Published in final edited form as:

Mamm Genome. 2016 April ; 27(3-4): 99–110. doi:10.1007/s00335-015-9618-3.

Variants in an *Hdac9* intronic enhancer plasmid impact *Twist1* expression *in vitro*

Tyler E. Siekmann¹, Madelyn M. Gerber², and Amanda Ewart Toland^{3,*}

¹Biomedical Sciences Program, The Ohio State University College of Medicine, Columbus, OH, 43210, USA

²Biomedical Sciences Graduate Program, The Ohio State University College of Medicine, Columbus, OH, 43210, USA

³Department of Molecular Virology, Immunology and Medical Genetics and the Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210 USA

Abstract

Skin tumorsusceptibility 5 (Skts5) was previously mapped to mouse chromosome 12 through linkage analysis of skin tumor susceptible *Mus musculus* (NIH/Ola-S) and skin tumor resistant outbred *Mus spretus* (SPRET/Out-R) mice. *Hdac9* was identified as a potential candidate for Skts5 based on conserved non-synonymous sequence variants and expression analyses. Studies by others identified an enhancer in human *HDAC9* that correlated with *TWIST1* expression. We identified 45 sequence variants between NIH/Ola-S and SPRET/Out-R mice from the orthologous region of the human *HDAC9* enhancer. Variants mapping to intron 18 differentially affected luciferase expression *in vitro*. NIH/Ola-S clones showed an approximate 1.7-fold increased luciferase expression relative to vector alone or the equivalent clones from SPRET/Out-R-R. Furthermore, cells transfected with a portion of the NIH/Ola-S intron induced 2.2-fold increases in *Twist1* expression, but the same region from SPRET/Out-R mice resulted in no up-regulation of *Twist1*. *In silico* transcription factor analyses identified multiple transcription factors predicted to differentially bind NIH/Ola-S and SPRET/Out-R polymorphic sites. Chromatin immunoprecipitation studies of two transcription factors, Gata3 and Oct1, demonstrated differential binding between NIH/Ola-S and SPRET/Out-R plasmids that corroborated the *in silico* predictions. Together these studies provide evidence that the murine orthologous region to a human *HDAC9* enhancer also acts as a transcriptional enhancer for mouse *Twist1*. As ectopic sequence variants between NIH/Ola-S and SPRET/Out-R differentially impacted luciferase expression, correlated with *Twist1* expression *in vitro* and affected Gata3 and Oct1 binding, these variants may explain part of the observed differences in skin tumor susceptibility at Skts5 between NIH/Ola-S and SPRET/Out-R.

*Correspondence: 998 Biomedical Research Tower, 460 W. 12th Avenue, Columbus, OH 43210; Phone: 614-247-8185; Fax: 614-688-8675; Amanda.toland@osumc.edu.

Conflict of Interest

The authors declare that they have no conflict of interest.

Keywords

Hdac9, *Twist1*; enhancer; cutaneous squamous cell carcinoma; Skts5

Introduction

Mus spretus are highly resistant to chemically-induced skin cancer whereas *Mus musculus* mice are highly susceptible (Nagase et al. 1995). More than 15 skin cancer susceptibility loci have been mapped using these and other mouse strains that differ in susceptibility to skin (Nagase et al. 1995; Angel et al. 2003; Ewart-Toland et al. 2003; Fujiwara et al. 2007; Mahler et al. 2008; Fujiwara et al. 2010; Okumura et al. 2012). One such locus, skin tumor susceptibility 5 (Skts5) that maps to mouse chromosome 12, was identified in linkage analyses of NIH/Ola-S × outbred *Mus spretus* (SPRET/Out-R) backcross mice [(SPRET/Out-R × NIH/Ola-S) × NIH/Ola-S]. In subsequent studies, we refined the peak linkage region of Skts5 to approximately 14 Mb (chr12:31.3-45.35 Mb; GRCm38/mm10) (Mahler et al. 2008). To identify candidate genes for Skts5 we performed extensive sequence analysis of all named genes mapping to Skts5 and expression studies of genes mapping to this region by both RNA-seq and qPCR analyses of normal skin RNA from the strains of mice used for the linkage analyses (Mahler et al. 2008; Skeeles et al. 2013). A number of potential candidate genes for Skts5 were identified by sequence and expression differences between SPRET/Out-R and NIH/Ola-S mice; however, as recent genome-wide association studies (GWAS) are demonstrating, a large proportion of genetic variants conferring susceptibility to disease reside in non-coding regions that are predicted to be regulatory and include promoters, enhancers, and non-coding RNAs (Maurano et al. 2012). Similarly, many disease-associated regions have been found to house expression quantitative trait loci (Nica et al. 2010). Thus, from the data emerging from human studies, variants in enhancers or other regulatory element should also be considered.

A study searching for exonic enhancers identified an enhancer in the human *HDAC9* gene spanning exons 18-19 that is correlated with *Twist1* expression in the limbs of mice carrying the transgene (Birnbaum et al. 2012). HDAC9 is a class II histone deacetylase that represses gene transcription through deacetylation of histones H3 and H4. HDAC9 has been implicated in cancer and is expressed in the hair follicle (Brockschmidt et al. 2011; Parra, 2015). In the mouse, both *Hdac9* (XM_006515263.1) and *Twist1* (NM_011658.2) map to the peak region of linkage for Skts5 (Mahler et al. 2008). *Twist1* is a transcription factor with a basic helix-loop-helix domain that homo- or hetero-dimerizes with partners to act as either a transcriptional activator or repressor (Qin et al. 2012). It has a documented role in metastasis of melanoma and basal cell carcinoma (Majima et al. 2012; Weiss et al. 2012). In humans, TWIST1 has been shown to suppress apoptosis, override senescence, induce angiogenesis, and increase cancer stem cell populations (Maestro et al. 1999; Mironchik et al. 2005; Ansieau et al. 2008; Mani et al. 2008). In a 7,12-dimethylbenz[a]anthracene (DMBA) and 12-O-tetradecanoylphorbol-13-acetate (TPA) mouse skin cancer model, inducible *Twist1* expression led to a higher conversion rate of papilloma to invasive squamous cell carcinoma (SCC) (Tsai et al. 2012). Conversely, a keratinocyte-specific knock-out of *Twist1* protects mice from SCC (Srivastava et al. 2015). From these studies, we

hypothesized that *Twist1* was a strong candidate gene for the skin cancer susceptibility locus Skts5, and that there might be strain specific differences in the region containing the enhancer between NIH/Ola-S and SPRET/Out-R mice that could impact susceptibility to skin cancer through regulation of *Twist1*.

