HHS Public Access

Author manuscript

Mol Carcinog. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

Mol Carcinog. 2020 January ; 59(1): 45–55. doi:10.1002/mc.23127.

Single nucleotide polymorphism rs13426236 contributes to an increased prostate cancer risk via regulating *MLPH* splicing variant 4

Fankai Xiao[1,2], **Peng Zhang**[2], **Yuan Wang**[2], **Yijun Tian**[2], **Michael James**[3], **Chiang-Ching Huang**[4], **Lidong Wang**[1,*], **Liang Wang**[2,5,*]

[1]Henan Key Laboratory for Cancer Research, The First Affiliated Hospital of Zhengzhou University, 40 Daxue Road, Zhengzhou, Henan 450052, China.

[2]Department of Pathology, MCW Cancer Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA.

[3]Department of Surgery, MCW Cancer Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA.

[4]Department of Biostatistics, University of Wisconsin, Milwaukee, Wisconsin 53201, USA

[5]Department of Tumor Biology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA

Abstract

A prostate cancer risk single nucleotide polymorphism (SNP), rs13426236, is significantly associated with melanophilin (*MLPH*) expression. To functionally characterize role of the rs13426236 in prostate cancer, we first performed splicing-specific expression Quantitative Trait Loci (eQTL) analysis and refined the significant association of rs13426236 allele G with an increased expression of *MLPH* splicing transcript variant 4 (V4) (P= 7.61E−5) but not other protein-coding variants (V1-V3) (P>0.05). We then performed an allele-specific reporter assay to determine if SNP-containing sequences functioned as active enhancer. Compared to allele A, allele G of rs13426236 showed significantly higher luciferase activity on the promoter of the splicing transcript V4 (P <0.03) but not on promoter of transcript V1 (P>0.05) in two prostate cancer cell lines (DU145 and 22Rv1). Cell transfection assays showed stronger effect of transcript V4 than V1 on promoting cell proliferation, invasion and anti-apoptotic activities. RNA profiling analysis demonstrated that transcript V4 overexpression caused significant expression changes in

[*]**Correspondence** Liang Wang, M.D., Ph.D., Department of Tumor Biology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA, Liang.Wang@moffitt.org, Phone: +1-813-7454955, Li-Dong Wang, M.D., Ph.D., Henan Key Laboratory for Cancer Research, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China, ldwang2007@126.com, Phone: +86-139-37183227.

Author contributions

FKX and YW performed experimental tests and initiated the manuscript. PZ, YJT, CCH and MJ performed additional data analysis. LDW and LW initiated and directed the study.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability Statement

RNA-seq data that support the findings of this study have been deposited in GEO database with accession number GSE135888.

glycosylation/glycoprotein and metal-binding gene ontology pathways (FDR<0.01). We also found that both transcripts V4 and V1 were significantly up-regulated in prostate adenocarcinoma (P≤2.49E-6) but only transcript V4 up-regulation was associated with poor recurrence free survival (P=0.028, HR=1.63, 95%CI=1.05–2.42) in The Cancer Genome Atlas (TCGA) data. This study provides strong evidence showing that prostate cancer risk SNP rs13426236 up-regulates expression of *MLPH* transcript V4, which may function as a candidate oncogene in prostate cancer.

Keywords

GWAS; eQTL; risk factor; isoform

1 INTRODUCTION

Prostate cancer is the most common non-skin cancer and second leading cause of cancer-related death in American men [1]. The 2018 National Comprehensive Cancer Network Prostate Cancer Guidelines recommend baseline screening for prostate cancer beginning at age 45. Although surgery and radiation therapy are effective for early and localized tumors [2], a significant number of prostate cancer patients eventually progress to advanced stages. Previous epidemiologic and twin studies have shown that genetics play an important role in the development of prostate cancer [3,4]. To elucidate the genetic causes of prostate cancer, genome-wide association studies (GWASs) have been performed and reported over 170 risk loci showing prostate cancer association [5,6]. However, more work is needed to understand molecular mechanisms of prostate cancer etiology.

It is well known that most disease risk single nucleotide polymorphisms (SNPs) are located in non-coding regions of the genome, many residing with some distance from nearby annotated genes. This result suggests that many of these SNPs (or their closely linked causal SNPs) likely reside in regulatory domains of the genome that control gene expression rather than in coding regions that directly affect protein function [7,8]. To identify causal SNPs in regulatory regions of the genome, epigenomic projects such as ENCODE [9] and REMC [10,11] have been initiated. These projects have applied high-throughput methods that detect open chromatin, specific histone modifications, and transcription factor (TF) binding sites and produced genome-wide maps of epigenomic events in a variety of cell types. These valuable databases have provided rich resources for molecularly characterizing cancer-associated regulatory SNPs. Additionally, to facilitate regulatory SNP discovery, several computational programs have been developed to integrate epigenomic landscapes with GWAS risk SNPs [12–18]. These databases and computational programs have been widely used and help identify functional SNPs for a given phenotype. However, functional validation of the knowledge-based SNP prediction remains a significant challenge.

