



Published in final edited form as:

Mol Cancer Res. 2014 July ; 12(7): 1029–1041. doi:10.1158/1541-7786.MCR-13-0286-T.

Transcriptomes and shRNA Suppressors in a TP53 Allele-specific Model of Early-onset Colon Cancer in African Americans

Charles C. Weige^{1,*}, Marc R. Birtwistle^{2,*}, Himel Mallick⁸, Nengjun Yi⁸, Zuzana Berrong³, Emily Cloessner³, Keely Duff³, Josephine Tidwell³, Megan Clendenning³, Brent Wilkerson⁴, Christopher Farrell⁵, Fred Bunz⁶, Hao Ji⁷, Michael Shtutman⁷, Kim E. Creek⁷, Carolyn E. Banister⁷, and Phillip J. Buckhaults^{1,#}

¹Department of Medicine, The University of Alabama at Birmingham, Birmingham, AL 35294

²Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY 10029

³Department of Biochemistry and Molecular Biology, Georgia Regents University, Augusta, GA 30912

⁴Department of Otolaryngology-Head and Neck Surgery, University of California, Davis, Sacramento, CA 95817

⁵Department of Pharmaceutical and Administrative Sciences School of Pharmacy, Presbyterian College, Clinton, SC 29325

⁶Department of Radiation Oncology and Molecular Radiation Sciences, The Johns Hopkins University School of Medicine, Baltimore, MD 21287

⁷Department of Drug Discovery and Biomedical Sciences, South Carolina College of Pharmacy, University of South Carolina, Columbia, SC 29208

⁸Department of Biostatistics, The University of Alabama at Birmingham, Birmingham, AL 35294

Abstract

African Americans are disproportionately affected by early-onset, high-grade malignancies. A fraction of this cancer health disparity can be explained by genetic differences between individuals of African or European descent. Here the wild-type Pro/Pro genotype at the *TP53*Pro72Arg (P72R) polymorphism (SNP: rs1042522) is more frequent in African Americans with cancer than in African Americans without cancer (51% vs 37%), and is associated with a significant increase in the rates of cancer diagnosis in African Americans. To test the hypothesis that p53 allele-specific gene expression may contribute to African American cancer disparities, p53 hemizygous knockout variants were generated and characterized in the RKO colon carcinoma cell line, which is wild-type for p53 and heterozygous at the *TP53*Pro72Arg locus. Transcriptome profiling, using RNAseq, in response to the DNA-damaging agent etoposide revealed a large number of p53-

#To whom correspondence should be addressed: phillip.buckhaults@gmail.com.

*These authors contributed equally to the work

The authors declare no conflicts of interest.

regulated transcripts, but also a subset of transcripts that were *TP53*Pro72Arg allele specific. In addition, a shRNA-library suppressor screen for p53 allele-specific escape from p53-induced arrest was performed. Several novel RNAi suppressors of p53 were identified, one of which, PRDM1 β (BLIMP-1), was confirmed to be an Arg-specific transcript. PRDM1 β silences target genes by recruiting H3K9 trimethyl (H3K9me3) repressive chromatin marks, and is necessary for stem cell differentiation. These results reveal a novel model for African American cancer disparity, in which the TP53 codon 72 allele influences lifetime cancer risk by driving damaged cells to differentiation through an epigenetic mechanism involving gene silencing.

Implications—TP53 P72R polymorphism significantly contributes to increased African American cancer disparity.

Keywords

TP53 polymorphism; colorectal cancer

INTRODUCTION

Colon cancer is the second leading cancer cause of death in the US (1). African Americans have a disproportionately high incidence and mortality rate from many epithelial cancers, and colon cancer specifically, when compared to European Americans (1). This increased risk is associated with an earlier age of onset (5 years on average) as well as more poorly differentiated tumor phenotypes (2). This disparity in relative risk is likely due in part to genetic factors, as many environmental factors are shared between these two groups within the US.

Differences in variant allele frequency within the TP53 gene may account for some of the disproportionate cancer risk associated with African ancestry. The TP53 gene is a tumor suppressor and is found to be mutated or lost in more than half of human malignancies (3). This gene codes for the p53 protein which is a regulator of many apoptosis, cell cycle arrest and DNA damage response pathways. There is a non-synonymous polymorphism in codon 72 (SNP ID= rs1042522) of the TP53 gene, one allele codes for Proline (Pro) at amino acid 72 and the other for arginine (Arg) (4). Multiple studies, including the 1000 genomes project(5–7), show that the Pro allele is more common in African Americans while the Arg allele is predominant in those of European descent. The Pro allele has demonstrated the ability to enhance the inflammatory response in a mouse knock-in model (8). Individuals exhibiting the Pro/Pro (P/P) genotype have a decreased mean age of onset in colorectal cancer (9). The P/P genotype is enriched in individuals over the age of 85, despite the increased cancer-related mortality (10, 11). In contrast, studies utilizing a temperature sensitive TP53 mutant suggest that the Arg allele has the ability to preferentially induce apoptosis(12, 13), and the Arg allele induces a stronger apoptotic response to chemotherapeutic treatments (14, 15).

Given this allele specific predisposition toward increased longevity (P/P) or apoptosis (Arg), individuals expressing the Arg allele may be protected from DNA damage by more efficient removal of stem cells with somatic mutations. However, epidemiologic results are not clear; in some cases the Arg allele appears protective while in others the Pro allele exhibits

protection (4). Furthermore, the increased frequency of the P/P genotype in African Americans does not account for the increased percentage of high grade, undifferentiated tumors found within this group. We therefore sought to identify a unifying mechanism that could explain the influence of TP53 codon 72 polymorphisms on both cancer and longevity.

The normal differentiation program of stem cells involves the p53 protein(16, 17). Mutations in TP53 are found with greater frequency in poorly differentiated tumors (18), and these tumors exhibit a gene expression signature in common with embryonic and induced pluripotent stem cells (19, 20). This signature includes the targeted silencing of genes by the Polycomb Repressor Complex 2 (PRC2), a key regulator of stem cell differentiation (20). TP53 deletion or knockdown enhances the ability of differentiated cells to revert back to a pluripotent state (21, 22). Although a precise mechanism is yet to be defined, TP53 is clearly linked to stem cell differentiation. TP53 can control the expression of chromatin modifying genes like histone-lysine N-methyltransferase (EZH2), a catalytic subunit of PRC2 (23). This interaction between TP53 and chromatin modifying genes expands the influence of TP53 beyond the list of genes for which TP53 has a direct transcriptional role in their expression.

We hypothesized that TP53 codon 72 polymorphism accentuates the previously described ethnic disparity in cancer risk by affecting stem cell differentiation. If a mutated stem cell were to undergo differentiation, subsequent daughter cells would still be affected by the mutations present, but if the same stem cell is allowed to maintain pluripotency the number of cells it can pass the mutations on to is increased and will have a greater ability to potentiate. This would lead to less differentiated tumors with a higher grade, as seen African Americans.

This hypothesis was tested by investigation of the TP53 codon 72 allele-specific gene expression profiles and screening for allele-specific shRNA suppressors of TP53 induced growth arrest. We used somatic cell knockout techniques (24) to create isogenic derivatives of RKO colon carcinoma cells that are hemizygous for each TP53 allele (Pro/- or Arg/-). RNA sequencing was then performed to characterize TP53-dependent transcripts. A second-site repressor screen for shRNAs that bypassed TP53-mediated growth arrest was also conducted, identifying candidate TP53 allele specific suppressors. One Arg-specific candidate identified in both assays was PR domain containing 1 β (PRDM1 aka BLIMP1). PRDM1 β recruits H3K9 and H3K27 trimethyl repressive chromatin marks and is required for adult intestinal stem cell differentiation(25, 26). We propose that this may be the mechanism by which TP53 enforces stem cell differentiation, protects from cancer, but causes aging. This would explain the disparity in the incidence of early onset colon cancer in TP53 P/P individuals, who are overrepresented in the African American population.

MATERIALS AND METHODS

Cell culture

Wild type TP53 and double knockout (KO) RKO colon cancer cells were obtained from B. Vogelstein (27), and were grown at 37°C/5% CO₂ in McCoy's 5A media (Fisher)

supplemented with 10% FBS and pen/strep (Fisher). Cells were passaged every 2–3 days to maintain subconfluency.

Genotyping of Colon Cancer Patients

The TP53Arg72Pro polymorphism resides in exon 4 of the TP53 gene. We used exon-spanning primers to amplify exon 4 by PCR: TP53E4F = 5'-CCTGGTCCTCTGACTGCTCT-3'; TP53E4R = 5'-GCCAGGCATTGAAGTCTCAT-3' (Integrated DNA Technologies, Coralville, IA). Sequences GCCTCCCTCGCGCCATCAG (454A) and GCCTTGCCAGCCCGCTCAG (454B) (Integrated DNA Technologies, Coralville, IA) were appended to the 5' and 3' ends of each PCR primer, and used as universal priming sites for subsequent sequencing of the PCR products. PCR products were prepared using KAPA SYBR® FAST Universal 2X qPCR Reagents and the following PCR protocol. A 20 µl reaction used 10 µl 2X Master Mix, 1 µl 5 µM primer F, 1 µl 5 µM primer R, 7 µl H₂O, and 1 µl (3 ng) of genomic DNA template (Kappa Biosystems, Woburn MA). Thermal cycling conditions were as follows: 98°C, 2 minutes, 98°C, 10 seconds, 63°C, 10 seconds, 70°C, 30 seconds, 3 repeats; 98°C, 10 seconds, 60°C, 10 seconds, 70°C, 30 seconds, 3 repeats; 98°C, 10 seconds, 57°C, 10 seconds, 70°C, 30 seconds, 3 repeats; 98°C, 10 seconds, 56°C, 10 seconds, 70°C, 30 seconds, and finally 49 repeats for a total of 58 cycles. A melt curve analysis was performed to verify the PCR products, starting at 55°C for 30 seconds, and then heating slowly at 0.5°C per minute until 95°C. Bi-directional amplicon sequencing was performed by Beckman Coulter Genomics (Danvers, MA). Sequences were aligned to the appropriate reference sequence using CLC-Genomics Workbench (Cambridge, MA) and all sequence trace files were inspected manually to determine genotype. Genotypes were determined from tumor and/or normal genomic DNA.