To test this hypothesis, we conducted sequence analysis of the exons and intron sequences near the described enhancer and evaluated variants for effects on expression *in vitro*. Here, we describe our finding that variants in *Hdac9* which correlate with expression of *Twist1* could be important in the skin cancer susceptibility differences observed in these mice.

Materials and Methods

Sequencing

Tails from NIH/Ola-S mice were provided by Dr. Allan Balmain and tails from SPRET/Out-R mice were provided by Hiroki Nagase as approved by the University of California, San Francisco Institutional Animal Care and Use Committee. DNA was isolated from tails using standard methods (Laird et al. 1991). Intronic and exonic sequences of the *Hdac9* gene orthologous to the published enhancer region in human were identified using the UCSC database build GRCm38/mm10. We designed PCR primers using Integrated DNA Technology's SciTools PrimerQuest web-based program (Coralville, IA). Genomic tail DNA was amplified using Qiagen's Taq polymerase kit, according to the manufacturer's protocol (Valencia, CA). Amplified products were analyzed by gel electrophoresis to confirm product size and quality. PCR products were treated with Exo/SAP-IT (USB; Cleveland, OH) to remove single stranded DNA. Automated sequencing of PCR products was conducted on an ABI 3700 by standard methods (Foster City, CA). Primers used for PCR were also used for the sequencing (Supplemental Table 1). Forward and reverse sequences were analyzed and compared using DNASTAR 3.0 (Madison, WI). The chromatograms were inspected visually whenever a nucleotide substitution was indicated.

Cloning

PCR products from NIH/Ola-S and SPRET/Out-R were TA cloned or cloned using Clontech's InFusion HD cloning kit (Mountain View, CA). For TA cloning a Promega pGL3 promoter vector (Madison, WI) was linearized at the promoter multiple cloning site using the restriction enzyme SmaI (New England Biolabs, Ipswich, MA). Thymine nucleotide overhangs were created using a modified protocol of Graham and Holton (1990). PCR products were ligated to the modified vector via complementary A-T base pairing. Ligation products were transformed into Stellar competent cells according to manufacturer's recommended conditions (Clontech, Protocol PT5055-2). Plasmid DNA was purified using GeneJET plasmid miniprep kit (ThermoScientific, Waltham, MA). Plasmids were sequence by Sanger Sequencing to verify the presence of the expected DNA inserts. Intron segments that were cloned into the pGL3 promoter empty vector by the In-Fusion method were PCR amplified from genomic DNA using primers specifically designed to recombine the product into the pGL3 promoter plasmid at sites flanking the SmaI site used for plasmid linearization. Recombined plasmids were transformed into Stellar competent cells in the same manner as TA-cloned plasmids. Clones that were positive for recombination of the

intron fragment into the pGL3 promoter vector were detected using colony PCR. Plasmid DNA was isolated from colonies and sequence was confirmed by Sanger sequencing.

Site-directed mutagenesis

Site-directed mutagenesis was performed to introduce specific sequence substitutions using the QuikChange Lightning Site-Directed mutagenesis kit according to manufacturer's protocol (Agilent, Santa Clara, CA). Variants were confirmed by Sanger sequencing. The sequences being altered and primers for introducing or abolishing the putative NIH/Ola-S Gata3 and SPRET/Out-R Oct1 binding sites are listed in Supplemental Materials (Supplemental Table 2).

Cell culture and transfections

A5, a murine cutaneous spindle cell line derived from a DMBA/TPA-treated SPRET/Out-R × CBA F1 mouse carcinoma, and C5N, a non-tumorigenic murine keratinocyte cell line isolated from a Balb/C mouse, were obtained from Allan Balmain (Zoumpourlis et al. 2003). These lines were chosen for study because they are of the relevant murine tissues to study candidates for Skts5 (keratinocyte or SCC derived). A5 was grown in Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. Transient transfections were performed using Lipofectamine LTX and Plus reagents (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. Cells were plated in triplicate and transfected at 60-80% confluency. Mock transfection and pGL3 promoter empty vector transfections were included as controls. For luciferase assays, cells were co-transfected with a Promega pRL-TK renilla reporter vector. To determine transfection efficiency, we performed co-transfection of constructs with a Clontech pEGFP NI vector and assessed for GFP-positive cells. Transfection efficiency was around 60% for all cell lines and plasmids evaluated.

Luciferase Assays

Cells were incubated at 37°C for 24 hours post-transfection. Cell lysates were prepared using M-PER (Pierce Biotechnology, Rockford, IL) and 30 µl of each sample were used for analysis. Luciferase assays were completed as described (Skeels et al. 2013). Firefly and renilla luciferase measurements were acquired 24 hours post-transfection using a Veritas Microplate Luminometer. Luciferase assays were done a minimum of three times with the exception of inserts that did not show differences which were repeated only once.

Quantitative PCR

RNA was harvested from transfected cells at 24 and 48 hours post-transfection using a modified RiboZol protocol (Amresco, Solon, OH). One microgram of RNA from each sample was reversed transcribed using iScript cDNA synthesis kit (Bio-Rad, Hercules, CA). Taqman probes for *Hdac9* (Mm00458454), *Twist1* (Mm00442036/Mm.PT.56a41778798) and *Hprt* (Mm00446968/Mm.PT.39122214828) were purchased from Life Sciences Technologies or Integrated DNA Technologies. To measure expression of *Hdac9* and *Twist1*, mRNA expression was measured at 24 hours post-transfection of the pGL3-enhancer construct in A5 cells. *Hprt* was used as a normalization control to calculate the relative

expression of each test gene using the delta Ct method. Standard deviations of fold differences of the test gene relative to the control were made by comparisons between different transfections. Experiments included no template and no reverse transcriptase controls for each gene. Each sample was measured in triplicate and each experiment was performed three times for genes showing differences in expression.

In silico transcription factor binding prediction

Multiple web-based prediction programs were used to predict transcription factor binding differences for *Hdac9* intron 18 based on the DNA sequence variations between NIH/Ola-S and SPRET/Out-R. These include TF Search (<http://www.cbrc.jp/research/db/>) (Akiyama, 1995), TRANSFAC ® 7.0 Public 2005 database produced by BIOBASE (<http://www.gene-regulation.com/cgi-bin/pub/databases/transfac/search.cgi>) (Heinemeyer et al. 1999), TFSiteScan Database (<http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl>) (Ghosh, 2000), and DBD (www.transcriptionfactor.org) (Wilson et al. 2008). TFSearch was the primary tool utilized, and the recommended threshold score of 85.0 was used as a cutoff of significance for predicted binding. Candidate transcription factors were prioritized for chromatin immunoprecipitation studies if they had high threshold scores for one strain but not the other and were predicted by multiple programs.

