The American Journal of Human Genetics, Volume *107*

Supplemental Data

eQTL Colocalization Analyses Identify *NTN4* as a Candidate Breast Cancer Risk Gene

Jonathan Beesley, Haran Sivakumaran, Mahdi Moradi Marjaneh, Wei Shi, Kristine M. Hillman, Susanne Kaufmann, Nehal Hussein, Siddhartha Kar, Luize G. Lima, Sunyoung Ham, Andreas Möller, Georgia Chenevix-Trench, Stacey L. Edwards, and Juliet D. French

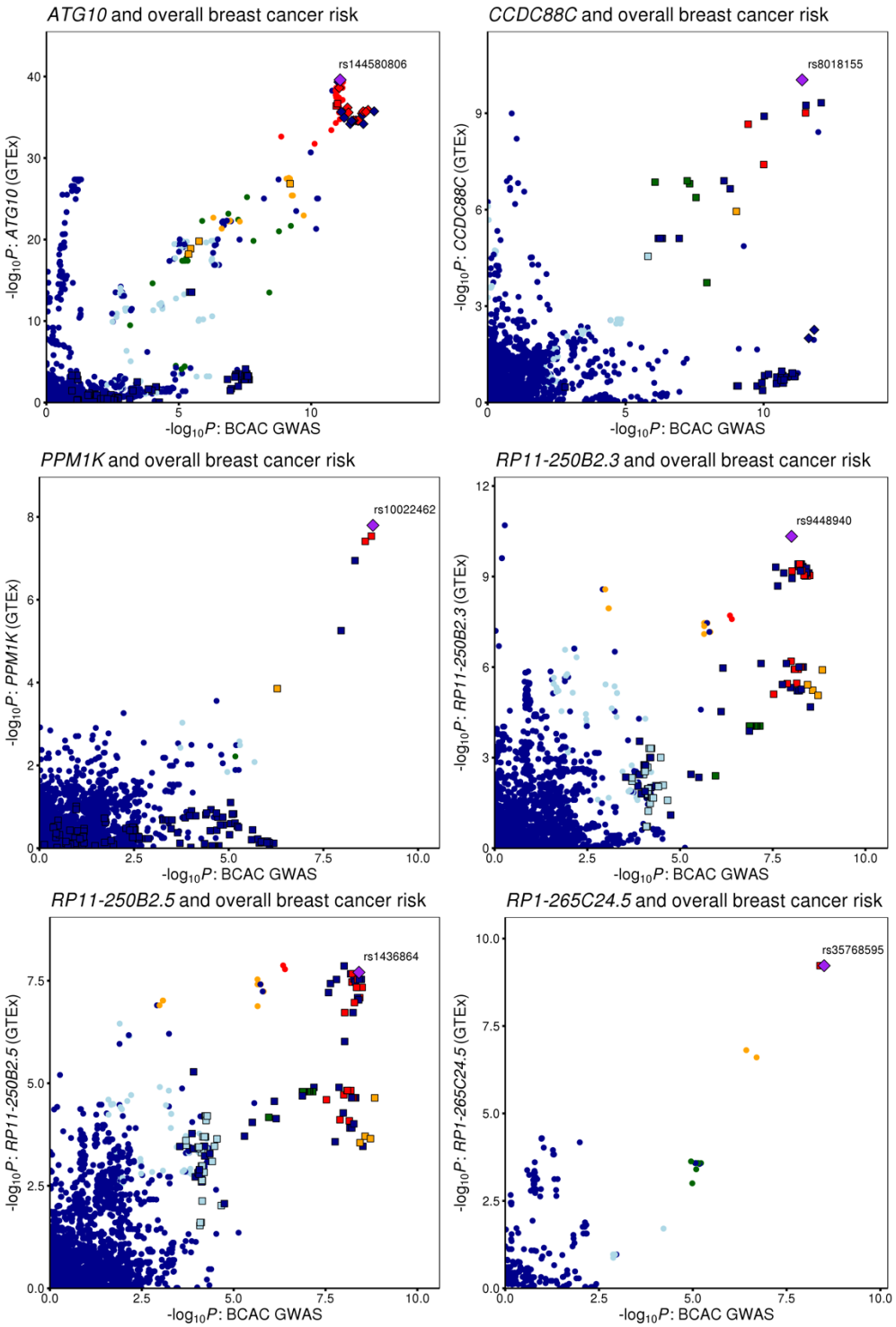

Figure S1. Comparison of BCAC moderate signals with GTEx v8 breast tissue eQTLs. LocusCompare plots[39] for six colocalized signals**.** Gene names and the relevant breast cancer phenotypes are shown in the plot headings. Points are coloured based on linkage disequilibrium (LD) bins relative to the candidate SNP prioritized by HyPrColoc (purple diamond labeled with rsID; red:≥0.8, orange: 0.6–0.8, green: 0.4–0.6, light blue: 0.2–0.4, and dark blue: <0.2). LD data from 1000 Genomes phase 3, version 5 were retrieved from the LDlink portal[40]. Strong CCVs for breast cancer risk are annotated as small diamonds and moderate CCVs as squares[14].