RKO P72R polymorphism clone construction

RKO human colon carcinoma cells are heterozygous Pro/Arg at p53 codon 72 (28). A TP53 knockout (–/–) line (RKO 249—KO) was constructed and previously described. Pro/– (8B7—Pro1 and 10C5—Pro2) and Arg/– (9D4—Arg1 and 10H2—Arg2) lines were created by knocking out either the Pro or the Arg allele or both using an existing adeno-associated TP53 targeting vector (27, 29). All lines are available upon request. Genotyping was performed using one primer in the neomycin resistance gene and one in the TP53 gene: NeoF = 5'-CTG GGA GAA GGT GCC ATG AT-3' and TP53R = 5'-TGT GCT GTG ACT GCT TGT AG-3' in a 20µL reaction. Thermal cycling conditions were as follows: 98°C, 2 minutes; 98°C, 20 seconds, 67.2°C, 30 seconds, 70°C, 120 seconds, 65 repeats; 55°C, 120 seconds; melt curve starting at 55°C increasing 0.5°C every 10 seconds for 81 repeats. The TP53 status of all cell lines was routinely verified by restriction digest using the BstU1 method (30). Moreover, western blotting did not detect any truncated forms of TP53 expressed from the deleted alleles (whereas it clearly showed the full length protein—Fig. 2B).

RNA-seq

RNA (4 µg) from all 6 RKO clones treated with either DMSO or 10 µM etoposide was prepared for sequencing using the TruSeq RNA Sample Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol. Briefly, polyA mRNA was purified and

fragmented. Then first and second strand cDNA synthesis was performed, ends were repaired and adenylated, adapters were ligated, and fragments were enriched with PCR amplification. The library was validated with an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

RNA-seq preparations from all 6 RKO clones treated with either DMSO or 10 μ M etoposide were paired ends, and read using Illumina HiSEQ2K platform using PhiX as a reference control (Illumina, San Diego, USA). The resultant FASTQ files were imported into CLC Genomics Workbench and aligned to the human hg19/B37 reference sequence. Total exon reads or transcript reads per kilobasepair per million (RPKM) were used as gene expression values.

We utilized the R/Bioconductor package DESeq, which provides methods to test for differential expression for sequence count data by the use of negative binomial distribution. We tested for differential expression for each gene w.r.t sample type after accounting for treatment effect and used the Benjamini and Hochberg (BH) method to correct the p-value for false discovery rate (FDR).

shRNA Screen

The second-site suppressor screen was done on all six RKO cell line variants. The experiment took 12 total days of tissue culture. On Day 1, RKO clones were seeded at 1.5×10^6 cells in one T-150 flask per cell line, and incubated at 37°C for 24 hrs. On Day 2, each cell line was infected with a lentiviral shRNA library at MOI~0.8 as described (31) with the addition of 8.0ug/mL Polybrene for 24 hrs. The library consists of 27,500 shRNAs targeting 5,043 genes (Decipher Project – Human Module 1, Signaling Pathway Targets). On Day 3, the lentiviral-containing media was changed with fresh growth media, and on Day 5, cell were treated with 5 μ M nutlin-3a, and left in this media for seven days. This achieved enrichment of cells infected with specific shRNAs enabling proliferation rather than senescence. After seven days of nutlin-3a selection, the surviving cells were collected and genomic DNA of each cell line was purified (Agentcourt DNAdvance; Beckman Coulter).

For high-throughput sequencing, shRNA integrated DNA was real time PCR amplified for each RKO cell line using KAPA HiFi HotStart (Kapa Biosystems) and then PCR purified (Agentcourt AMPure XP). Primers, Gex_F and Gex_R are part of the lentiviral plasmid construct (Decipher Project - pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro vector). Primers were designed with A and B linkers to compliment 454 A and B sequencing primers. The following Gex-tagged A and B primers were used:

TagA_Gex_F

GCCTCCCTCGCGGATCAGCAAGCAGAAGACGGCATACGAGA

TagB_Gex_R

GCCTTGCCAGCCCGCTCAGAATGATACGGCGACCACCGAGA

Data were collected with a Roche 454 sequencer at the *engencore* of the University of South Carolina, and analyzed using CLC Genomics Workbench.

Western Blotting

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed at 200 V in a 10% polyacrylamide gel containing SDS (Criterion TGX, BioRad). The proteins in the gel were then transferred to a polyvinylidene difluoride (PVDF) membrane (Immun-Blot PVDF Membrane, 0.2µm, BioRad) using a Criterion Blotter (BioRad). After blocking with 5% Bovine Serum Albumin (BSA—Fisher) in Tris-buffered saline with Tween-20 (TBST—Fisher), proteins were incubated with specific primary antibodies (monoclonal anti-p53 DO-1; Santa Cruz, dilution 1:1000 and monoclonal anti-PARP 46D11; Cell Signalling, dilution 1:2000) overnight at 4°C with gentle rocking. Membranes were washed 3 times in TBST for 7 minutes, incubated in the appropriate secondary antibody (anti-mouse IgG HRP-labeled; Vector Laboratories, dilution 1:10,000 or anti-rabbit IgG HRP-labeled; Santa Cruz Biotechnology, dilution 1:10,000) and then washed 3 times in TBST for 7 minutes again, all with gentle rocking. Bands were visualized with Pierce ECL Western Blotting Substrate (Thermo Scientific) according to the manufacturer's instructions. β-Actin was used as a loading control. To probe for this, membranes were first incubated in stripping buffer (Glycine/Tween/pH 2.2) for 15 min at 20°C, blocked with 5% BSA in TBST, and then probed with anti-β-actin (goat polyclonal, C11 Santa Cruz, dilution 1:1000). Following washing as above, the membrane was incubated in the secondary antibody, anti-goat IgG HRP-labeled (Santa Cruz Biotechnology, dilution 1:10,000), then washed and visualized as above.

qPCR

Total RNA from RKO clones treated with 0, 10, 20 or 50 µM etoposide or 5, 10 or 20 µM (–)-nutlin3a was isolated using Trizol according to manufacturer's protocol. RNA quality was assayed with an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Random decamers (Invitrogen, Carlsbad, CA) and 125 ng of RNA was used to make 20 µL of cDNA using SuperScriptII (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. The cDNA was diluted 1:100 prior to qPCR. PCR products were prepared using KAPA SYBR® FAST Universal 2X qPCR Reagents and the following PCR protocol. A 20 µL reaction used 10 µL 2X Master Mix, 1 µL 5 µM primer F, 1 µL 5 µM primer R, and 2 µL of etoposide treated diluted cDNA or ChIP preparations or 5 µL of nutlin3a treated diluted cDNA template (Kappa Biosystems, Woburn MA). Thermal cycling conditions were as above for TP53 genotyping. Thermal cycling conditions for ChIP amplification were as follows: 98°C, 2 minutes, 94°C, 10 seconds, 63°C, 15 seconds, 3 repeats; 94°C, 10 seconds, 60°C, 15 seconds, 3 repeats; 94°C, 10 seconds, 57°C, 15 seconds, 3 repeats; 98°C, 10 seconds, 56°C, 15 seconds, 49 repeats for a total of 58 cycles. A melt curve analysis was performed to verify the PCR products, starting at 95°C for 30 second, and then decreasing slowly at 0.1°C per 15 seconds until 65°C. The following cDNA primers were used:

cDNA_PRDM1_F CTTCCCTCCTTTGCATTGAA,

cDNA_PRDM1_R GAACCTTGCCTTTTTGTGGA;

GDF15_F2 AAGATTCTGAACACCGACC,
GDF15_R2 GAGAGATACGCAGGTGCAG.

ChIP primers were as follows:

PRDM1_ChIP_bs4_F CTTCCCTCCTTTGCATTGAA,
PRDM1_ChIP_bs4_R GAACCTTGCCTTTTTGTGGA.