Chromatin immunoprecipitation (ChIP) studies

Chromatin immunoprecipitation (ChIP) studies were completed using a modified Millipore protocol (www.millipore.com). In brief, A5 and C5N cells used for transfection were grown to a count of 1×10^6 cells in a 10 cm dish. Cells were transfected with pGL3 promoter vector only, pGL3-NIH/Ola-S Insert 1, or pGL3-SPRET/Out-R Insert 1 constructs. At 24 hours post-transfection, cells were treated with a 1% formaldehyde solution in supplemented DMEM (10% FBS, 1% penicillin/streptomycin) for 10 min at 37°C to crosslink DNA and proteins. Harvested cells were suspended in 500 µl of SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1) and sonicated at 4°C using a Fisher Scientific Sonic Dismembrator Model 500 under the following conditions: 12 total cycles of 30 s sonication and 60 s rest at 40% amplitude. Sonication resulted in sheared DNA of 200-400 bp in length as determined by gel electrophoresis. Samples were precleared of non-specific proteins by incubating for 30 min. at 4°C in a 50% Salmon Sperm DNA/Protein A agarose slurry. Antibodies for ChIP targeted Oct1 (A301-716A, Bethyl Laboratories, Montgomery, TX) and Gata3 (sc-268, Santa Cruz Biotechnologies, Santa Cruz, CA). Antibody concentrations for ChIP were chosen based on manufacturer's recommended conditions for ChIP and were 0.2 µg/µl for Gata3 and 1 µg/µl for Oct1. Precleared crosslinked DNA/protein was incubated with Oct1 or Gata3 antibodies or controls (no antibody) for 16 hours at 4°C in ChIP dilution buffer per Millipore recommended ChIP conditions (Darmstadt, Germany). Samples were washed using recommended wash conditions and solutions. Crosslinked proteins were eluted from the DNA using elution buffer (1% SDS, 0.1M NaHCO₃) followed by reversal of crosslinks by treatment of supernatant with 5M NaCl for 4 hours at 65°C and removal of protein by Proteinase K treatment. Residual protein was removed by phenol/chloroform/isoamyl extraction and DNA was ethanol precipitated. DNA pellets were resuspended in 40 µl of dH₂O. We performed both semi-quantitative PCR and quantitative real-time PCR (qPCR). One µl of precipitated DNA for each condition and control step was used for semi-

quantitative PCR. Thermocycler parameters were as follows: 95°C for 10 minutes followed by 35 cycles of 94°C for 40s, 56°C for 40s, 72°C for 40s and a final extension step of 72°C for 10 minutes. Reactions contained 1 Unit Qiagen Taq polymerase, 250µM dNTPs, 0.5µM PCR Primers (Supplemental Material, Table 3), Q solution, and 1X Qiagen PCR buffer. PCR products were run on 1% agarose gels, stained with ethidium bromide and visualized using an Alpha Imager (Protein Simple, San Jose, CA). Semi-quantitation of PCR products was performed using Alpha Imager software for pixel intensity. Pixel density for NIH/Ola-S and SPRET/Out-R plus Gata3 or Oct1 antibody lanes were compared to vector-only transfected cells, no antibody and/or GFP antibody controls. Reactions for qPCR contained 0.8 µl ChIP DNA or input DNA (for normalization), 1X SYBR Green Mastermix (Bio-rad) and 1 µM of each primer. Each sample was run in triplicate.

Statistical analysis

When more than 2 values were being compared, which was for the majority of experiments, a one-way ANOVA was utilized. Comparisons resulting in a p-value corresponding to the F-statistic of the one-way ANOVA < 0.05 were then assessed by a post-hoc Tukey HSD test to determine which groups were significantly different from the others. For each comparison, a p-value for the observed Q-value was generated and used to determine significance.

Results

Identification of sequence variants between NIH/Ola-S and SPRET/Out-R at Hdac9 intron 18

Skts5 is a susceptibility locus for chemically-induced skin cancer that maps to mouse chromosome 12. From our studies in mouse and human samples, *Hdac9* emerged as a potential candidate gene for skin tumor susceptibility at Skts5 (Mahler et al. 2008; Fleming et al. 2014). To test whether strain-specific differences between NIH/Ola-S and SPRET/Out-R mice in the orthologous region of a predicted human *HDAC9* enhancer impacted potential enhancer activity and could therefore be considered as a candidate for the skin tumor susceptibility mapping to Skts5, we first sequenced approximately 5000 bp of *Hdac9* intron 18, corresponding to the described human *HDAC9* enhancer, in NIH/Ola-S and SPRET/Out-R (Figure 1). Forty-five single nucleotide polymorphisms or small indels were identified between the strains (Supplementary Table 4). In previous studies, we identified no coding changes in the corresponding mouse *Hdac9* exons 18 or 19 (Mahler et al. 2008).

Conservation of Hdac9 intron 18 across species

To assess conservation of this region across a more diverse group of species, we performed an nBlast analysis for the first 999bp of mouse intron 18. Interestingly, the only species showing a high degree of homology ($>80\%$) were primate species (Figure 1b and data not shown). We then performed species specific Blast using this region and identified strong homology for rat, dog, and other mammalian species (Figure 1b and data not shown). The only non-mammalian species showing homology was chicken (*Gallus gallus*), but only for the first 400 bp of intron 18. Sequence alignments for multiple species were done using Multiple Sequence Alignment (Clustal Omega) (<http://www.ebi.ac.uk/Tools/msa/>).

As it appears from the literature that the region containing both mouse and human *HDAC9* intron 18 show evidence of enhancer activity (Birnbaum et al. 2012), we were curious if any of the polymorphisms between NIH/Ola-S and SPRET/Out-R were present in humans. By nBlast analysis, the first 999bp and last 959bp of intron 18 are highly conserved (homologies of 87% and 89% respectively) between mouse and human. In the first 1000 bp of human intron 18 there are 21 variants listed in dbSNP137; most of these are extremely rare. In comparison of these with the mice, none of the variants observed in the human were found to be different between the strains of mice and *vice versa*. We assessed this region in the UCSC genome browser and queried for functional relevance of the human variants by HaploReg3 (Broad Institute) to obtain *in silico* evidence that this region might be relevant to the skin or cancer. ENCODE results indicate H3K27ac histone modifications for cell lines H1-hESC, NT2-D1, and K562. Several promoter marks for fat and skin were noted to be present throughout the region by HaploReg3 and several variants were noted to map to regions of “weak enhancer” activity. Multiple transcription factors including GATA2, GATA3, FOXP2, CEBP3, p300 and RAD21 were shown to bind to this locus by ChIP-seq analyses of other tissue types. The human *in silico* data, although not definitive because the exact tissues used for this study were not specifically analyzed, are supportive of this region being an enhancer.