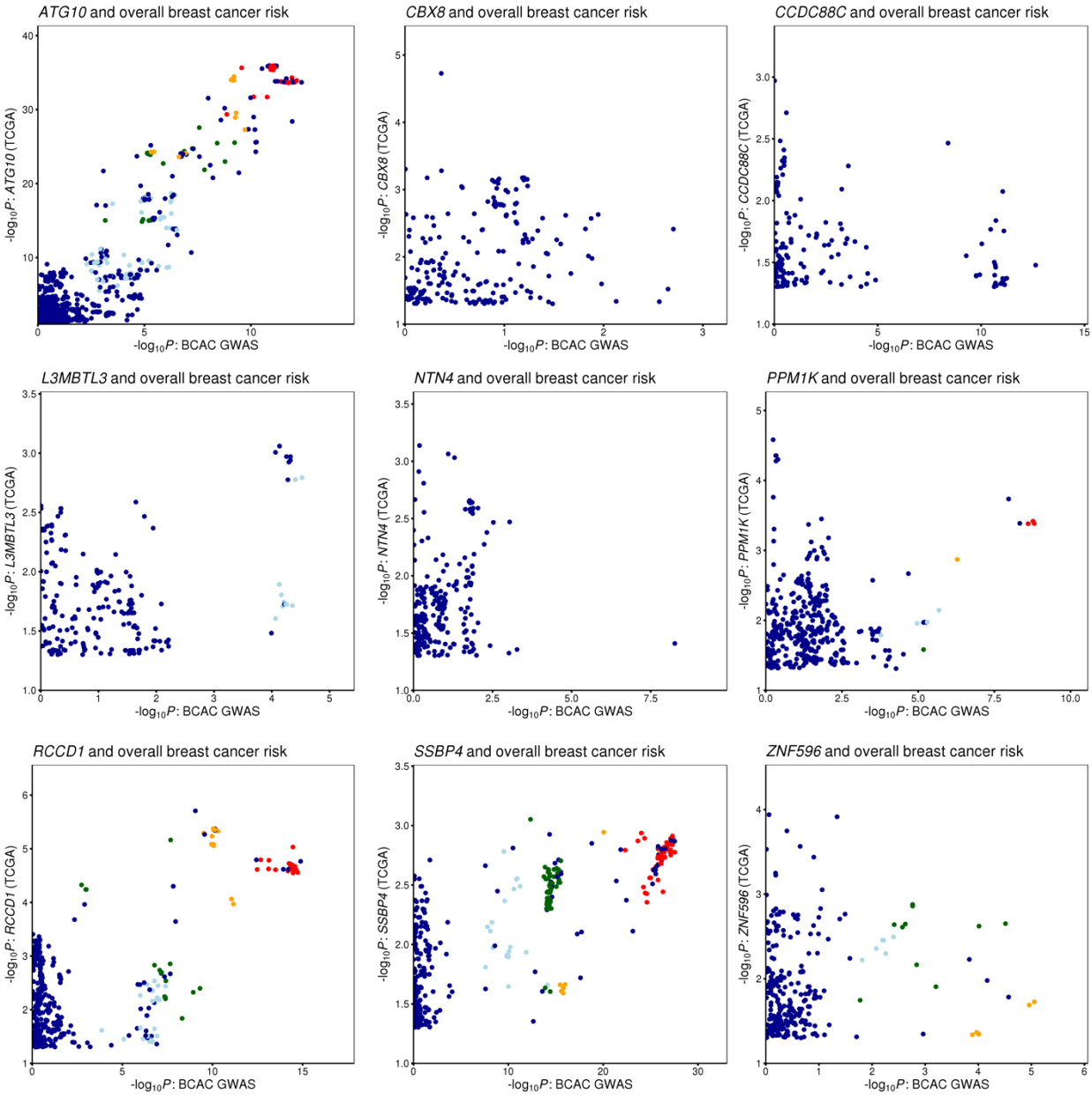

Figure S2. Comparison of BCAC signals with TCGA breast tumor tissue eQTLs. LocusCompare plots[39] for nine signals**.** Gene names and the relevant breast cancer phenotypes are shown in the plot headings. Points are coloured based on linkage disequilibrium (LD) bins relative to the candidate SNP prioritized by HyPrColoc (purple diamond labeled with rsID; red:≥0.8, orange: 0.6–0.8, green: 0.4–0.6, light blue: 0.2–0.4, and dark blue: <0.2). LD data from 1000 Genomes phase 3, version 5 were retrieved from the LDlink portal[40].

