132–147 *Nucleic Acids Research, 2012, Vol. 40, No. 1* *Published online 13 September 2011*
doi:10.1093/nar/gkr610

An insulator loop resides between the synthetically interacting elements of the human/rat conserved breast cancer susceptibility locus *MCS5A/Mcs5a*

Bart M. G. Smits, Benjamin D. Traun, Thomas L. Devries, Ann Tran, David Samuelson, Jill D. Haag and Michael Gould*

McArdle Laboratory for Cancer Research, Department of Oncology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI 53706, USA

Received February 11, 2011; Revised July 7, 2011; Accepted July 12, 2011

ABSTRACT

Many low-penetrance breast cancer susceptibility loci are found to be located in non-protein-coding regions, suggesting their involvement in gene expression regulation. We identified the human/rat-conserved breast cancer susceptibility locus *MCS5A/Mcs5a.* This locus has been shown to act in a non-mammary cell-autonomous fashion through the immune system. The resistant *Mcs5a* allele from the Wistar–Kyoto (WKy) rat strain consists of two non-protein-coding genetic elements that must be located on the same chromosome to elicit the phenotype. In this study, we show the presence of a conserved higher order chromatin structure in *MCS5A/Mcs5a* located in between the synthetically interacting genetic elements. The looped elements are shown to be bound by CTCF and cohesin. We identify the downregulation of *Fbxo10* expression in T cells as a strong candidate mechanism through which the interacting genetic elements of the resistant *Mcs5a* allele modulate mammary carcinoma susceptibility. Finally, we show that the human *MCS5A* polymorphisms associated with breast cancer risk are located at both sides of the looped structure and functionally interact to downregulate transcriptional activity, similar to rat *Mcs5a*. We propose a mechanistic model for *MCS5a/Mcs5a* in which a CTCF-mediated insulator loop encompassing the *TOMM5/Tomm5* gene, resides in between and brings into closer physical proximity the synthetically and functionally interacting resistant genetic variants.

INTRODUCTION

The risk of developing breast cancer involves the interaction between a woman's inherited genetics and her environment. The genetic component of risk for most common forms of breast cancer defines it as a complex trait consisting of numerous susceptibility alleles and their interactions. Approximately 25 of such alleles have been identified thus far using genome-wide association studies (GWAS) and comparative genetics and each is associated with a low relative risk (1–10). Currently, a major open question regarding these low-penetrance susceptibility alleles is defining their function in regard to breast cancer etiology. The great majority of these alleles are non-protein-coding, which implies that they may modulate gene expression regulation. For example, the alleles of a breast cancer risk variant in the second intron of *FGFR2* have been shown to differentially regulate its expression (11). For the vast majority of low-penetrance breast cancer susceptibility alleles, it remains unclear if these are involved in gene expression regulation and what their target genes may be. Understanding mechanistically how these alleles modulate breast cancer susceptibility could go beyond population-based screening and lead to intervention strategies for cancer prevention and therapy.

The non-protein-coding locus *Mcs5a*/*MCS5A* modulates breast cancer susceptibility in both rats and women (5). We identified this locus via a comparative genetics approach consisting of linkage analysis in the backcross progeny of the mammary carcinoma susceptible Wistar–Furth (WF) rat strain and the resistant Wistar–Kyoto (WKy) rat strain (12), subsequent genetic fine-mapping using congenic recombinant rat lines (13,14), and a case–control association study using variants in the human orthologous locus (5). When the resistant allele of *Mcs5a* from the WKy rat strain is introgressed into the

*To whom correspondence should be addressed. Tel: +1 608 263 6615; Fax: +1 608 262 2824; Email: gould@oncology.wisc.edu
Present address:
David Samuelson, University of Louisville School of Medicine, Department of Biochemistry and Molecular Biology, Louisville, KY 40292, USA.

susceptible genetic background from the WF rat strain, a ~50% reduction in mammary carcinoma multiplicity is observed (5). The resistant allele acts in a non-mammary cell-autonomous manner, during early mammary carcinoma progression through the immune system (15). We have also shown that the presence of the resistant *Mcs5a* allele is associated with T-cell homeostasis and functions (15). Specifically, T cells of the *Mcs5a* resistant congenic line show an increased mitogen-induced proliferation potential and Th1 cytokine production, suggesting that the *Mcs5a* allele exerts its effect on mammary carcinoma susceptibility through T cells (15). *Mcs5a* consists of two non-protein-coding, synthetically interacting elements (*Mcs5a1*, *Mcs5a2*), which have to be located on the same chromosome to elicit the resistance phenotype (5). The variants associated with breast cancer risk in women are located in the human orthologous loci, *MCS5A1* and *MCS5A2*, respectively, at a distance of ~60 kb of each other (5). Although identified by comparative genetics, the *MCS5A* variants resemble GWAS-identified risk alleles as the risk-associated polymorphisms are non-protein-coding, are common in the population and display low-genetic penetrance. The risk-associated variant in *MCS5A1* marked by single nucleotide polymorphism (SNP) rs6476643 has four correlated polymorphisms. These are located in a ~5.7-kb region around the CpG island that is associated with a predicted promoter of *FBXO10*, a gene encoding the uncharacterized E3 ubiquitin ligase F-box protein 10. Similarly, the *MCS5A2* variant marked by SNP rs2182317 encompasses a total of 15 correlated polymorphisms spanning ~26.1 kb. These are located around the CpG island that is associated with the predicted promoter of the FERM and PDZ domain containing 1 gene, *FRMPD1*. We have shown previously that between susceptible congenic control and *Mcs5a* resistant congenic rats, *Fbxo10* and *Frmpd1* are differentially expressed in the thymus and spleen, respectively, and not in the mammary gland (5). A third gene in the *MCS5A* locus, the translocase of the outer mitochondrial membrane5 gene, *TOMM5*, is not directly associated with the breast cancer risk variants of *MCS5A1* and *MCS5A2*, but is located in the 60 kb of sequence that separates them. SNP rs2182317 in *MCS5A2* has recently been shown to be among the most significant of a list of 710 candidate breast cancer risk alleles and to modify risk to both Estrogen Receptor-positive and -negative breast cancer (4).

The availability of both human cell based and congenic rat models allows us to address several mechanistic questions regarding the *Mcs5a*/*MCS5A* breast cancer susceptibility alleles. First, we mechanistically address the observation that the mammary carcinoma resistance phenotype requires two genetic *Mcs5a* elements, separated by *Tomm5*, to synthetically interact on the same chromosome. Next, we show that the synthetic interaction is also required for downregulation of the transcript levels of *Fbxo10* in the thymus and in various sorted T-cell populations. Finally, we show that the resistant alleles of the human breast cancer risk-associated variants could functionally interact to downregulate transcriptional activity, resembling transcriptional regulation of *Fbxo10* by the interacting genetic elements of the resistant rat *Mcs5a* allele.

MATERIAL AND METHODS

Animals

The congenic rat lines were established and maintained in an AAALAC-approved facility as previously published (14). All animal protocols were approved by the University of Wisconsin Medical School Animal Care and Use Committee. Congenics are defined as genetic lines developed on a WF,susceptible genetic background and carrying the selected WKy,resistant *Mcs5a* alleles. The resistant congenic line WF.WKy-*Mcs5a* ('*Mcs5a*', line WW) with a decreased 7,12- dimethylbenz(a)anthracene (DMBA)-induced mammary carcinoma susceptibility phenotype is WKy-homozygous at the entire *Mcs5a* locus (5). The susceptible congenic control line WF.WKy ('susc.') is WF-homozygous at *Mcs5a* and all other identified *Mcs* loci. The susceptible congenic lines WF.WKy-*Mcs5a1* ('*Mcs5a1*', line B3) and WF.WKy-*Mcs5a2* ('*Mcs5a2*', Line LL) are WKy-homozygous at *Mcs5a1* and *Mcs5a2*, respectively (5). For all experiments, only female animals were used.

Flow cytometry, magnetic cell sorting and T-cell activation

For the gene expression studies, cells were prepared for immunostaining by submerging fresh thymus and/or spleen tissues in cold flow cytometry solution (FCS: PBS with 2% FBS and 25 mM HEPES). Tissues in 1 ml cold FCS were dispersed using a 50 μM Medicon (BD Biosciences) filter grinders and a BD Medimachine for ~1 min. The resulting cell suspension was filtered with 70 μM Filcon (BD Biosciences) and centrifuged to pellet cells. In sterile 12 ml polystyrene round-bottom Falcon tubes (BD Biosciences), cells were resuspended in cold FCS (25×10^6 cells/ml) and immunostained at 4°C with 0.5 μg antibody-conjugate per 10^6 cells (25 μg each antibody) for 30 min. Monoclonal antibodies from BD Biosciences included anti-rat CD3 (clone 1F4), CD45RA (clone OX-33), CD4 (clone OX-35) and CD8a (clone OX-8). Cells were washed 1× with 1 ml cold FCS and filtered prior to flow cytometry sorting on a FACSVantageSE (BD Biosciences) at the University of Wisconsin Flow Cytometry Core Facility. For chromatin studies, the enrichment of $CD3^+$ T cells was done by magnetic cell sorting. Single cell suspensions (in cold FCS) of anonymous cryopreserved human peripheral blood mononuclear cell samples and rat (red blood cell-reduced) splenocytes were incubated with CD3 and pan-T cell (OX-52) microbeads (Miltenyi), respectively. The labeled cell suspension was loaded onto a pre-equilibrated MACS cell separation column (Miltenyi), washed and eluted according to the manufacturer's recommendations. $CD3^+$ T-cell purity was estimated by flow cytometry to be >92%. For T cell culturing and activation, 2 million lymphocytes from spleen were stimulated with 1 μg/ml Concanavalin A

(conA; Calbiochem/EMD Chemicals) and were cultured in 2 ml RPMI medium containing 10% FBS for 24 h at 37°C in a 95% air/5% CO_2 atmosphere. Unstimulated cells served as a control. Splenocytes were stained with anti-T-cell microbeads (OX52, Miltenyi) and T cells were sorted by positive selection using Easysep® magnet (Stem Cell Technologies). Purity of T cells was >90% as determined by CD3 labeling and flow cytometry.

Primary T-cell nucleofections with siRNAs

Nucleofections were carried out by following the Mouse T-Cell Nucleofector Kit (Lonza) protocol. Briefly, 2×10^6 T cells (from a pool of sorted splenic T cells with equal contribution from four female *Mcs5a* resistant congenic rats) were pelleted at low speed (10 min, 90*g*) and resuspended in 100 µl supplemented Nucleofector Solution. Two nanomoles of the desired siRNAs (ON-TARGETplus Non-Targeting, and ON-TARGETplus SMARTpool CTCF; Dharmacon) were mixed in, followed by nucleofection using program X-001 of the Nucleofector II Device. The cells were cultured in Nucleofection Media for 64–68 h at 37°C in a 95% air/5% CO_2 atmosphere.