Identification of segments showing enhancer activity

To determine if the *Hdac9* intron 18 showed enhancer activity in mouse skin cell lines, we divided the intron into nine regions (inserts) of approximately 700 bp each (Figure 1). Polymorphisms between the strains were not distributed randomly across the intron as the distal portion contained no sequence differences. We cloned Inserts 1, 2, 3, 4, 3+4, 5, 6, and 7 from both NIH/Ola-S and SPRET/Out-R into a pGL3 promoter reporter vector (empty vector) and transfected them into C5N, a non-tumorigenic mouse keratinocyte cell line, and A5, a spindle SCC mouse cell line (Zoumpourlis et al. 2003). Insert 7 was chosen as a control as this region did not contain any polymorphisms between the strains, and were therefore not expected to show strain-specific differences. We were most interested in results for Insert 1 and Insert 2 as they contained 6 and 4 variations between the strains respectively (Figure 1). At 24 hours post-transfection, cells containing Insert 1 from NIH/Ola-S had a 1.4- to 1.8-fold increased luciferase expression relative to vector-only transfected A5 cells and a 1.3- to 1.5-fold increased luciferase expression relative to SPRET/Out-R Insert 1-transfected A5 cells (p-values 0.001 and 0.001, respectively; Figure 2A). C5N cells transfected with NIH/Ola-S Insert 1 showed a similar increase of luciferase expression over empty vector and SPRET/Out-R Insert 1 of 1.7- to 2-fold and 1.3- to 1.5-fold (p-values 0.001 and 0.011, respectively; Figure 2B). Insert 2 from NIH/Ola-S also showed a 1.3- to 1.4-fold increased increase in luciferase expression in both cell lines over both empty vector only and SPRET/Out-R Insert 2-transfected A5 cells (p-values 0.014 and 0.017, respectively; Figure 2A) and a modest 1.3-fold increased expression over empty vector and SPRET/Out-R Insert 2-transfected C5N cells (p-values 0.03 and 0.059, respectively, Figure 2B). As expected, Insert 7 showed no evidence of differences in enhancer activity between NIH/Ola-S and SPRET/Out-R or between Insert 7 and empty vector only controls (Figures 2A and B). None of the additional Inserts (3, 4, 3 +4, 5 or 6) showed any differences in enhancer activity between the strains and vector only controls (Figures 2A and B). These

results suggest that the proximal portion of *Hdac9* intron 18, like the orthologous region at human *HDAC9* intron 18, exhibits enhancer activity *in vitro*.

Hdac9 intronic enhancer upregulates Twist1 expression

To determine if the intronic regions of *Hdac9* intron 18 influence expression of nearby genes, we performed qPCR studies of *Hdac9* and *Twist1* in A5 and C5N cells. At the time of initiating experiments we did not know if the *Hdac9* intronic sequence in the plasmid would be able to affect expression of an endogenous gene as it would be outside of its normal genomic position relative to the *Twist1* promoter. However, the studies which originally identified the *Hdac9* enhancer used transgenic zebrafish with sections of the human enhancer integrated into the zebrafish genome and were able to document enhancer activity (Birnbaum et al. 2012). Thus, we theorized that if this DNA was acting in some kind of looping mechanism to enhance expression of nearby genes it might retain enhancer activity without being dependent on a specific genomic location. *Hdac9* was chosen for expression studies as some intronic enhancers have been found to affect expression of the genes in which they map (Su et al. 2013). *Twist1* is a neighboring gene to *Hdac9* and in previous studies showed similar expression patterning in transgenic mice with lacZ under control of the human *Hdac9* enhancer (Birnbaum et al. 2012). Thus, *Twist1* was a compelling gene to assess as a putative target.

Hdac9 was expressed at very low levels in A5 and was below detection limits in C5N cells. No consistent differences in *Hdac9* expression were observed for transfections in A5 of either NIH/Ola-S or SPRET/Out-R Insert 1 constructs or Insert 7 constructs (Figure 3A). However, *Twist1* showed an approximate 2.2-fold increase in expression for the NIH/Ola-S Insert 1 reporter relative to empty vector and SPRET/Out-R Insert 1 reporter transfections in A5 cells (p-values 0.009 and 0.012, respectively; Figure 3B). There was a similar increase in *Twist1* expression of 2.2-fold in A5 cells transfected with NIH/Ola-S Insert 2 pGL3 plasmid DNA (p-values 0.002 and 0.003 respectively, Figure 3B). C5N cells transfected with empty vector or Inserts 1, 2, and 7 had extremely low levels of *Twist1* expression which was below reliable levels for quantitative comparisons (data not shown). As expected, for Insert 7, no increase in expression was seen for A5 cells transfected with the plasmid containing Insert 7 for NIH/Ola-S or SPRET/Out-R relative to mock-transfected cells (Figure 3B).

Oct1 and Gata3 show strain specific binding to Hdac9 intron 18 insert 1

We utilized several *in silico* programs that predict transcription factor binding to determine which potential proteins might be involved in the differential enhancer activity observed between NIH/Ola-S and SPRET/Out-R. We first determined whether any of the sequence variants identified between NIH/Ola-S and SPRET/Out-R across the entire intron 18 were predicted to affect transcription factor binding using TFSearch as our primary search tool (Akiyama, 1995). Twelve transcription factor binding sites were predicted to be disrupted by a variant in NIH/Ola-S or SPRET/Out-R (Table 1). Because variants in NIH/Ola-S Insert 1 resulted in the largest increase in luciferase expression over both empty vector and SPRET/Out-R, we specifically were interested in further analysis of the transcription factor binding sites in Insert 1 using different *in silico* programs. Six transcription factor binding sites for five different transcription factors were identified that preferentially bound to either NIH/

Ola-S or SPRET/Out-R Insert 1 DNA for one or more prediction programs (Table 1). These included Oct1, CdxA, Gata3, Gfi-1, and AP-1. Because six of the 12 variants predicted to disrupt/introduce transcription factor binding sites mapped to Insert 1, we focused on this region in our subsequent studies.

