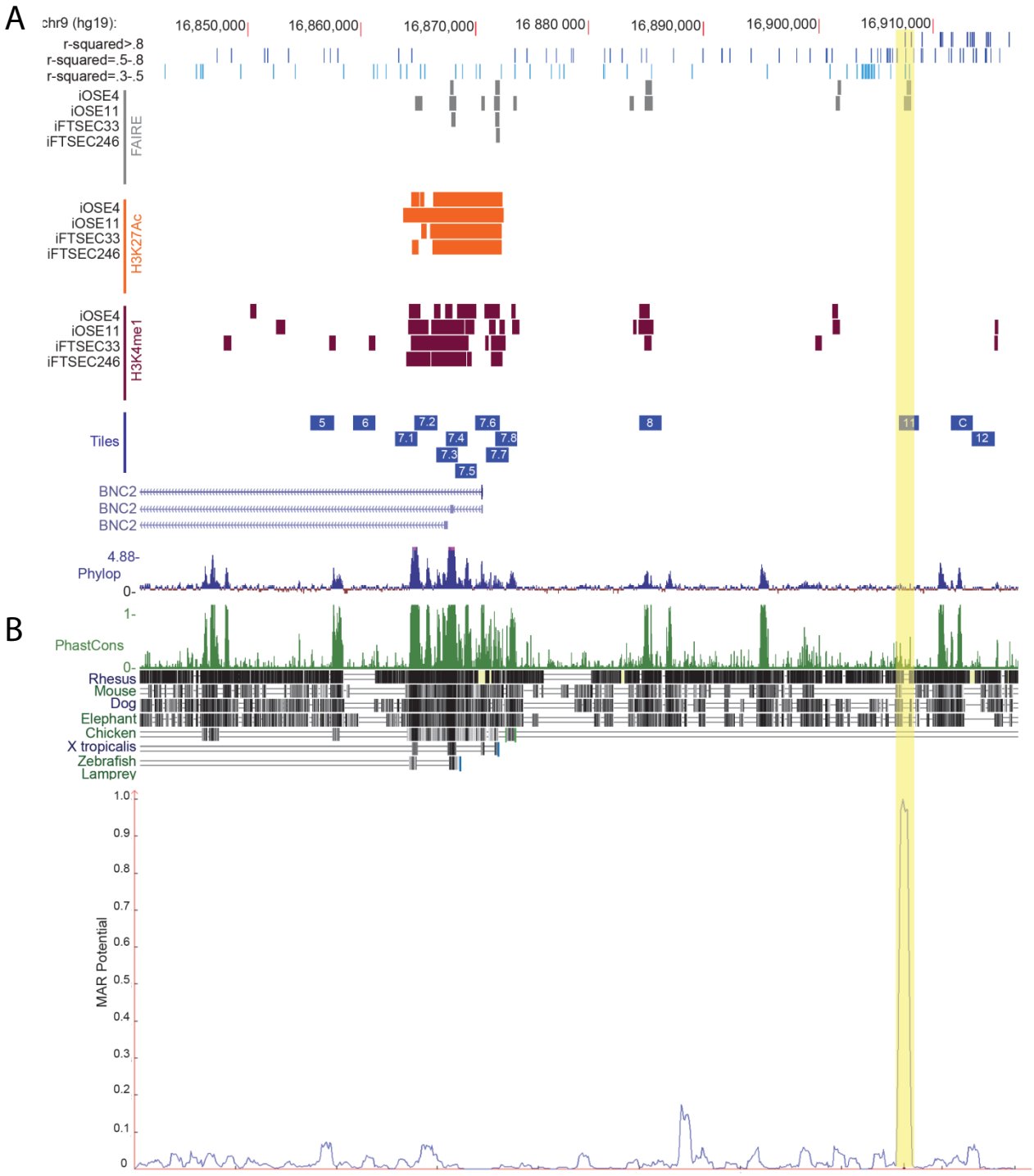
SUPPLEMENTARY FIGURES

EXTENDED DATA AND METHODS (BNC2 ChIP-Seq)

