Published in final edited form as:

Mol Carcinog. 2014 April ; 53(4): 300–313. doi:10.1002/mc.21977.

Human *MCS5A1* Candidate Breast Cancer Susceptibility Gene *FBXO10* Is Induced by Cellular Stress and Correlated with Lens Epithelium-Derived Growth Factor (LEDGF)

Xin Xu[1], **David W. Powell**[1,2], **Courtney J. Lambring**[1], **Aaron H. Puckett**[1], **Lucas Deschenes**[1], **Russell A. Prough**[1,3], **Eric M. Poeschla**[4], **David J. Samuelson**[1,3,*]

[1]Department of Biochemistry and Molecular Biology, University of Louisville Health Sciences Center, Louisville, Kentucky

[2]Department of Medicine/Nephrology, University of Louisville Health Sciences Center, Louisville, Kentucky

[3]Center for Environmental Genomics and Integrative Biology, University of Louisville, Louisville, Kentucky

[4]Department of Molecular Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota

Abstract

Genetic variation and candidate genes associated with breast cancer susceptibility have been identified. Identifying molecular interactions between associated genetic variation and cellular proteins may help to better understand environmental risk. Human *MCS5A1* breast cancer susceptibility associated SNP *rs7042509* is located in *F-box protein 10* (*FBXO10*). An orthologous *Rattus norvegicus* DNA-sequence that contains SNV *ss262858675* is located in rat *Mcs5a1*, which is part of a mammary carcinoma susceptibility locus controlling tumor development in a non-mammary cell-autonomous manner via an immune cell-mediated mechanism. Higher *Fbxo10* expression in T cells is associated with *Mcs5a* increased susceptibility alleles. A common DNA–protein complex bound human and rat sequences containing *MCS5A1/Mcs5a1 rs7042509/ss262858675* in electrophoretic mobility shift assays (EMSAs). Lens epithelium-derived growth factor (LEDGF), a stress-response protein, was identified as a candidate to bind both human and rat sequences using DNA-pulldown and mass spectrometry. LEDGF binding was confirmed by LEDGF-antibody EMSA and chromatin immunoprecipitation (ChIP). Ectopic expression of LEDGF/p75 increased luciferase activities of co-transfected reporters containing both human and rat orthologs. Over-expressed LEDGF/p75 increased endogenous *FBXO10* mRNA levels in Jurkat cells, a human T-cell line, implying LEDGF may be involved in increasing *FBXO10* transcript levels. Oxidative and thermal stress of Jurkat cells increased FBXO10 and LEDGF expression, further supporting a hypothesis that LEDGF binds to a regulatory region of *FBXO10* and increases expression during conditions favoring

*Correspondence to: Department of Biochemistry & Molecular Biology, Center for Genetics & Molecular Medicine, University of Louisville, School of Medicine, HSC-A Building, Room 708, 319 Abraham Flexner Way, Louisville, KY 40292.

Additional supporting information may be found in the online version of this article.

carcinogenesis. We conclude that *FBXO10*, a candidate breast cancer susceptibility associated gene, is induced by cellular stress and LEDGF may play a role in expression of this gene.

Keywords

MCS5A ; *rs7042509* ; non-protein-coding genetic variants; complex diseases; comparative genetics

INTRODUCTION

Breast cancer is a complex disease with genetic, epigenetic, and environmental components. Low penetrance breast cancer susceptibility associated genetic variants identified to date have common population allele frequencies and non-protein-coding locations. Many breast cancer susceptibility loci are thought to function as regulatory elements of nearby or distant transcribed genes. Identification of complex disease associated susceptibility alleles may be useful to better evaluate risk. Furthermore, understanding mechanisms underlying these susceptibility alleles may lead to additional avenues of breast cancer prevention [1]. Translating this genetic information to breast cancer risk prediction and prevention will benefit from knowledge of any breast cancer susceptibility gene and environment interactions that might exist. Efforts to establish associations between low penetrance breast cancer susceptibility alleles and environmental factors using epidemiological approaches or animal studies have been successful [2-5]. Additionally, genetics and molecular biology have been used to uncover an interaction between the Sp1 transcription factor and a variant at the human *Mdm2, p53 E3 ubiquitin protein ligase* homolog (*MDM2*) gene locus. This interaction is associated with age of cancer onset and tumors at multiple sites in Li-Fraumeni syndrome patients [6].

Extension of complex disease associated risk alleles to causal variants is another important area of research. Low penetrance breast cancer susceptibility associated single-nucleotide polymorphism (SNP) *rs7042509* is one of four genetic variants that are in linkage disequilibrium within human *mammary carcinoma susceptibility 5A1* (*MCS5A1*), a species conserved susceptibility locus that contains exonic, intronic, and 5′ intergenic sequence of *F-box protein 10* (*FBXO10*), a ubiquitin ligase encoding and candidate breast cancer susceptibility gene [7]. Human SNP *rs7042509* is located approximately 3.2 kb upstream from a T-cell preferred *FBXO10* transcriptional start site [8]. A rat single-nucleotide variant (SNV), *ss262858675*, which is potentially orthologous to *rs7042509*, is located within rat *Mcs5a1*, a human *MCS5A1*-concordant non-protein-coding carcinoma susceptibility locus that contains part of the rat *Fbxo10* gene [7,9]. The *Mcs5a1* resistance allele synthetically interacts with the adjacent *Mcs5a2* resistance allele to control rat mammary carcinoma susceptibility in a non-mammary cell-autonomous manner that is mediated by immune cells [7,9]. Mammary carcinoma susceptible *Mcs5a* females have higher *Fbxo10* mRNA levels in thymus tissue and T cells compared to resistant females [7,9]. Further, the rat *Mcs5a* decreased-susceptibility allele is correlated with an increased number of mammary gland $\gamma\delta TCR^+$ T cells in carcinogen exposed mammary glands [9]. How mammary gland T cells might contribute to reduced breast cancer susceptibility is unknown. One potential

mechanism is mammary gland γδTCR$^+$ T cells may be involved in immunosurveillance against aberrant epithelial cells [10-13].

Breast cancer susceptibility modifier genes may be controlled by inherited genetic variation, environmental risk factors, or both. Here, we used Jurkat cells, a human T-cell line, as an in vitro system to study *FBXO10* expression and rat thymocytes to identify DNA-binding proteins that interact with DNA sequences containing human *rs7042509* and rat *ss262858675*. We extend our findings to show that cellular stress increases *FBXO10* expression in Jurkat cells.

MATERIALS AND METHODS

Genomics and Bioinformatics

UCSC genome browser (http://genome.ucsc.edu/) was used to compare human SNP *rs7042509* and predicted orthologous sequences. Genome assembly versions used were *Homo sapiens* GRCh37/hg19, *Pan troglodytes* CGSC 2.1/panTro2, *Rhesus macaque* MGSC Merged 1.0/rheMac2, *Bos taurus* Baylor 4.0/bosTau4, *Canis familiaris* Broad/canFam2, *Equus caballus* Broad/equCab2, *Rattus norvegicus* Baylor 3.4/rn4, and *Mus musculus* NCBI37/mm9.

Resequencing

Human and rat *MCS5A1/Mcs5a1* genomic regions containing *FBXO10/Fbxo10* 5′ respective promoter regions spanning *rs7042509/ss262858675* were published previously [7]. This segment was resequenced in Jurkat cell DNA for this study using primers listed in Supplementary Table S1. Resequencing was performed using BigDye Terminator v3.1 Cycle Sequencing kit (Life Technologies, Carlsbad, CA) by the University of Louisville Center for Genetics and Molecular Medicine.

Cell Culture and Tissue Samples

Jurkat cells (ATCC #TIB-152, Manassas, VA) were cultured in RPMI 1640 medium with 10% fetal bovine serum (FBS) and penicillin/streptomycin in 5% CO_2 atmosphere at 37°C. Procedures involving rats followed approved University of Louisville Animal Care and Use Committee protocols. Rat thymus tissue used in electrophoretic mobility shift assay (EMSA) and chromatin immunoprecipitation (ChIP) was from Wistar–Furth (WF) and Wistar–Kyoto (WKY) female rats at 12 wk of age. Fresh tissue was submerged in cold flow cytometry sorting (FCS) buffer, which contained 2% FBS, 25 mM HEPES in sterile PBS without Mg^{2+} and Ca^{2+} on ice. Cells were disaggregated from tissues using 50 μM Medicons (BD Biosciences, San Jose, CA) and a BD Medimachine (BD Biosciences) for 1 min. Cell suspensions were filtered using 70 μM Filcons (BD Biosciences). After centrifugation at 1000 rpm and 4°C, cells were resuspended and treated with red blood cell lysis solution (Miltenyi Biotec, Cambridge, MA). After washing twice with FCS buffer, cells were used to extract nuclear proteins or for ChIP assays.

Plasmid Construction and siRNA

Luciferase reporters containing human major and minor alleles of SNP *rs7042509* and orthologous rat *ss262858675* variants were constructed by ligating two copies of 41 bp of respective human genome reference DNA (SNP ± 20 bp) into the pGL3-Promoter vector [National Center for Biotechnology Information (NCBI) Accession U47298; Promega, Madison, WI] using *Mlu*I and *Xho*I restriction enzyme cutting sites. Specific sequences used for cloning are shown in Supplementary Table S2. All constructs were sequenced by Sanger sequencing at the University of Louisville Center for Genetics and Molecular Medicine to confirm insertion of correct sequences. Control reporter vector pRL-TK *Renilla* was from Promega. The pEGFP-LEDGF/p75 and pEGFP-C1 control vectors were provided by Dr. T. Shinohara (University of Nebraska, Omaha, NE). Short-interfering RNAs (siRNAs) targeting the lens epithelium-derived growth factor (LEDGF) encoding gene *PSIP1* (ON-TARGETplus SMARTpool siRNA J-015209-05) were from Thermo Scientific (Lafayette, CO). A scramble siRNA (ON-TARGETplus negative control D-001810-10-05) that was designed to not target known human, mouse, or rat genes was used for comparison.