Chromosome conformation capture assay

Rat templates were prepared from T cells of six susceptible congenic control, and six *Mcs5a* resistant congenic animals. Human templates were prepared from three T-cell samples. Approximately 5×10^6 T cells (or 2×10^6 for the siRNA experiments) were diluted in 40 ml of PBS and 1.7 ml of 37% formaldehyde was added to fix chromatin. The solution was allowed to fix for 10 min at room temperature. To quench the reaction, 2.7 ml of 2 M glycine was added to each tube and incubated for 5 min at room temperature, after which the tubes were placed on ice for 15 min. To burst the cells and expose the fixed chromatin, the cells were centrifuged for 10 min at 800 *g* and resuspended in 0.5 ml ice-cold lysis buffer, containing 10 mM Tris–HCl pH 8.0, 10 mM NaCl, 0.2% Nonidet P-40 (NP-40), 1× protease inhibitors (Roche Applied Science). The reaction was allowed to proceed on ice for 10 min. The cells were gently lysed via dounce homogenization using a tight pestle for 1 min on ice. The reaction sat on ice for 1 min and homogenization was repeated for 30 s. Next, the lysis buffer was removed, the nuclei were washed in 0.5 ml of 1× restriction enzyme buffer (NEB3; New England Biolabs) and resuspended in 400 µl of 1× restriction enzyme buffer containing 0.1% SDS. This mixture was incubated at 65°C for 10 min. To sequester SDS, Triton X-100 was thoroughly mixed in to a final concentration of 1%. To digest the DNA and obtain cross-linked DNA fragments, 400 U of BglII restriction enzyme (New England Biolabs) was added. The reaction was incubated overnight at 37°C under slow rotation. The following day, SDS was added to a concentration of 2%. The mixture was incubated at 65°C for 30 min. The ligation was done in a final volume of 7.5 ml, in the presence of 50 mM of Tris–HCl (pH 7.5), 10 mM of DTT, 10 mM of $MgCl_2$, 1% Triton X-100, 0.11 mg/ml of BSA, 1.07 mM of ATP and 4000 U of T4 DNA ligase (New England Biolabs). The reactions were incubated at 16°C for 4 h. To remove any residual proteins, 33 µl of Proteinase K (15 mg/ml; Roche Applied Science) was added to each tube and incubated at 65°C overnight. An additional 33 µl of Proteinase K was added and the mixture was incubated for 2 h at 42°C. Finally, the ligated DNA products were isolated using two phenol:chloroform (1:1) extractions, and one chloroform extraction in 7.5 ml volume, followed by ethanol precipitation. The pellets were resuspended in 1 ml of dH_2O containing RNAse A (Roche Applied Science) in a final concentration of 10 µg/ml. Following a 30 min incubation at room temperature, final phenol:chloroform and chloroform extractions were performed, followed by ethanol precipitation. The pelleted DNA was washed five times with 70% ethanol and dissolved in 400 µl of dH_2O. The concentration of template from each individual tissue sample was determined by agarose gel electrophoresis, quantifying the template band by comparing it to the known amount of the highest band of a 1-kb DNA ladder (NEB). Equal amounts of each template were pooled. All pools were diluted to the same concentration. The amount to use in a PCR detection reaction (1 µl) was determined empirically. For preparing the control DNA templates, Bacterial Artificial Chromosome (BAC) DNA templates corresponding to the human *MCS5A* and rat *Mcs5a* regions (human: RP11-656G18; rat: CH230-298P15 and CH230-303N12) were obtained from Children's Hospital Oakland Research Institute (CHORI). The BAC DNA was isolated using a Plasmid Maxi kit (QiaGen) and 500 ng was digested overnight at 37°C using 5 U of BglII restriction enzyme (NEB). The restriction fragments were recovered using phenol:chloroform and chloroform extractions, followed by ethanol precipitation. The digested BAC DNA was religated by adding 5 µl of 10× ligation buffer and 2 U of T4 DNA Ligase (Roche Applied Science). The mixture was incubated overnight at 16°C. The religated fragments were recovered by phenol:chloroform and chloroform extraction, followed by ethanol precipitation. The DNA concentration was measured by nanodrop and diluted to 10 ng/µl. A working stock of 1:1000 dilution was prepared. The amount to use in a reaction (0.02 µl) was empirically determined. Empirically determined DNA template was added to each reaction well and stored on ice. The PCR mixture was prepared on ice, and mixed into the pre-cooled DNA templates. The reaction was done in the presence of 1× Herculase reaction buffer, 0.2 mM of each NTP, 0.4 µM of each primer, 0.3 µl of Herculase Enhanced polymerase (5 U/µl; Stratagene) in a total volume of 25 µl. Primer sequences are listed in Supplementary Table S1. The PCR was performed using a Peltier Thermocycler (PTC-225; MJ Research), using the following cycling conditions: 95°C for 1 min, 35 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 20 s, followed by a final extension of 72°C for 8 min. When the program was complete, 5 µl of 6× loading buffer (15% w/v Ficoll-400, 0.25% w/v Xylene Cyanol FF) was added to the reaction and 12 µl of each reaction was analyzed by agarose gel electrophoresis. The gels were stained by soaking in a 1 µg/ml ethidium bromide solution for 10 min, and

destained in dH_2O for 15 min. An image was taken under UV-exposure using the Gel Doc XR system (Biorad). PCR band intensity was quantified using ImageQuant software (GE Healthcare).

Chromatin immunoprecipitation

The JURKAT (clone E6-1; $CD4^+$ leukemic T-lymphocytic cell line) cell line was obtained from ATCC and grown in GIBCO RPMI-1640 media (Invitrogen) supplemented with 10% decomplemented Hyclone Fetal Bovine Serum (FBS; Thermo Scientific) and 1% Penicillin/Streptomycin (Invitrogen). JURKAT cells were cultured to a density of $\sim 3 \times 10^6$ cells/ml. Roughly 2×10^6 JURKAT cells or splenic T cells from female susceptible congenic control and *Mcs5a* resistant congenic rats were fixed using formaldehyde to a final concentration of 1% at room temperature for 10 min. Fixation was followed by centrifugation (5 min, 800 g) to pellet the cells. The cells were washed twice with PBS. The chromatin immunoprecipitation (ChIP) assay was done using the ChIP Assay Kit (Upstate/Millipore) following manufacturer's recommendations with one exception: the SDS lysis buffer was reconstituted to contain 0.1% of SDS instead of 1%. For one assay 4×10^6 cells were resuspended in 200 µl of SDS-lysis buffer. Chromatin was sheared to 200- to 1000-bp fragments by sonication using three pulses of 10 s on the third power level of a Branson sonifier (Cell disruptor 185; Branson). Roughly 1% of the sonicated chromatin was set aside as the input (positive control) sample. Immunoprecipitations were done on the remainder of the sonicated chromatin using the following antibodies: rabbit monoclonal to CTCF (07–729; Millipore), rabbit polyclonal to Rad21 (cohesin; ab992; Abcam), and rabbit control IgG (negative control; ab46540; Abcam). Immunoprecipitations and input DNA were recovered by phenol:chloroform and chloroform extractions, followed by ethanol precipitation. Input- and immunoprecipitation-recovered DNAs were tested in a dilution series as templates in detection PCRs. Empirically determined DNA template was added to each reaction well and stored on ice. The PCR mixture was prepared on ice, and mixed into the pre-cooled DNA templates. The reaction was done in the presence of 1× Herculase reaction buffer, 0.2 mM of each NTP, 0.4 µM of each primer, 0.3 µl of Herculase Enhanced polymerase (5 U/µl) in a total volume of 20 µl. The PCR was performed using a Peltier Thermocycler, using the following cycling conditions for the human primers: 95°C for 1 min, 32 cycles of 95°C for 30 s, 58°C for 30 s, 72°C for 15 s, followed by a final extension of 72°C for 3 min, and the following cycling conditions for the rat primers: 95°C for 1 min, 32 cycles of 95°C for 30 s, 56°C for 30 s, 72°C for 15 s, followed by a final extension of 72°C for 3 min. Primer sequences are listed in Supplementary Table S1. When the program was complete, 5 µl of 6× loading buffer (15% w/v Ficoll-400, 0.25% w/v Xylene Cyanol FF) was added to the reaction and 10 µl of each reaction was analyzed by agarose gel electrophoresis. The gels were stained by soaking in a 1 µg/ml ethidium bromide solution for 10 min, and destained in dH_2O for 10 min. Image was taken under UV exposure.

RNA extraction and quantitative real-time PCR

Congenic female rats (11 weeks of age) of line WF.WKy (susc.), line WW (*Mcs5a*), line B3 (*Mcs5a1*) and line LL (*Mcs5a2*) were used as tissue donors. RNA was extracted from snap-frozen tissues and sorted cell samples using the MagMax-96 Total RNA isolation kit (Ambion) according to manufacturer's directions. To synthesize cDNA from 500 ng total RNA treated with TURBO-free DNase (Ambion), the reverse transcriptase Superscript II kit (Invitrogen) was used according to manufacturer's directions. Quantitative real-time PCR was used to quantify transcript levels. TaqMan quantitative PCR primers and probes (available in Supplementary Table S1) were either designed using Primer Express v 2.0 (ABI/Applied Biosystems) according to developer's specifications, or ordered as made-to-order assays (*Fbxo10* Rn01439970_m1; *Tomm5* Rn01436622_g1; *Mcart1* Rn01764629_g1; *Dcaf10* Rn01764617_m1; ABI/Applied Biosystems). A microliter of cDNA (~12.5 ng RNA equivalent cDNA) was used in a 16 µl TaqMan QPCR reaction. Reaction components were 1× TaqMan Buffer A (Applied Biosystems), 5.5 mM $MgCl_2$, 400 µM each dATP, dCTP, dGTP, dTTP and 0.4 U Taq Gold DNA Pol (ABI). For the custom designed assays, 500 nM each experimental primer, 200 nM TaqMan experimental probe, 60 nM each *Gapdh* primer, 120 nM rodent *Gapdh* probe, were used per reaction. For the made-to-order assays, 0.8 µl of the experimental assay and 0.4 µl of the *ActB* standard assay were used per reaction. Cycling conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. FAM (*Fbxo10, Frmpd1, Tomm5* probe) and VIC (Rodent *Gapdh or ActB* probe, ABI) fluorescence values were measured in real-time on an ABI 7900 real-time PCR machine using ABI SDS v2.2 software. Quantities of transcripts were measured by comparison of C_t values with a standard curve calculated from serial dilutions made from reverse transcriptase (RT) reactions that contained 2 µg of total RNA. Sample measurements are an average of four replicates. Sample measurements were normalized by dividing the gene-specific transcript quantity over the rodent *Gapdh* or *ActB* transcript quantity detected in the same reaction. Ratios were scaled to the ratio of the susceptible congenic control sample from the same experiment. Data were analyzed using Kruskal–Wallis or Mann–Whitney non-parametric tests. Transcript levels were only statistically compared if the samples were run in the same assay plate.

5′ RNA ligase mediated-rapid amplification of cDNA ends (5′ RLM-RACE)

To localize the exact transcriptional start site (TSS) of the *Fbxo10* transcripts, the 5′ RLM-RACE assay was performed using the FirstChoice RLM-RACE kit (Ambion) following manufacturer's recommendations. Thymic or splenic total RNA from five susceptible congenic

control and five *Mcs5a* resistant congenic rats was used. Equal amounts of RNA (as determined by Nanodrop quantification) were pooled and 10 μg of each pool was used as input RNA into the reaction. To identify the human *FBXO10* TSS, 10 μg total RNA from breast, thymus or spleen (Ambion) were used as input. To amplify the 5′-ends of the *Fbxo10*/*FBXO10* transcripts, a nested PCR was performed with the gene-specific primers (available in Supplementary Table S1) annealing to the translational start codon (ATG)-containing exon that is presumably shared by all *Fbxo10*/*FBXO10* transcripts. A typical PCR product yielded multiple bands on an agarose gel, indicative of multiple TSS. The products were blunt-cloned using the ZEROBLUNT vector system (Invitrogen) for subsequent resequencing. In total, 150 individual rat thymus or spleen TSS products and 54 individual human thymus, spleen or breast TSS products were analyzed by resequencing. The subcloned sequences were identified by performing BLAT searches using the UCSC genome browser.