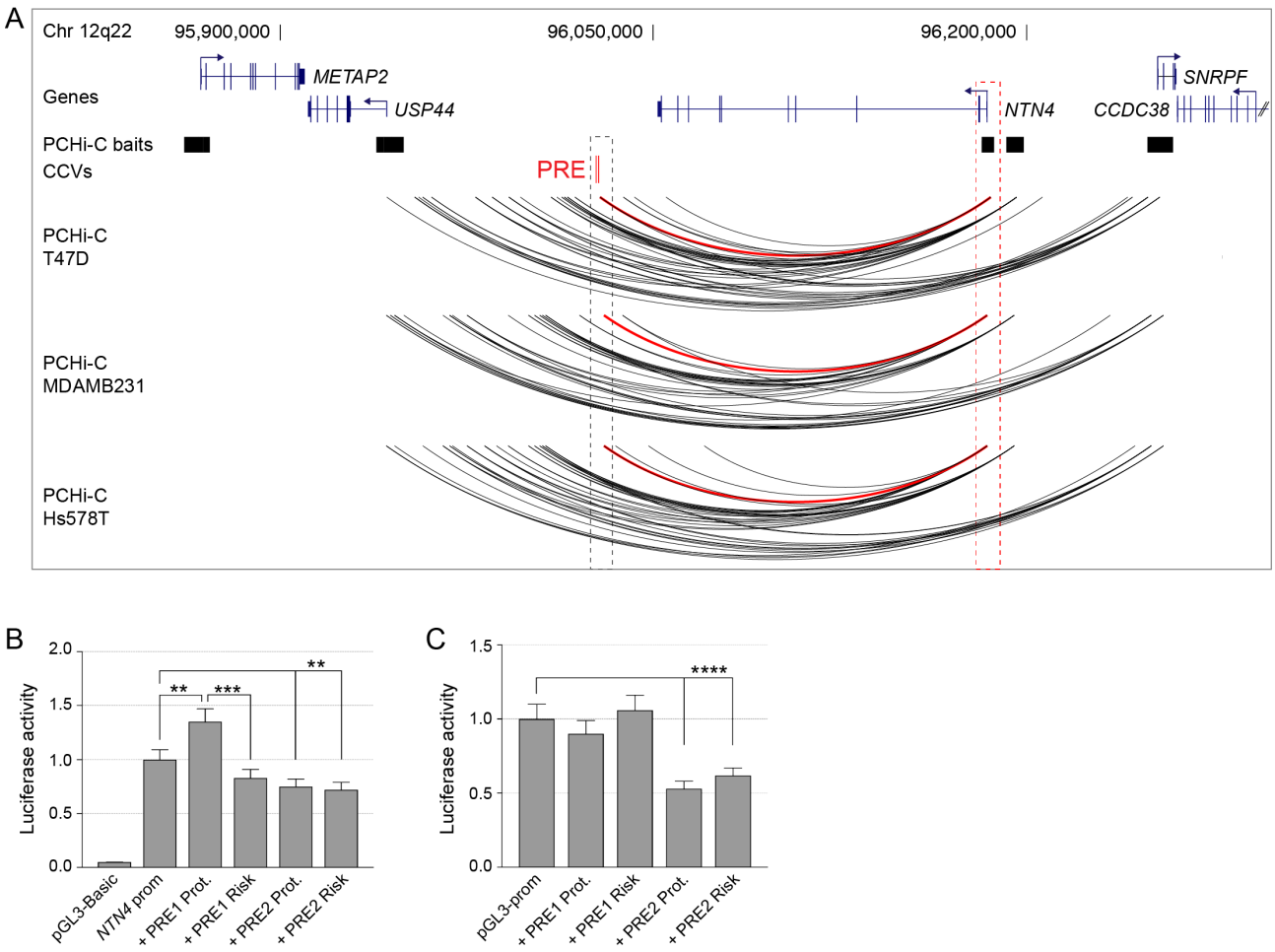

Figure S3. Capture HiC and reporter assays. (A) WashU genome browser showing GENCODE annotated coding genes (blue). The promoter capture Hi-C (PCHi-C) baits are depicted as black boxes. The putative regulatory element (PRE) containing the CCVs is shown as red colored vertical lines. PCHi-C chromatin interactions are shown as black arcs. Red arcs depict chromatin looping between CCVs and the *NTN4* promoter region. (**B**) Luciferase reporter assays following transient transfection of Bre80-hTERT1 breast cells. A PRE1 containing the protective (Prot.) or risk allele of rs61938093 and a PRE2 containing the protective (Prot.) or risk allele of rs1735907 were cloned into *NTN4*-promoter driven luciferase constructs. Error bars, SEM (n=3). *P*-values were determined by two-way ANOVA followed by Dunnett's multiple comparisons test (**p < 0.01, ***p < 0.001). (**C**) Luciferase reporter assays following transient transfection of MCF10A breast cells. A PRE1 containing the protective (Prot.) or risk allele of rs61938093 and a PRE2 containing the protective (Prot.) or risk allele of rs1735907 were cloned into pGL3-promoter luciferase constructs. Error bars, SEM (n=3). *P*-values were determined by two-way ANOVA followed by Dunnett's multiple comparisons test (****p < 0.0001).