Chromatin immunoprecipitation

RKO clones grown in 10 cm plates were treated with 10 μ M (–)-nutlin3a or DMSO for 24 hours. Chromatin preparations were made with ChIP-It Express Enzymatic kit (Active Motif, Carlsbad, CA) according to the manufacturer's protocol. Enzymatic shearing was performed for 30 minutes to produce short fragments for sequencing. Immunoprecipitation was performed with DO1 monoclonal antibody (Santa Cruz Biotechnology, Santa Cruz, CA) or IgG (Santa Cruz Biotechnology, Santa Cruz, CA) as a control overnight at 4°.

shRNA Inhibition of PRDM1 and TP53

A panel of shRNAs for PRDM1 (Invitrogen V2LHS_94736,94739,94740,94741 and V3LHS_407846), TP53 (Invitrogen V3LHS_333919 and 333920) and non-silencing (NS, Invitrogen) were transfected into RKO clone cells. Day 1–800,000 cells were plated on 60mm culture dishes. Day 2–Cells were transfected using Lipofectamine Reagent (Invitrogen). Day 3–Cells were plated in T-75 flasks and treated with DMSO, Puromycin 5 μ g/ml, or Nutlin-3a 10 μ M. After one week cells were trypsinized and pelleted. Pelleted cells were resuspended in PBS + 1% BSA and counted on a FACScan XP5 (Becton Dickinson).

Analysis of TCGA data

TCGA COAD bamfiles assembled to hg19 (<https://cghub.ucsc.edu/>) were imported into CLC genomics workbench (<http://www.clcbio.com/>) and variants were called using a probabilistic variant detector with 99.9% likelihoods threshold used as a cutoff. PRDM1 expression data (RNA-seq, total exon reads, normalized, log(10) transformed) was obtained from the UCSC cancer genome browser (<https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/> (32)). COAD clinical data was obtained from TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm>). TP53 genotype counts for European and African American individuals without cancer were obtained from the 1000 genomes project (<http://www.1000genomes.org/>) Statistical associations between genotype frequency, PRDM1 expression, ethnicity, gender, cancer, and age of diagnosis were determined using Partek Genomics Suite (<http://www.partek.com/>) and Graphpad Prism (<http://www.graphpad.com/scientific-software/prism/>)

RESULTS

Role of TP53Pro72Arg in Colon Cancer

We first investigated TP53Pro72Arg genotype frequencies in colon cancer patients and normal controls of both European and African ancestry. We determined the TP53Pro72Arg genotype of 93 colon cancer patients with known ethnic ancestry and age at diagnosis and

combined these data with that obtained from 97 colon cancer patients from the Cancer Genome Atlas Colorectal Adenocarcinoma dataset (33) (Supplemental Table 1). There is a positive association between the P/P genotype and a diagnosis of colon adenocarcinoma, but only in African Americans (P/P 51% to not P/P 37%, Fisher's exact $p=0.0505$). We next hypothesized that the P/P genotype may be associated with a lower mean age of diagnosis for colon cancer patients. The P/P genotype was significantly associated with a decrease in the mean age of diagnosis by 6.84 years ($p=0.0107$) (Fig. 1A), whereas being African American only reduced the mean age of diagnosis by 6.26 years ($p=0.0076$) (Fig. 1B). When both genotype and ethnicity are analyzed together, African Americans with the P/P genotype have a decrease in mean age of diagnosis of 9.12 years ($p=0.0337$) (Fig. 1C). Non-African Americans with the P/P genotype had a reduced mean age of diagnosis of 3.03 years, but this was not statistically significant ($p=0.3915$) (Fig. 1C). We considered two possible mechanisms by which the P/P genotype could accelerate colon cancer progression: either by being linked to a fewer number of sequential somatic mutations required, or by increasing the rate at which these mutations occur. To determine which mechanism is most consistent with the observed data, we fit the incidence data to a progression model (34)(Fig. 1D—solid lines) given by

$$\log S = Ct^{n_M} \quad (1)$$

where S is the fraction of patients remaining undiagnosed, C is the rate coefficient, t is time, and n_M is the number of mutations minus one. We found that for both P/P and non-P/P individuals $n_M=6$, which is consistent with other predictions for the number of genetic events required for the development of colon cancers (29). This analysis revealed that the difference between P/P and non-P/P individuals was in the rate at which the mutations occurred; African American P/P individuals accumulated cancer-causing mutations more than twice as fast ($C_{AA-P/P}/C_{AA-NotP/P} = 2.35$). Thus, this analysis suggests that individuals with a P/P genotype are at elevated risk for early-onset colon cancer due to having an accelerated rate at which colon cancer-causing mutations accumulate (number of mutations per **person** per year). We next identified all somatic mutations in TCGA COADREAD tumors using the UCSC Cancer Genome Browser (TCGA_COADREAD_mutation dataset), and compared the numbers of somatic mutations present within tumors derived from AA, EA, P/P and Not-P/P individuals. There was no significant difference in the numbers of somatic mutations observed or in the rate at which they accumulated (number of mutations per **tumor** per year). This contrasts with the significant increase in the number of mutations per person per year observed in African Americans with P/P genotypes, and indirectly suggests a mechanism of increased numbers of at risk cancer progenitor (stem) cells within African Americans of the P/P genotype. Taken together, these results suggest that the TP53 Pro72Arg P/P genotype is a significant risk factor for colon cancer development in African Americans.

TP53 Allele-Specific Gene Expression Profiling

Because p53 acts primarily as a transcription factor, we reasoned that the tendency of P/P individuals to exhibit accelerated colon cancer progression may be related to TP53Pro72Arg allele-specific gene expression patterns. To investigate Pro and Arg allele-specific

expression, we made use of a colon carcinoma cell line, RKO, which is heterozygous for the TP53 codon 72 polymorphism (denoted WT), and somatic cell knockout techniques (see Fig. 2 and Methods (24)) to derive four independent hemizygous knockout derivative cell lines, two having only a single Pro allele, and two having only a single Arg allele. We denote these single-allele cell lines as Pro1, Pro2, Arg1, and Arg2. We compared these single-allele derivatives to the starting WT cell line and to a knockout derivative in which both alleles had been disrupted using the same technology (27, 35). TP53-dependent gene expression was observed by treatment of the panel with the known TP53-inducers etoposide and (-)nutlin-3a (the (-) denotes the active enantiomer), as compared to treatment with the DMSO vehicle control (etoposide) and the inactive nutlin-3a enantiomer ((+)nutlin-3a) control (Fig. 2B). Relative to the controls, wild-type RKO (Pro/Arg) and all four hemizygous knockout derivatives demonstrated equivalently robust increases in TP53 levels 24 hrs post drug treatment. The Arg variant of TP53 clearly showed a higher electrophoretic mobility, consistent with previous reports (36), while the heterozygous, WT cell line showed a wider band spanning the effective molecular weights of both the Pro and Arg variants.

To probe allele-specific function using these six cell lines, we performed two sets of global experiments. First, we treated the cells with 10 μ M etoposide for 24 hours (to induce DNA damage and thus a TP53 response) or a control (DMSO), and analyzed the gene expression responses via RNA sequencing (RNA-seq). Second, we performed a second-site suppressor screen, where all six cell lines were infected with a lentiviral library consisting of ~27,000 shRNAs targeting 5,043 mRNAs (Decipher project—Signaling Pathway Targets Set). The cells were then treated with the active enantiomer of nutlin-3a for seven days, which inhibits the MDM2-p53 interaction and thereby stabilizes p53, inducing cell cycle arrest and a senescence-like state. By performing colony formation assays on uninfected cells, we noted that at this concentration of nutlin, all TP53 containing RKO derivatives underwent permanent arrest, whereas the double knockout continued to proliferate unchecked. Infected cells that incorporated into their genomes shRNAs that enable proliferation in the presence of TP53 would outgrow those cells that remained arrested. We reasoned that allele-specific transcripts overlapping with allele-specific shRNA suppressors would be highly relevant as functionally important TP53 allele-specific target genes. Below we present analyses of these two global experiments.

RNA-seq Analysis

We performed hierarchical cluster analysis of moderate depth (20 million reads per sample) gene expression data obtained from cells treated with etoposide (E) or vehicle control (C) (Fig 3A, Supplemental Table S2). The TP53-containing cell lines showed robust TP53-response gene activation, while TP53-knockout cells were largely unresponsive to treatment. In the presence of etoposide, RKO clones clustered according to genotype, with Arg-containing clones activating a distinct subset of genes compared to Pro/- derivatives. To identify TP53 and allele-specific transcripts, we had a biostatistician analyze the data using the R/Bioconductor package DESeq, which provides methods to test for differential expression for sequence count data by the use of negative binomial distribution. The method improves upon the existing methods on the issue of over dispersion by providing shrinkage estimator for the distribution's variance (37). We analyzed TP53-dependent, but allele

independent behavior and identified a list of genes including the TP53 effectors *bcl2a1*, *btg2*, and *gdf15* (Figure 3B–D). For the WT and ARG and PRO vs KO comparison, we detected 3657 significant differentially expressed genes after adjusting for FDR (Figure 4A–B, Supplemental Table S3). Example expression profiles are shown for the ARG vs. PRO comparison, we detected 22 genes, which showed significant differential expression after adjusting for FDR (Figure C–D, Supplemental Table S4). To verify these results we performed deep sequencing (>200 million reads per sample) of one Arg^{-/-} and one Pro^{-/-} clone in the presence of etoposide, querying the expression levels of 18,302 genes (Figure 4E–F). We detected 125 genes that were significantly differentially expressed between Arg^{-/-} and Pro^{-/-} genotypes (FDR-corrected p-value <0.05). TP53 responsive genes (allele specific and allele independent) are summarized in Supplemental Table S5.