Transient Transfection, Dual Luciferase Assay, and Cell Sorting

Jurkat cells were transfected with plasmids or siRNA using different transfection reagents depending on the downstream methods used. For dual luciferase assays, LTX transfection reagent (Life Technologies) was used to transfect plasmids into Jurkat cells according to the manufacturer's protocol. For siRNA transfection of Jurkat cells, Amaxa Cell Line Nucleofector Kit V (Lonza, Walkersville, MD) was used according to the manufacturer's protocol. Jurkat cells (1×10^6 cells/well) were transfected with siRNA (100 nM) against LEDGF/p75 or scrambled negative control siRNA in six-well plates. For cell sorting, Jurkat cells were transfected with either pEGFP-LEDGF/p75 or pEGFP vector using TransIT Jurkat transfection reagent (Mirus, Madison, WI) according to the manufacturer's protocol. Plasmids (1 μg) indicated were transfected into cells contained in 24-well plates. Twenty-four hours after transfection, cells were washed three times with cold PBS. Cell lysates were prepared for dual luciferase assays according to the manufacturer's protocol (Promega). Luciferase activities in 10 μl of each cell lysate were measured using a Biotek Synergy 2 microplate reader (Biotek, Winooski, VT). Each experiment was repeated at least three times. Changes in *Firefly* luciferase activity were calculated and plotted after standardizing for *Renilla* luciferase activity in the same sample. Fluorescence-activated cell sorting (FACS) and acquisition of green fluorescence positive and negative cell populations was performed 48 h after transfection using a MoFlo Cell Sorter (Beckman Coulter, Miami, FL) by the University of Louisville Flow Cytometry Core Facility.

Quantitative PCR

Total RNA was extracted from Jurkat cells using TRI Reagent (Molecular Research Center, Inc., Cincinnati, OH). One microgram of total RNA from each sample was reverse transcribed to cDNA using Invitrogen Superscript III reverse transcriptase (Life Technologies). Synthesized cDNA served as template in a 20-μl quantitative PCR (QPCR) reaction. QPCR was performed using TaqMan methods (Life Technologies). The reaction contained 1 × Taq 1000 RXN gold with buffer A (Life Technologies),

3.5 mM $MgCl_2$, 0.2 mM dNTP, 500 nM each forward and reverse primer, 200 nM target probes, 0.025 U/μl Taq Gold DNA pol (Life Technologies), and 50 ng RNA equivalent cDNA in a StepOnePlus real-time PCR system. The QPCR cycling conditions were 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Primers for *FBXO10* QPCR were: forward 5′-CTGCCGGCATAGCAGTGAAT-3′; reverse 5′-CCACACCTCCCCACTGATTCT-3′; and probe 6FAM-CTCATCACAGAAAATGTMGBNFQ. Primers for FERM and PDA containing 1 (*FRMPD1*) QPCR were: forward 5′-GTGAGCGTCGTCAAAGTGTATCTT; reverse 5′-GCTAGGTCTTTTGCACTGTTGGA; and probe 6FAM-AGGACGTCAAGGTTC-MGBNFQ. Primers and probe for *LEDGF* QPCR were premade assay *HS00253515.m1* (Life Technologies). Expression of target transcript was standardized to *GAPDH* (Life Technologies) expression levels after determining that *GAPDH* levels were not significantly different between treatments. Relative concentrations of each target template were calculated according to a comparative C_t method using StepOne software (V2.0; Life Technologies). Results from three independent experiments were analyzed.

Electrophoretic mobility shift assays were performed using Pierce Lightshift Chemiluminescent EMSA Kit according to the manufacturer's protocol (Pierce, Rockford, IL). Forward and reverse DNA sequence of 41 bp human *rs7042509* major/minor and rat *ss262858675* WF/WKY alleles were annealed and labeled with biotin using a biotin 3′ end DNA labeling kit (Pierce). Oligonucleotide sequences are shown in Supplementary Table S3. Rat (WKY) thymus tissue was first separated into single cell suspensions using 50 μm Medicons as described above. Nuclear extracts (NE) were collected as previously described [14]. Labeled oligonucleotides (20 fmol) and 2 μg nuclear proteins were used in EMSA. A 4.5% non-denaturing polyacrylamide gel (4.5% 29:1 acrylamide/bisacrylamide, 0.5× TBE, 1% glycerol, 0.075% APS, and 0.1% TEMED) was used to resolve protein–DNA complexes in 0.5× TBE buffer. Gel supershift of protein–DNA complexes was performed by incubating nuclear protein and 2 μg LEDGF antibodies (BD Biosciences) or a negative IgG in EMSA buffer prior to adding biotin-labeled probes.

DNA-pulldown experiments were performed as described previously [15]. Oligonucleotides containing human *rs7042509* major allele and *ss262858675* WKY allele were synthesized by Integrated DNA Technologies with 5′-prime biotin attached (Coralville, IA) and annealed to complementary sequences to obtain double biotinylated oligonucleotides. Sequences of oligonucleotides were the same as those used in EMSA (Supplementary Table S3). Biotinylated oligonucleotides (2 μg) were conjugated to streptavidin-coated magnetic particles (Roche, Indianapolis, IN) as described in the manufacturer's protocol. Rat thymus nuclear protein was incubated with 5′biotin-labeled human major and rat WKY oligonucleotides for half an hour at 4°C. DNA-pulldown proteins were collected using a magnetic ring stand (Life Technologies) and analyzed using mass spectrometry (MS).

Mass Spectrometric (MS) Analysis

Protein–DNA complexes of interest, excised from polyacrylamide gels or DNA-pulldown, were, respectively, subjected to in-gel or direct trypsin digestion as previously reported [16].

The resulting peptides were analyzed using a previously described LC-MS/MS approach [17]. Database searching for relevant peptides was performed with tandem mass spectra extracted by ReAdW (Ionsource.com) and converted to mzXML format. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using SequestSorcerer (Sage-N). Sequest (University of Washington, Seattle, WA) was set up to search a FASTA formatted rat protein database (rat RefSeq), with a fragmentation mass tolerance of 1.00 Da and a parent ion tolerance of 1.2 Da. Iodoacetamide derivatization of cysteine residues was specified in Sequest as a fixed modification. Oxidation of methionine was specified as a variable modification. Scaffold (version Scaffold 3.0.00; Proteome Software, Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at >95.0% probability as specified by the Peptide Prophet algorithm [18]. Protein probabilities were assigned by the Protein Prophet algorithm [19]. Protein identifications were accepted if they could be established at >95.0% probability. Proteins that contained similar peptides and, thus, could not be differentiated based on MS/MS analysis alone were grouped to satisfy principles of parsimony.

Chromatin immunoprecipitation (*ChIP*), *ChIP-QPCR*, and *ChIP-coimmunoprecipitation*—were performed according to a protocol by Michael et al. [20]. Primary rat thymus cells were obtained as described above. LEDGF/p75 antibody (A300-848A; Bethyl, Montgomery, TX) was used to precipitate LEDGF. RNA polymerase II (RNAPII) antibody (sc-9001, Santa Cruz Biotechnology, Santa Cruz, CA) and IgG (BD Pharmingen, Sparks, MD) was used as a control. Oligonucleotide primers used to amplify genomic DNA for ChIP, and oligonucleotides primers and Taqman probes (Life Technologies) for ChIP-QPCR are shown in Supplementary Table S1. Relative DNA abundance was as described previously with a slight variation [21]. The modified formula was: ΔC_t (sample) = C_t (precipitated DNA) – C_t (5%input); $\Delta\Delta C_t = \Delta C_t$ (sample) – ΔC_t (IgG); Relative abundance = $2^{-\Delta\Delta C_t}$. Data were obtained from three independent experiments. For performing ChIP-co-immunoprecipitation (CoIP), DNA–protein complexes from LEDGF, RNAPII, and IgG pulldown were collected after immunoprecipitation wash step and before the DNA purification step of ChIP. ChIP samples (5% of input) were used as a positive control. LEDGF and RNAPII blots were detected using primary antibodies (LEDGF 611714, BD Biosciences; RNAPII sc-9001, Santa Cruz, CA) and secondary antibodies (ECLplex goat-α-rabbit IgG, GE Healthcare Life Sciences, Piscataway, NJ). Fluorescence from immune complexes was determined using a Typhoon 7900 (GE Healthcare Life Sciences).

Western Blotting

Jurkat cells were collected and whole cell extracts were prepared using RIPA lysis buffer (Cell Signaling, Danvers, MA) with addition of protease inhibitor mixture. Samples were separated by SDS–PAGE (12% gel for GAPDH and 8% gel for FBXO10) and transferred onto Hybond-LFP membrane (GE Healthcare Life Sciences). Primary antibody for GAPDH was from Bethyl Laboratories (Montgomery, TX) and primary antibody for FBXO10 was from Sigma (St. Louis, MO). GAPDH antibody was used at 1:4000 and FBXO10 antibody was used at 1:2000. The anti-rabbit secondary antibody was from Cell Signaling. Chemiluminescence was detected using Pierce Supersignal West Pico

Chemiluminescent Substrate. Relative protein abundance was calculated based on the density of each representative band using ImageJ (http://imagej.nih.gov/ij/) and GAPDH intensity for standardization.

Statistical Analysis

Quantitative results are reported as mean ± standard deviation. To correct for multiple comparisons, all groups within an experiment were compared using a non-parametric Kruskal–Wallis test followed by non-parametric Mann–Whitney post hoc tests between two groups. Comparisons with a *P*-value ≤0.05 were considered statistically significant.