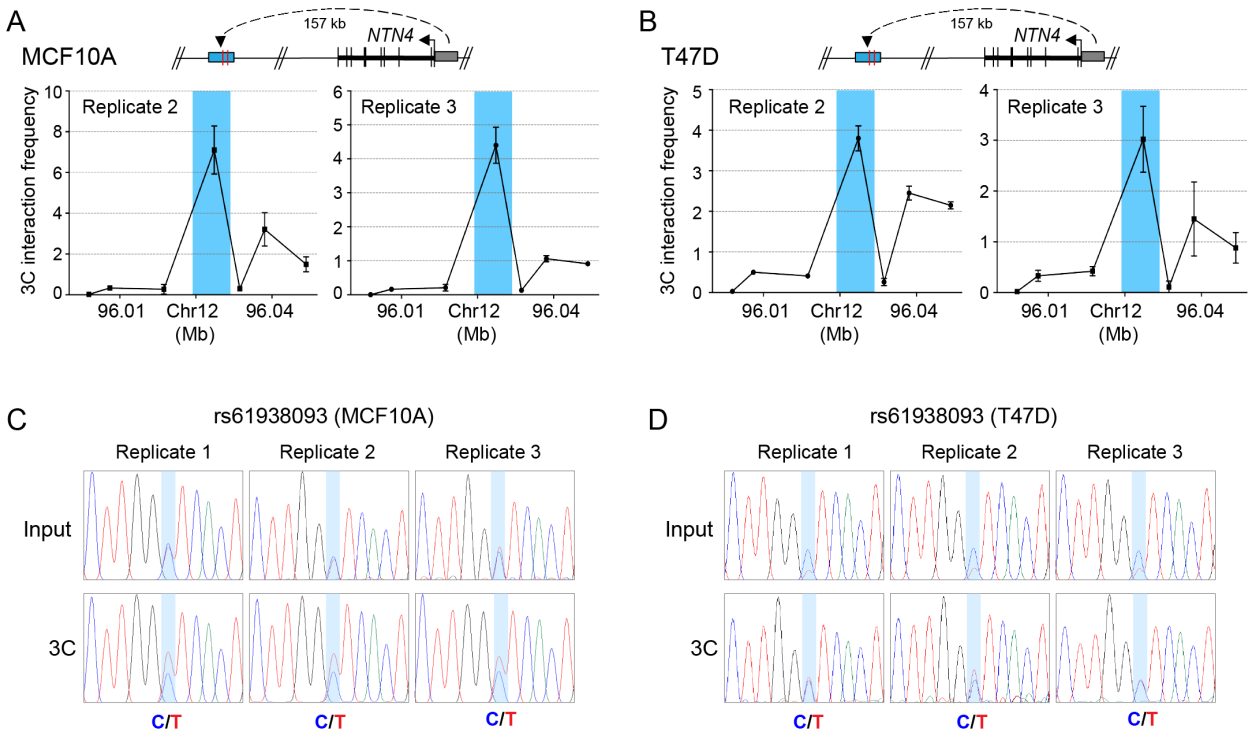

Figure S4. Allele-specific 3C. (**A,B**) Replicate 3C interaction profiles between the *NTN4* promoter and the genomic region containing the PRE in MCF10A and T47D breast cells. 3C libraries were generated with HindIII, with the anchor point set at the *NTN4* promoter. A physical map of the region interrogated by 3C is shown above, the blue shading represents the position of the putative regulatory element (PRE). Error bars, SD (n=3). (**C,D**) Sanger sequencing chromatograms of MCF10A and T47D genomic input DNA versus 3C PCR product confirming allele-specific looping at rs61938093 (Primer set 1; **Table S2**).