Vector construction, transient transfection, luciferase assays and Bradford assays

Genomic fragments with the appropriate polymorphisms were amplified from *MCS5A1* and *MCS5A2* heterozygous genomic DNA using the Platinum HiFi Taq polymerase (Invitrogen) and primers containing the restriction sites (BamHI, SalI) used for cloning (available in Supplementary Table S1). First, the *FBXO10*-TSS constructs were made by inserting the BamHI-, SalI-digested genomic *FBXO10*-TSS fragment using the XhoI and BglII site located upstream of Luc^{+} in pGL3-Basic (Promega). The S and R versions of the *FBXO10*-TSS construct were digested with BamHI and SalI to insert the BamHI-, SalI-digested genomic fragments containing the *MCS5A2* polymorphisms. To ensure integrity, all constructs underwent visual inspection by restriction enzyme digestion with NotI and SacI (Supplementary Figure S2), as well as resequencing of both inserts. The JURKAT (clone E6-1; $CD4^{+}$ leukemic T-lymphocytic cell line) cell line was obtained from ATCC and grown in GIBCO RPMI-1640 media (Invitrogen) supplemented with 10% decomplemented Hyclone Fetal Bovine Serum (FBS; Thermo Scientific) and 1% Penicillin/Streptomycin (Invitrogen). JURKAT cells were cultured to a density of $\sim 3 \times 10^6$ cells/ml. The cells were diluted into fresh prewarmed RPMI/FBS media to a density of 0.5×10^6 cells/ml. A total of 500 μl cells (approximately 250 000 cells) were distributed into each well of a 24-wells culture plate. The cells were allowed to rest for at least 3 h. Transient transfection was performed using LTX Lipofectamine reagent (Invitrogen) according to manufacturer's recommendations. The DNA concentration of each construct was verified before each transfection. All DNA concentrations were between 90 and 110 ng/μl. For each transfection, 500 ng of construct DNA was diluted with a mixture of 100 μl GIBCO OPTI-MEM reduced serum media (Invitrogen), 5 μl of PLUS reagent (Invitrogen) and 5 ng of Renilla internal control vector pRL-TK (Promega). Following a 5-min incubation at room temperature, 5 μl of LTX Lipofectamine was added. The transfection solution was mixed by pipetting up-and-down, followed by a 30-min incubation at room temperature. A total of 100 μl of the transfection solution was added to each well containing the cells. The cells and the transfection solution were mixed by gently shaking the plate in multiple directions. The cells were cultured for 44–48 h. The Luciferase assay was performed using the Dual-Luciferase Reporter Assay kit (Promega) according to the manual. The cells were harvested, spun down and washed with PBS. The cell pellet was lysed in 100 μl 1× Passive Lysis Buffer and incubated at room temperature for 15 min. An amount of 20 μl of the lysate was measured in a 12 × 75 mm disposable culture tube (Fisher Scientific). Each transfection was measured three times, using a Luminometer Monolight 3010 (BD Biosciences). The Luminometer was set as follows: inject 100 μl of the Luciferase substrate, 2 s delay, 10 s of measuring time, inject 100 μl of the Stop-and-Glo solution, 2 s delay, 10 s measuring time. The remainder of each lysates was used to quantify total protein content using Bradford assays. To determine total protein content of the lysate, 5 μl of the lysate was mixed with 250 μl 1× QuickStart Bradford Dye reagent (Biorad) in a 96-well flat bottom titer plate. The plate was incubated for 15 min at room temperature. The optical density (OD) at 560 nm wavelength was determined using a Spectrophotometer for microtiter plates. The OD was correlated to a standard curve derived from a dilution series using Bovine Serum Albumine (BSA). Relative Luciferase/Renilla activity of each construct was determined by dividing the Luciferase/Bradford ratio over the Renilla value. Each construct was subsequently normalized against the relative Luciferase activity of the pGL3-control construct present in each 24-well plate. The SS and RR version of each construct were always present in the same 24-well plate. The normalized values for the SS and RR version of each construct from at least six transfections were tested for a significant difference using the Wilcoxon rank sum test.

RESULTS

The Chromosome conformation capture assay identifies a conserved higher order chromatin loop within *Mcs5a*/*MCS5A*

To investigate if chromatin looping could provide a mechanistic basis for the synthetic (genetic) interaction between the genetic elements in *Mcs5a1* and *Mcs5a2*, we applied the chromosome conformation capture (3C) assay to the rat and human *Mcs5a*/*MCS5A* locus (Figure 1A). The 3C assay measures the relative (looping) interaction frequency of crosslinking captured, enzymatically digested (BglII), ligated chromatin fragments (16). For the rat *Mcs5a* locus, the fixed fragment was initially chosen to encompass the genomic region orthologous to the 5.7-kb region in the human genome harboring the correlated polymorphisms in *MCS5A1* that are associated with breast cancer risk (5). This fixed fragment also contained the

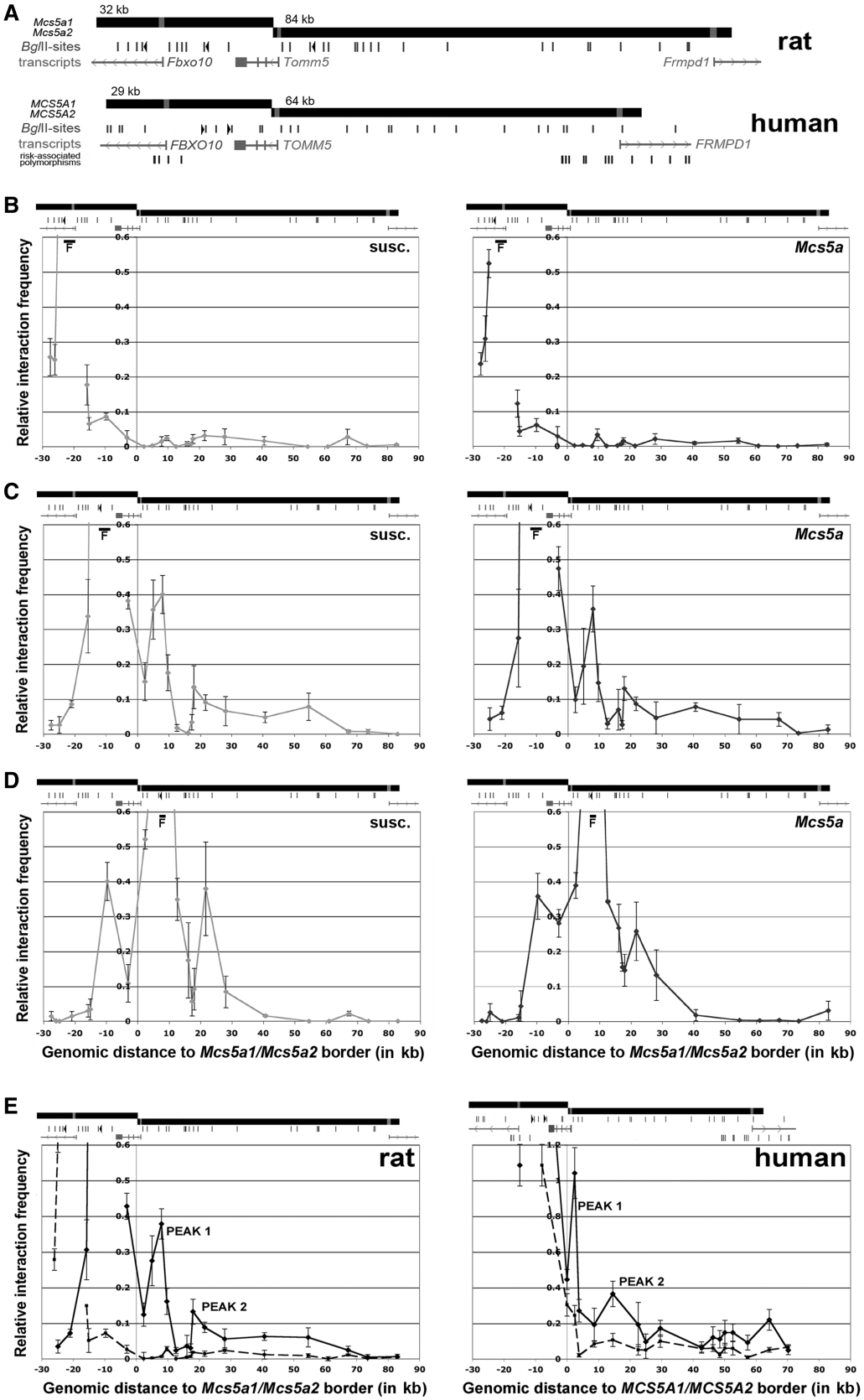

Figure 1. Analysis of the higher order chromatin structure of rat and human *Mcs5a*/*MCS5A*. (**A**) Maps of the rat and human *Mcs5a*/*MCS5A* locus. The relative position of *Mcs5a1*/*MCS5A1*, *Mcs5a2*/*MCS5A2*, all BglII restriction site, the transcripts *Fbxo10*/*FBXO10*, *Tomm5*/*TOMM5* and *Frmpd1*/*FRMPD1*, and the human breast cancer risk-associated variants is indicated. The primers matching the fixed fragments used to generate panels (**B–E**) are indicated by black vertical triangles. (B–D) 3C profiles of the *Mcs5a* locus in T cells of susceptible congenic control (susc.)

(continued)

CpG island associated with a predicted *Fbxo10* promoter. Relative interaction frequencies with the majority of BglII fragments in *Mcs5a* were determined. Restriction fragments located close to the fixed fragment yielded a high relative interaction frequency that decreased with increasing genomic distance, indicative of random ligation, but no looping (Figure 1B). Similarly, taking a nearby BglII fragment (two BglII fragments equaling ~5 kb upstream of the one containing the predicted *Fbxo10* promoter) as the fixed fragment yielded no looping (Supplementary Figure S1A and B). However, when the fixed fragment was shifted further upstream (five BglII fragments equaling ~10 kb) of the one containing the predicted *Fbxo10* promoter, two areas of locally increased relative interaction frequency were identified, indicative of looping (Figure 1C). Both looping areas are located in the *Mcs5a2* region, close to the *Mcs5a1–Mcs5a2* border. For confirmation, a BglII fragment in the first looping area was chosen as the fixed fragment. The profile showed the same areas of looping, namely the previous fixed fragment in *Mcs5a1* and the second looping area in *Mcs5a2* (Figure 1D). Furthermore, none of the loops appeared to be influenced by the susceptible or resistant *Mcs5a* genotype as the 3C profiles show the same trends. All 3C experiments described above were performed on primary T cells, as we consider this a specific cell type of *Mcs5a* activity (15). The spatial organization of the *Mcs5a* locus in the mammary gland was found to be similar to the profile in T cells, as the same regions were found to be involved in the chromatin looping interactions (Supplementary Figure S1C and D).

The 3C profile of the orthologous human *MCS5A* locus was found to be highly similar to rat *Mcs5a* (Figure 1E). No looping was found with the fixed fragment containing the *MCS5A1* correlated polymorphisms that associate with breast cancer risk (Figure 1E). This fixed fragment also contained the CpG island associated with the predicted *FBXO10* promoter. With the fixed fragment shifted upstream (two BglII fragments equaling ~10 kb) of the CpG island associated with the predicted *FBXO10* promoter, two areas of looping were found to be located in *MCS5A2*, close to the *MCS5A1–MCS5A2* border (Figure 1E). The similarities in the 3C profiles of the rat and human *Mcs5a*/*MCS5A* suggest that the higher order chromatin structure of *Mcs5a*/*MCS5A* is evolutionary conserved.