RESULTS

Conserved Human MCS5A1 Breast Cancer Susceptibility Associated SNP rs7042509

Non-protein coding low-penetrance breast cancer susceptibility alleles may function as transcriptional regulators of nearby or distant genes by binding relevant regulatory proteins. Human and rat *MCS5A1/Mcs5a1* are concordant breast and mammary cancer susceptibility alleles [7,22]. Gould and colleagues identified four breast cancer risk associated common variants within *MCS5A1* that were highly correlated based on pairwise comparison of haplotypes from population-based samples of European descent [7]. We noted that one of these variants, SNP *rs7042509*, was contained in both human and rat conserved sequences. Many functional *cis*-regulatory DNA elements are evolutionarily conserved; therefore, we first compared the sequence containing human *MCS5A1* SNP *rs7042509* and predicted orthologs in different species. We chose a 41 bp (SNP ± 20 bp flanking sequence) human reference genome DNA sequence containing *rs7042509* to identify potential orthologous loci in various species using the convert option of UCSC genome browser [23]. Human *rs7042509* containing sequence was found to be conserved between different species (Figure 1A).

Interestingly, rat DNA sequence predicted to be orthologous using the UCSC genome browser convert function did not contain variants between *Mcs5a1*-susceptible (WF) and -resistance (WKY) alleles. However, there was a variant between *Mcs5a1* alleles from these strains at approximately the same physical position relative to an orthologous *FBXO10/Fbxo10* 5′UTR (untranslated region) and a conserved CpG island (Figure 1B and C). This rat SNV has the NCBI assay ID *ss262858675* [7]. The increased-risk minor allele and population reference major allele of *rs7042509* in humans contain G and T, respectively, whereas rat *ss262858675* WF susceptible and *Mcs5a1* resistance associated variants are G and A, respectively. Thus, G in human and rat alleles is associated with increased breast and mammary cancer susceptibility, while T in human and A in rat is associated with population reference breast cancer risk and decreased rat mammary cancer susceptibility, respectively.

A Common EMSA Shift Was Identified Between Human rs7042509 and Rat ss262858675 Sequences

Due to the respective physical locations within *MCS5A1/Mcs5a1* we speculated that rat sequence containing SNV *ss262858675* might function similarly to human sequence containing *rs7042509*. We sought to identify nuclear proteins binding to these conserved

sequences using EMSA and DNA-pulldown assays followed by MS. Considering similarities between human SNP *rs7042509* and rat *ss262858675* containing sequences, we hoped to identify DNA binding proteins in common between these species. We performed EMSA experiments using a respective 41-bp double-stranded (DS) oligonucleotide based on species respective genomic DNA with a SNV sequence identical to the human major or minor allele of *rs7042509* and rat *Mcs5a1*WF-susceptible or WKY-resistance allele of *ss262858675*. We used NE of primary thymus cells isolated from 12-wk-old virgin female rats, as this is an age that *Mcs5a1* resident candidate gene *Fbxo10* was found to be differently expressed between mammary cancer susceptible and resistant rat strains in thymus tissue and T cells [7-9]. We speculated that from rat thymus NE we might identify candidate regulatory proteins binding to human *rs7042509*. We anticipated that by incorporating rat *ss262858675* in EMSA experiments we would narrow our candidate list of DNA–protein complexes. The shift marked by an arrow in Figure 2A was in common between human and rat alleles. Note that, in addition to this band, there were also DNA–protein complexes of apparent different molecular weights between human and rat alleles. While these may be important, we did not pursue these DNA–protein complexes.

We performed competitive assays to determine if the DNA–protein complex marked with an arrow might be caused by a similar protein or protein complex. We used unlabeled DS-oligonucleotides of human *rs7042509* to compete with biotin-labeled oligonucleotides of rat *ss262858675* and, vice versa (Figure 2B). The shifts from human and rat oligonucleotides marked by the arrow were of similar size between species, and were both competed by unlabeled DS-oligonucleotide sequences of either species. The upper human DNA–protein complexes in Figure 2B were competed for by unlabeled human, but not rat DS-oligonucleotides. Whereas, both human and rat unlabeled DS-oligonucleotides competed with the labeled DS-oligonucleotides of rat *ss262858675* for the upper rat DNA–protein complex in Figure 2B. These results suggested that these human and rat DNA sequences were capable of forming common DNA–protein complexes between them.

LEDGF Was Identified as a Candidate Protein Binding to Human and Rat Variants

To identify specific proteins binding to both human *rs7042509* and rat *ss262858675* variant containing sequences, we used MS. Protein for MS was collected from EMSA gels by excising the specific bands marked with an arrow in Figure 2. Specifically, these were DNA–protein complexes representing the human major and rat WKY-resistance alleles. Additionally, protein for MS was collected by DNA-pulldown using 41 bp human/rat major/WKY-resistant DS-oligonucleotides and rat WKY thymus NE. We performed MS analysis using proteins from both EMSA gels and DNA pulldown samples. From each MS run, multiple candidates were identified; however, by combining the MS results of two approaches, we narrowed candidates to six proteins (Table 1).

LEDGF Was Confirmed to Bind Human rs7042509 and Rat ss262858675 by LEDGF-Antibody EMSA and ChIP

Among six candidates identified using MS, LEDGF was a likely candidate to bind both *rs7042509* and *ss262858675* containing sequences. LEDGF is ubiquitously expressed in many tissues including thymus [24]. Based on the LEDGF-specific peptide sequences in our

MS analysis, the protein observed was LEDGF/p75, and less likely LEDGF/p52, a splice variant of LEDGF having a different C-terminus than LEDGF/p75 [25]. We repeated EMSA analysis including antibodies against LEDGF/p75 and found that adding anti-LEDGF, but not IgG control, led to a supershift of the band of interest when DS-oligonucleotides based on human *rs7042509* major allele were used, and a decrease in or partial-clearing of the DNA–protein complex of interest when DS-oligonucleotides based on rat *ss262858675* WKY allele were used (Figure 3A). A partial-clearing of the DNA-LEDGF-containing complex also occurred in a positive control LEDGF-antibody EMSA using the mouse *AOP2* classical stress-response element (STRE; Supplementary Figure S1). We confirmed by ChIP that LEDGF interacted with respective genomic regions containing human *rs7042509* and rat *ss262858675* using Jurkat cell and WKY rat thymus DNA (Figure 3B and C). LEDGF was detected in both human and rat chromatin regions suggesting that it is involved in the transcription of *FBXO10*.

Similar Enrichment of LEDGF Was Detected at the WF and WKY Rat Genomic Region Containing Variant ss262858675

One-way genetic variants might contribute to differential expression of *FBXO10* is by having different affinities for a transcription factor. Based on EMSA we did not notice differential binding of LEDGF to human minor/major alleles or to rat WF/WKY alleles. It was possible that our in vitro approach was not sensitive enough to detect differential binding of these variants by LEDGF. Thus, we used ChIP-QPCR to compare in vivo LEDGF binding at WF and WKY *Mcs5a1*-resistance genome sites containing rat *ss262858675* variants (Figure 4A). Thymocytes from 12-wk-old WF ($n = 3$) and WKY ($n = 3$) females were analyzed. We also performed ChIP-QPCR of this region for RNAPII binding. Further, *Cdkn1b* and *Gapdh* were included as respective LEDGF binding positive and negative controls [26]. A Kruskal–Wallis test between all *Fbxo10* ChIP-QPCR groups was not statistically significant ($P = 0.0752$); however, the Kruskal–Wallis test for *Cdkn1b* groups was significant ($P = 0.0434$). The WKY *Cdkn1b* region had a higher occupancy of LEDGF than WF *Cdkn1b* ($P < 0.05$, Mann–Whitney's *U*-test). Results depicted in Figure 4A show that there was no detectable difference between WF and WKY alleles in the quantity of chromatin occupied by LEDGF at this *Fbxo10* locus. Thus, we did not detect a difference in enrichment of LEDGF/p75 bound to WF and WKY rat genomic sequence containing *ss262858675* variants. Interestingly, there was a trend toward a lower quantity of RNAPII in WKY than WF *Fbxo10* chromatin. Although additional work will be required to confirm, a lower level of RNAPII would be consistent with lower expression of *Fbxo10* from the WKY *Mcs5a* allele [7]. We confirmed by ChIP/Co-IP that LEDGF/p75 and LEDGF/p52 co-existed with RNAPII at these rat genomic sequences (Figure 4B). We concluded from EMSA and ChIP-QPCR that LEDGF bound to sequences containing *rs7042509* and rat *ss262858675*; however, under the conditions tested, these variants had no effect on LEDGF quantities bound.

Over-Expression of LEDGF/p75 Increased SNP rs7042509 Luciferase Reporter Activities and Endogenous FBXO10 mRNA Levels in Jurkat cells

We determined whether ectopic expression of LEDGF/p75 had an effect on luciferase activities of reporter vectors containing variants of human *rs7042509* or rat *ss262858675*.

Two copies of each respective SNV ± 20 bp of genomic DNA sequence used in EMSAs and DNA pulldowns were inserted upstream of the SV40 promoter contained in the pGL3-Promoter vector (NCBI Accession U47298). Jurkat cells, a human cell line originated from a patient with T-cell leukemia, were independently transfected with each cloned variant. All cloned variants produced luciferase activities that were lower than the pGL3-Promoter without an insert, but higher than results from the pGL3-Basic vector (NCBI Accession U47295), which lacked insert and promoter sequences (Figure 5A). This implies that sequences containing human *rs7042509* or rat *ss262858675* may have repressive effects on transcription from certain promoters.

Next, Jurkat cells were co-transfected with each SNV luciferase reporter and either pEGFP-LEDGF/p75 or pEGFP expression vectors. A Kruskal–Wallis test comparing all transfection groups that received a vector containing either human *rs7042509* or rat *ss262858675* sequence was significant ($P = 0.006$). We observed that ectopic LEDGF/p75 expression resulted in modest increases in relative luciferase activities of both human *rs7042509* variants and the WF *ss262858675* allele that were statistically significant (Figure 5B, *P*-values <0.05, Mann–Whitney's *U*-test). Ectopic LEDGF/p75 expression did not significantly increase luciferase activity driven by the WKY allele ($P = 0.1266$, Mann–Whitney's *U*-test). We determined that ectopic expression of LEDGF had no effect on pGL3-Promoter luciferase activity in the absence of human *rs7042509* and rat *ss262858675* containing sequences (Figure 5B). These results suggested that human *rs7042509* and rat *ss262858675* may be contained in gene regulatory DNA-sequence and provided functional evidence that LEDGF may be involved.