Second-site Suppressor Screen Identifies PRDM1 as an important Allele-Specific Target

The RNA-seq experiments described above allowed us to identify TP53 allele-specific target transcripts. However, it remained unclear which members of these lists may contribute to TP53 allele-specific phenotypes. Nutlin induces growth arrest of RKO colon cancer cells, in a TP53 dependent manner. We used this phenotype as a proxy for stem cell differentiation, and proceeded to perform a second-site suppressor screen whereby all six cell lines were infected with 27,500 individual shRNAs targeting 5043 mRNAs and subsequently put under selection with nutlin-3a. Nutlin-3a binds to the E3 Ubiquitin ligase MDM2 and inhibits its ability to downregulate p53 levels, thereby increasing the stability of the p53 protein and raising its expression levels (38). Thus, the targets of the shRNAs that enable cell growth in the presence of both TP53 and nutlin-3a may be thought of as mediating the effects of p53-mediated cell cycle arrest. We isolated genomic DNA of each cell line after seven days of nutlin-3a selection, prepared PCR products spanning the integrated lentiviral genomes, and then identified and enumerated the shRNAs via deep amplicon sequencing. Since nutlin-3a normally causes growth arrest, any shRNA that blocks the effects of nutlin-3a to allow cell proliferation would be significantly enriched after seven days of growth. Any overlap between the allele-specific lists generated above and lists generated by this screen would yield interesting candidate genes for further analysis.

We first analyzed TP53 allele-independent function by ranking shRNA gene targets by their counts in TP53-containing cell lines vs. the KO and corresponding p-value (see Methods; Supplemental Table S6 and Fig. 5A). One of the top hits was TP53, as would be expected, providing confidence in the proposed screening method. We next analyzed TP53 allele-specific function by ranking shRNA gene targets by their cumulative counts in the Pro1 and Pro2 cell lines vs. the Arg1 and Arg2 cell lines, and vice versa (see Methods; Supplemental Tables S7–S8 and Fig. 5B). The Polycomb-related proteins EED (39) and PHF1 (40), which enforce transcriptional repression and differentiated states, were on the Arg list. This is consistent with the idea that Arg may have an allele-specific differentiation-related function. Most notably, the only gene, which is both on the allele-specific transcript RNA-seq and on the second site suppressor screen lists, is PRDM1. We therefore wanted to analyze PRDM1 in greater detail as an Arg-specific transcript.

Validation of PRDM1 as an Arg-Specific Transcript

The RNAseq data suggested that the β isoform of PRDM1 is an Arg-Up transcript (Fig. 6A). To validate PRDM1 β as an Arg-Up transcript, we performed several qPCR experiments. First, we treated all cell lines with various doses of etoposide and observed PRDM1 β mRNA levels (Fig. 6B). Second, we treated all cell lines with various doses of Nutlin-3a and observed PRDM1 β mRNA levels (Fig. 6C). These results provide strong evidence that PRDM1 β mRNA expression levels are only etoposide-responsive in cell lines containing the Arg allele. This case is made even stronger by the fact that in these same samples, significant drug-induced responses of three canonical TP53 allele-independent response genes (CDKN1A, GDF15, and MDM2) are seen for both the Pro and Arg cell lines (Fig. 6D–I).

An obvious hypothesis for explaining these data is that the Arg variant of TP53 binds more strongly to the promoter of PRDM1 β than does the Pro variant. This is particularly relevant because of a recently reported p53 binding motif upstream of exon 4 of the PRDM1 β isoform, which contains the start site for PRDM1 β (41). To test this hypothesis, we performed chromatin immunoprecipitation using a p53 antibody in the presence or absence of Nutlin-3a, and then looked for CDKN1A (an allele-independent control) and PRDM1 (Fig. 6J). We found that there was greatly enhanced binding of p53 to the promoters of both genes in all cell lines relative to the KO, but there was not significant allele-specificity. Thus, it seems that differential promoter binding cannot explain the p53 allele-specific behavior of PRDM1 β . This is consistent with previous work that didn't show any differential DNA binding capacity of these p53 variants (36). Finally, the data collected from the PRDM1 shRNA transfection panel (Fig. 6K) show that the PRDM1 shRNAs offer a survival advantage to the WT and 9D4 Arg cell lines (2.32–3.82 fold increase over NS), though the effect is less robust in the 10H2 cells (1.15–1.81), and are ineffective in the KO and Pro cells (no significant change from NS). This further supports the hypothesis that the Arg allele affects expression and function of PRDM1.

To confirm PRDM1 expression is associated with the Arg/Arg (R/R) genotype at the TP53P72R locus, we performed variant calls on TCGA COAD-READ normal bamfiles to identify genotypes, and analyzed the TCGA COAD-READ gene expression data set (TCGA_COADREAD_exp_HiSeqV2_exon) for PRDM1 expression (RNA-seq, total exon reads, normalized, log₁₀ transformed). The mean PRDM1 expression level for 99 R/R individuals (8.84) was significantly higher ($p=0.032$) than that found in 18 P/P individuals (8.55). P-value was determined by three-way ANOVA using genotype, gender and ethnicity as factors.

DISCUSSION

We have investigated the TP53Pro72Arg polymorphism in colon cancer from both an epidemiological and molecular biology perspective. Epidemiologically, we found that the P/P genotype is significantly associated with accelerated cancer progression, and is likely a major factor in the early age of onset in African Americans (Figure 1). Our data are consistent with the findings published by Katkooori et al. demonstrating that the P/P genotype is a significant risk factor for colon cancer in African Americans (42). Zhang et al., likewise

found that the Pro allele and P/P genotype were enriched in Asian gastric cancer patients compared to non-cancer control individuals (43). Sullivan et al. reported that Pro72 cells primarily enter a G1-arrest with minimal apoptosis at chemotherapeutic doses that result in a strong apoptotic response in Arg72 cells (14). Jeong et al. found that the Arg allele increased expression of proapoptotic genes, such as *PUMA* and the Pro allele favored expression of cell cycle regulators, like *p21*(12). However, not all studies have been in agreement over the function of the Pro allele; Frank et al. found that the Pro allele enhanced apoptotic response to ionizing radiation and increases transactivation of inflammatory response genes (8). The differential function of the Arg and Pro alleles between these studies may be due to species differences in the model systems deployed, or differences in the genetic backgrounds of human studies, as we and others have shown that the Pro allele is a colon cancer risk factor for humans of African, but not European descent. In our view, the preponderance of data suggests that the Pro allele favors a temporary cell cycle arrest and DNA repair response over a more permanent response like apoptosis or differentiation in the case of stem cells. Further analysis suggests that the P/P genotype-related health disparity is based on an overall two-fold faster accumulation of mutations, rather than a reduction in the number of mutations necessary for transformation. Our analysis of the mutation rate within tumors of P/P and not P/P individuals found no significant difference in the mutation rate due to TP53 codon 72 polymorphism within the tumor. Multiple factors are influential on the mutation rate, but individuals with the P/P genotype having a larger stem cell population would increase the probability that driver mutations could accumulate to be passed on. If this is the case, our analyses would predict that P/P individuals could have as many as twice the number of colonic crypt stem cells. This would help explain the increased longevity of P/P individuals (10, 11) and propensity for high grade undifferentiated tumors in African Americans (2).

Using a panel of single allele hemizygous derivatives having only the Arg or only the Pro polymorphism, along with deep sequencing technology, we generated lists of TP53-dependent, allele-independent and TP53 codon 72 allele-dependent transcripts. We went on to perform an shRNA-based second-site suppressor screen for generating resistance to TP53-induced growth arrest. We took a similar approach as with the RNA-seq data: first establish relevance with a TP53 allele-independent study, then move onto allele-specific analysis. Predictably, shRNAs against TP53 showed up at the top of the TP53 allele-independent study, and other identified targets are of interest for future study. The chromatin-modifying and differentiation-related factors EED, PHF1, and PRDM1 showed up as Arg-specific mediators of TP53 function, supporting the ideas developed from the RNA-seq data analysis that TP53^{72Arg} has a novel, but as yet unidentified differentiation function.

The overlap between the RNA-seq studies and the shRNA screen was PRDM1, and we confirmed that the β isoform is specifically upregulated by TP53^{72Arg}. PRDM1 binds to a currently uncharacterized array of promoters and recruits a Histone H3 K4 Demethylase and a Histone H3 K9 methyltransferase (G9a) (25). PRDM1 is therefore able to silence active chromatin, and some of its targets may include PCNA and Ki67, well-known proliferation genes (44). Moreover, PRDM1 is critical for enforcing terminal B-cell differentiation (45,

46), is implicated in differentiation of T cells (47) and functions as a tumor suppressor (48, 49). It is interesting to speculate about a potential Arg-related aging mechanism, by which DNA damage induces a PRDM1 response, leading to a transient chromatin silencing activity whose products persist throughout an individual's life. The build-up of these products would eventually increase the propensity of stem cells to differentiate and/or simply lose their pluripotent function, thereby reducing the size of the stem cell pool. A similar TP53-dependent mechanism has indeed been described in Fanconi's anemia for controlling the size of the hematopoietic stem cell pool (50). Such a mechanism is consistent with the possible tendency of P/P individuals to have a larger stem cell pool, thereby exhibiting increased longevity at the expense of accelerated cancer progression. More experiments, however, will of course be needed to more fully explore this hypothesis.