As our luciferase data and *Twist1* qPCR showed higher expression with ectopic NIH/Ola-S plasmid DNA relative to SPRET/Out-R, we postulated that the transcription factors binding to this region could either act to enhance expression, in the case of NIH/Ola-S, or potentially repress expression, in the case of SPRET/Out-R. We conducted a literature review of the five transcription factors predicted to differentially bind to variants between the strains in order to identify those that were described as transcriptional activators, repressors or both. In addition, we assessed the literature to identify factors that correlated with *Twist1* expression and/or those that have been reported to be active in the skin, skin cancer or other cancers. Based on the literature review, two transcription factors of high interest emerged: Gata3 and Oct1 (Kaufman et al. 2003; Kurek et al. 2007; Shakya et al. 2011; Masse et al. 2014). Depending on the context, Oct1 and Gata3 are associated with both transcriptional activation and repression in the skin (Kurek et al. 2007; Shakya et al. 2011; Masse et al. 2014). Gata3 is a critical transcription factor for epidermal and hair follicle differentiation and is overexpressed in basal cell and squamous cell carcinomas (Kaufman et al. 2003; Riker et al. 2008; Masse et al. 2014). Oct1 is expressed at high levels in normal keratinocytes of the epidermis and is also reported to be important in melanoma (Sturm et al. 1993; Jang et al. 2000). Thus, Gata3 and Oct1 show evidence of being critical transcription factors in the skin and in skin-related cancers.

From our *in silico* transcription factor binding analyses, Oct1 is predicted to preferentially bind SPRET/Out-R *Hdac9* intron 18 Insert1 at two sites and Gata3 is predicted to preferentially bind NIH/Ola-S at one site (Table 1 and data not shown). There is a second Gata3 site in Insert 1 which is not polymorphic between the strains and as such is predicted to bind both equally (Figure 1b). To determine whether Gata3 and Oct1 bound to NIH/Ola-S versus SPRET/Out-R Insert 1 as predicted, we performed ChIP studies in A5 and C5N cell lines comparing the binding of these to the ectopic plasmid DNA (Figure 4; Supplementary Figure 1). To separate the two predicted Oct1 and Gata3 binding sites, Insert 1 was further divided into two regions of about 200 bp each (Supplemental Table 3). Quantitative ChIP for Oct1 in A5 cells showed that Oct1 preferentially bound SPRET/Out-R plasmid DNA approximately 1.7-fold higher for region 1 relative to NIH/Ola-S and vector only (p-values 0.02 and 0.008, respectively; Figure 4A) and ~1.7-fold higher for region 2 (p-values 0.01 and 0.009, respectively, Figure 4C). Gata3 bound both SPRET/Out-R plasmid DNA and NIH/Ola-S plasmid DNA at a similar level for Insert 1 region 1, which was not predicted to be different as there was not a polymorphism in the Gata3 binding site between the strains (Figure 4B). In Insert 1 region 2, Gata3 preferentially bound to the NIH/Ola-S plasmid DNA compared to SPRET/Out-R and vector only DNA with approximately 1.5-fold higher binding (p-values 0.01 and 0.001, respectively, Figure 4D). Very similar results were observed for C5N cells by semi-quantitative PCR (Supplementary Figure 1). These results are consistent with the luciferase data and *in silico* predictions and suggest that variants in the Insert 1 region of *Hdac9* intron 18 affect transcription factor binding.

Variant effects on luciferase activity

To provide additional evidence that the variants predicted to bind Gata3 and Oct1 were influencing luciferase and *Twist1* expression, we substituted the nucleotide observed in one strain for that of the other using site-directed mutagenesis of these sites in NIH/Ola-S and SPRET/Out-R Insert 1 containing plasmids. We expected that alteration these sites (from NIH/Ola-S to SPRET/Out-R or SPRET/Out-R to NIH/Ola-S sequence) would reverse the luciferase effects observed if the Gata3 and/or Oct1 binding sites in the vector DNA were impacting luciferase expression and/or *Twist1* expression. Alterations to the Gata3 site unique to NIH/Ola-S decreased both luciferase expression and *Twist1* expression 1.3-fold from wild-type NIH/Ola-S plasmid DNA (p-values 0.001 and 0.007, respectively; Figure 5). Alteration of the second Oct1 site in SPRET/Out-R increased luciferase and *Twist1* expression 1.2- and 1.5-fold over wildtype SPRET/Out-R Insert 1, but this was not statistically significant (Figures 5A and 5B). These sequence substitutions did not fully rescue the expression as the altered NIH/Ola-S still showed higher luciferase (not statistically significant) and *Twist1* expression relative to altered SPRET/Out-R (1.2 and 3-fold, respectively; p-values 0.07 and 0.001, respectively; Figure 5), suggesting that other variants within NIH/Ola-S Insert 1 may also contribute to the observed increase in luciferase and *Twist1* expression.

Discussion

Our study shows that plasmid variants in the mouse orthologous locus to a human enhancer in *HDAC9* impact *Twist1* expression *in vitro*. Furthermore, we identified two transcription factors, Gata3 and Oct1, which differentially bind to variants present in skin cancer susceptible NIH/Ola-S or skin cancer resistant SPRET/Out-R mice. Recent studies show that GWAS variants for complex disease are enriched in large evolutionarily conserved enhancer regions suggesting that variants in the long, conserved *HDAC9/Hdac9* enhancer evaluated in this study may be at an increased likelihood to influence phenotypes (Parker et al. 2013). Thus, variants in this enhancer may partly explain the skin cancer susceptibility differences between these mice.

The HDAC9 enhancer and TWIST1 expression

The first evidence that *HDAC9* houses an enhancer that may affect *TWIST1* expression came from a ChIP-seq study looking for enhancer marks that overlapped with exons in human cell lines and mouse developmental tissues (Birnbaum et al. 2012). *HDAC9* exons 18 and 19 were found to contain enhancer marks, yet *HDAC9* was not expressed in the tissues used for the ChIP-seq studies. *In vivo* studies showed that the human *HDAC9* exon18 through exon19 enhancer drove LacZ expression in the limb buds of E11.5 mice. *Twist1* expression exhibited very similar patterning to the *HDAC9* enhancer driving LacZ expression suggesting that this enhancer might affect neighboring *Twist1* expression. Interestingly, the genomic distance and relationship of *HDAC9* and *TWIST1* to each other is conserved from fish through humans suggesting that the enhancer may be conserved across vertebrates (Birnbaum et al. 2012). Further refinement of the human *HDAC9* enhancer locus showed that exonic sequences plus neighboring intronic sequences augmented the enhancer expression (Birnbaum et al. 2012). As we did not have any sequence variants in the mouse

equivalent exons to human exon 18 or 19, we focused on the intervening intronic regions for this study. Results from our study build upon the initial findings by Birnbaum et al. as we provide evidence that the NIH/Ola-S version of the *Hdac9* intron 18 enhancer increases *Twist1* expression more than the SPRET/Out-R version.