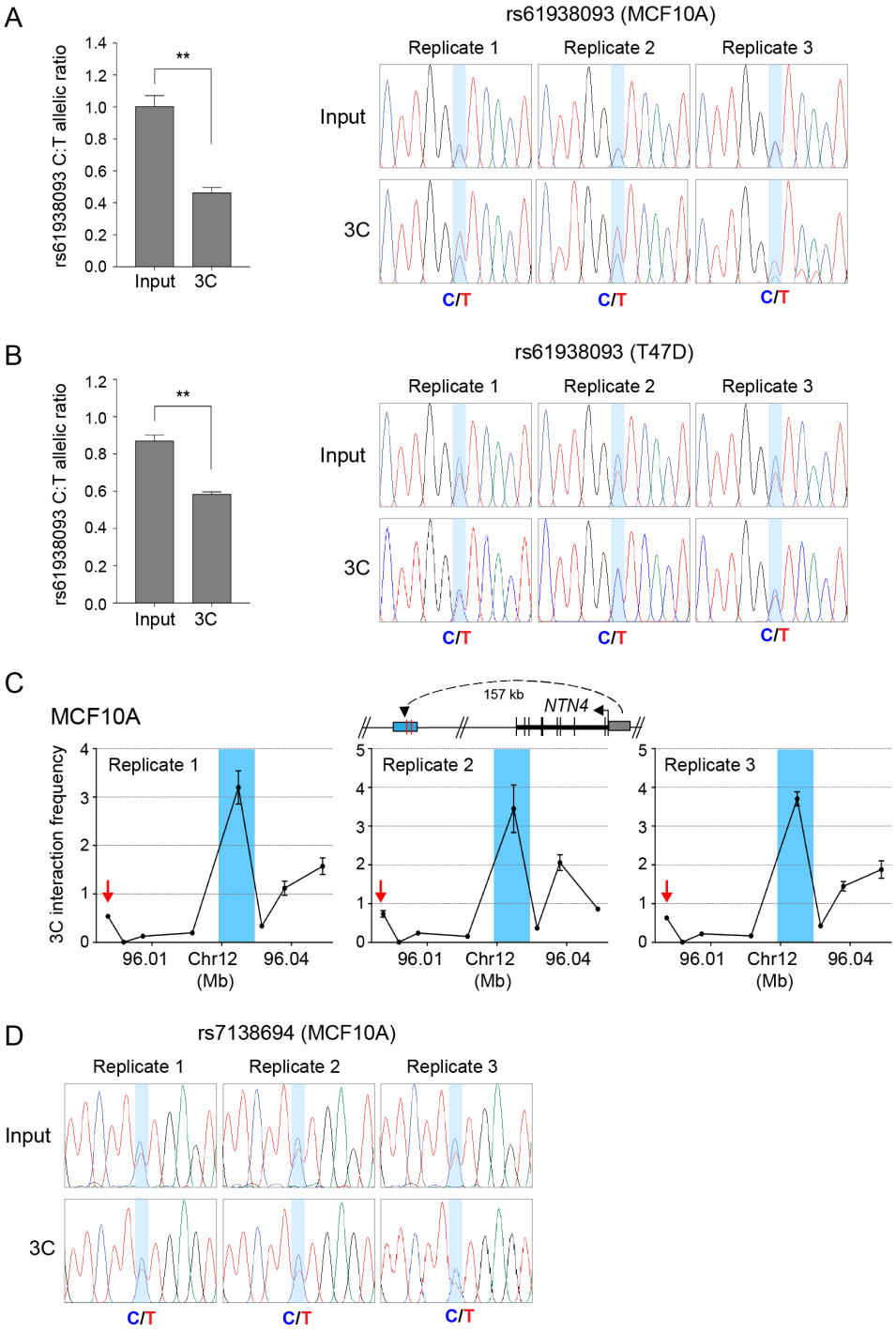

Figure S5. Allele-specific 3C. (**A,B**; *left panels*) Allele-specific qPCR using primer set 2 (**Table S2**) and Taqman SNP assay to quantify the allelic ratio at CCV rs61938093 in MCF10A and T47D breast cells. Error bars, SEM (n=3). *P*-values were determined using a Student's t-test (**p < 0.01). (**A,B**; *right panels*) Sanger sequencing chromatograms of MCF10A and T47D genomic input DNA versus 3C PCR product confirming allele-specific looping at rs61938093. (**C**) 3C interaction profiles between the *NTN4* promoter and the genomic region containing the PRE in MCF10A breast cells. 3C libraries were generated with HindIII, with the anchor point set at the *NTN4* promoter. A physical map of the region interrogated by 3C is shown above, the blue shading represents the position of the putative regulatory element (PRE). The red arrow shows the HindIII fragment assessed by allele-specific 3C as a negative control. Error bars, SD (n=3). (**D**) Sanger sequencing chromatograms of MCF10A genomic input DNA versus 3C PCR product at rs7138694.