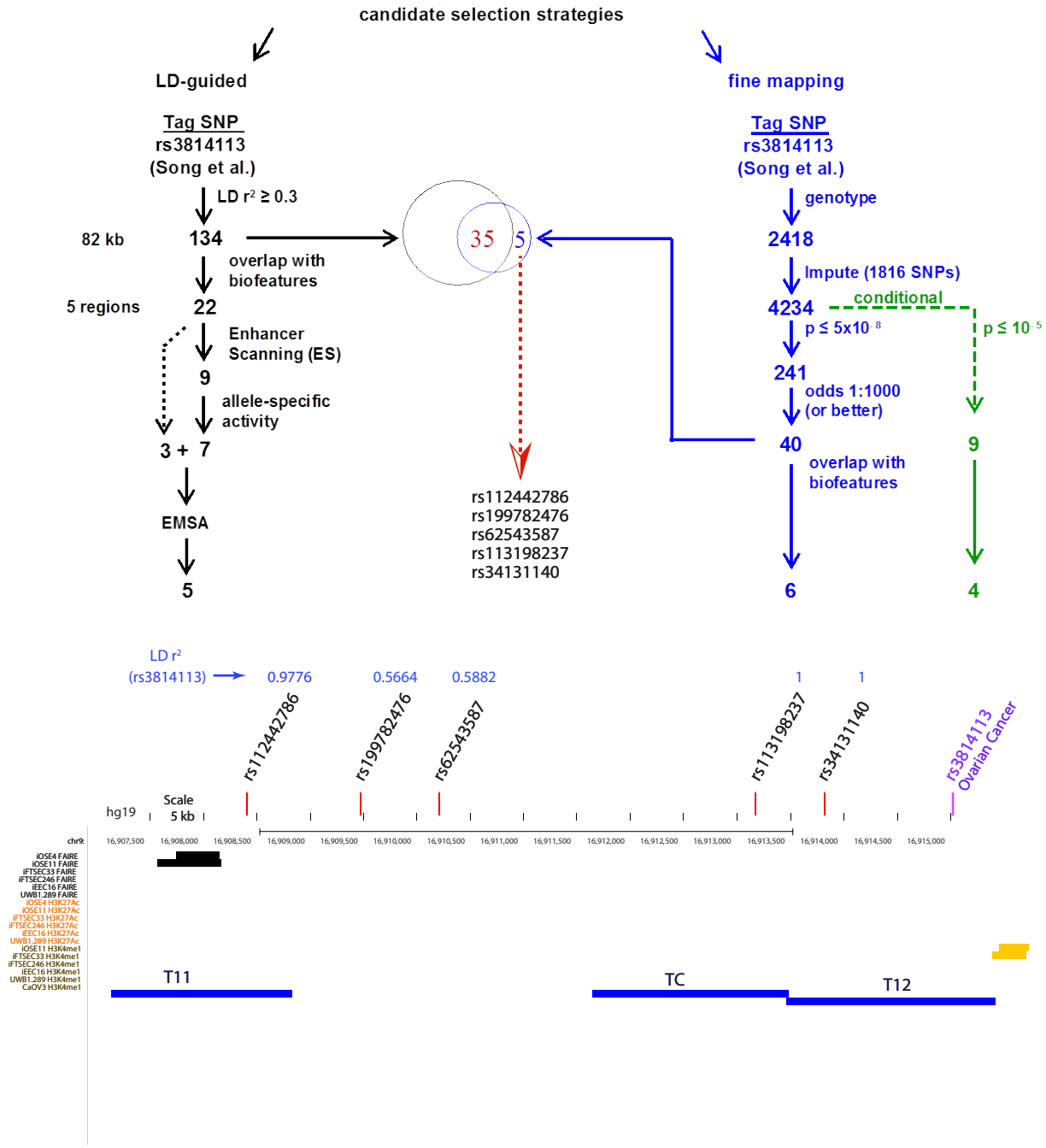
All populations (1000 Genomes Project)