HDAC9 as a candidate gene/locus for skin cancer susceptibility

HDAC9/Hdac9 is a class II histone deacetylase that represses gene transcription through deacetylation of histones H3 and H4 (Parra, 2015). HDAC9/Hdac9 has been implicated in cardiac growth, T-regulatory cell function, cancer, and muscle differentiation and has been shown to be important in homologous recombination (Kotian et al. 2011; Parra, 2015). On its own merits, *Hdac9* is a candidate gene for Skts5; it contains amino acid and expression differences between the strains of mice used for linkage (Mahler et al. 2008; Skeeles et al. 2013). In human cutaneous SCCs (cSCC), we previously found that variants in *HDAC9* show evidence of allele-specific gains in cSCC samples (Fleming et al. 2014). GWAS studies for male pattern baldness have identified SNPs mapping near *HDAC9* (Brockschmidt et al. 2011). Thus, there is evidence that SNPs mapping in or near *HDAC9* have a role in hair follicle biology and multiple types of cancer. At this time, it is unknown whether the main effect of these variants is on HDAC9 expression or function or if they are in linkage disequilibrium with enhancer variants and their primary effects are on *TWIST1* regulation. It is possible that there is a combinatorial effect of SNPs impacting HDAC9 expression and/or function as well as the *HDAC9* associated- enhancer leading to increased *TWIST1* expression in skin cancer. Future studies will be important to see if the SNPs associated with skin or lung cancer or male-pattern baldness are in linkage disequilibrium with variants in the *HDAC9* enhancer element.

Transcription regulation of TWIST1

Multiple transcriptional activators and repressor of *TWIST1* in tumors have been identified (Qin et al. 2012; Hwang-Verslues et al. 2013). OCT1 was recently identified as a negative regulator of *TWIST1* expression, but only in the presence of PER2 in a hypoxic state. Hwang-Verslues et al. (2013) showed evidence that a PER2 transcription repressor complex consisting of EZH2, SUZ12 and HDAC2 bound to OCT1 at five described POU2F1 binding sites in the *TWIST1* promoter. It is unknown whether Per2 binds to Oct1 at the mouse *Hdac9* intron 18 enhancer locus to facilitate repression of *Twist1* or if there is looping involved between *Hdac9* intron 18 and the Oct1 binding sites in the proximal *Twist1* promoter. Our study suggests that distal enhancers play a part in regulation of *Twist1* in cell lines derived from keratinocytes and cSCCs. Future studies will evaluate the role of Per2 and hypoxia in this process.

Study Limitations

There are limitations to this study. First, we only assessed genes nearby to the *Hdac9* enhancer for expression differences. As we only measured expression of two of over 65 genes within the peak locus for Skts5 (Mahler et al., 2008), it is possible that this enhancer may modify additional genes at Skts5 or elsewhere in the genome which could be critical in understanding its impact on skin tumor susceptibility. Chromosome conformation capture (3C) or circularized chromosome conformation capture (4C) studies could be used in the

future to determine if this enhancer affects additional genes or functions in other tissue types which would be important for understanding functional implications of variants at this locus. We do not know if the *Hdac9* enhancer behaves similarly *in vivo* as it does *in vitro*. The *Twist1* expression studies were done using a plasmid containing intron 18 DNA outside of its normal genomic context which may act differently than when it is in its typical genomic context. To do studies using this entire region *in vitro* would be difficult as we would need to include at least 270 kb of sequence to include the *Hdac9* intron 18 through the promoter region of *Twist1*. *In vitro* or *in vivo* genomic editing of the specific sites by CRISPR-cas9 or other techniques could confirm these findings. The variants we identified in *Hdac9* intron 18 that affect Gata3 and Oct1 binding are important candidate variants for skin tumor susceptibility associated with this locus. However, this study does not rule out the possibility that other genes or regulatory regions in *Skts5* may impact skin cancer susceptibility alone or in combination with these variants.

In conclusion, we identified sequence variants in a possible enhancer of the *Twist1* gene between SPRET/Out-R and NIH/Ola-S that impact luciferase expression, *Twist1* expression, and Oct1 and Gata3 binding *in vitro*. As *Hdac9* and *Twist1* map to *Skts5*, a locus for susceptibility to chemically-induced skin cancer, these are strong candidates for *Skts5*. Future studies are essential to determine if this enhancer affects expression of other genes and if differential *Twist1* expression, due to these variants, leads to increased susceptibility to chemically-induced skin tumors in mice. This work provides an example of variants in a regulatory region that may influence cancer susceptibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the Parvin laboratory for use of their Fisher Scientific Sonic Dismembrator Model 500. The OSU Comprehensive Cancer Center Nucleic Acids Shared Resource provided support for Sanger Sequencing and qPCR. This work was supported in part by the National Institutes of Health (1R01 CA134461-01), the American Cancer Society (RSG-07-083-1-MGO), and the OSU Comprehensive Cancer Center. TES was supported by a Pelotonia Undergraduate Research Fellowship, and MMG was supported by a Pelotonia Graduate Research Fellowship.

References

- Akiyama, Y. TFSEARCH: Searching Transcription Factor Binding Sites. 1995. <http://www.rwcp.or.jp/papia/> (9 June 2014, date last accessed)
- Angel JM, Caballero M, DiGiovanni J. Identification of novel genetic loci contributing to 12-O-Tetradecanoylphorbol-13-acetate skin tumor promotion susceptibility in DBA/2 and C57BL/6 mice. *Cancer Res.* 2003; 63:2747–2751. [PubMed: 12782578]
- Ansieau S, Bastid J, Doreau A, Morel AP, Bouchet BP, Thomas C, et al. Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. *Cancer Cell.* 2008; 14:79–89. [PubMed: 18598946]
- Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* 2012; 22:1059–1068. [PubMed: 22442009]
- Brockschmidt FF, Heilmann S, Ellis JA, Eigelshoven S, Hanneken S, Herold C, et al. Susceptibility variants on chromosome 7p21.1 suggest HDAC9 as a new candidate gene for male-pattern baldness. *Br J Dermatol.* 2011; 165:1293–1302. [PubMed: 22032556]