Human *rs7042509* and rat *ss262858675* are in potential regulatory regions that we anticipate have an effect on *FBXO10* transcript levels. We determined if ectopic over-expression of LEDGF/p75 had an effect on endogenous *FBXO10* mRNA expression levels in Jurkat cells, which are homozygous for the population-reference major allele of *rs7042509* (Supplementary Figure S2). Expression vectors for pEGFP-LEDGF/p75 or pEGFP were transiently transfected into Jurkat cells. Due to the documented low transfection efficiency of this cell line and our desire to detect even small increases in *FBXO10* expression, cells were sorted into green-fluorescence-positive and -negative populations by FACS and acquired for QPCR comparison. Over-expression of LEDGF/p75 increased *FBXO10* mRNA levels approximately twofold to threefold compared to pEGFP vector control when green fluorescence-positive cells were compared (Figure 6A, $P < 0.05$, Mann–Whitney's *U*-test). We confirmed that *LEDGF* mRNA was over-expressed in green-fluorescence-positive cells compared to -negative cells using QPCR (Figure 6B, $P < 0.05$, Mann–Whitney's *U*-test).

We determined the effect of LEDGF knockdown in Jurkat cells on *FBXO10* mRNA levels using siRNA specific for *LEDGF*. We found that transient knockdown of LEDGF had no effect on *FBXO10* mRNA levels in Jurkat cells (Figure 6C, $P = 0.2454$, Mann–Whitney's *U*-test). We confirmed that *LEDGF* mRNA levels were lower in Jurkat cells with target siRNA compared to scrambled siRNA (Figure 6D, $P = 0.0209$, Mann–Whitney's *U*-test). In confirmation, *FBXO10* levels in a Jurkat cell line with stable knockdown of LEDGF/p75 [27] were not different from cells with LEDGF/p75 expression re-established (Supplementary Figure S3). Considered together, these results imply that LEDGF is

involved in the induction of *FBXO10*, but diminished levels of LEGDF are sufficient for low levels of basal *FBXO10* transcript production.

Known Stimulators of LEDGF Activity, Oxidative and Heat Stress, Increased FBXO10 mRNA, and Protein Levels

LEDGF/p75 is a stress-induced protein [28]. Therefore, we sought to determine whether stresses that are known to induce LEDGF/p75 activity might also increase FBXO10 expression. We chose two conditions known to induce LEDGF/p75 activity, namely, heat stress and oxidative stress [29,30]. We exposed Jurkat cells to 0.2 mM H_2O_2 and collected total RNA and whole cell extract at: 0, 1, 3, 6, 9, 12, and 24 h. In a separate experiment, Jurkat cells were heat stressed at 42°C for 1, 3, and 6 h. To test the effect of recovery from heat stress, cells were heat stressed for 6 h and allowed to recover at 37°C for 3, 6, or 18 h. Total RNA and whole cell extract at each time point were also collected. Transcript levels of *FBXO10* increased in both H_2O_2 exposed and heat stressed cells (Figure 7A and B). Compared to 0 time, *FBXO10* mRNA levels increased in Jurkat cells after 3 h of H_2O_2 exposure ($P < 0.05$, Mann–Whitney's *U*-test), peaked at fourfold over baseline at 6 h ($P < 0.05$, Mann–Whitney's *U*-test), and continued to be significantly elevated after 9 h ($P < 0.05$, Mann–Whitney's *U*-test). *FBXO10* mRNA levels increased fourfold in Jurkat cells that were heat stressed for 6 h and allowed to recover for 3 h at 37°C ($P = 0.0039$, Mann–Whitney's *U*-test). *FBXO10* mRNA levels remained significantly above baseline after 6 and 18 h of recovery from heat stress (*P*-values = 0.0039 and 0.0163, respectively, Mann–Whitney's *U*-tests). Protein levels of FBXO10 and LEDGF also increased in H_2O_2 exposed and heat stress recovering Jurkat cells in a pattern similar to *FBXO10* mRNA levels (Figure 7C,D). In addition, LEDGF induction by H_2O_2 and heat stress occurred earlier than increases in FBXO10 levels suggesting that LEDGF could mediate increased FBXO10 expression during cell stress. In further confirmation, the stress of serum starvation also increased *FBXO10* transcript and protein levels in Jurkat cells (Supplementary Figure S4).

Smits et al. [8] discovered that *MCS5A1* and *MCS5A2* are maintained in close physical proximity to each other through a chromatin looping mechanisms involving CCCTC-binding factor (CTCF). Breast cancer risk associated genetic polymorphisms in both regions are located outside of this chromatin loop. In addition, comparative rat genetic and human in vitro data suggest *MCS5A1/Mcs5a1* and *MCS5A2/Mcs5a2* breast/mammary cancer risk associated genetic variants functionally interact to control gene expression and cancer risk [7,8]. We hypothesized that since cellular stress had an effect on *FBXO10*, which is partly located in *MCS5A1*, it may also have an effect on *FRMPD1*, which is partly located in adjacent *MCS5A2*. We measured *FRMPD1* levels in Jurkat cells exposed to 0.2 mM H_2O_2 for 0, 1, 3, 6, 9, 12, and 24 h (Figure 7A). A Kruskal–Wallis test comparing all *FRMPD1* QPCR groups was statistically significant ($P = 0.0220$). Compared to zero time, *FRMPD1* mRNA levels decreased in Jurkat cells at 3 and 6 h after H_2O_2 exposure (*P*-value < 0.05, Mann–Whitney's *U*-tests). After 6 h *FRMPD1* levels began to increase and peaked significantly above baseline at 12 h after H_2O_2 exposure ($P < 0.05$, Mann–Whitney's *U*-test), at which time *FRMPD1* levels began to return toward baseline. This suggests that at least one stress-responding gene-regulatory element or combination of elements exits that controls expression of *MCS5A1* and *MCS5A2* resident transcripts.

DISCUSSION

Our data support a hypothesis that human breast cancer susceptibility associated SNP *rs7042509*, located in human *MCS5A1*, and SNV *ss262858675*, located in rat *Mcs5a1*, are functional orthologs that may be involved in human/rat *FBXO10/Fbxo10* transcription. Sequences flanking these non-protein coding SNVs are conserved between multiple species, which further suggests that these human breast and rat mammary cancer risk variant-containing sequences may be *cis*-regulatory elements. It is important to point out that the rat sequence predicted to be the conserved site using a publicly available comparative alignment algorithm did not contain genetic variants between the two rat strain alleles of interest. The rat orthologous SNV to human *rs7042509* was found to be in a similar physical position with respect to species conserved gene elements. This distinction may have broad implications for testing potential functional complex disease associated variants in other comparative experimental organisms.

Non-protein coding variants located at promoters of nearby genes may mediate differential protein binding and be related to gene expression. For example, a SNP (*SNP309*) located at the *MDM2* promoter altered Sp1 transcriptional factor binding. Further, this SNP was demonstrated to be associated with *MDM2* over-expression [6]. In another example, a six-nucleotide insertion-deletion polymorphism in a *CASP8* promoter led to the abolishment of an Sp1-binding site and was associated with both *CASP8* transcription and susceptibility to multiple cancers [31]. We discovered that both sequences containing human *rs7042509* and rat *ss262858675* bound LEDGF; however, we did not detect significant differential binding of LEDGF to these sequences. Further, results of ChIP-QPCR assays indicated that LEDGF did not bind differentially to rat chromatin containing variants of *ss262858675*. We observed that ectopic expression of LEDGF/p75 in Jurkat cells activated reporters containing human major/minor *rs7042509* and rat WF *ss262858675* alleles similarly. Based on these results, we propose that LEDGF is part of a transcription factor complex bound to chromatin containing these SNV sites, but LEDGF may not bind differentially to allelic variants of *rs7042509* and *ss262858675*. Differential LEDGF binding may be revealed if the entire human *MCS5A* and rat *Mcs5a* regions are considered. It is also possible that other DNA-binding proteins interact synergistically with LEDGF. Furthermore, these sequence variants may differentially bind to other DNA-interacting proteins.

Although LEDGF was identified to bind to *MCS5A1/Mcs5a1* sequences containing human *rs7042509* and rat *ss262858675*, there are at least three additional breast cancer risk associated variants in human *MCS5A1*, which are *rs6476643, rs10758441*, and *Indel 138-9899* [7]. These polymorphisms are in linkage disequilibrium with *rs7042509*. Human *rs7042509* is logical candidate causal variant as it is in conserved sequence where the rat ortholog contains a variant between mammary cancer susceptible and *Mcs5a* resistant rats. We chose to test these orthologous human *MCS5A1* and rat *Mcs5a1* breast and mammary carcinoma risk associated variants in the context of the widely used SV40 promoter. The possibility that *MCS5A1/Mcs5a1* associated variants act in a promoter-specific manner should be considered in future studies. This hypothesis was not tested in this study as we were unaware of which human or rat DNA-segment to use as a promoter. Smits et al. [8] has

recently shed light on which *MCS5A1/Mcs5a1* orthologous regions to focus studies of T-cell *FBXO10* transcriptional regulation.

We addressed *FBXO10* regulation in this study by focusing on human *rs7042509* and rat *ss262858675* orthologous *MCS5A1/Mcs5a1* variants. Alternatively, one of the *MCS5A1* breast cancer susceptibility associated polymorphisms in a non-conserved sequence or a combination of associated polymorphisms may be causal. Under these scenarios the human and rat causal sequence variants may have arisen and evolved independently to affect molecular mechanisms that have similar effects on the same pheno-types. It is likely that the entire *MCS5A/Mcs5a* region may be required for environmental regulation of *FBXO10*, just as it is required for the genetically determined differential expression of *Fbxo10* between T cells from mammary carcinoma resistant and susceptible rats [8]. Thus, testing the hypothesis that these variants may be promoter specific may require cloning long species-specific DNA sequences of variable lengths. However, to fully appreciate the environmental regulation of *FBXO10* each *MCS5A1* breast cancer risk associated variant may need to be investigated independently and in the context of *MCS5A2* breast cancer susceptibility associated polymorphisms.