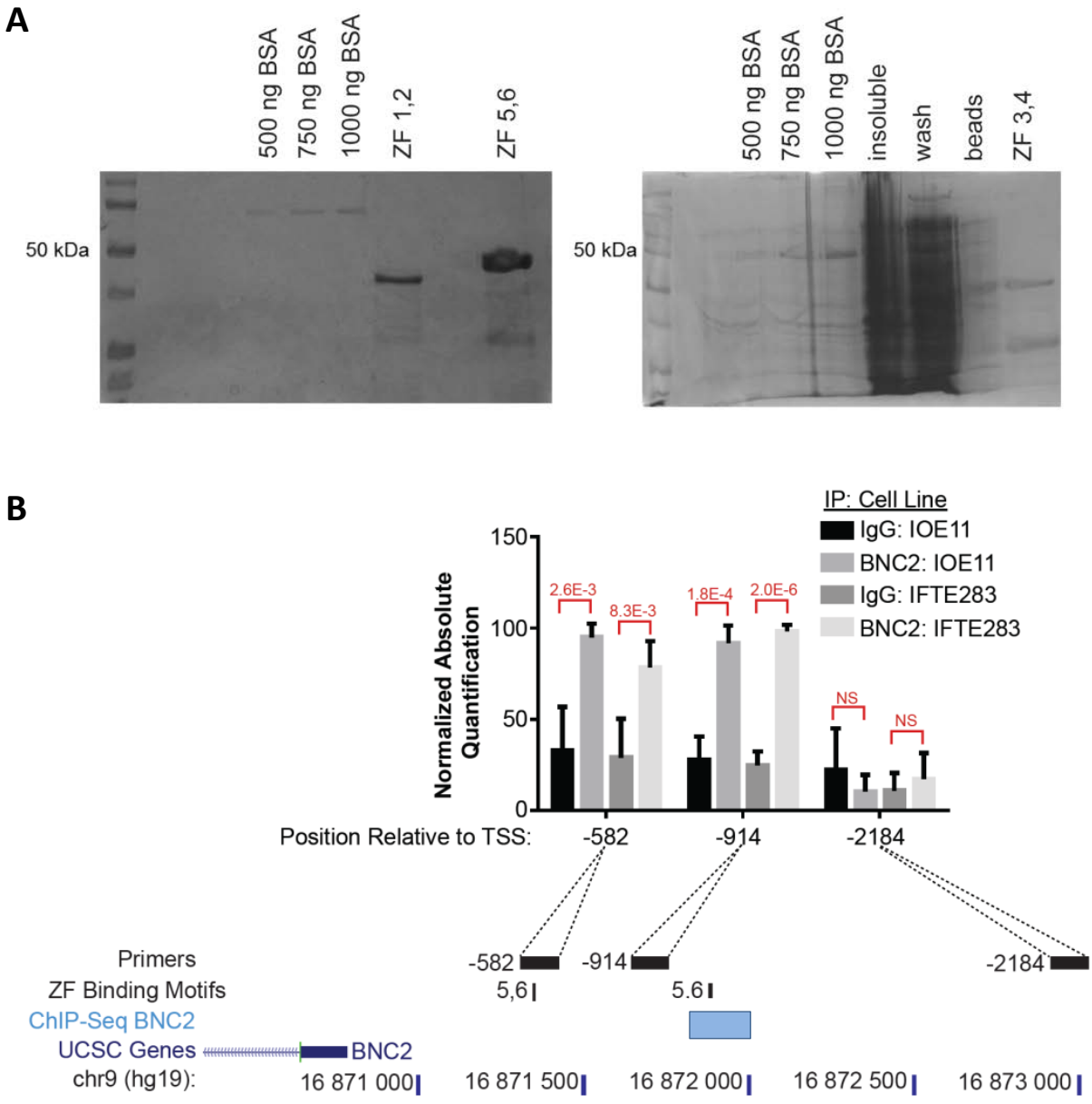
Supplementary Figure 1 (Related to Figure 2): Haplotype frequency for tiles Enhancer Scanning times with multiple SNPs. Haplotype frequencies obtained from LDLink (V 2.0; December 2, 2016; All Populations).



Supplementary Figure 2: (Related to Figure 3) Conservation and S/MAR predicted sequences within the locus. A. This snapshot from the genome browser for the region containing linked SNPs includes tracks for Phylop, PhastCons scoring for conservation and alignment of DNA sequences among several vertebrates. Interestingly, region 7 and 8 have peaks of conservation for both scoring systems while regions 6 and 11 lack a conservation signal. **B.** Region 11 contains sequences highly predicted by MAR-Wiz to attach to the nuclear scaffold/matrix compared to the rest of the locus. The intensity of prediction was driven by the Origin of Replication Rule and the A-T Richness Rule.