CTCF and cohesin binding is confirmed in all interacting chromatin looping fragments

The position of the looping elements at either side of the *Tomm5*/*TOMM5* gene, thus spatially isolating this gene, suggests a function as an insulator loop. The CCCTC -binding factor (CTCF) protein is widely known for its role as a vertebrate insulator protein (17). Additionally, CTCF has been shown to be essential for long-distance enhancer–promoter looping (18). Recently, the cohesin protein complex has also been shown to be essential for long-range chromatin looping in the developmentally controlled *IFNG* locus (19). We sought to investigate whether CTCF and/or cohesin binding could also underlie the observed higher order chromatin structure of *MCS5A*/*Mcs5a*. Using a ChIP assay with antibodies against CTCF and Rad21, a DNA-binding subunit of the cohesin complex, association of both CTCF and cohesin to all three interacting chromatin looping fragments was confirmed in both a human T-lymphocytic cell line (JURKAT, Figure 2A) and rat primary T cells (Figure 2B). No evidence of CTCF or cohesin binding was found in a location outside of the looping fragments (Figure 2A and B). CTCF binding was confirmed on a site in the *H19* locus previously shown to bind CTCF in human (20) and rat (21). The location of the CTCF-binding site in the looping fragment *MCS5A1* and the first looping fragment in *MCS5A2* coincides with CTCF binding previously identified in a genome-wide ChIP-seq study (22).

To test if CTCF is necessary for the higher order chromatin structure, a short-interfering RNA (siRNA) approach was undertaken to deplete primary rat T cells with CTCF. Nucleofection of siRNA against CTCF resulted in a significant reduction of CTCF transcript level (measured after 64–68 h) as compared with nucleofection with control siRNAs (Figure 2C). Interestingly, the transcript level of *Tomm5*, the gene residing within the looped structure was also significantly reduced after nucleofection with siRNAs against CTCF (Figure 2D). The transcript levels of a housekeeping gene, *Gapdh* and two genes close to the *Mcs5a* locus, *Mcart1* and *Dcaf10* were not affected by CTCF depletion (Figure 2D). Using the ChIP assay a reduction of enrichment of the *Mcs5a1* 3C looping fragment in the cells nucleofected with siRNAs against CTCF as compared with the cells nucleofected with control siRNAs was observed, consistent with CTCF depletion (Figure 2E).

Figure 1. Continued
and *Mcs5a* resistant congenic (*Mcs5a*) rats. The position of the fixed fragments is indicated with a black bar and F, and the primer within the fixed fragment is indicated with a vertical triangle. In (B) the fixed fragment contains the CpG island associated with the predicted promoter of *Fbxo10* in *Mcs5a1*. In (C) the fixed fragment is located five BglII restriction fragments (~10 kb) upstream of the previous fixed fragment that contains the CpG island in *Mcs5a1*. In (D) the fixed fragment was located in the first looping fragment in *Mcs5a2*. (**E**) Comparison of the rat and human *Mcs5a*/*MCS5A* 3 C profile. Dashed lines represent the 3C profile with the fixed fragment containing the predicted *Fbxo10*/*FBXO10* promoter in *Mcs5a1*/*MCS5A1* and does not display looping to *Mcs5a2*/*MCS5A2*. In the human, this fixed fragment also contains the four correlated *MCS5A1* polymorphisms associated with breast cancer risk. Solid lines represents the 3C profile with the fixed fragment located five (rat) or two (human) BglII restriction fragments (~10 kb) upstream of the BglII fragment containing the predicted *Fbxo10*/*FBXO10* promoter in *Mcs5a1*/*MCS5A1*. The chromatin fragments in *Mcs5a2*/*MCS5A2* looping to the fixed fragment in *Mcs5a1*/*MCS5A1* are indicated with PEAK 1 and PEAK 2. Graphed are the average ± SEM relative interaction frequencies of a fixed BglII fragment with other BglII fragments in *Mcs5a*/*MCS5A*. The genomic distance (in kilobases) represents the distance of the midpoint of a BglII fragment to the *Mcs5a1*/*MCS5A1*-*Mcs5a2*/*MCS5A2* border.

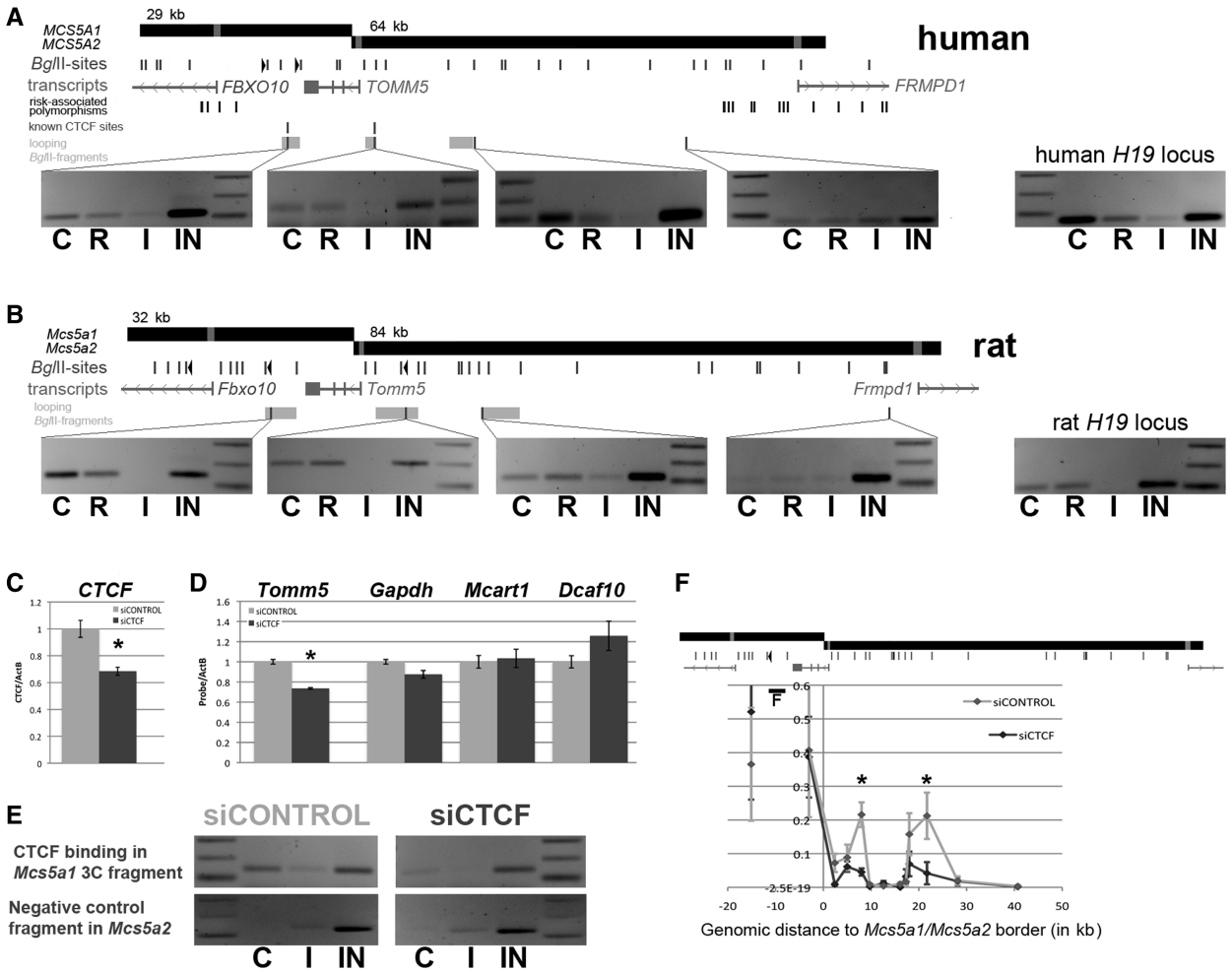

Figure 2. CTCF and cohesin bind to the looping fragments in *Mcs5a*/*MCS5A* and CTCF is necessary for the higher order chromatin structure. (**A** and **B**) The *Mcs5a1*/*MCS5A1* and *Mcs5a2*/*MCS5A2* loci are depicted as black lines. The light gray bars within the black lines are the CpG islands associated with the promoters of the *Fbxo10*/*FBXO10*, *Tomm5*/*TOMM5* and *Frmpd1*/*FRMPD1* genes, respectively. The locations of the three interacting chromatin looping fragments identified in the 3C assay are indicated by light gray blocks. In the human, the location of known CTCF sites from a genome-wide CTCF ChIP-seq study is indicated (22). Amplicons within and outside the looping *Mcs5a*/*MCS5A* fragments, as well as an amplicon in the *H19* locus (as a positive control) were analyzed by PCR on CTCF (**C**), cohesin (R; Rad21) and IgG (I; negative control) antibody immunoprecipitated chromatin samples, and an input (IN, positive control) sample, prepared from JURKAT cells (human) or primary rat splenic T cells (rat). Each gel image is accompanied by a 100-bp DNA ladder of which the lower three bands (100, 200, 300 bp) are shown. CTCF and cohesin binding was found in all looping fragments and the *H19* positive control locus, but not in the *Mcs5a2* fragment not located in a looping fragment. (**C** and **D**) Transcript level of *CTCF*, *Tomm5, Gapdh, Mcart1 and Dcaf10* (normalized to *ActB*) measured 64–68 h after nucleofection of rat primary T cells with control siRNAs (siCONTROL; light gray bars) and siRNAs against CTCF (siCTCF; dark gray bars). The transcript level of *CTCF* and *Tomm5* was significantly reduced (indicated with an asterisk) in the siCTCF-treated T cells. The transcript levels of the housekeeping gene *Gapdh* and genes *Mcart1* and *Dcaf10* located adjacent to the *Mcs5a* locus were not affected. (**E**) ChIP analysis of CTCF binding to the looping fragment in *Mcs5a1* and the negative control fragment in *Mcs5a2* in primary rat T-cells 64–68 h after nucleofection with control siRNAs and siRNAs against CTCF. CTCF binding to the *Mcs5a1* looping fragment was reduced in the siCTCF-treated T cells. (**F**) 3C analysis of the *Mcs5a* locus in primary rat T cells 64–68 h after nucleofection with control siRNAs (light gray line) and siRNAs against CTCF (dark gray line). The looping fragment in *Mcs5a1* was taken as the fixed fragment (indicated with a black bar and F). Looping of the fixed fragment to the looping fragments in *Mcs5a2* was significantly reduced (indicated with an asterisk) in the siCTCF-treated T cells.

Finally, using the 3C assay a significant reduction in relative interaction frequency between the *Mcs5a1* and first *Mcs5a2* looping fragments and between the *Mcs5a1* and second *Mcs5a2* looping fragments was observed in the cells nucleofected with siRNAs against CTCF, as compared with cells nucleofected with control siRNAs (Figure 2F). We conclude that CTCF is necessary for the higher order chromatin structure of *Mcs5a*. Disruption of this structure resulted in lower transcript level of *Tomm5*.