There are likely other germline variants that contribute to the accelerated cancer progression observed in African Americans, but the high frequency of P/P individuals within this population is clearly a contributing factor. The theorized stem cell differentiation induced by the Arg allele removing damaged stem cells from the population. The model predicts that with successive waves of DNA damaging signals, each transient activation of p53 would trigger the accumulation of additional H3K9Me3 repressive marks at genes required for stem-cell maintenance, thus slowly eroding the overall replicative potential of the stem cell pool but protecting from cancer in the process. Our hypotheses based on these data would help explain the disparity in the incidence of early onset colon cancers in TP53 Pro72/Pro72 individuals, who are overrepresented in the African American population. Finally, the novel cell lines described herein will be valuable reagents for the scientific community to use in pursuit of allele-specific pharmacologic intervention strategies to mitigate inherited cancer risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the NIH for funding (Grant # U01CA158428). Additional support was provided by core facilities supported by NCI funding (Grant #P30-AR048311 and P30-AI027767). Data from the Cancer Genome Atlas Colorectal Adenocarcinoma dataset was included in our analyses, as such the authors would like to thank the TCGA Research Network. PJB designed research and wrote the paper. MRB and CCW performed research, analyzed data, and wrote the paper. HM and NY analyzed data. JT, KD, ZB, MC, EC, CF and BW performed research and analyzed data. KEC, CEB, MS, and HJ contributed reagents.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin.* 2013; 63:11–30. [PubMed: 23335087]
2. Karami S, Young HA, Henson DE. Earlier age at diagnosis: another dimension in cancer disparity? *Cancer Detect Prev.* 2007; 31:29–34. [PubMed: 17303347]
3. Vazquez A, Bond EE, Levine AJ, Bond GL. The genetics of the p53 pathway, apoptosis and cancer therapy. *Nature reviews Drug discovery.* 2008; 7:979–987.
4. Pietsch EC, Humbey O, Murphy ME. Polymorphisms in the p53 pathway. *Oncogene.* 2006; 25:1602–1611. [PubMed: 16550160]

5. Sjalander A, Birgander R, Saha N, Beckman L, Beckman G. p53 polymorphisms and haplotypes show distinct differences between major ethnic groups. *Hum Hered.* 1996; 46:41–48. [PubMed: 8825462]
6. Shi H, Tan SJ, Zhong H, Hu W, Levine A, Xiao CJ, et al. Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. *Am J Hum Genet.* 2009; 84:534–541. [PubMed: 19344876]
7. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–1073. [PubMed: 20981092]
8. Frank AK, Leu JI, Zhou Y, Devarajan K, Nedelko T, Klein-Szanto A, et al. The codon 72 polymorphism of p53 regulates interaction with NF- κ B and transactivation of genes involved in immunity and inflammation. *Mol Cell Biol.* 2011; 31:1201–1213. [PubMed: 21245379]
9. Jones JS, Chi X, Gu X, Lynch PM, Amos CI, Frazier ML. p53 polymorphism and age of onset of hereditary nonpolyposis colorectal cancer in a Caucasian population. *Clin Cancer Res.* 2004; 10:5845–5849. [PubMed: 15355915]
10. Bojesen SE, Nordestgaard BG. The common germline Arg72Pro polymorphism of p53 and increased longevity in humans. *Cell Cycle.* 2008; 7:158–163. [PubMed: 18256523]
11. van Heemst D, Mooijaart SP, Beekman M, Schreuder J, de Craen AJ, Brandt BW, et al. Variation in the human TP53 gene affects old age survival and cancer mortality. *Exp Gerontol.* 2005; 40:11–15. [PubMed: 15732191]
12. Jeong BS, Hu W, Belyi V, Rabadan R, Levine AJ. Differential levels of transcription of p53-regulated genes by the arginine/proline polymorphism: p53 with arginine at codon 72 favors apoptosis. *FASEB J.* 2010; 24:1347–1353. [PubMed: 20019240]
13. Pim D, Banks L. p53 polymorphic variants at codon 72 exert different effects on cell cycle progression. *Int J Cancer.* 2004; 108:196–199. [PubMed: 14639602]
14. Sullivan A, Syed N, Gasco M, Bergamaschi D, Trigiante G, Attard M, et al. Polymorphism in wild-type p53 modulates response to chemotherapy in vitro and in vivo. *Oncogene.* 2004; 23:3328–3337. [PubMed: 15077186]
15. Bonafe M, Salvioli S, Barbi C, Mishto M, Trapassi C, Gemelli C, et al. p53 codon 72 genotype affects apoptosis by cytosine arabinoside in blood leukocytes. *Biochem Biophys Res Commun.* 2002; 299:539–541. [PubMed: 12459171]
16. Li M, He Y, Dubois W, Wu X, Shi J, Huang J. Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells. *Mol Cell.* 2012; 46:30–42. [PubMed: 22387025]
17. Molchadsky A, Rivlin N, Brosh R, Rotter V, Sarig R. p53 is balancing development, differentiation and de-differentiation to assure cancer prevention. *Carcinogenesis.* 2010; 31:1501–1508. [PubMed: 20504879]
18. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proceedings of the National Academy of Sciences of the United States of America.* 2005; 102:13550–13555. [PubMed: 16141321]
19. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet.* 2008; 40:499–507. [PubMed: 18443585]
20. Mizuno H, Spike BT, Wahl GM, Levine AJ. Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. *Proceedings of the National Academy of Sciences of the United States of America.* 2010; 107:22745–22750. [PubMed: 21149740]
21. Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, et al. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature.* 2009; 460:1132–1135. [PubMed: 19668191]
22. Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, et al. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature.* 2009; 460:1149–1153. [PubMed: 19668189]

23. Tang X, Milyavsky M, Shats I, Erez N, Goldfinger N, Rotter V. Activated p53 suppresses the histone methyltransferase EZH2 gene. *Oncogene*. 2004; 23:5759–5769. [PubMed: 15208672]
24. Rago C, Vogelstein B, Bunz F. Genetic knockouts and knockins in human somatic cells. *Nat Protoc*. 2007; 2:2734–2746. [PubMed: 18007609]
25. Gyory I, Wu J, Fejer G, Seto E, Wright KL. PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing. *Nature immunology*. 2004; 5:299–308. [PubMed: 14985713]
26. Bikoff EK, Morgan MA, Robertson EJ. An expanding job description for Blimp-1/PRDM1. *Curr Opin Genet Dev*. 2009; 19:379–385. [PubMed: 19592232]
27. Bunz F, Dutriaux A, Lengauer C, Waldman T, Zhou S, Brown JP, et al. Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science*. 1998; 282:1497–1501. [PubMed: 9822382]
28. Bergamaschi D, Samuels Y, Sullivan A, Zvelebil M, Breysens H, Bisso A, et al. iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53. *Nat Genet*. 2006; 38:1133–1141. [PubMed: 16964264]
29. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:18545–18550. [PubMed: 20876136]
30. Siddique MM, Balram C, Fiszer-Maliszewska L, Aggarwal A, Tan A, Tan P, et al. Evidence for selective expression of the p53 codon 72 polymorphs: implications in cancer development. *Cancer Epidemiol Biomarkers Prev*. 2005; 14:2245–2252. [PubMed: 16172238]
31. Shtutman M, Maliyekkel A, Shao Y, Carmack CS, Baig M, Warholc N, et al. Function-based gene identification using enzymatically generated normalized shRNA library and massive parallel sequencing. *Proc Natl Acad Sci U S A*. 2010; 107:7377–7382. [PubMed: 20368428]
32. Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, et al. The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Res*. 2013; 41:D949–D954. [PubMed: 23109555]
33. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–337. [PubMed: 22810696]
34. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971; 68:820–823. [PubMed: 5279523]
35. Sur S, Pagliarini R, Bunz F, Rago C, Diaz LA Jr, Kinzler KW, et al. A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc Natl Acad Sci U S A*. 2009; 106:3964–3969. [PubMed: 19225112]
36. Thomas M, Kalita A, Labrecque S, Pim D, Banks L, Matlashewski G. Two polymorphic variants of wild-type p53 differ biochemically and biologically. *Mol Cell Biol*. 1999; 19:1092–1100. [PubMed: 9891044]
37. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biology*. 2010; 11:R106. [PubMed: 20979621]
38. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. 2004; 303:844–848. [PubMed: 14704432]
39. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ 3rd, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. 2009; 461:762–767. [PubMed: 19767730]
40. Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol*. 2008; 28:2718–2731. [PubMed: 18285464]
41. Yan J, Jiang J, Lim CA, Wu Q, Ng HH, Chin KC. BLIMP1 regulates cell growth through repression of p53 transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:1841–1846. [PubMed: 17264218]
42. Katkoori VR, Jia X, Shanmugam C, Wan W, Meleth S, Bumpers H, et al. Prognostic significance of p53 codon 72 polymorphism differs with race in colorectal adenocarcinoma. *Clin Cancer Res*. 2009; 15:2406–2416. [PubMed: 19339276]
43. Zhang Q, Ma YY, Wang HJ, Shao CM, Zhang J, Ye ZY. Meta-analysis of the association between P53 codon 72 polymorphisms and gastric cancer. *Journal of surgical oncology*. 2013; 107:360–366. [PubMed: 22886602]