- Ewart-Toland A, Briassouli P, de Koning JP, Mao JH, Yan J, Chan F, et al. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat Genet.* 2003; 34:403–412. [PubMed: 12881723]
- Fleming JL, Dworkin AM, Allain DC, Fernandez S, Wei L, Peters SB, et al. Allele-specific imbalance mapping identifies HDAC9 as a candidate gene for cutaneous squamous cell carcinoma. *Int J Cancer.* 2014; 134:244–248. [PubMed: 23784969]
- Fujiwara K, Igarashi J, Irahara N, Kumura M, Nagase H. New chemically induced skin tumour susceptibility loci identified in a mouse backcross between FVB and dominant resistant PWK. *BMC Genet.* 2007; 8:30. [PubMed: 17559648]
- Fujiwara K, Wie B, Elliot R, Nagase H. New outbred colony derived from *Mus musculus castaneus* to identify skin tumor susceptibility loci. *Mol Carcinog.* 2010; 49:653–661. [PubMed: 20564342]
- Ghosh D. Object-oriented transcription factors database (ooTFD). *Nucleic Acids Res.* 2000; 28:308–310. [PubMed: 10592257]
- Graham MW, Holton TA. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* 1990; 19:1156. [PubMed: 2020554]
- Heinemeyer T, Chen X, Karas H, Kel AE, Kel OV, Liebich I, et al. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.* 1999; 27:318–322. [PubMed: 9847216]
- Hwang-Versluis WW, Chang PH, Jeng YM, Kuo WH, Chiang PH, Chang YC, et al. Loss of corepressor PER2 under hypoxia up-regulated OCT1-mediated EMT gene expression and enhances tumor malignancy. *Proc Natl Acad Sci USA.* 2013; 110:12331–12336. [PubMed: 23836662]
- Jang SI, Karaman-Jurukovska N, Morasso MI, Steinert PM, Markova NG. Complex interactions between epidermal POU domain and activator protein 1 transcription factors regulate the expression of the profilaggrin gene in normal human epidermal keratinocytes. *J Biol Chem.* 2000; 275:15295–15304. [PubMed: 10809764]
- Kaufman CK, Zhou P, Pasolli HA, Rendl M, Bolotin D, Lim KC, et al. GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev.* 2003; 17:2108–2122. [PubMed: 12923059]
- Kotian S, Liyanarachchi S, Zelent A, Parvin JD. Histone deacetylases 9 and 10 are required for homologous recombination. *J Biol Chem.* 2011; 286:7722–7726. [PubMed: 21247901]
- Kurek D, Garinis GA, van Doorninck JH, van der Wees J, Grosveld FG. Transcriptome and phenotypic analysis reveal Gata3-dependent signaling pathways in murine hair follicles. *Development.* 2007; 134:261–272. [PubMed: 17151017]
- Laird PW, Zijderfeld A, Linders K, Rudnicki MA, Jaenisch R, Burns A. Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* 1991; 19:4293. [PubMed: 1870982]
- Maestro R, Dei Tos AP, Hamamori Y, Krasnokutsky S, Sartorelli V, Kedes L, et al. Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* 1999; 13:2207–2217. [PubMed: 10485844]
- Mahler KL, Fleming JL, Dworkin AM, Gladman N, Cho HY, Mao JH, et al. Sequence divergence of *Mus spretus* and *Mus musculus* across a skin cancer susceptibility locus. *BMC Genomics.* 2008; 9:1–12. [PubMed: 18171476]
- Majima Y, Hirakawa S, Kito Y, Suzuki H, Koide M, Fukamizu H, et al. Twist1 as a possible biomarker for metastatic basal cell carcinoma. *Acta Derm Venereol.* 2012; 92:621–622. [PubMed: 22965800]
- Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell.* 2008; 133:704–715. [PubMed: 18485877]
- Masse I, Barbolat-Boutrand L, Kharbili ME, Berthier-Vergnes O, Aubert D, Lamartine J. GATA3 inhibits proliferation and induces expression of both early and late differentiation markers in keratinocytes of the human epidermis. *Arch Dermatol Res.* 2014; 306:201–208. [PubMed: 24346062]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, et al. Systemic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–1195. [PubMed: 22955828]

- Mironchik Y, Winnard PT Jr, Vesuna F, Kato Y, Wildes F, Pathak AP, et al. Twist overexpression induces in vivo angiogenesis and correlated with chromosomal instability in breast cancer. *Cancer Res.* 2005; 65:10801–10809. [PubMed: 16322226]
- Nagase H, Bryson S, Cordell H, Kemp CJ, Fee F, Balmain A. Distinct genetic loci control development of benign and malignant skin tumors in mice. *Nat Genet.* 1995; 10:424–429. [PubMed: 7670492]
- Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* 2010; 6:e1000895. [PubMed: 20369022]
- Okumura K, Sato M, Saito M, Miura I, Wakana S, Mao JH, et al. Independent genetic control of early and late stages of chemically induced skin tumors in a cross of a Japanese wild-derived inbred mouse strain, MSM/Ms. *Carcinogenesis.* 2012; 33:2260–2268. [PubMed: 22843548]
- Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci USA.* 2013; 110:17921–17924. [PubMed: 24127591]
- Parra M. Class IIa HDACs- new insights into their functions in physiology and pathology. *FEBS J.* 2015; 282:1736–1744. [PubMed: 25244360]
- Qin Q, Xu Y, He T, Qin C, Xu J. Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell Res.* 2012; 22:90–106. [PubMed: 21876555]
- Riker A, Enkemann SA, Fodstad O, Liu S, Ren S, Morris C, et al. The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC Med Genomics.* 2008; 1:13. [PubMed: 18442402]
- Shakya A, Kang J, Chumley J, Williams MA, Tantin D. Oct1 is a switchable, bipotential stabilizer of repressed and inducible transcriptional states. *J Biol Chem.* 2011; 286:450–459. [PubMed: 21051540]
- Skeels LE, Fleming JL, Mahler KL, Toland AE. The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. *PLoS One.* 2013; 8:e58609. [PubMed: 23472213]
- Srivastava J, Rho O, Youssef RM, DiGiovanni J. Twist1 regulates keratinocyte proliferation and skin tumor promotion. *Mol Carcinog.* 2015 [Epub ahead of print].
- Sturm RA, Thomson JA, Parsons PG. In vivo and in vitro expression of octamer binding proteins in human melanoma metastasis, brain tissue and fibroblasts. *Pigment Cell Res.* 1993; 6:13–22. [PubMed: 8502621]
- Su MY, Steiner LA, Bogardus H, Mishra T, Schulz VP, Hardison RC, et al. Identification of biologically relevant enhancers in human erythroid cells. *J Biol Chem.* 2013; 288:8433–8444. [PubMed: 23341446]
- Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell.* 2012; 22:725–736. [PubMed: 23201165]
- Weiss MB, Abel EV, Mayberry MM, Basile KJ, Berger AC, Aplin AE. TWIST1 is an ERK1/2 effector that promotes invasion and regulates MMP-1 expression in human melanoma cells. *Cancer Res.* 2012; 72:6382–6392. [PubMed: 23222305]
- Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. DBD- taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* 2008; 36:D88–92. [PubMed: 18073188]
- Zoumpourlis V, Solakidi S, Papathoma A, Papaevangelou D. Alterations in signal transduction pathways implicated in tumour progression during multistage mouse skin carcinogenesis. *Carcinogenesis.* 2003; 24:1159–1165. [PubMed: 12807763]

Figure 1. Map of *Hdac9* Intron 18.