Mcs5a controls the transcript level of *Fbxo10* in T cells

We previously reported that the non-protein-coding resistant *Mcs5a* allele is associated with lower *Fbxo10* and higher *Frmpd1* transcript levels in thymus and spleen, respectively (5), and that the resistance phenotype is mediated by the immune system (15). Here, we explored by quantitative real-time PCR using TaqMan assays, if the difference in transcript levels of these two genes, like the mammary carcinoma resistance phenotype, is controlled by the two synthetically interacting genetic

elements of *Mcs5a*. Hence, thymic *Fbxo10* and splenic *Frmpd1* transcript levels were compared between the susceptible congenic control line (WF-homozygous at *Mcs5a*), the *Mcs5a* resistant congenic line (WKy-homozygous at *Mcs5a*) and the *Mcs5a1* and *Mcs5a2* susceptible congenic lines (WKy-homozygous at *Mcs5a1* or *Mcs5a2*, respectively). The *Fbxo10* transcript level was significantly reduced ($P = 0.01$) in the thymus of the *Mcs5a* congenic resistant animals and was not significantly different in thymus tissue from either the *Mcs5a1* or *Mcs5a2* susceptible congenic lines when compared to the susceptible congenic control line (Figure 3A). Thus, having lower thymic *Fbxo10* transcript levels required the same synthetic interaction between *Mcs5a1* and *Mcs5a2* as was necessary for the mammary carcinoma resistance phenotype. Conversely, higher *Frmpd1* transcript levels in the spleen did not require the synthetic interaction, as the *Mcs5a* resistant congenic line, and both *Mcs5a1* and *Mcs5a2* susceptible congenic lines had higher splenic *Frmpd1* transcript levels as compared with the susceptible congenic control line (Figure 3A).

Subsequently, *Fbxo10* transcript levels were monitored in various classes of sorted and unsorted thymocytes (Figure 3B). The reduction in *Fbxo10* transcript levels associated with the resistant *Mcs5a* allele was observed in unsorted ($P = 0.03$), $CD4^+CD8^+$ ($P = 0.02$), $CD4^+CD8^-$ ($P = 0.03$) and $CD4^-CD8^+$ ($P = 0.002$), but not in $CD4^-CD8^-$ thymocytes ($P = 0.77$). The *Fbxo10* transcript level difference was also found in $CD3^+$ T cells sorted from the spleen ($P = 0.001$), whereas other classes of sorted and unsorted splenocytes did not display the differential transcript levels (Figure 3C). The *Fbxo10* transcript level difference in splenic T cells persisted upon culturing ($P = 0.0003$) and concanavalin A (conA) stimulation ($P = 0.01$; Figure 3D).

C9ORF105/*C9orf105* is a spliced transcript located in the 60-kb sequence between the human *MCS5A1* and *MCS5A2* risk-associated polymorphisms on the rat *Mcs5a1–Mcs5a2* border and was recently annotated as the translocase of outer mitochondrial membrane 5 gene, *TOMM5*/*Tomm5*. We checked if the transcript level of *C9orf105* is associated with the resistant rat *Mcs5a* allele in the immune system. *Tomm5* transcript levels were not different between susceptible congenic control and *Mcs5a* resistant congenic rats in whole spleen ($P = 0.18$) and thymus ($P = 0.40$) tissues (Figure 3E). Similarly, in primary T cells sorted from the spleen, *Tomm5* was not differentially expressed ($P = 0.94$; Figure 3E). In cultured T cells from susceptible congenic control and *Mcs5a* resistant congenic rats, the transcript levels of *Tomm5* were found to be different ($P = 0.008$; Figure 3F). In conA-stimulated cultured T-cells *Tomm5* transcript levels were not found to be different ($P = 0.95$) between the *Mcs5a* genotypes. Interestingly, in conA-stimulated T cells, *Tomm5* transcript levels were found to be strongly increased (~3-fold) when compared to unstimulated cultured T-cell samples from the rat lines harboring both *Mcs5a* genotypes ($P < 0.0001$ for both the susceptible and the resistant congenic line; Figure 3F). For comparison, *Fbxo10* transcript levels in conA-stimulated T cells were not different ($P = 0.45$) when compared to the unstimulated cultured T-cell samples from the susceptible congenic control line, but were increased (~1.5-fold, $P = 0.003$) in the resistant congenic line (Figure 3D). It should also be noted that *Frmpd1* failed to amplify in TaqMan assays on primary, cultured unstimulated and cultured activated T-cell samples (data not shown).

Taken together, the gene expression data implicate that lower *Fbxo10* transcript levels, and not *Frmpd1* and *Tomm5* transcript levels in the immune system are associated with the *Mcs5a* resistant allele. Furthermore, these data suggest that *Tomm5* and not *Fbxo10* and *Frmpd1* may play a role in T-cell activation.

The TSS cluster of *FBXO10*/*Fbxo10* is located in *MCS5A1*

To understand the regulation of the *Fbxo10*/*FBXO10* gene it is important to know the exact TSS. This facilitates localization of the putative promoter. Two areas of transcriptional initiation of the *Fbxo10*/*FBXO10* gene in rats and humans are annotated in the UCSC genome browser, namely the CpG island in *Mcs5a1*/*MCS5A1* and an area close to the *Mcs5a1*/*MCS5A1–Mcs5a2*/*MCS5A2* border (Figure 4A). To identify the TSS of the *Fbxo10* gene, we performed the 5′ RLM-RACE assay (RNA Ligase Mediated-Rapid Amplification of cDNA Ends). The assay makes use of an RNA adapter that is ligated to the de-CAP-ped 5′-end of transcripts, followed by a RT reaction to make cDNA. To amplify just the 5′-ends of the *Fbxo10* gene, a nested PCR reaction was performed with primers annealing to the translational start codon (ATG)-containing exon and universal primers annealing to the 5′ RNA-adapter. The assay was done on pools of thymus or spleen RNA from four susceptible congenic control and four *Mcs5a* resistant congenic rats and on human RNA samples from thymus, spleen and breast tissue. A typical PCR product yielded multiple bands on an agarose gel, indicative of multiple TSSs. In total, 150 rat and 54 human TSS clones were sequenced to elucidate the exact start position of the *Fbxo10*/*FBXO10* transcripts. In the rat, 14 TSS positions were found, all located in the CpG island of the *Mcs5a1* locus (Figure 4B). In the human, three TSS positions were found, again, located in the CpG island of the *MCS5A1* locus (Figure 4B). It should be noted that in the human breast RNA sample the *FBXO10* TSS cluster was found to be located close to the *MCS5A1-MCS5A2* border, indicating that the assay was able to identify such cases. Identification of the *Fbxo10*/*FBXO10* TSS cluster in the CpG island of *Mcs5a1*/*MCS5A1* suggests that potential regulatory elements (e.g. promoter or enhancers) may be located in the same region. The human TSS cluster is located within 150 bp of SNP rs6476643 (Figure 4B), which is one of the breast cancer risk-associated SNPs found in *MCS5A1* (5).

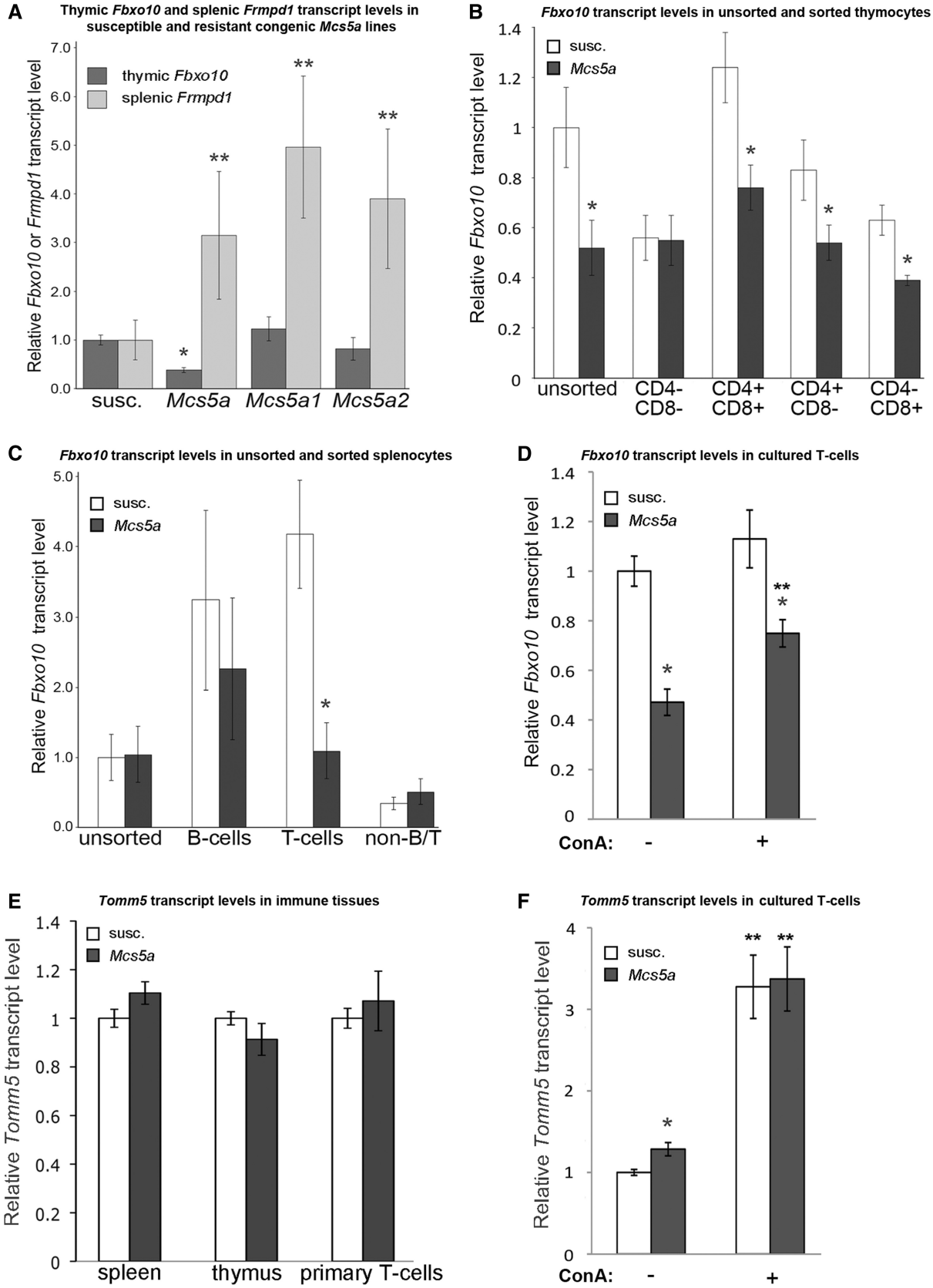

Figure 3. Lower *Fbxo10* transcript level in T cells is associated with the mammary carcinoma resistance phenotype mediated by *Mcs5a*. (**A**) The thymic *Fbxo10*, and not splenic *Frmpd1* transcript level is associated with the resistant *Mcs5a* allele. Graphed are average ± SEM thymic *Fbxo10* (dark gray) and splenic *Frmpd1* (light gray) transcript levels (normalized to *Gapdh*) scaled to the thymic *Fbxo10* or splenic *Frmpd1* transcript level of the susceptible congenic control line. Significantly ($P < 0.05$) different *Fbxo10* and *Frmpd1* transcript levels as compared with the susceptible congenic

(continued)

Reporter constructs identify functional interactions between the *MCS5A1* and *MCS5A2* risk-associated polymorphisms

In rat T cells, the *Fbxo10* transcript level is under control of the synthetically interacting resistant *Mcs5a* genetic elements. This observation led to the hypothesis that the polymorphisms of the resistant risk-associated alleles of *MCS5A1* and *MCS5A2* could interact similarly to downregulate *FBXO10* transcript levels in human T cells. To investigate this, Luciferase reporter assays were performed using a series of constructs containing selected *MCS5A1* and *MCS5A2* alleles. All constructs described below were visually inspected by restriction enzyme digest to ensure integrity (Supplementary Figure S2). A human T-lymphocytic cell line (JURKAT) was transiently transfected with each construct in the presence of a Renilla expressing vector as a control for transfection efficiency.