First isolated from a human lens epithelial cell (LEC) cDNA library, LEDGF/p75 is a chromatin binding protein [32] and cell survival factor [33] encoded by *PC4 and SFRS1 interacting protein 1* (*PSIP1*). It has a modular domain structure and interacts with several cellular proteins [reviewed in Ref. [34]]. LEDGF/p75 also interacts with lentivirus integrase proteins [35] and is a cofactor for viral integration into host DNA [27,36-38]. A wide range of environmental stresses, including thermal and oxidative stresses, induce LEDGF/p75 activity [29]. Subsequently, LEDGF/p75 can activate a series of stress-related genes, including *Hsp27*, *alphaB-crystallin*, and *AOP2* via STREs, in what is thought to be a protective role [39-41]. Further, LEDGF/p75 is considered an oncogenic protein in that it has been reported to be highly expressed in prostate and breast cancers [42,43]. An involvement of LEDGF/p75 in cancer may stem from its anti-apoptotic and DNA repair roles [43-45].

Based on this background, we extended our finding that LEDGF may play a role in *FBXO10* transcription to demonstrate that cellular stress increases *FBXO10* transcript and protein levels in Jurkat cells. Therefore, in addition to demonstrating that both sequences containing human *rs7042509* and rat *ss262858675* are bound by LEDGF, we further report that ectopically expressed LEDGF/p75 and known stimulators of LEDGF activity have effects on *FBXO10* transcript levels in Jurkat cells. We showed that over-expression of LEDGF/p75 in Jurkat cells increased *FBXO10* mRNA levels, whereas knockdown of LEDGF had no effect on *FBXO10* mRNA levels. Jurkat cells have low basal levels of *FBXO10*; therefore, a potential explanation of these results is that increasing LEDGF produced a detectable increase in *FBXO10*, but a decrease in LEDGF did not produce a change in *FBXO10* levels detectable by QPCR. Another explanation is that only low levels of LEDGF are required to produce basal levels of *FBXO10* in Jurkat cells.

Interestingly, the TGG(T/G)G consensus sequence of a responsive element containing human *rs7042509* is similar to the AGGGG containing sequences of a classical STRE

[46]. Thus, it is possible that sequence containing this SNP is important for responses to environmental stressors. Interestingly, transient increases in *FBXO10* mRNA and protein levels occurred within hours after exposure to each stress. Thus, there are likely signaling pathways mediating rapid induction of *FBXO10* following cellular stress. For example, in *Saccharomyces cerevisiae*, MAP kinase and Ras-cAMP pathways are environmentally regulated pathways that have been demonstrated to regulate STREs [46]. A MAP kinase cascade was also reported to be activated under conditions of heat stress [33]. Cellular stress is known to induce many protective proteins, including heat shock proteins and LEDGF. Many of these proteins are degraded following exposure of cells to stress. As a ubiquitin ligase it is possible that FBXO10 is responsible for regulating levels of a specific protein during or following environmental exposures that result in cell stress.

In conclusion, our results demonstrate the utility of comparative approaches between species having concordant phenotypes to inform upon regulation of uncharacterized breast cancer susceptibility genes. Human and rat *MCS5A1/Mcs5a1* SNVs *rs7042509* and *ss262858675* are orthologs contained in DNA sequence that is bound by LEDGF. We propose that *MCS5A1* candidate breast cancer susceptibility gene *FBXO10* is stress-induced. Future studies will determine whether environmental exposures associated with breast cancer risk and cellular stress also increase T-cell expression of FBXO10 in the context of breast cancer susceptibility alleles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract grant sponsor: National Institutes of Health (partial support) (to D.J.S., D.W.P., E.M.P.); Contract grant numbers: CA137052, DK176743, AI77344; Contract grant sponsor: Juvenile Diabetes Research Foundation International (to D.W.P.); Contract grant number: 1-2011-588; Contract grant sponsor: UofL Center for Environmental Genomics and Integrative Biology (to D.J.S., KS Ramos, PI); Contract grant number: P30ES01443; Contract grant sponsor: UofL School of Medicine Research Office (to D.J.S.).

Abbreviations:

MDM2	Mdm2 p53 E3 ubiquitin protein ligase homolog
SNP	single-nucleotide polymorphism
MCS5A1	mammary carcinoma susceptibility 5A1
FBXO10	F-box protein 10
SNV	single-nucleotide variant
FBS	fetal bovine serum
EMSA	electrophoretic mobility shift assay
ChIP	chromatin immunoprecipitation

WF	Wistar–Furth
WKY	Wistar–Kyoto
FCS	flow cytometry sorting
NCBI	National Center for Biotechnology Information
siRNAs	short-interfering RNAs
FACS	fluorescence-activated cell sorting
QPCR	quantitative PCR
FRMPD1	FERM and PDA containing 1
NE	nuclear extracts
MS	mass spectrometry
RNAPII	RNA polymerase II
DS	double stranded
STRE	stress-response element
CTCF	CCCTC-binding factor
LEC	lens epithelial cell

REFERENCES

1. Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. N Engl J Med 2010;362:986–993. [PubMed: 20237344]
2. Travis RC, Reeves GK, Green J, et al. Gene-environment interactions in 7610 women with breast cancer: Prospective evidence from the Million Women Study. Lancet 2010;375:2143–2151. [PubMed: 20605201]
3. Milne RL, Gaudet MM, Spurdle AB, et al. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: A combined case-control study. Breast Cancer Res 2010;12:R110. [PubMed: 21194473]
4. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J Natl Cancer Inst 2011;103:1252–1263. [PubMed: 21791674]
5. Kim H, Hall P, Smith M, et al. Chemoprevention by grape seed extract and genistein in carcinogen-induced mammary cancer in rats is diet dependent. J Nutr 2004;134:3445S–3452S. [PubMed: 15570052]
6. Bond GL, Hu W, Bond EE, et al. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell 2004;119:591–602. [PubMed: 15550242]
7. Samuelson DJ, Hesselson SE, Aperavich BA, et al. Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. Proc Natl Acad Sci USA 2007;104:6299–6304. [PubMed: 17404222]

8. Smits BMG, Traun BD, Devries TL, et al. An insulator loop resides between the synthetically interacting elements of the human/rat conserved breast cancer susceptibility locus MCS5A/Mcs5a. Nucleic Acids Res 2012;40:132–147. [PubMed: 21914726]
9. Smits B, Sharma D, Samuelson D, et al. 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 2011;13:R81. [PubMed: 21846333]
10. Sordillo LM, Streicher KL. Mammary gland immunity and mastitis susceptibility. J Mammary Gland Biol Neoplasia 2002;7:135–146. [PubMed: 12463736]
11. Miescher S, Schreyer M, Barras C, Capasso P, von Fliedner V. Sparse distribution of gamma/delta T lymphocytes around human epithelial tumors predominantly infiltrated by primed/memory T cells. Cancer Immunol Immunother 1990;32:81–87. [PubMed: 2149671]
12. Girardi M, Oppenheim DE, Steele CR, et al. Regulation of cutaneous malignancy by gammadelta T cells. Science 2001;294:605–609. [PubMed: 11567106]
13. Girardi M, Glusac E, Filler RB, et al. The distinct contributions of murine T cell receptor (TCR)gammadelta+ and TCRalphabeta+ T cells to different stages of chemically induced skin cancer. J Exp Med 2003;198:747–755. [PubMed: 12953094]
14. Wadman IA, Osada H, Grutz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J 1997;16:3145–3157. [PubMed: 9214632]
15. Wang X, Peng Y, Ma Y, Jahroudi N. Histone H1-like protein participates in endothelial cell-specific activation of the von Willebrand factor promoter. Blood 2004;104:1725–1732. [PubMed: 15150074]
16. Powell DW, Rane MJ, Chen Q, Singh S, McLeish KR. Identification of 14-3-3zeta as a protein kinase B/Akt substrate. J Biol Chem 2002;277:21639–21642. [PubMed: 11956222]
17. Cummins TD, Barati MT, Coventry SC, Salyer SA, Klein JB, Powell DW. Quantitative mass spectrometry of diabetic kidney tubules identifies GRAP as a novel regulator of TGF-beta signaling. Biochim Biophys Acta 2010;1804:653–661. [PubMed: 19836472]
18. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74:5383–5392. [PubMed: 12403597]
19. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75:4646–4658. [PubMed: 14632076]
20. Carey MF, Peterson CL, Smale ST. Chromatin immunoprecipitation (ChIP). Cold Spring Harb Protoc 2009;2009:pdb prot 5279.
21. Qiu C, Yu M, Shan L, Snyderwine EG. Allelic imbalance and altered expression of genes in chromosome 2q11-2q16 from rat mammary gland carcinomas induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. Oncogene 2003;22:1253–1260. [PubMed: 12606953]
22. Mavaddat N, Dunning AM, Ponder BAJ, Easton DF, Pharoah PD. Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. Cancer Epidemiol Biomarkers Prev 2009;18:255–259. [PubMed: 19124506]
23. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. Nucleic Acids Res 2003;31:51–54. [PubMed: 12519945]
24. Maertens G, Cherepanov P, Pluymers W, et al. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem 2003;278:33528–33539. [PubMed: 12796494]
25. Singh DP, Kimura A, Chylack LT Jr., Shinohara T. Lens epithelium-derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. Gene 2000;242:265–273. [PubMed: 10721720]
26. Yokoyama A, Cleary ML. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell 2008;14:36–46. [PubMed: 18598942]
27. Ciuffi A, Llano M, Poeschla E, et al. A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 2005;11:1287–1289. [PubMed: 16311605]
28. Shinohara T, Singh DP, Fatma N. LEDGF, a survival factor, activates stress-related genes. Prog Retin Eye Res 2002;21:341–358. [PubMed: 12052388]