44. Desai S, Maurin M, Smith MA, Bolick SC, Dessureault S, Tao J, et al. PRDM1 is required for mantle cell lymphoma response to bortezomib. *Mol Cancer Res.* 2010; 8:907–918. [PubMed: 20530581]
45. Shapiro-Shelef M, Lin KI, McHeyzer-Williams LJ, Liao J, McHeyzer-Williams MG, Calame K. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity.* 2003; 19:607–620. [PubMed: 14563324]
46. Cattoretti G, Angelin-Duclos C, Shaknovich R, Zhou H, Wang D, Alobeid B. PRDM1/Blimp-1 is expressed in human B-lymphocytes committed to the plasma cell lineage. *J Pathol.* 2005; 206:76–86. [PubMed: 15772984]
47. Oestreich KJ, Mohn SE, Weinmann AS. Molecular mechanisms that control the expression and activity of Bcl-6 in TH1 cells to regulate flexibility with a TFH-like gene profile. *Nat Immunol.* 2012; 13:405–411. [PubMed: 22406686]
48. Kucuk C, Iqbal J, Hu XZ, Gaulard P, De Leval L, Srivastava G, et al. PRDM1 is a tumor suppressor gene in natural killer cell malignancies. *P Natl Acad Sci USA.* 2011; 108:20119–20124.
49. Iqbal J, Kucuk C, Deleeuw RJ, Srivastava G, Tam W, Geng H, et al. Genomic analyses reveal global functional alterations that promote tumor growth and novel tumor suppressor genes in natural killer-cell malignancies. *Leukemia.* 2009; 23:1139–1151. [PubMed: 19194464]
50. Ceccaldi R, Parmar K, Mouly E, Delord M, Kim JM, Regairaz M, et al. Bone marrow failure in Fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell.* 2012; 11:36–49. [PubMed: 22683204]

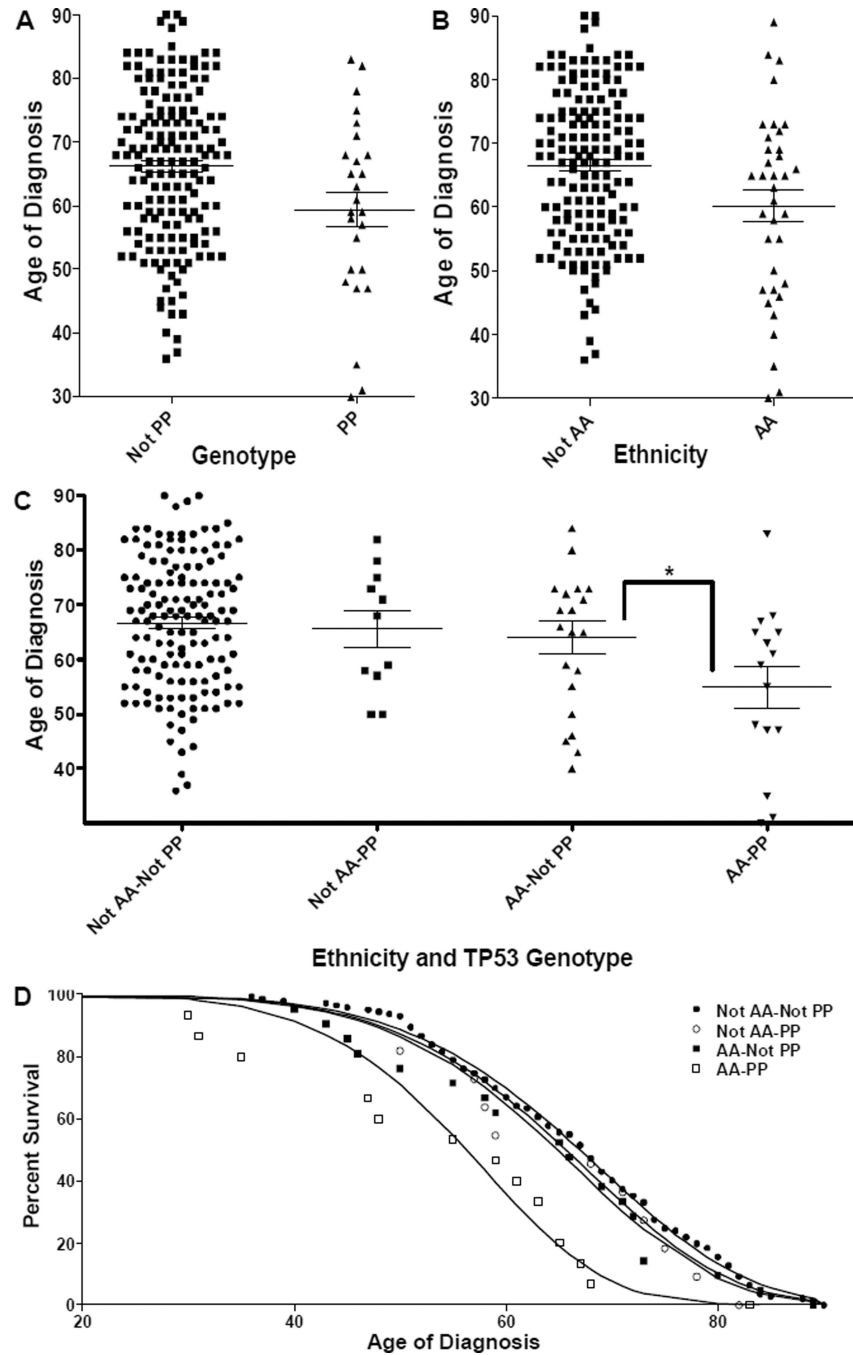


Figure 1. Effect of Race and the P/P Genotype on Mean Age of Colon Cancer Diagnosis
 In all Panels, error bars correspond to standard error. (A–C) The mean age of diagnosis is plotted for individuals having either a (A) P/P phenotype or not ($p=0.0107$), (B) being African American (AA) or not ($p=0.0076$), or (C) the combination of phenotype and ethnicity (the * indicates the difference between AA-P/P and AA-Not P/P is significant $p=0.0337$). (D) The % of undiagnosed patients is plotted against the age of diagnosis for P/P or not P/P individuals. The solid lines correspond to a least squares best-fit of Eq. 1 using $nM=6$ (the number of mutations minus 1). The C values for AA-P/P, AA-Not P/P, Not AA-

P/P and Not AA-Not P/P populations are $-2.19 \times 10^{-11} \text{ yr}^{-1}$, $9.35 \times 10^{-12} \text{ yr}^{-1}$, $8.67 \times 10^{-12} \text{ yr}^{-1}$, and $7.70 \times 10^{-12} \text{ yr}^{-1}$.

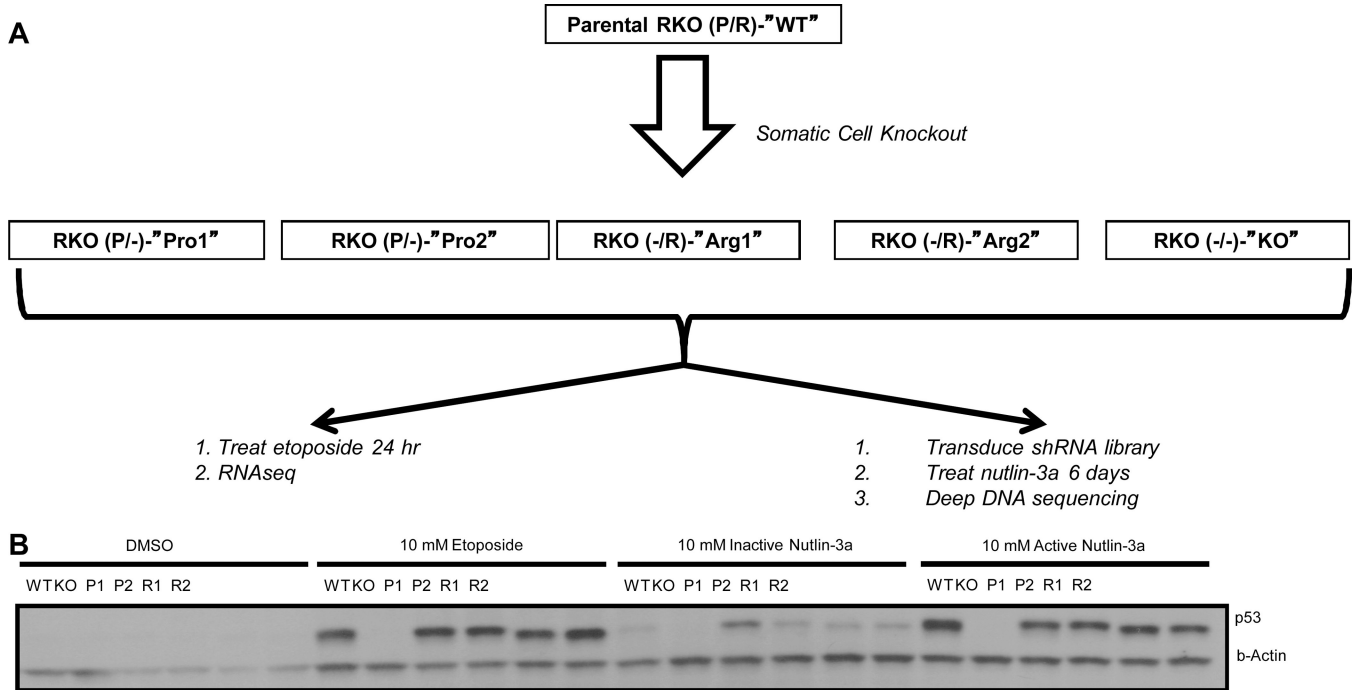


Figure 2. Generation of Cell Lines and Experimental Workflow

(A) The parental RKO colon carcinoma cell line is heterozygous for the TP53 codon 72 polymorphism. Using somatic cell knockout techniques, four hemizygous cell lines were created, two having only the Pro allele, and two having only the Arg allele. A knockout cell line was also created. All six of these cell lines were (i) treated with etoposide or DMSO (control), and then subjected to RNAseq analysis and (ii) transduced with an shRNA library, treated with nutlin-3a, and the subjected to deep DNA sequencing. (B) Western blot analysis for p53 expression levels of all six cells lines treated with either DMSO, 10 μ M etoposide, 10 μ M inactive nutlin-3a, or 10 μ M active nutlin-3a. Note the large p53 response in all cell lines with the exception of the KO, and the slight mobility differences in Arg vs. Pro.