Supplementary Figure 3: Overlap between the set of 134 SNPs obtained through the LD-guided analysis and the set of 40 SNPs obtained through the fine mapping-guided approach. Thirty-five out of 40 SNPs obtained through the fine mapping-guided approach were present in the set of 134 SNPs and where functionally assessed. The five remaining SNPs emerged due to updated imputation and were separately assessed (red dashed arrow and lower panel). These five SNPs were visualized in the Human Genome Browser and their LD r^2 to the rs3814113 (tag SNP) is shown on top of the browser. Locations of SNPs are indicated by a thin red line. Enhancer scanning tiles tested in Figure 2 are shown as blue bars.

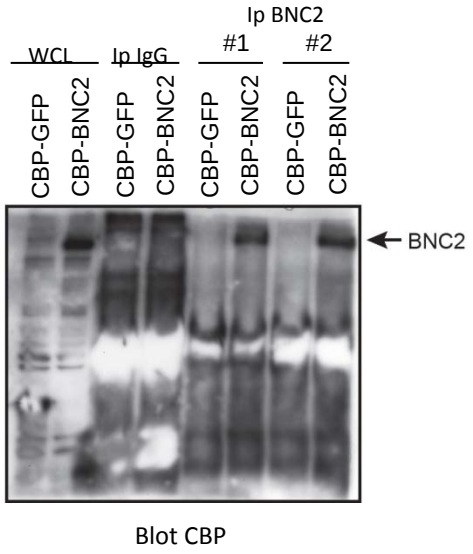


Supplementary Figure 4: (Related to Figure 5) BNC2 binds to its own promoter. **A.** Coomassie stain of protein purification of GST tagged BNC2 ZF pairs: 1,2; 3,4; and 5,6. **B.** Chromatin immunoprecipitation (ChIP) indicates that BNC2 binds to its own promoter. Potential ZF 5,6 binding sites within the *BNC2* promoter are indicated with black lines. Black boxes indicate location of amplicons analyzed with ChIP qPCR. In iOSE11 and iFTSEC283 cells there is a signal that BNC2 is indeed binding to those sites (bar graph). ChIP-seq data for BNC2 in iOSE11 cells replicate the binding at the -914 position (blue bar).

EXTENDED DATA AND METHODS (BNC2 ChIP-Seq)

Antibody

Chromatin immunoprecipitations were performed as previously described [1] using a validated Prestige® BNC2 antibody (Sigma Atlas; cat.no. HPA018525). Prestige Antibodies were developed supported by the Human Protein Atlas (proteinatlas.org). According to the manufacturer Prestige Antibodies are tested in a series of validation steps. The BNC2 antibody was able to immunoprecipitate ectopically expressed CBP-tagged BNC2 showing that it can also specifically recognize native BNC2 (Extended Data Figure 1).



Extended Data Figure 1: GFP and CBP tagged BNC2 were over expressed in 293FT cells. Lysates of these cells were immunoprecipitated with either Rabbit IgG or the Prestige antibody for BNC2 (Sigma). Immunoprecipitates (IP; #1, 1µg of BNC2 antibody; #2, 2µg of BNC2 antibody) undergo Western Blot for CBP. A band for BNC2 between the 150 kDA and 250 kDA mark appears in the input and BNC2 IP for over expressed BNC2 but not in the input and BNC2 IP for over expressed GFP nor in the IgG IP.

ChIP-Seq

ChIP-Seq was performed on the endogenous BNC2. In brief, iOSE11 or iFTSEC283 cells at 70% confluence were cross-linked with 1% Formaldehyde in PBS. Crosslinking was quenched by adding Glycine to a concentration of 0.125 M. After washing, cells were collected in Szaks' RIPA buffer [150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris HCl pH8, 5 mM EDTA, Protease Inhibitors, 50 mM NaF, 0.2 mM sodium orthovanadate, 0.5 mM PMSF] and

the lysate was brought to approximately 1 mg/mL. The lysate was then sonicated in Biogenode Sonicating Water Bath for 12 cycles of 30 sec on and 30 sec off for 8 min. One mg of protein was then mixed with 40 µL of 50% slurry protein A/G agarose beads (Santa Cruz) previously washed in Szaks' RIPA buffer and pre-cleared for 1-2 h at 4°C. We prepared one lysate per cell line, referred to as OSE_input and FTE_input.