29. Sharma P, Singh DP, Fatma N, Chylack LT Jr., Shinohara T. Activation of LEDGF gene by thermal-and oxidative-stresses. Biochem Biophys Res Commun 2000;276:1320–1324. [PubMed: 11027629]
30. Cohen B, Addadi Y, Sapoznik S, et al. Transcriptional regulation of vascular endothelial growth factor C by oxidative and thermal stress is mediated by lens epithelium-derived growth factor/p75. Neoplasia 2009;11:921–933. [PubMed: 19724686]
31. Sun T, Gao Y, Tan W, et al. A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. Nat Genet 2007;39:605–613. [PubMed: 17450141]
32. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, Poeschla EM. Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. J Mol Biol 2006;360:760–773. [PubMed: 16793062]
33. Singh DP, Ohguro N, Chylack LT Jr., Shinohara T. Lens epithelium-derived growth factor: Increased resistance to thermal and oxidative stresses. Invest Ophthalmol Vis Sci 1999;40:1444–1451. [PubMed: 10359326]
34. Llano M, Morrison J, Poeschla EM. Virological and cellular roles of the transcriptional coactivator LEDGF/p75. Curr Top Microbiol Immunol 2009;339:125–146. [PubMed: 20012527]
35. Cherepanov P, Maertens G, Proost P, et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem 2003;278:372–381. [PubMed: 12407101]
36. Llano M, Saenz DT, Meehan A, et al. An essential role for LEDGF/p75 in HIV integration. Science 2006;314:461–464. [PubMed: 16959972]
37. Shun M-C, Raghavendra NK, Vandegraaff N, et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 2007;21:1767–1778. [PubMed: 17639082]
38. Vandekerckhove L, Christ F, Van Maele B, et al. Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. J Virol 2006;80:1886–1896. [PubMed: 16439544]
39. Singh DP, Fatma N, Kimura A, Chylack LT Jr., Shinohara T. LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes. Biochem Biophys Res Commun 2001;283:943–955. [PubMed: 11350077]
40. Fatma N, Singh DP, Shinohara T, Chylack LT Jr.. Transcriptional regulation of the antioxidant protein 2 gene, a thiol-specific antioxidant, by lens epithelium-derived growth factor to protect cells from oxidative stress. J Biol Chem 2001;276:48899–48907. [PubMed: 11677226]
41. Shin JH, Piao CS, Lim CM, Lee JK. LEDGF binding to stress response element increases alphaB-crystallin expression in astrocytes with oxidative stress. Neurosci Lett 2008;435:131–136. [PubMed: 18343576]
42. Daniels T, Zhang J, Gutierrez I, et al. Antinuclear autoantibodies in prostate cancer: Immunity to LEDGF/p75, a survival protein highly expressed in prostate tumors and cleaved during apoptosis. Prostate 2005;62:14–26. [PubMed: 15389814]
43. Daugaard M, Kirkegaard-Sorensen T, Ostenfeld MS, et al. Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer. Cancer Res 2007;67:2559–2567. [PubMed: 17363574]
44. Mediavilla-Varela M, Pacheco FJ, Almaguel F, et al. Docetaxel-induced prostate cancer cell death involves concomitant activation of caspase and lysosomal pathways and is attenuated by LEDGF/p75. Mol Cancer 2009;8:68. [PubMed: 19715609]
45. Daugaard M, Baude A, Fugger K, et al. LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat Struct Mol Biol 2012;19:803–810. [PubMed: 22773103]
46. Martinez-Pastor MT, Marchler G, Schuller C, Marchler-Bauer A, Ruis H, Estruch F. The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). EMBO J 1996;15:2227–2235. [PubMed: 8641288]

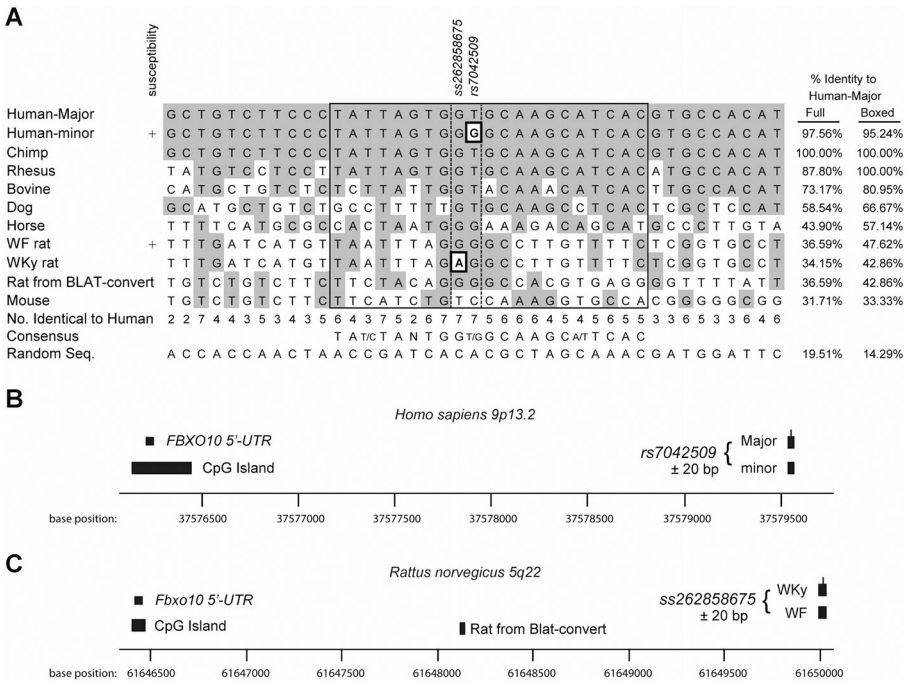

Figure 1.

Comparative genomic analysis of human breast cancer risk associated SNP *rs7042509*. (A) Alignment of human genomic DNA sequence containing SNP *rs7042509* with orthologous sequences from different species. Human genomic DNA sequence containing *rs7042509* major allele was compared to genomic reference sequences of other mammalian species to determine the percentage (%) of base pairs that were identical within the full and boxed sequences shown. Orthologous sequences were found by aligning the human major allele sequence to the human reference genome (GRCh37/hg19) using the UCSC genome browser BLAT function followed by the convert function. Rat genomic DNA containing WF and WKY variants of SNV *ss262858675*, which is in the same relative genomic position compared to human SNP *rs7042509*, was found to have similar sequence identity to human sequence containing SNP *rs7042509*. (B) Human genome (GRCh37/hg19) position of *rs7042509* relative to a *FBXO10* 5′-UTR. (C) Rat genome (Baylor 3.4/rn4) position of *ss262858675* relative to an *Fbxo10* 5′-UTR. Both human and rat variants are approximately 3.5 kb from respective 5′-UTRs and CpG islands. The rat sequence identified using BLAT-convert is located approximately 1.7 kb from an *Fbxo10* 5′-UTR and does not contain dimorphisms between susceptible WF and *Mcs5a1*-resistant WKY rat alleles.

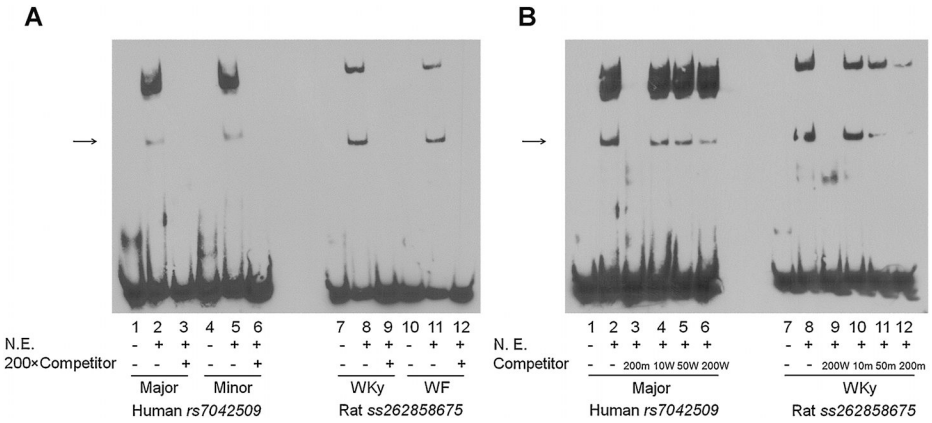

Figure 2.

Identification of nuclear protein–DNA interactions using EMSA. (A) Protein–DNA complexes identified in EMSA using 41 bp oligonucleotides containing human *rs7042509* ± 20 bp flanking sequence, and 41bp oligonucleotides containing rat *ss262858675* ± 20 bp flanking sequence with rat thymus NE. Biotin-labeled oligonucleotides (20 fmol) were incubated with 2 μg NE. The shifts marked with an arrow showed commonality between human and rat alleles. Unlabeled DS-oligonucleotides (200×) competed with labeled oligonucleotides for NE binding. (B) Human *rs7042509* and rat *ss262858675* alleles competed for binding to NE. Excess unlabeled 41 bp DS-oligonucleotides containing human *rs7042509* major allele competed with labeled 41 bp oligonucleotides containing rat *ss262858675* WKY allele for binding to rat thymus NE and vice versa. Lower-case letter m indicates human major allele of *rs7042509*. Upper case W indicates rat WKY allele of *ss262858675*. The number before letters m and W is the fold (×) excess of unlabeled oligonucleotides.