A. A map with approximate locations of *Hdac9* exons (black rectangles) near intron 18 are shown. Genomic coordinates are based on the UCSC GRCm38/mm10 Build. A magnified view of intron 18 illustrating the nine inserts used for sequencing and cloning is drawn approximately to scale. The 45 sequence variants identified between NIH/Ola-S and SPRET/Out-R are indicated by vertical lines. The three variants predicted to disrupt Oct1 or Gata3 sites are indicated by an asterisk *. B. Sequence alignments for the first 520bp of intron 18 in five mammalian species and chicken were generated using Multiple Sequence Alignment (CLUSTAL Omega). Variants between NIH/Ola-S and SPRET/Out-R are bolded and highlighted in gray. Oct1 and Gata3 sites are underlined and indicated. Variants conserved for all 5 species are marked with a * for the first 416 bp and for the 4 mammalian species for the last 104 bp.. MM, *Mus musculus*; HS, *Homo sapiens*; PT, *Pan troglodytes*; RN, *Rattus norvegicus*; CLF, *Canis lupus familiaris*; GG, *Gallus gallus*.

Figure 2. Effects of *Hdac9* intron 18 inserts on luciferase expression

The fold luciferase expression of *Hdac9* intron 18 Inserts 1, 2, 3, 4, 3+4, 5, 6 and 7 relative to empty vector transfected cell lines for A5 (A) and C5N (B) from representative experiments are shown. Transfections include mock transfected cells (pale gray), pGL3 promoter empty vector alone (medium gray), vector plus SPRET/Out-R DNA insert (black) and vector plus NIH/Ola-S insert DNA (dark gray). Errors bars represent standard deviation. Experiments for Inserts 1 and 2 were done a minimum of three times and each sample was run in triplicate. A representative experiment is shown. A one-way Anova test was performed to determine if any groups were different. P-values were calculated using a Tukey's HSD test. *, p-value 0.05; ** p-value 0.01; *** p-value 0.001.

Figure 3. Effects of *Hdac9* inserts on *Hdac9* and *Twist1* expression

Hdac9 (A) and *Twist1* (B) expression was measured by qPCR in A5 cells and was normalized to control gene *Hprt*. Expression is plotted as fold expression relative to mock transfected A5 cells. Expression of mock (pale gray), pGL3 promoter vector (medium gray), SPRET/Out-R insert DNA (black) and NIH/Ola-S insert DNA (dark gray) and transfected A5 cells are shown for *Hdac9* intron 18 Inserts 1 and 7 for *Hdac9* and 1, 2 and 7 for *Twist1*. Errors bars represent standard deviation. Experiments were done multiple times and each sample was run in triplicate. A representative experiment is shown. M, mock; EV, empty vector; S, SPRET/Out-R and N, NIH/Ola-S insert DNA. One way ANOVA analyses were conducted to determine significance across groups and a Tukey's HSD test was performed for groups showing significant differences. *, p-value = 0.05; ** p-value = 0.01; *** p-value = 0.001.

Figure 4. ChIP of A5 cells for *Hdac9* intron 18 Insert 1

Insert 1 of intron 18 was divided into 2 regions for quantitative ChIP analysis. Experiments were done three times and each sample was run in triplicate. A representative image of the PCR products for a semi-quantitative ChIP experiment is shown to the left of a quantitative PCR graph. The fold difference in DNA abundance relative to empty-vector of transfected cells and standard deviations across three PCR reactions as measured by qPCR are indicated for a representative experiment. Abundance was normalized to input DNA. Oct1 (A) and Gata3 (B) ChIP for A5 cells transfected with Insert 1, Region 1. ChIP Oct1 (C) and Gata3 (D) ChIP for A5 cells transfected with Insert 1, Region 2. L, 100 bp ladder; EV, empty vector; N, NIH/Ola-S insert DNA; S, SPRET/Out-R insert DNA. Light gray, Empty vector; Black, SPRET/Out- R; Dark gray, NIH/Ola-S. Significance of differences between groups was determined by one-way ANOVA. p-values were calculated using Tukey's HSD tests. *, p-value 0.05; **, p-value 0.01; ***, p-value 0.001.

Figure 5. Effect of NIH/Ola-R and SPRET/Out-S insert1 sequence variants at Oct1 and Gata3 sites on luciferase and *Twist1* mRNA expression

The fold luciferase expression (A) and *Twist1* mRNA expression (B) of *Hdac9* intron 18 Insert 1 relative to mock transfected cell lines for A5 cells are shown. Expression is plotted relative to mock transfected cells. Transfections include mock transfected cells (M, pale gray), pGL3 promoter empty vector alone (EV, medium gray), vector plus SPRET/Out-R Insert 1 DNA (S, black), vector plus SPRET/Out-R Insert 1 altered DNA (Sx, black), vector plus NIH/Ola-S Insert 1 DNA (N, dark gray) and vector plus altered NIH/Ola-S Insert 1 DNA (Nx, dark gray). Errors bars represent standard deviation. Experiments were done three times and each sample was run in triplicate. A representative experiment with p-values is shown. A one way ANOVA was used to determine significance between groups. P-values were calculated using a Tukey's HSD test. *, p-value 0.05; ** p-value 0.01; *** p-value 0.001.

Table 1

Predicted Transcription Factor Binding Differences between NIH/Ola-S and SPRET/Out-R

Transcription Factor	DNA Binding Site ¹	Polymorphism (NIH/SPRET)	Binding Strain	Hdac9 Insert	Significance Score ²
Oct1	TAA <u>ACC</u>	T/C	Spretus	1	88.0
Oct1	ATATA <u>CA</u> CT	G/C	Spretus	1	89.7
Gata3	AATCA <u>CG</u>	C/T	NIH	1	85.9
CdxA	ATATA <u>G</u>	G/C	NIH	1	85.0
Gfi-1	ATAGTT <u>G</u> TGA T	A/G	Spretus	1	85.9
AP-1	GAGTGATTA <u>A</u>	A/G	NIH	1	86.4
Mzf1	<u>T</u> CCCC <u>T</u> TTT	GC/T-	Spretus	2	88.7
Pbx-1	TAC <u>AA</u> TCA	C/T, A/T	NIH	2	86.3
Brn2	<u>T</u> TCAAAATG	A/T	Spretus	3	86.1
Srebp	<u>G</u> TCA <u>CC</u> CC <u>A</u> T	G/A, A/G	NIH	4	85.9
Aml-1a	AC <u>CC</u> CA <u>T</u> C	A/G	NIH	4	87.4
Rora	ACC <u>T</u> -TAGAA	-/A	NIH	5	87.5

¹ Underlined nucleotides are those within the predicted binding site that vary between strains,², the significance score was determined by TF Search.