Next, pre-cleared lysate was mixed with 5 µg of BNC2 antibody (Sigma Atlas) and 40 µL of 50% slurry protein A/G agarose beads previously washed in Szaks' RIPA buffer and saturated with 1 mg/mL BSA. The mix was incubated overnight at 4°C while rotating. Beads were then washed twice with Szaks' RIPA Buffer, four times with Szaks' IP wash buffer [100 mM Tris HCl pH 8.5, 500 mM LiCl, 1% NP-40, 1% deoxycholate], twice again with Szak' RIPA Buffer and twice with cold TE. Immunocomplexes were eluted by incubating samples at 65°C for 10 min in 1.5X Talianidis Elution Buffer [70 mM Tris HCl pH 8, 1 mM EDTA, 1.5% SDS]. Crosslinks were reversed by bringing samples to 200 mM NaCl solution and incubating at 65°C for 5 h. DNA was purified by phenol-chloroform extraction and re-suspended in 50 µL 10 mM Tris pH 8.0.

For BNC2 ChIP-Seq four individual ChIP samples (from each input lysate) were pooled for each cell line (iOSE11 and iFTSEC283) in two biological replicates, referred to as OSE1, OSE2, FTE1, and FTE2. Immunoprecipitated DNA was used to generate a sequencing library using the NuGEN Ovation Ultralow Library System with indexed adapters (NuGEN, Inc., San Carlos, CA). The library was PCR amplified and size-selected using AxyPrep Fragment Select beads (Corning Life Sciences – Axygen Inc., Union City, CA). The size and quality of the library was evaluated using the Agilent BioAnalyzer, and the library was quantitated with the Kapa Library Quantification Kit (Kapa Biosystems, Woburn MA). Each enriched DNA library was then sequenced on an Illumina HiScan SQ sequencer to generate 100-base paired-end reads. The raw sequence data was de-multiplexed using the Illumina CASAVA 1.8.2 software (Illumina,

Inc., San Diego, CA) and binding sites were identified using the MACS2 software [2] using input DNA as a control and callpeak function without building the shifting model, minimum FDR as 0.01.

The .bam and .wig files were visualized and inspected using the UCSC genome browser[3]. The number of reads for each sample and their quality metrics are shown in Extended Data Figure 2. All samples had >70% of reads with Q30 or better and 2% or less of duplicates. For peak calling $-\log_{10}$(q value) > 2 (corresponding to an 1% FDR) was used as a cut-off. The number of paired end reads ranged from ~50M to ~69M per sample above the ENCODE minimum requirement of 20M for point-source (ChIP-Seq) experiments [4].