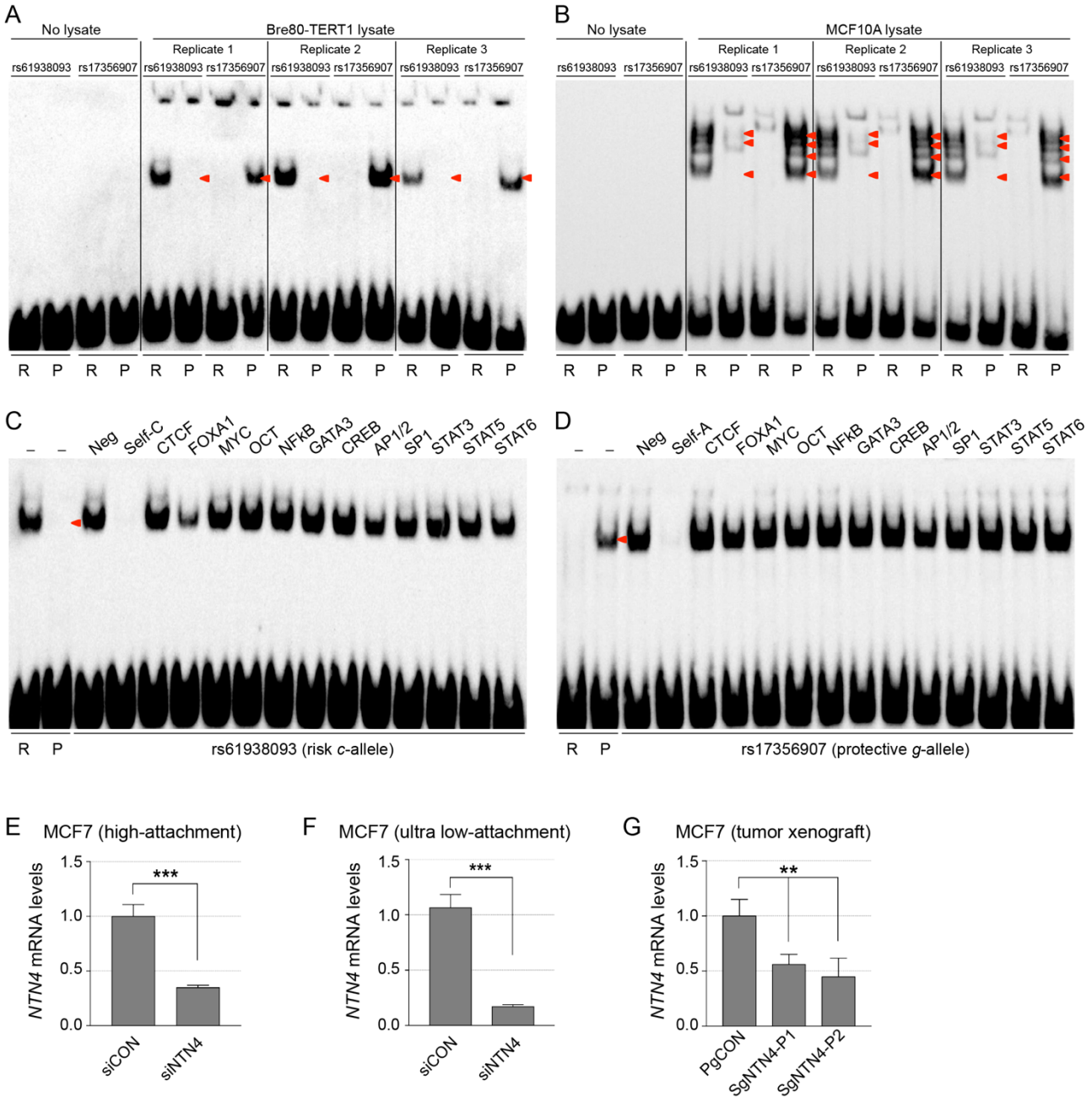

Figure S6. EMSAs and qPCR validation of *NTN4* depletion. (**A,B**) EMSAs for oligonucleotide duplexes containing CCVs rs61938093 or rs17356907 with the risk allele (R) or protective allele (P), assayed using no lysate, Bre80-TERT1 (**A**) or MCF10A (**B**) nuclear extracts. Red arrowhead indicates band mobility differences between alleles. (**C,D**) Competitive EMSAs for oligonucleotide duplexes containing CCVs rs61938093 (**C**) or rs17356907 (**D**) with the risk allele (R) or protective allele (P), assayed using Bre80-TERT1 nuclear extracts. Note: the first four lanes of each EMSA are shown in Figure 3F. Competitor oligonucleotides are listed above each panel and were used at 100-fold molar excess: (-) no competitor; (Neg) a non-specific competitor; (Self) an identical oligonucleotide with no biotin label; (transcription factor) consensus binding site. Red arrowheads indicate band mobility differences between alleles. (**E, F**) *NTN4* depletion after transient transfection of a non-targeting control (siCON) or *NTN4* (siNTN4) siRNAs. *NTN4* levels were measured by qPCR and normalized to *beta-glucuronidase (GUSB)*. Error bars, SEM (n=3). *P*-values were determined with a student's *t*-test (***p<0.001). (**G**) *NTN4* depletion in MCF7-control (PgCON) or