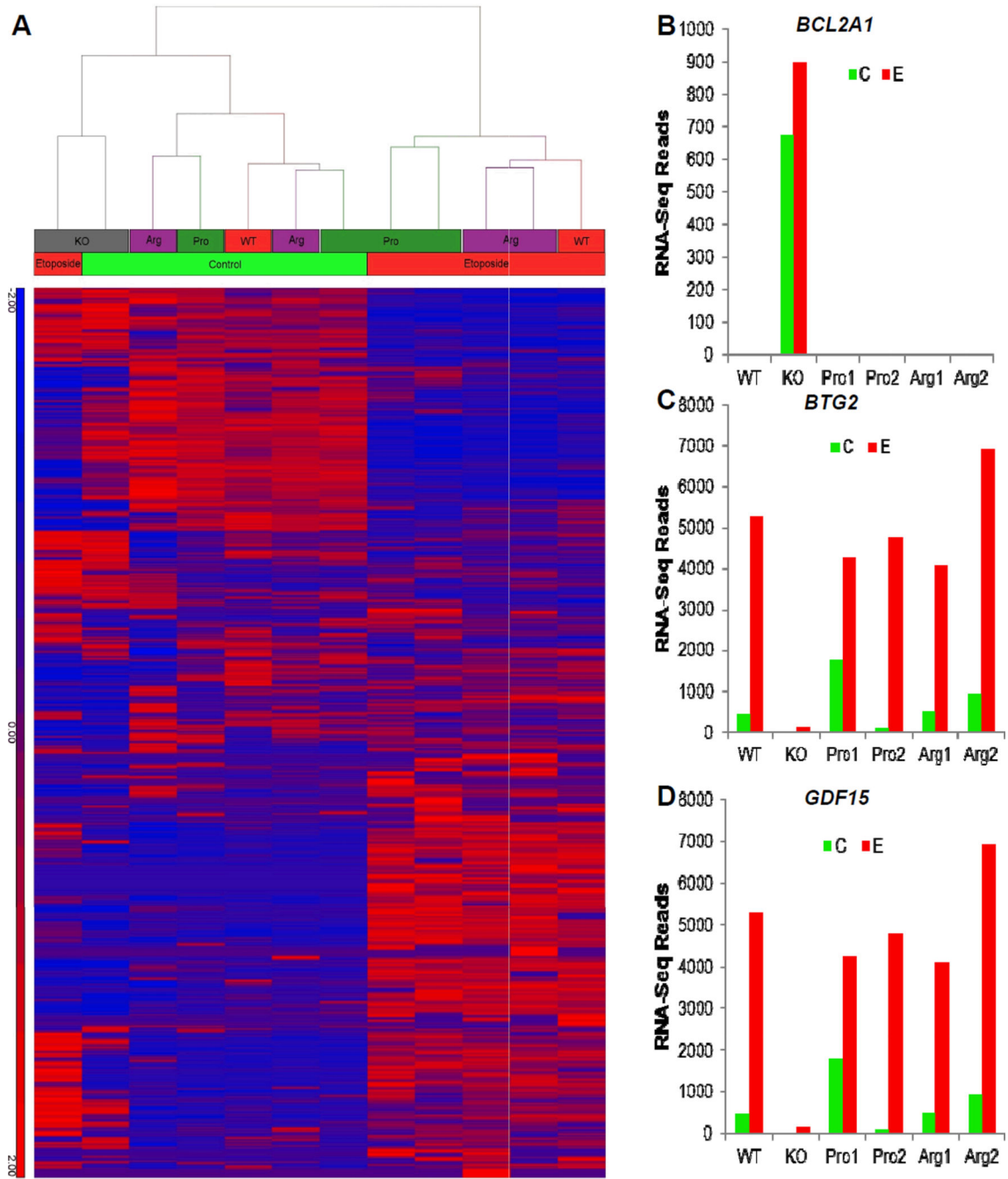


Figure 3. Overview of the RNAseq Data
(A) After alignment using the CLC Genomics Workbench, the number of reads per exon for each gene used as an input for hierarchical clustering analysis. Blue denotes low counts, whereas red denotes high counts. (B–D) RNA-Seq transcript expression profile for TP53 effectors *bcl2a1* (B), *btg2* (C), and *gdf15* (D). The x-axis indicates cell line, and the y-axis denotes the transcript expression levels.

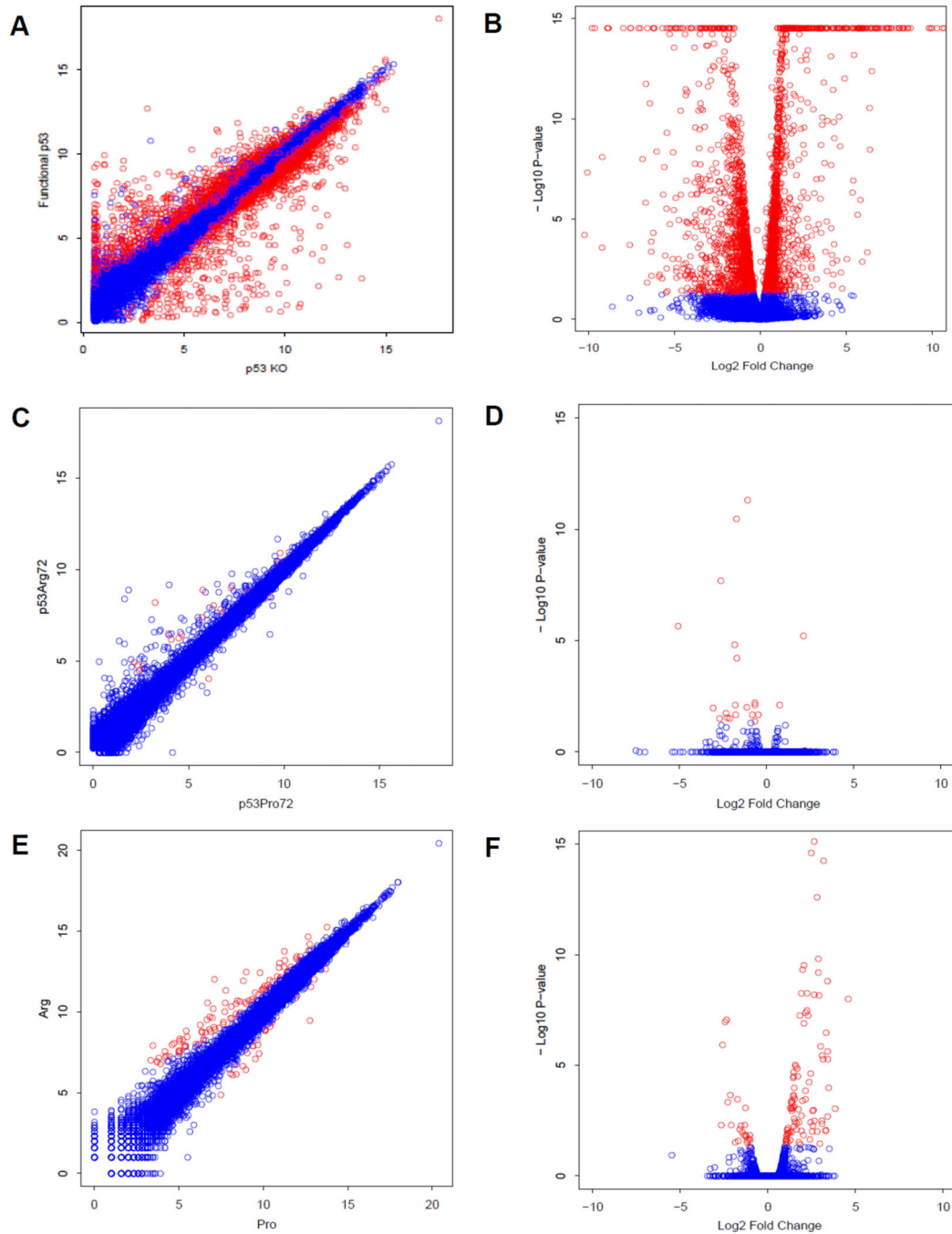


Figure 4. RNaseq Transcript Expression Profiles

(A–B) Scatter plot (A) and volcano plot (B) of expression profiles for TP53-dependent, allele independent genes. Points colored in red were significantly different (FDR adjusted p-value < 0.05). (C–D) Scatter plot (C) and volcano plot (D) of expression profiles for TP53 codon 72 dependent genes. Points colored in red were significantly different (FDR adjusted p-value < 0.05). (E–F) Scatter plot (E) and volcano plot (F) of expression profiles for etoposide-treated Arg and Pro samples subjected to ultra-deep sequencing. Points colored in

red were significantly different (FDR adjusted p-value <0.05 , and greater than twofold different between the two samples.