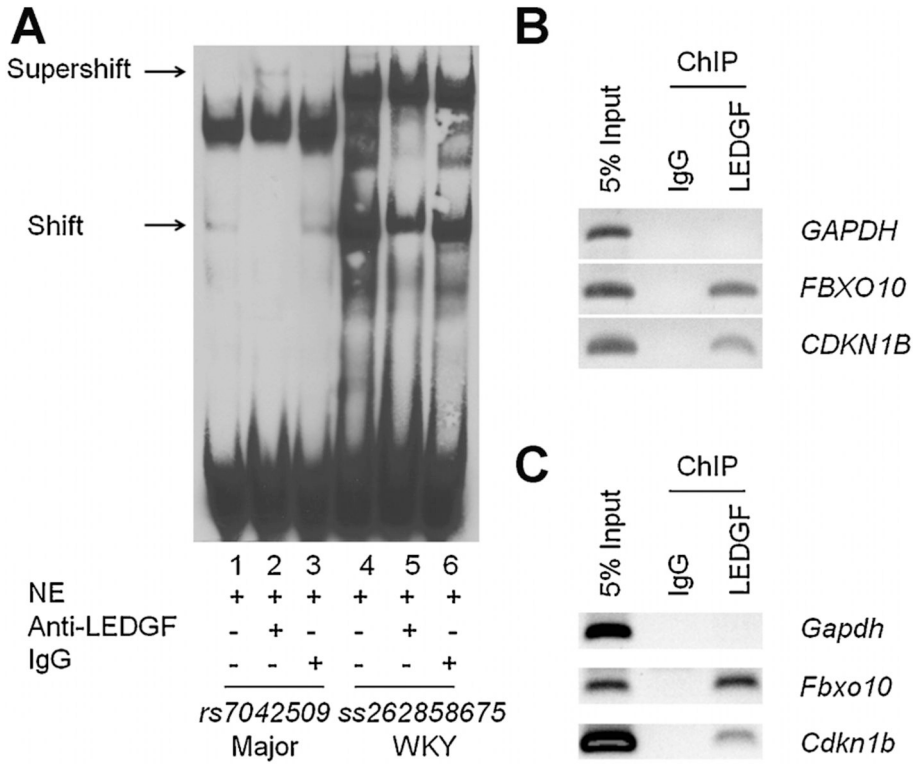

Figure 3.

LEDGF interacted with DNA sequence containing human breast cancer risk associated SNP *rs7042509* and rat *ss262858675*. (A) LEDGF was confirmed to bind human *rs7042509* and rat *ss262858675* containing oligonucleotides using LEDGF-antibody compared to an IgG control. Human *rs7042509* major and rat *ss262858675* WKY oligonucleotides were used to bind proteins in WKY rat thymus NE. Anti-LEDGF (2 μg) or IgG control antibodies were added to reaction mixtures prior to the addition of labeled oligonucleotides. (B and C) LEDGF was confirmed to bind human *rs7042509* major and rat *ss262858675* WKY oligonucleotides by chromatin immunoprecipitation. Chromatin immunoprecipitation was performed using both human Jurkat cell (B) and WKY rat thymus (C) genomic DNA. Human/rat *FBXO10/Fbxo10* regions of interest were amplified using species respective primers. Human/rat *GAPDH/Gapdh* and *CDKN1B/Cdkn1b* loci were included as respective negative and positive LEDGF binding controls.

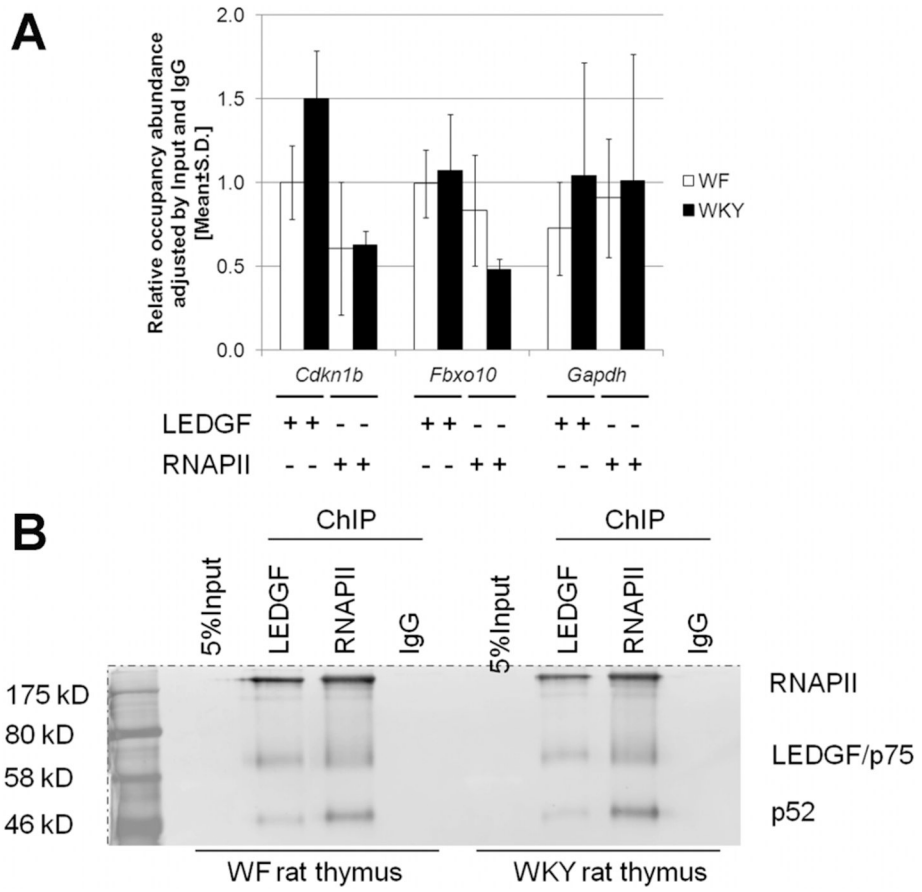

Figure 4.

Genomic sequences containing rat *Mcs5a1* SNV *ss262858675* susceptible WF and resistant WKY alleles bound similarly by LEDGF. (A) ChIP-QPCR results showing LEDGF enrichment at WF and WKY rat genomic regions containing SNV *ss262858675*. Rat thymus chromatin was used in chromatin immunoprecipitation (ChIP) of LEDGF, RNAPII, or IgG bound DNA. Locus-specific primers were used in QPCR to compare relative protein binding occupancy between strains. *Mcs5a1* SNV *ss262858675* was contained within the *Fbxo10* amplicon. Rat *Cdkn1b* and *Gapdh* loci were included as respective positive and negative LEDGF binding controls. There were no statistically significant differences between *Fbxo10* ChIP-QPCR groups ($P = 0.0752$, Kruskal–Wallis test). A Kruskal–Wallis test was significant for *Cdkn1b* ChIP-QPCR groups ($P = 0.0434$). LEDGF was enriched more at WKY *Cdkn1b* compared to WF ($P < 0.05$, Mann–Whitney's *U*-test). (B) LEDGF interacted with RNA polymerase II on both WF and WKY rat chromatin in ChIP-CoIP experiments. Chromatin complexes enriched for LEDGF, RNAPII, or IgG were collected after an immunoprecipitation wash step and before the DNA purification step of ChIP. Amounts of 5% input from initiating ChIP samples were also included. The band-labeled p52 is likely LEDGF/p52.

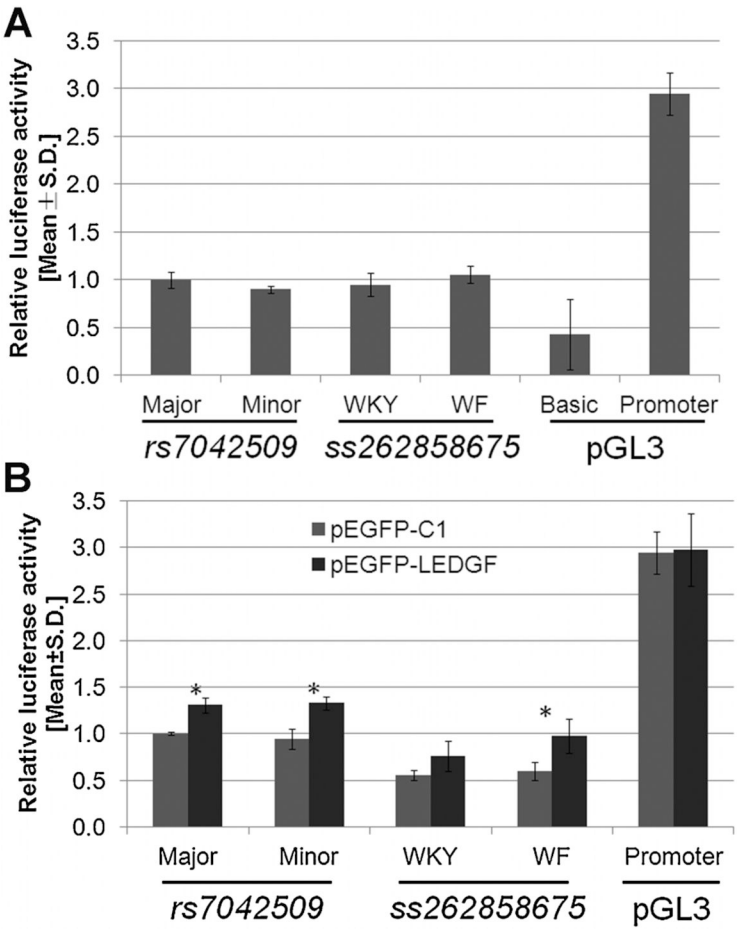

Figure 5.

Ectopic expression of LEDGF/p75 in Jurkat cells increased luciferase activities of reporters containing human SNP *rs7042509* and rat SNV *ss262858675* variants. Two copies each of human major, human minor, rat WKY and rat WF, respectively, were cloned into the pGL3-Promoter vector. Control reporter vector pRL-TK *Renilla* was used as a transfection control. Data are presented as mean ± SD of relative luciferase activity for each co-transfection group. (A) Each variant resulted in luciferase activities that were lower than the pGL3-Promoter parent vector, but higher than the promoterless pGL3-Basic vector. (B) Variant *rs7042509* and *ss262858675* respective luciferase reporters were co-transfected with pEGFP-LEDGF/p75 expression or pEGFP-C1 control vectors. Kruskal–Wallis test comparing all experimental groups was statistically significant ($P = 0.0061$). Asterisks indicate P-values < 0.05 of independent Mann–Whitney's U-tests comparing each LEDGF co-transfection to the respective control. Neither co-transfection with pEGFP-LEDGF/p75 nor pEGFP-C1 vector affected pGL3-Promoter luciferase activity.