MCF7-dCas9-KRAB *NTN4* depleted cells (SgNTN4-P1/P2). *NTN4* levels were measured by qPCR and normalized to *GUSB*. Error bars, SEM (n=3). *P*-values were determined by one-way ANOVA followed by Dunnetts multiple comparisons test (**p<0.01).

Table S1. Bioinformatic predictions of target genes.

Gene name	INQUISIT[3] score	TWAS/eQTL prediction	References
NTN4	1	Yes	5
CBX8	2	Yes	5,
L3MBTL3	1	Yes	5,6,7,19
MARCH11	Not applicable	Yes	5
PIDD1	1	Yes	5,7
RCCD1	1	Yes	5,7
SSBP4	2	Yes	5,7
ZNF596	Not applicable	No	Not applicable
PRC1-AS1	2	Yes	5
RP11-53O19.1	2	Yes	6
RP5-855D21.3	Not applicable	No	Not applicable
ATG10	2	Yes	5,6,7
CCDC88C	2	Yes	5,
PPM1K	Not applicable	Yes	5
RP11-250B2.3	Not applicable	No	Not applicable
RP11-250B2.5	Not applicable	Yes	18,19
RP1-265C24.5	Not applicable	No	Not applicable

Table S2. Oligonucleotides and genomic coordinates used in this study.

3C assay validation primer sequences

Primer name	Sequence (5' to 3')
NTN4 promoter bait	CTGTTAGTGGCTGCCAGACTGTAGACACTCTCC
NTN4_Up4F	GCCAGGCGTAATGTTGTGTTAATGTGG
NTN4_Up3F	AACTCACGTTTTACTATCTGCTCTTCCTCCCACC
NTN4_Up2F	GTTCTGCCTCCACCATGTTAATAACCATCTGG
NTN4_Up1R	GCTAATGTCCAAAGTTTTCTGTGGCACAGTGC
NTN4_EnhR	TGAGCAAAGACTGAAGGAAATTAGGGGAGAGG
NTN4_Dn1R	TGTGATAACCAGACTGGATTACTGTGCTCAAGAGG
NTN4_Dn2F	CAAACGGGTGATTTGGAGGATGCACC
NTN4_Dn3R	CTCCCTGCCTTATTTGTCATGAACTGCAAGC