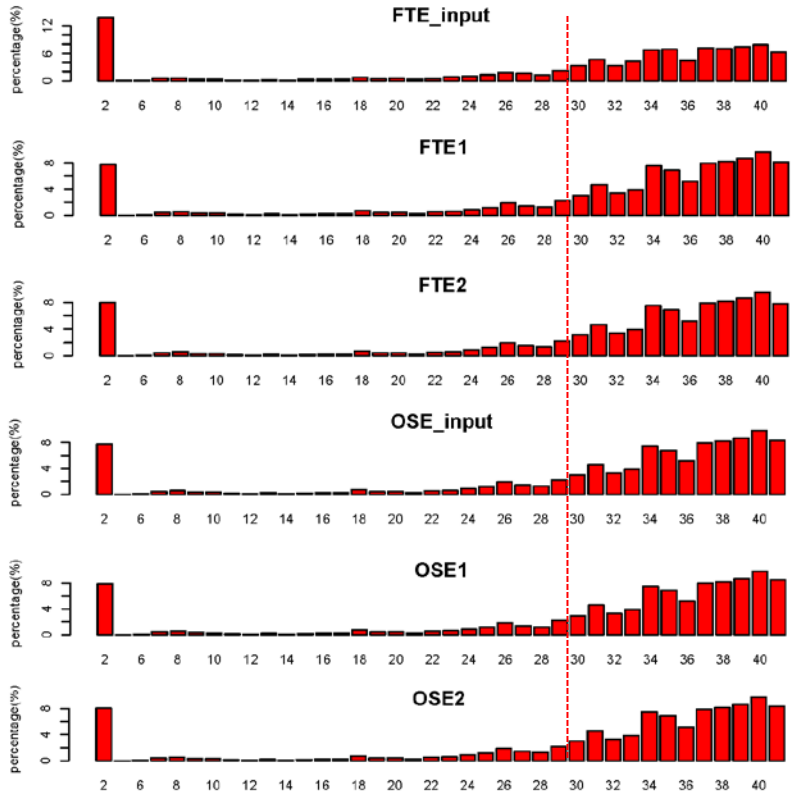
iFTSEC283 cells had a total of 5,687 (FTE1) and 5,730 (FTE2) peaks with 3,396 overlapping peaks and iOSE11 cells had a total of 5,492 (OSE1) and 9,818 (OSE2) with 3,205 overlapping peaks. Peaks used for identification of potential target genes had an intensity greater than 0.05 (reads/length), number of reads greater than 50, and a fold change compared to the input greater than 10 for a total of 2,012 peaks for iFTSEC283 cells and 544 peaks for iOSE11 cells. Median enrichment ranged from 5.2 to 6.9 considered within the norm for ENCODE experiments[4]. Typical peaks are illustrated in Extended Data Figure 3.

Extended Data Figure 2. Sample quality metrics for ChIP-Seq experiment. A. Sample description and quality metrics. Rawnum and cleanreadsnum, number of raw and clean map reads, respectively. Paired and paired_perc, number and percent of paired mapped reads. ≥Q30(%), percent of map reads Q30 and above. **B.** Distribution of map reads according to their quality metrics (Q bins). Red dashed line indicates threshold of ≥Q30. **C.** Peak overlaps between replicates of the same cell lines and overlapped between the two samples.

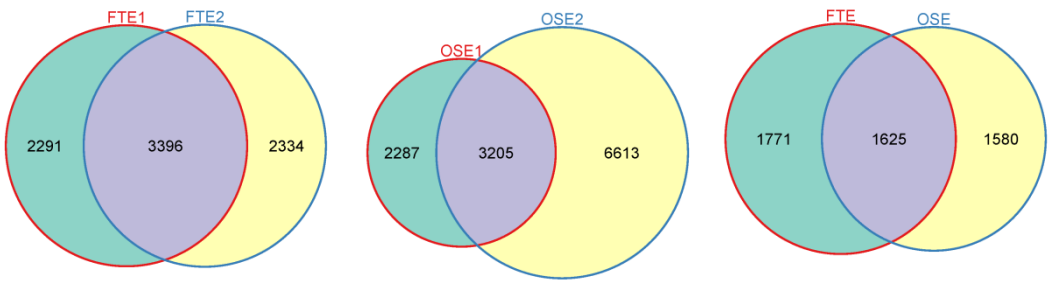
A

sample	rawnum	cleanreadsnum	paired	paired_perc(%)	singleton	≥Q30(%)	Duplication(%)	Unique(%)
FTE1	71,314,272	60,150,320	59,956,210	99.67	194,110	76.59	1.08	99.6
FTE2	62,325,344	50,518,833	50,334,218	99.63	184,615	75.9	0.86	99.53
FTE_input	57,373,894	36,646,199	36,475,486	99.53	170,713	70.9	2.02	99.4
OSE1	46,248,654	35,419,803	35,274,898	99.59	144,905	76.51	0.44	99.47
OSE2	80,421,386	64,143,808	63,883,442	99.59	260,366	75.88	0.84	99.47
OSE_input	80,228,438	69,061,743	68,754,354	99.55	307,389	76.15	0.68	99.41