First, a 1464-bp fragment of *MCS5A1* including the previously identified *FBXO10* TSS cluster was inserted upstream of the *Luciferase* (Luc^+) reporter gene in the pGL3-basic vector (Figure 5A). Two versions of the construct were made (Figure 5B), one containing the resistant (R) allele of rs6476643 (G) and one containing the susceptible (S) allele of rs6476643 (T). The transcriptional activity of the R and S version of the *FBXO10* TSS-Luciferase construct is not significantly different ($P = 0.45$; Figure 5C).

Subsequently, the R and S *FBXO10* TSS-Luciferase constructs were modified to harbor *MCS5A2* fragments containing the resistant or susceptible allele of one or two *MCS5A2* SNPs (Figure 5A). The *MCS5A2* fragments were inserted downstream of Luc^+ (Figure 5B). The 15 correlated *MCS5A2* polymorphisms were integrated into 10 *MCS5A2* fragments of on average ~990 bp in length (Supplementary Table S2). This yielded a series of 10 constructs, each present in two versions, namely SNP rs6476643 susceptible plus *MCS5A2* susceptible (SS) and SNP rs6476643 resistant plus *MCS5A2* resistant (RR). The transcriptional activity of the entire series of 10 RR constructs is significantly lower compared to the entire series of 10 SS combination constructs (Wilcoxon rank test $P < 10^{-48}$; Figure 5D).

Next, the transcriptional activity of the SS and RR versions was compared for each construct (Figure 5D). The transcriptional activity was not significantly different between the SS and RR versions of constructs 1, 2, 5, 7, 8 ($P > 0.05$, Supplementary Table S2). However, the activity of constructs 3, 4, 6, 9 and 10 were significantly lower in the RR version compared to the SS version ($P < 0.05$, Supplementary Table S2). In addition, the *MCS5A2* SNPs rs62534439 (from construct 4), and rs62534443 and rs62534444 (from constructs 5 and 6) were also represented in constructs 4a and 5a, respectively, again resulting in lower activities for the RR versions compared to the SS versions (Supplementary Figure S3), although only significant for construct 4a ($P < 0.05$, Supplementary Table S2).

The *MCS5A2* fragments in construct 4, 6 and 10 (yielding lower activity in the RR combination constructs) were also tested on their own for transcriptional activity by transferring them to the cloning site upstream of Luc^+ (instead of the *FBXO10* TSS-containing fragment). All of these constructs had lower transcriptional activity as compared with the *FBXO10* TSS-containing fragment (Supplementary Figure S4). The S and R version of these constructs were not significantly different (Supplementary Figure S4), suggesting that the R allele of the polymorphisms in these fragments act only when combined with the R allele of the *MCS5A1* SNP rs6476643 in the *FBXO10* TSS-containing fragment. The observation that the *MCS5A1* and *MCS5A2* R alleles functionally interact to downregulate transcriptional activity is consistent with our observation of lower *Fbxo10* transcript levels in T cells of *Mcs5a* resistant congenic animals compared to susceptible congenic control animals.

Figure 3. Continued
control line are indicated with one and two asterisks, respectively. Sample sizes were: susc., $n = 9$; *Mcs5a*, $n = 6$; *Mcs5a1*, $n = 9$; *Mcs5a2*, $n = 8$. (**B**) Lower *Fbxo10* transcript level associated with the resistant *Mcs5a* allele is observed in unsorted and sorted thymocytes, except for double negative ($CD4^-CD8^-$) thymocytes. Graphed are average ± SEM *Fbxo10* transcript levels (normalized to *Gapdh*) in unsorted and sorted thymocytes, scaled to the average *Fbxo10* transcript level for the unsorted thymocytes of the susceptible congenic control line. Sample sizes for the susceptible congenic control line and *Mcs5a* congenic resistant line were, respectively: unsorted, $n = 16$ and $n = 9$; $CD4^-CD8^-$, $n = 4$ and $n = 4$; $CD4^+CD8^+$, $n = 17$ and $n = 19$; $CD4^+CD8^-$, $n = 15$ and $n = 18$; $CD8^+CD4^-$, $n = 18$ and $n = 17$. (**C**) Lower *Fbxo10* transcript level associated with the resistant *Mcs5a* allele is observed in splenic T cells, but not in other splenocytes. Graphed are average ± SEM *Fbxo10* transcript levels (normalized to *Gapdh*) in unsorted and sorted splenocytes, scaled to the *Fbxo10* transcript level for the unsorted splenocytes of the susceptible congenic control line. Sample sizes for the susceptible congenic control line and *Mcs5a* resistant congenic line were, respectively: unsorted, $n = 16$ and $n = 12$; B cells, $n = 9$ and $n = 8$; T cells, $n = 16$ and $n = 13$; non-B-/non-T cells, $n = 3$ and $n = 5$. (**D**) Lower *Fbxo10* transcript level associated with the resistant *Mcs5a* allele is observed in cultured unstimulated and conA-stimulated T cells. Graphed are average ± SEM *Fbxo10* transcript levels (normalized to *ActB*) in cultured T cells, unstimulated (−) and stimulated (+) with conA, scaled to the average *Fbxo10* transcript level of the unstimulated sample of the susceptible congenic control line. Sample sizes for the susceptible congenic control line and *Mcs5a* resistant congenic line were, respectively: unstimulated, $n = 11$ and $n = 10$; conA-stimulated, $n = 11$ and $n = 12$. (**E**) *Tomm5* transcript levels in the immune system are not associated with the resistant *Mcs5a* allele. Graphed are average ± SEM *Tomm5* transcript levels (normalized to *ActB*) in spleen, thymus and primary T cells, scaled to the average *Tomm5* transcript level of the susceptible congenic control line. Sample sizes for the susceptible congenic control line and *Mcs5a* resistant congenic line were, respectively: spleen, $n = 7$ and $n = 7$; thymus, $n = 8$ and $n = 8$; primary T cells, $n = 12$ and $n = 9$. (**F**) *Tomm5* transcript levels in cultured T cells are strongly increased after conA-stimulation. Graphed are average ± SEM *Tomm5* transcript levels (normalized to *ActB*) in cultured T cells, unstimulated and stimulated with conA, scaled to the average *Tomm5* transcript level of the unstimulated sample of the susceptible congenic control line. Sample sizes for the susceptible congenic control line and *Mcs5a* resistant congenic line were, respectively: unstimulated, $n = 11$ and $n = 10$; conA-stimulated, $n = 11$ and $n = 12$. In B–F, significantly different transcript level ($P < 0.05$) between the susceptible congenic control line (susc.; open bars) and the *Mcs5a* resistant congenic line (*Mcs5a*; filled bars) are indicated with an asterisk. In D and F, significantly different *Fbxo10* or *Tomm5* transcript level ($P < 0.05$) between samples without and with conA stimulation is indicated with two asterisks.

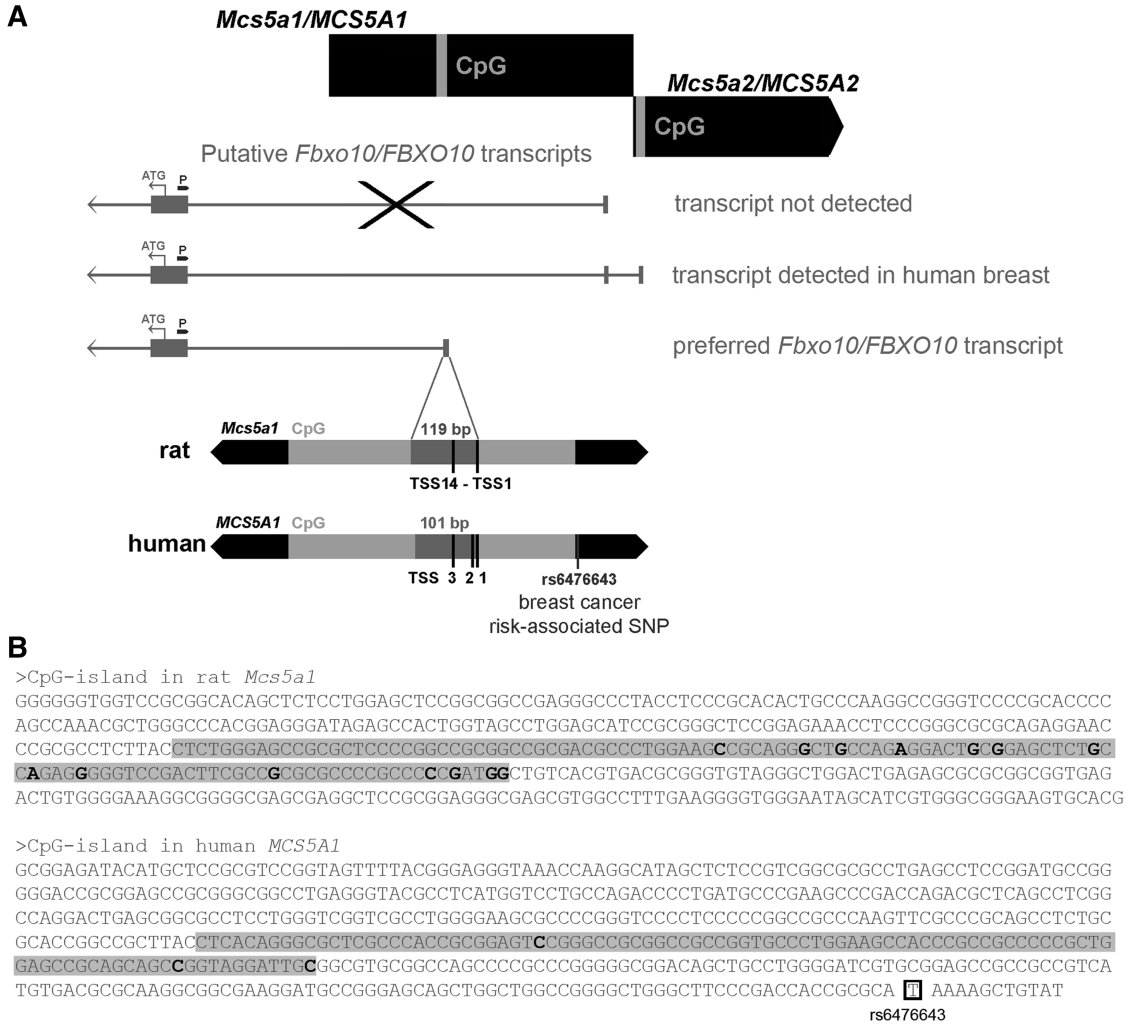

Figure 4. *Fbxo10*/*FBXO10* TSS analysis by the 5′ RLM-RACE assay. (**A**) Schematic representation of three putative *Fbxo10*/*FBXO10* transcripts annotated in the UCSC genome browser, indicated relative to the position of the *Mcs5a1*/*MCS5A1* locus (in black). The location of the *Fbxo10*/*FBXO10* specific primers in the first coding exon is indicated (P). The 5′ RLM-RACE revealed a cluster of 14 rat *Fbxo10* TSSs within the *Mcs5a1* CpG island (indicated in light gray). A cluster of three human *FBXO10* TSSs in the orthologous *MCS5A1* CpG island (indicated in light gray) was identified. The position of the first non-coding exon of *Fbxo10*/*FBXO10* within the CpG island is indicated in dark gray. The human TSS cluster was found to be located at 150-bp distance from breast cancer risk-associated SNP rs6476643. The transcripts not identified in rat or human immune tissue are indicated with a X. (**B**) Sequence information of the rat and human *Mcs5a1*/*MCS5A1* CpG island, first non-coding *Fbxo10*/*FBXO10* exon (highlighted in gray) and TSS identified (in bold). In the human sequence, the breast cancer risk-associated SNP rs6476643 (T/G) is indicated.