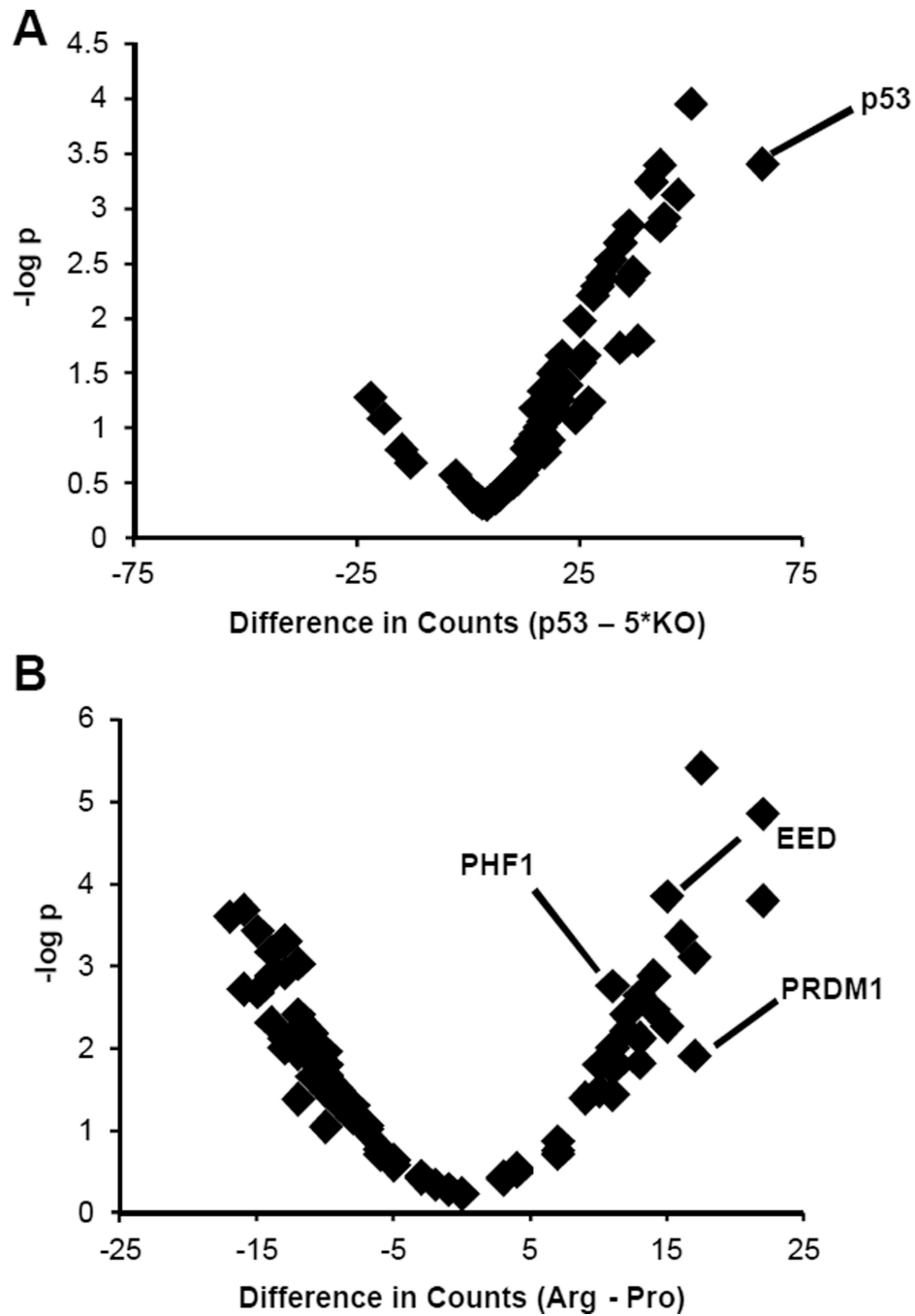


Figure 5. Volcano Plots of the shRNA Screen Results

(A) The x-axis is the difference in counts between the sum of those found in any TP53 containing cell line (WT, Pro1, Pro2, Arg1, or Arg2) minus five times those found in the KO. A count corresponds to a sequenced shRNA against a particular gene. The y-axis is the negative log of the p-value of obtaining the result, which was calculated using the binomial distribution, with the probability of success corresponding to an shRNA being equally likely to be observed in any cell line, and the number of trials as the total number of counts across all cell lines for a particular gene. (B) This plot is similar to (A), with the exception that the

x-axis corresponds to the difference in counts between Arg-containing (Arg1 and Arg2) and Pro-containing (Pro1 and Pro2) cell lines.

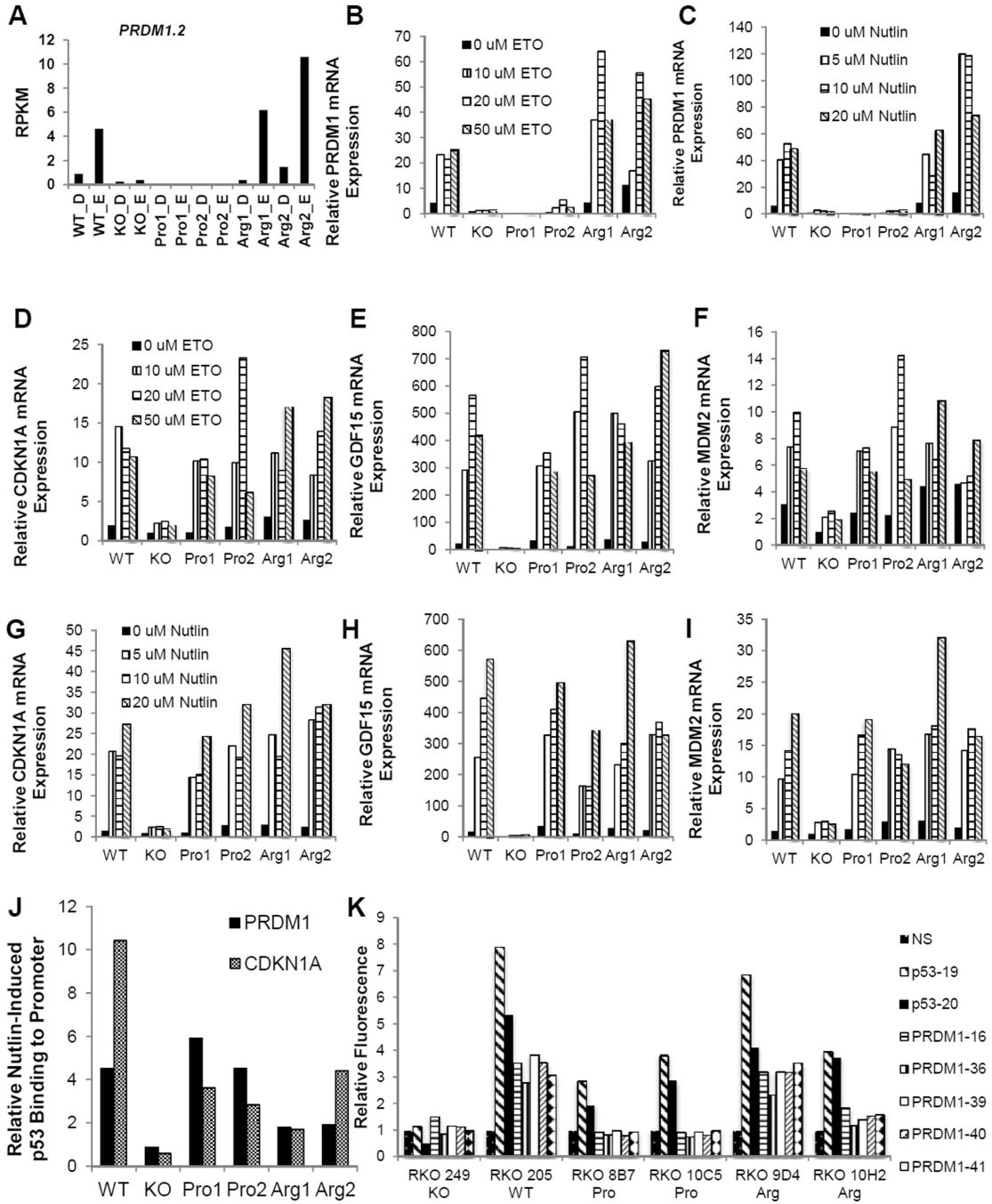


Figure 6. Validation of the Arg-specific Transcript PRDM1 β
 (A) RNAseq transcript expression profile for PRDM1 β . (B–C) PRDM1 β expression induced by various doses of etoposide (B) or Nutlin-3a (C) in different cell lines as measured by qPCR. (D–I) Expression of TP53 positive controls (CDKN1A, GDF15, MDM2) induced by various doses of etoposide (D–F) or Nutlin-3a (G–I) in different cell lines, as measured by qPCR. (J) Amount of PRDM1 or CDKN1A DNA pulled down by p53-ChIP as measured by qPCR. Cells were treated with Nutlin-3a or control, and the ratio of measurements obtained for these two treatment conditions is plotted for each gene and cell line. A value of 1

indicates that Nutlin-3a did not induce any additional p53 binding to the gene of interest. **(K)** Cells containing shRNA after 1 week treatment with Nutlin-3a, measured by FACS analysis selecting for GFP expression. Data shown as fold change compared to NS shRNA.