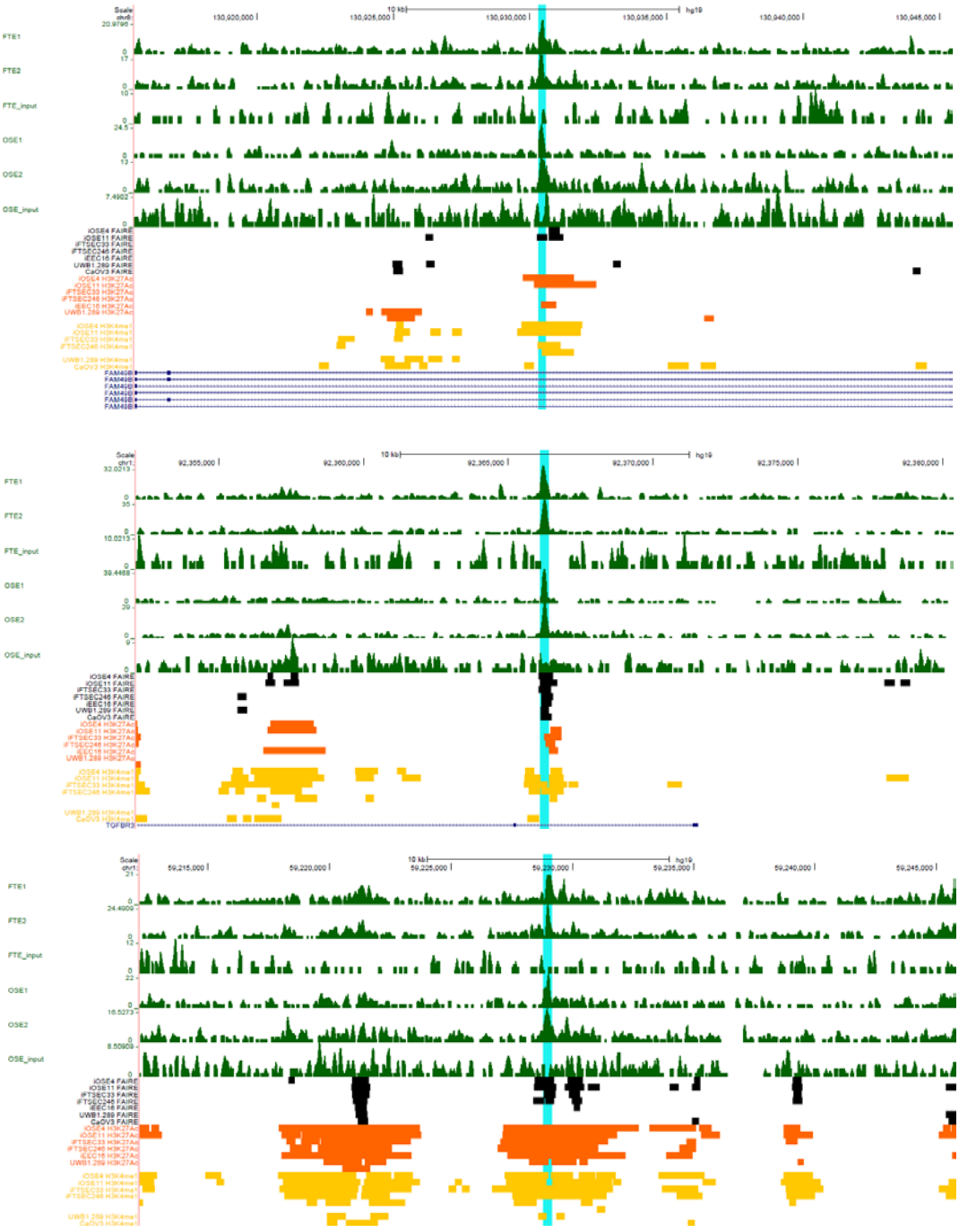
B



C



Extended Data Figure 3. Examples of ChIP-Seq data from the Human Genome Browser. A. FAM49B peak. **B.** TGBR3 peak. **C.** Jun peak. Peaks located in these regions and found both in FTE and OSE samples. The total length of the called peak is shown as the blue highlight.



References

1. Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes & development* **20**, 601-612 (2006).

2. Zhang Y*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).

3. Kent WJ*, et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).

4. Landt SG*, et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813-1831 (2012).