DISCUSSION

MCS5A/*Mcs5a* is a human/rat-conserved breast cancer susceptibility locus. In the rat, the resistant *Mcs5a* allele has previously been shown to consist of two interacting genetic elements that must be located on the same chromosome to elicit mammary carcinoma resistance (5). The carcinoma susceptibility phenotype of *Mcs5a* is not transferable from donor to recipient animals in a mammary gland transplantation assay, indicating that *Mcs5a* acts in a non-mammary cell-autonomous fashion (15). We have also demonstrated that carcinoma development is influenced by the immune system, as *Mcs5a* resistant congenic rats reconstituted with the susceptible congenic control line's immune system develop more mammary carcinomas as compared with *Mcs5a* resistant congenic rats reconstituted with their own immune system. Finally, we have observed that T-cell homeostasis and functions are under control of the *Mcs5a* locus (15).

In this study, we begin to translate the association of genetic variants with the complex trait breast cancer susceptibility into molecular mechanisms. As the genetic variants underlying the *MCS5A*/*Mcs5a* locus are found in non-protein-coding sequences, we sought to investigate the gene regulatory function of *MCS5A*/*Mcs5a*. We showed previously that the expression level of genes located within 1 Mb surrounding the locus, including *Fbxo10*, *Frmpd1* and *Tomm5* that are (partially) located within *Mcs5a*, were not differentially expressed in the

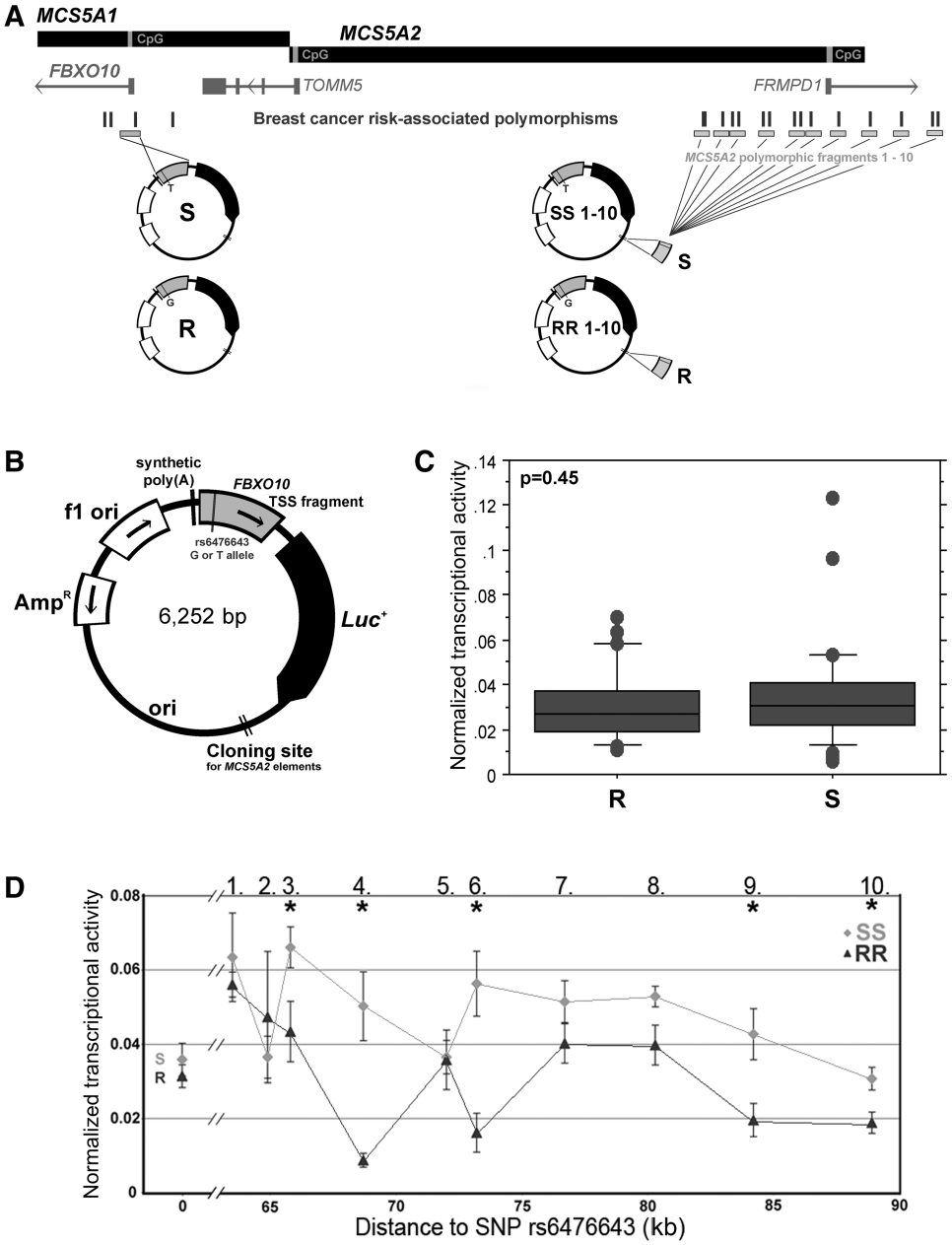

Figure 5. Transcriptional activity analysis of the human breast cancer risk-associated susceptible and resistant alleles in Luciferase reporter assays. (**A**) Schematic representation of the position of the genomic fragments derived from the *MCS5A1* and *MCS5A2* loci combined into the Luciferase constructs. The *MCS5A1* and *MCS5A2* loci are shown as black lines. The light gray bars within the black lines represent the CpG islands located in the locus. The three genes *FBXO10*, *FRMPD1* and *TOMM5* are shown in dark gray. The breast cancer risk-associated polymorphisms are represented as vertical gray lines. The fragments subcloned into the reporter constructs are indicated as horizontal light gray bars. The susceptible alleles of the *MCS5A2* polymorphisms were combined with the susceptible allele of the *MCS5A1* SNP rs6476643 (SS constructs 1–10). Similarly, the resistant alleles were combined (RR constructs 1–10). (**B**) Map of the *FBXO10* TSS-Luciferase reporter construct. A *MCS5A1* fragment containing the *FBXO10* TSSs and the risk-associated SNP rs6476643 was inserted upstream of the Luciferase reporter gene (Luc^+) in reverse genomic orientation. Two versions of the construct were created, namely having the susceptible (S; T allele) or resistant (R; G allele) of SNP rs6476643. Other features of the construct include the ampicillin resistance gene (Amp^R), origin of replication derived from filamentous phage (f1 ori), origin of replication in *E. coli* (ori), a synthetic poly(A) signal/transcriptional pause site for background reduction [synthetic poly(A)] and a cloning site downstream of Luc^+. (**C**) Box plot of the relative Luciferase activity of the R and S constructs. $n = 30$ transient transfection assays. (**D**) Average ± SEM relative Luciferase activity of constructs SS 1–10 and RR 1–10. On the *x*-axis, the genomic distance of the *MCS5A2* polymophisms to the *MCS5A1* SNP rs6476643 is plotted. The data points at genomic distance 0 correspond to the *FBXO10* TSS-Luciferase constructs S and R. The measurements indicated with an asterisk are significantly different between SS and RR ($P < 0.05$). The entire series of RR1–10 is significantly lower as compared with the entire series of SS 1–10 ($P < 10^{-48}$). $n = 6$ or more transient transfection assays per data point.

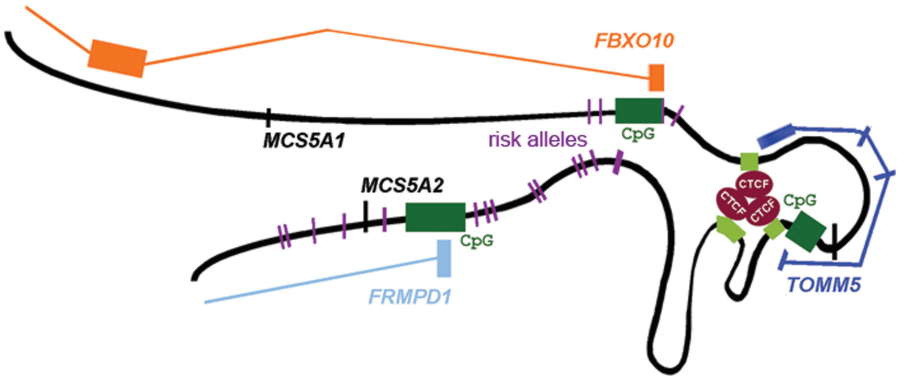

Figure 6. A model of the *MCS5A* locus in a folded configuration. The first (non-coding, 5′-UTR) and second (coding, ATG-containing) exons of the *FBXO10* transcript are displayed in orange. The *TOMM5* transcript is indicated in dark blue. The first (non-coding, 5′-UTR) exon of the *FRMPD1* transcript is shown in light blue. The CpG islands associated with their promoters are indicated in dark green. The correlated polymorphisms that associate with breast cancer risk are depicted as purple bars. The looping fragments are shown in light green and are shown to be bound by CTCF. In this model, the breast cancer susceptibility locus *MCS5A* harbors an insulator loop encompassing the *TOMM5* gene and regulatory region. The looped structure involves three elements that may loop simultaneously (as depicted) or loop in specific combinations in a single nucleus. The polymorphisms associated with breast cancer risk in *MCS5A1* and *MCS5A2* are located at both sides of the looped structure and are in closer proximity as may be derived from a linear genome view. We hypothesize that the *MCS5A1* and *MCS5A2* breast cancer risk variants interact to regulate the transcript levels of *FBXO10*.

mammary gland of susceptible congenic control and *Mcs5a* resistant congenic rats. Two of these genes, however, were found to be differentially expressed in the immune system, namely *Fbxo10* in the thymus, and *Frmpd1* in the spleen (5). Here, we have shown that the genetic interaction that is necessary for the mammary carcinoma resistance phenotype, is also essential to downregulate *Fbxo10* transcript levels in the thymus. The transcript levels of *Frmpd1* and *Tomm5* were not differentially associated with the *Mcs5a* alleles. Lower *Fbxo10* transcript levels in the resistant congenic rat line as compared with the susceptible congenic control line were detectable in $CD4^+CD8^-$ thymocytes, $CD8^+CD4^-$ thymocytes, $CD4^+CD8^+$ thymocytes and $CD3^+$ primary, cultured unstimulated and cultured activated T cells from the spleen. Analysis of some aspects of the gene regulatory mechanism mediated by the *Mcs5a*/*MCS5A* locus revealed striking similarities between rats and humans. First, the TSSs of the *Fbxo10*/*FBXO10* transcripts were found to be located in the CpG island in *Mcs5a1*/*MCS5A1* in both rat and human thymus and spleen RNA samples. The human *FBXO10* TSS cluster was found to be located 150-bp away from a breast cancer risk-associated SNP (rs6476643). Second, both the rat and human *Mcs5a*/*MCS5A* locus display a similar profile of chromatin looping and CTCF/cohesin binding. Third, functional assessment of the transcriptional regulatory properties of the resistant (R) and susceptible (S) allele of rs6476643 in combination with the R and S allele of each of the 15 *MCS5A2* risk-associated SNPs, revealed an overall significantly lower transcriptional activity for the RR combinations compared to the SS combinations, thus mimicking the resistant *Mcs5a1*–*Mcs5a2* interaction controlling *Fbxo10* transcript levels and mammary carcinoma resistance in the rat. Interestingly, the R allele of multiple *MCS5A2* SNPs combined with the R allele of rs6476643 can independently lower transcriptional activity compared to the combined SS alleles, perhaps underlying a haplotype effect between the *MCS5A1* and *MCS5A2* breast cancer risk-associated polymorphisms, which are separated by >60 kb in the genome. Physical interactions between the fragments containing the *MCS5A1* and *MCS5A2* risk-associated polymorphisms, however, were not directly observed, likely due to limited sensitivity of the 3C assay and the transient, stochastic and/or structurally diffuse nature of the putative interactions.