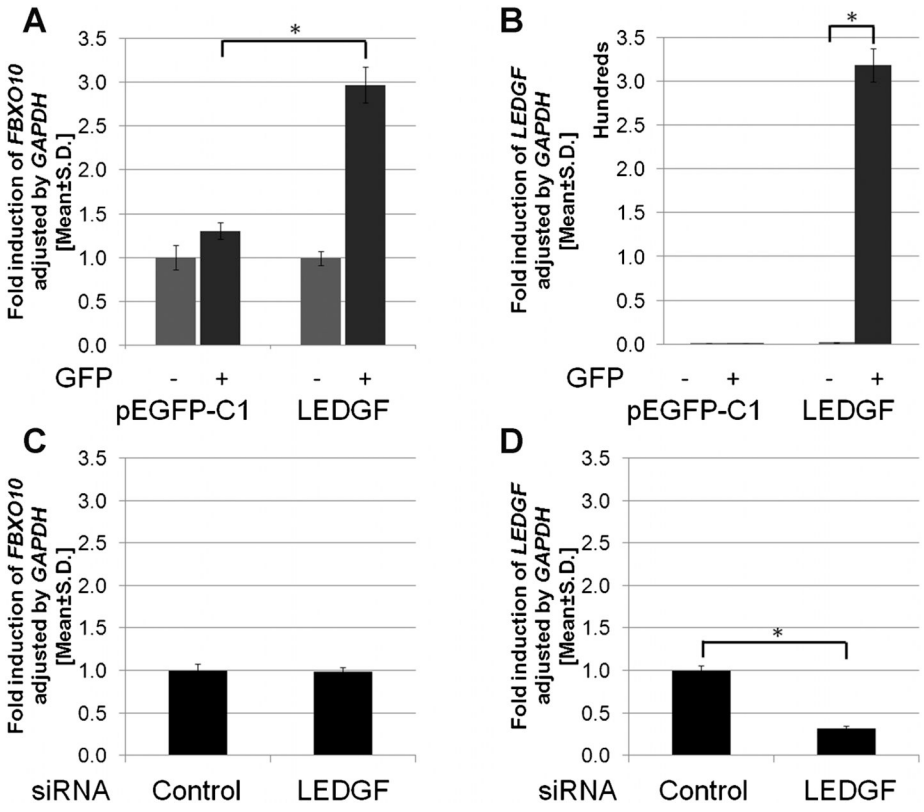

Figure 6.

Ectopic over-expression of LEDGF/p75 in Jurkat cells increased *FBXO10* levels, but knockdown of LEDGF/p75 had no effect on *FBXO10*. (A and B) Jurkat cells were transfected with pEGFP-C1 empty or pEGFP-LEDGF/p75 expression plasmids. Following transfection (48 h), cells were sorted and acquired based on green fluorescence using fluorescence activated cell sorting (FACS). Total RNA was used to synthesize cDNA for QPCR to determine *FBXO10* (A) and *LEDGF/p75* (B) transcript levels. Graphed are mean *Target/GAPDH* mRNA ± SD for each transfection and green fluorescence sorted group. (A) Transcript levels of *FBXO10* were significantly higher in green fluorescence positive cells transfected with pEGFP-LEDGF compared to pEGFP-C1 (Kruskal–Wallis test $P = 0.0249$ followed by a Mann–Whitney's *U*-test $P < 0.05$). (B) Increased *LEDGF* transcript levels were confirmed in Jurkat cells transfected with pEGFP-LEDGF that were green fluorescence positive compared to green fluorescence negative (Kruskal–Wallis test $P = 0.0249$ followed by a Mann–Whitney's *U*-test $P < 0.05$). (C and D) Knockdown of LEDGF/p75 in Jurkat cells using siRNA had no effect on *FBXO10* levels. Total RNA was collected for cDNA synthesis 48 h after transfection with LEDGF/p75 specific or scramble control siRNAs. Quantitative PCR was performed to determine *FBXO10* (C) and *LEDGF* (D) mRNA levels. The data presented are expressed as the mean *Target/GAPDH* mRNA ± SD for each experimental group. Jurkat cell *FBXO10* transcript levels were not changed by LEDGf knockdown ($P = 0.2482$, Mann–Whitney's *U*-test). Transcript levels of *LEDGF* in LEDGF siRNA transfected cells were 30% of levels in control siRNA transfected cells ($P = 0.0209$, Mann–Whitney's *U*-test).

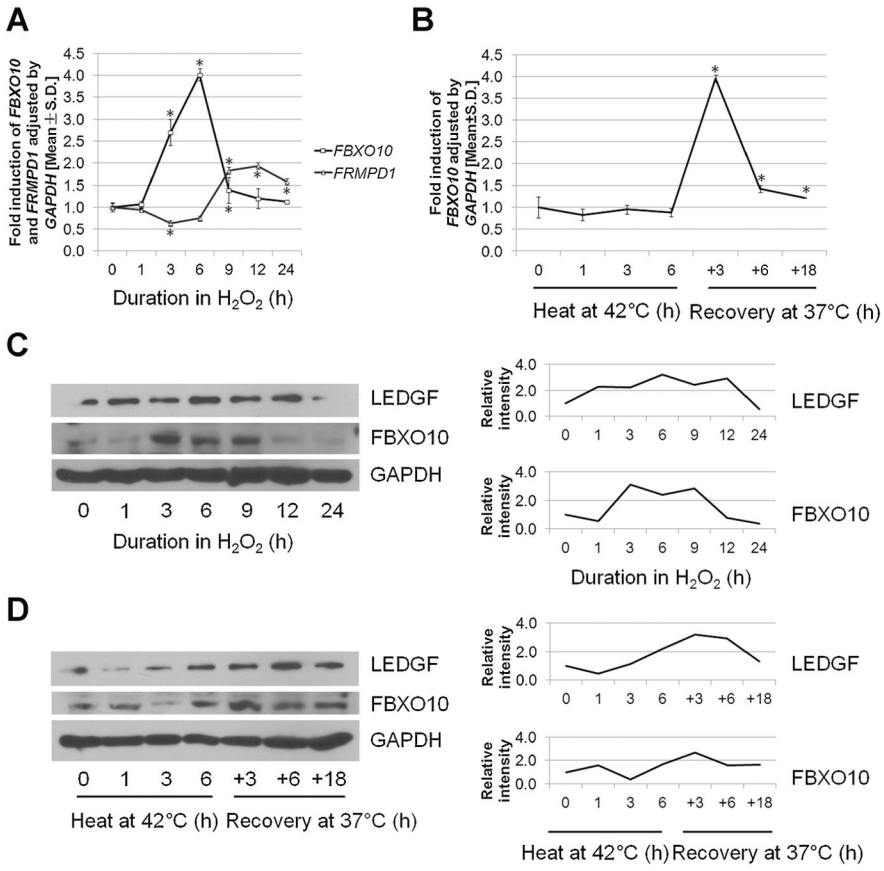

Figure 7.

Oxidative stress and recovery from heat stress resulted in transient increases in *FBXO10* mRNA and protein levels. (A) Oxidative stress (0.2 mM H_2O_2) resulted in increased *FBXO10* levels in Jurkat cells at 3, 6, and 9 h after initial exposure. Oxidative stress (0.2 mM H_2O_2) resulted in decreased *FRMPD1* levels in Jurkat cells at 3 and 6 h, and increased *FRMPD1* at 12 h after initial H_2O_2 exposure. Cells were collected at times (hours) shown on the *x*-axis after initial H_2O_2 exposure. *GAPDH* was used to standardize *FBXO10* mRNA expression. Results are presented as the mean ± SD of *Target/GAPDH* mRNA levels. Asterisks indicate *P*-values < 0.05 of Mann–Whitney's *U*-tests comparing *FBXO10* or *FRMPD1* between the respective time-point and 0 h following a significant Kruskal–Wallis test ($P = 0.0155$, *FBXO10*; and $P = 0.0220$, *FRMPD1*). B. Jurkat cells recovering at 37°C from 42°C heat stress had increased *FBXO10* mRNA levels. Jurkat cells were heat stressed for the times (hours) indicated on the *x*-axis. There was no change in *FBXO10* transcript levels during 6 h of heat stress; however, *FBXO10* increased in Jurkat cells recovering from 6 h of heat stress. Asterisks indicate *P*-values < 0.05 of Mann–Whitney's *U*-tests comparing *FBXO10* between the respective time-point and 0 h following a significant Kruskal–Wallis test ($P < 0.001$). (C and D) FBXO10 and LEDGF protein levels increased in H_2O_2 exposed and heat stress recovering Jurkat cells, respectively. Relative protein abundance standardized to GAPDH was measured by densitometry and these are shown adjacent to each respective Western blot.

Table 1.

Human *MCS5A1* Breast Cancer Risk Associated SNP *rs7042509* and Rat *Mcs5a1* Mammary Cancer Associated SNV *ss262858675* Sequence-Binding Proteins Identified Using Mass Spectrometry

Mass spectrometry identified protein	NCBI accession	Molecular weight (kDa)
Cytoplasmic beta-actin[a]	13592133	42
Heterogeneous nuclear ribonucleoprotein D[b]	13242324	38
PREDICTED: similar to Histone H2B 291B[c]	62663604	37
LEDGF[d]	28461167	60
PREDICTED: upstream binding transcription factor, RNA polymerase I	34873894	89
PREDICTED: similar to TCF7L2 protein	62642501	262

NCBI, National Center for Biotechnology Information.

[a]Also named actin, cytoplasmic 1.

[b]Also known as heterogeneous nuclear ribonucleoprotein D0 isoform a.

[c]*Rattus norvegicus* histone cluster 1, H2bm.

[d]Also named PC4 and SFRS1 interacting protein 1.

Mol Carcinog. Author manuscript; available in PMC 2022 December 10.