Allele-specific 3C PCR and sequencing primers

Primer name	Sequence (5' to 3')
Primer set 1: 3CgDNA1	AGATACCACATGGGTCACTTTCTTCCCTCTCC
Primer set 1: 3Cprom1	CTGTTAGTGGCTGCCAGACTGTAGACACTCTCC
Primer set 1: 3Cenh1	GGTGTACGTCCATGGTCCTAGTTACTTGGAAGG
Primer set 2: 3CgDNA2	CTTAGGCCCTTCCCTTGATATTCCCTCTGC
Primer set 2: 3Cprom2	CCTGTCACAGGTAGAAACCCTGAAGAGACAGC
Primer set 2: 3Cenh2	TGGGTGGGATGGTGTACGTCCATGG
Sequencing primer	CCCTAATTTCCTTCAGTCTTTGC
Neg primer: 3CgDNA3	GGGACATAACAACATGACTAGTTTGAACACACC
Neg primer: 3Cprom3	CTGTTAGTGGCTGCCAGACTGTAGACACTCTCC
Neg primer: 3Cenh3	GCCAGGCGTAATGTTGTGTTAATGTGG
Neg sequencing primer	GCGTAATGTTGTGTTAATGTGG

Genomic coordinates for luciferase assay constructs

DNA element	Genomic coordinates (hg19)
NTN4 Promoter	chr12:96,184,432-96,185,576
NTN4 PRE1-rs61938093	chr12:96,026,561-96,027,570
NTN4 PRE2-rs17356907	chr12:96,027,565-96,028,547

EMSA oligonucleotide sequences

Name	CCV-allele	Sequence (5' to 3'); *5'-biotinylated
5BIONTN4rs6193RiskF	rs61938093-c	*GTGGCACAATCTTGGCTCACTGCAGCCTCCG
5BIONTN4rs6193RiskR	rs61938093-c	*CGGAGGCTGCAGTGAGCCAAGATTGTGCCAC
5BIONTN4rs6193ProtF	rs61938093-t	*GTGGCACAATCTTGGTTCACTGCAGCCTCCG
5BIONTN4rs6193ProtR	rs61938093-t	*CGGAGGCTGCAGTGAACCAAGATTGTGCCAC
5BIONTN4rs1735RiskF	rs17356907-a	*TGGGGATTAGATGGTACCAAAATGACAGTGG
5BIONTN4rs1735RiskR	rs17356907-a	*CCACTGTCATTTTGGTACCATCTAATCCCCA
5BIONTN4rs1735ProtF	rs17356907-g	*TGGGGATTAGATGGTGCCAAAATGACAGTGG
5BIONTN4rs1735ProtR	rs17356907-g	*CCACTGTCATTTTGGCACCATCTAATCCCCA
CEBP		TGCAGATTGCGCAATCTGCA
NRF1		CTGCTAGCCCGCATGCGCGCGCACCTTA
STAT1		CATGTTATGCATATTCCTGTAAGTG
YY1		CGCTCCCCGGCCATCTTGGCGGCTGGT
CTCF		AAGAAACCGCTAGGGGGCCTACT
FOXA1		CTGGTCTTAAAGGTGTTTACCTTGTCTGAT
MYC		TCAGACCACGTGGTCGGG

OCT		TGTCGAATGCAAATCACTAGAA
NFkB		AGTTGAGGGGACTTTCCCAGGC
GATA3		CACTTGATAACAGAAAGTGATAACTCT
CREB		AGAGATTGCCTGACGTCAGAGAGCTAG
AP1		CGCTTGATGACTCAGCCGGAA
AP2		GATCGAACTGACCGCCCGCGGCCCGT
SP1		ATTCGATCGGGGCGGGGCGAGC
STAT3		GATCCTTCTGGGAATTCCTAGATC
STAT5		AGATTTCTAGGAATTCAATCC
STAT6		GTATTTCCCAGAAAAGGAAC

CRISPRi sgRNA sequences and genomic target coordinates

sgRNA name	Genomic coordinates (hg19)	sgRNA spacer sequence (5' to 3')
pgCON	Non-targeting control	GACCAGGAUGGGCACCACCC
sgEnh1	chr12:96027762-96027781	AAAAUGACAGUGGUCUCUGC
sgEnh2	chr12:96027763-96027782	AGCAGAGACCACUGUCAUUU
sgNTN4-P1	chr12:96184481-96184500	AAAAGCGGAGGAGGACGCCC
sgNTN4-P2	chr12:96184525-96184544	UCCGUCCCGUCCUUCUCCAC