Mcs5a/*MCS5A* harbors a higher order chromatin structure, which spatially isolates the *Tomm5*/*TOMM5* gene and its flanking regulatory region by locating them within the loop (Figure 6). The chromatin loop also functionally isolates *Tomm5* as its transcript level was found to be strongly upregulated in T-cell activation, whereas *Fbxo10* or *Frmpd1* transcript levels were not. The higher order chromatin structure is thought to be an insulator loop. There appear to be three looping fragments involved that are all bound by CTCF and cohesin. CTCF has been recognized as a DNA-binding protein with diverse functions (23), for example insulator loop formation (24), and enhancer–promoter looping, e.g. in the β-globin locus (25). CTCF has also been associated with boundary/barrier elements, which demarcate separately regulated chromatin domains (26). The transcription of a growing number of genes has been found to be dependent on CTCF- and/or cohesin-binding and -looping. For example, gene expression in the developmentally regulated HoxA cluster is under control of CTCF-mediated heterochromatin partitioning, with involvement of Oct4 and cohesin loading at the CTCF site in undifferentiated embryonic stem cells and differentiated neuronal progenitor cells, respectively (27). Another example of CTCF-mediated looping and cooperation of a transcription factor (T-bet) in cell type specific gene transcription is the Th1 cell-specific expression of *Ifng* (28). Ablation of

the *Mcs5a* loop by depletion of CTCF yielded a lower *Tomm5* transcript level (whereas the transcript levels of two other genes adjacent to *Mcs5a* were unaffected), possibly through misregulation by the adjacently located repressed chromatin (upstream of the untranscribed *Frmpd1* gene), which would classify the insulator loop as a boundary element/barrier loop. Since the looped fragments were identified in all cell types and both (resistant and susceptible) genotypes studied, it is currently unclear if the loops occur simultaneously in a nucleus or if combinations of looping partners have distinct functions, such as enhancer looping to the *Tomm5*/*TOMM5* promoter, or condition-dependent insulator looping.

We propose a model in which the risk-associated alleles in *MCS5A1* and *MCS5A2* are located on either sides of the loop, in closer physical proximity than may be implied from a linear genome view, allowing them to interact. It is currently unknown if the independently associating human *MCS5A1* and *MCS5A2* alleles genetically interact, as they do in the rat to modulate breast cancer risk. This is due to the technical inability to phase these alleles over the 60 kb that separates them, in the double heterozygous human case–control samples, which is the majority of the samples. However, there is linkage disequilibrium between the alleles, with the RR and SS double homozygotes being overrepresented in the population compared to the expected co-occurrences according to Hardy–Weinberg equilibrium, suggesting a synthetic interaction (5).

While GWAS have successfully identified many common variants associated with breast cancer susceptibility, implications concerning their mechanisms of action are limited. Two studies have implicated such breast cancer risk loci in expression regulation of nearby genes, namely loci close to *FGFR2* (11) and *COX11* (1). The difficulties in identifying causative polymorphisms from GWAS as well as their functions could be due to limited structural and mechanistic knowledge regarding the *cis*-acting polymorphisms and their possible interactions within a locus (e.g. synthetic interactions underlying a haplotype effect). Furthermore, the lack of knowledge of which cell type(s) to use for functional studies can compound these issues. Extending high-resolution comparative genomic studies to additional alleles, cell types, and traits will provide useful models to help interpret the results of GWAS findings. The results of this study identify *Fbxo10* as a strong candidate for driving the mammary carcinoma resistance phenotype of the *Mcs5a* locus likely *via* activities in T cells. In depth analysis of the activities of specific T-cell populations regarding the normal and transformed mammary epithelium of congenic rat models and genetic manipulation of *Fbxo10* will in the future elucidate more cellular and molecular mechanisms underlying the *Mcs5a* breast cancer susceptibility locus.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Job Dekker of the University of Massachusetts-Medical School, Worchester, MA, for help establishing the 3C technology. The authors thank Kathy Schell, Joel Puchalski and Dagna Sheerar of the Paul P. Carbone Comprehensive Cancer Center Flow Cytometry Core for their expertise, assistance and service. The authors thank Dr Deepak Sharma for help obtaining activated T cells.

FUNDING

National Institutes of Health, National Cancer Institute (R01-CA123272 to M.N.G.); Department of Defense Breast Cancer Research Program Postdoctoral Award (W81XWH-07-1-0404 to B.M.G.S.); Department of Defense Breast Cancer Research Program Postdoctoral Fellowship (DAMD17-03-1-0280 to D.J.S.). Funding for open access charge: Credit card of the McArdle Lab for Cancer Research, University of Wisconsin–Madison, Department of Oncology.

Conflict of interest statement. None declared.

REFERENCES

1. Ahmed,S., Thomas,G., Ghoussaini,M., Healey,C.S., Humphreys,M.K., Platte,R., Morrison,J., Maranian,M., Pooley,K.A., Luben,R. *et al.* (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.*, **41**, 585–590.
2. Easton,D.F., Pooley,K.A., Dunning,A.M., Pharoah,P.D., Thompson,D., Ballinger,D.G., Struewing,J.P., Morrison,J., Field,H., Luben,R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.
3. Hunter,D.J., Kraft,P., Jacobs,K.B., Cox,D.G., Yeager,M., Hankinson,S.E., Wacholder,S., Wang,Z., Welch,R., Hutchinson,A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870–874.
4. Mavaddat,N., Dunning,A.M., Ponder,B.A., Easton,D.F. and Pharoah,P.D. (2009) Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. *Cancer Epidemiol. Biomarkers Prev.*, **18**, 255–259.
5. Samuelson,D.J., Hesselson,S.E., Aperavich,B.A., Zan,Y., Haag,J.D., Trentham-Dietz,A., Hampton,J.M., Mau,B., Chen,K.S., Baynes,C. *et al.* (2007) Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. *Proc. Natl Acad. Sci. USA*, **104**, 6299–6304.
6. Stacey,S.N., Manolescu,A., Sulem,P., Rafnar,T., Gudmundsson,J., Gudjonsson,S.A., Masson,G., Jakobsdottir,M., Thorlacius,S., Helgason,A. *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **39**, 865–869.
7. Stacey,S.N., Manolescu,A., Sulem,P., Thorlacius,S., Gudjonsson,S.A., Jonsson,G.F., Jakobsdottir,M., Bergthorsson,J.T., Gudmundsson,J., Aben,K.K. *et al.* (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **40**, 703–706.
8. Thomas,G., Jacobs,K.B., Kraft,P., Yeager,M., Wacholder,S., Cox,D.G., Hankinson,S.E., Hutchinson,A., Wang,Z., Yu,K. *et al.* (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat. Genet.*, **41**, 579–584.
9. Zheng,W., Long,J., Gao,Y.T., Li,C., Zheng,Y., Xiang,Y.B., Wen,W., Levy,S., Deming,S.L., Haines,J.L. *et al.* (2009)

Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, **41**, 324–328.
10. Turnbull,C., Ahmed,S., Morrison,J., Pernet,D., Renwick,A., Maranian,M., Seal,S., Ghoussaini,M., Hines,S., Healey,C.S. *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, **42**, 504–507.
11. Meyer,K.B., Maia,A.T., O'Reilly,M., Teschendorff,A.E., Chin,S.F., Caldas,C. and Ponder,B.A. (2008) Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol.*, **6**, e108.
12. Lan,H., Kendziorski,C.M., Haag,J.D., Shepel,L.A., Newton,M.A. and Gould,M.N. (2001) Genetic loci controlling breast cancer susceptibility in the Wistar-Kyoto rat. *Genetics*, **157**, 331–339.
13. Samuelson,D.J., Aperavich,B.A., Haag,J.D. and Gould,M.N. (2005) Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5. *Cancer Res.*, **65**, 9637–9642.
14. Samuelson,D.J., Haag,J.D., Lan,H., Monson,D.M., Shultz,M.A., Kolman,B.D. and Gould,M.N. (2003) Physical evidence of Mcs5, a QTL controlling mammary carcinoma susceptibility, in congenic rats. *Carcinogenesis*, **24**, 1455–1460.
15. Smits,B.M.G., Sharma,D., Samuelson,D.J., Woditschka,S., Mau,B., Haag,J.D. and Gould,M.N. The non-protein coding breast cancer susceptibility locus Mcs5a acts in a non-mammary cell-autonomous fashion through the immune system and modulates T-cell homeostasis and functions. *Breast Cancer Res*, in press.
16. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
17. Bell,A.C., West,A.G. and Felsenfeld,G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, **98**, 387–396.
18. Splinter,E., Heath,H., Kooren,J., Palstra,R.J., Klous,P., Grosveld,F., Galjart,N. and de Laat,W. (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.*, **20**, 2349–2354.
19. Hadjur,S., Williams,L.M., Ryan,N.K., Cobb,B.S., Sexton,T., Fraser,P., Fisher,A.G. and Merkenschlager,M. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, **460**, 410–413.
20. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482–485.
21. Manoharan,H., Babcock,K. and Pitot,H.C. (2004) Changes in the DNA methylation profile of the rat H19 gene upstream region during development and transgenic hepatocarcinogenesis and its role in the imprinted transcriptional regulation of the H19 gene. *Mol. Carcinog.*, **41**, 1–16.
22. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
23. Phillips,J.E. and Corces,V.G. (2009) CTCF: master weaver of the genome. *Cell*, **137**, 1194–1211.
24. Gerasimova,T.I., Byrd,K. and Corces,V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell*, **6**, 1025–1035.
25. Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F. and de Laat,W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
26. Cuddapah,S., Jothi,R., Schones,D.E., Roh,T.Y., Cui,K. and Zhao,K. (2009) Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.*, **19**, 24–32.
27. Kim,Y.J., Cecchini,K.R. and Kim,T.H. (2011) Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl Acad. Sci. USA*, **108**, 7391–7396.
28. Sekimata,M., Perez-Melgosa,M., Miller,S.A., Weinmann,A.S., Sabo,P.J., Sandstrom,R., Dorschner,M.O., Stamatoyannopoulos,J.A. and Wilson,C.B. (2009) CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity*, **31**, 551–564.