Currently, to experimentally characterize functional SNPs for regulatory potential, the widely used assays include electrophoretic mobility shift assays (EMSA) and reporter assays. EMSA is applied to identify SNPs that had the potential to alter DNA–protein interactions [19]. A gene reporter assay can analyze the effect of a SNP on promoter or

enhancer activity [20]. Additionally, DNA editing strategies involving CRISPR/Cas9-based methods has also been used to evaluate the effect of a variant in cell lines or animal model [21–23]. Due to low throughput nature, these assays are not practical to screen hundreds to thousands of candidate SNP sites. To address this issue, we have developed single-nucleotide polymorphisms sequencing (SNPs-seq) technology to examine potential functional SNPs. In combination with prostate-specific expression quantitative trait loci (eQTL) dataset [8], we have reported 20 candidate functional SNPs and their associated genes at prostate cancer GWAS loci [24]. Among these SNP-gene association pairs, however, little is known about functional role of risk SNP rs13426236 and its target gene *MLPH* in prostate cancer. In this study, we performed functional analysis of the SNP-gene pair and molecularly characterized regulatory role of the rs13426236 in controlling *MLPH* expression. We also demonstrated a functional role of *MLPH* in prostate cancer proliferation and reported an association of its expression with clinical outcomes.

2 MATERIALS AND METHODS

2.1 *MLPH* transcript-specific eQTL analysis

Based on RefSeq gene database, the *MLPH* has four protein-coding splicing variants, denoted as transcript V1 (NM_024101), V2 (NM_001042467), V3 (NM_001281473) and V4 (NM_001281474), respectively. To associate the candidate SNP with each of the four transcripts, we designed transcript-specific TaqMan-based quantitative PCR assays. The transcript-specific primers and reference control *GAPDH* primers are listed in Table 1.

2.2 Cell lines and cell culture

Cell culture dishes and plates were purchased from Corning Inc (Corning, NY, USA). Human prostate cancer cell lines 22Rv1 and DU145 were purchased from American Type Culture Collection (Manassas, VA, USA). 22Rv1 was cultured in RPMI 1640 media and DU145 was in DMEM at 37°C and 5% CO2. All media contain 10% fetal bovine and 1% penicillin and streptomycin (Gibco, Carlsbad, CA, USA). The genotypes of rs13426236 in DU145 and 22Rv1 are A/G and A/A, respectively.

2.3 *MLPH* promoter-containing plasmid construction and allele-specific reporter assay

Promoter sequences of transcripts V1 and V4 were derived from Genecopoeia (www.genecopoeia.com). The transcript V1 promoter is located at chr2 from 238394621 to 238396015 (1395bp) and transcript V4 promoter is located at chr2 from 238393618 to 238395238 (1621bp) (hg19). The two promoters have 618bp overlap. We designed two separate PCR primer pairs to amplify the two promoters. Transcript V1 promoter sequence was amplified using primer pair F/R:

AGACACTAGAGGGTAGGGCCTGAAATACCCTGATT/
CCATGGTGGCTTTACGGGGCAAGGCTGGATAAT. Transcript V4 promoter sequence was amplified using primer pair F/R:
AGACACTAGAGGGTACAGACGTTGCAGTAAGCCGAGATCA/CCATGGTGGCTTTAC
GGTGGTTCCAGGACCGAGGACGC. At 5’ end of each primer, we added 15bp long tail homologous to plasmid sequence for Gibson-based vector construction. We inserted these

amplified promoter sequences into pGL4 plasmid (Promega, Madison, WI, USA) by replacing the minimal promoter.

We also inserted the predicted enhancer sequences containing SNP rs13426236 with either allele A or G into the plasmids containing either *MLPH* promoter transcripts V1 or V4. The primer sequences were CTGTCTAAGGTCAAGTGTTGC (forward) and GGCAGGCTTTAACTGTTGTG (reverse). A diagram illustrating the constructed plasmid along with locations of *MLPH* promoter and SNP-dependent enhancer sequence was shown in Supplementary Figure 1. To determine allele-dependent enhancer activity on transcript-specific promoters, we applied dual-luciferase reporter assay by co-transfecting cell lines with Renilla luciferase control vector (Promega) and *MLPH* promoter-containing plasmid in 96-well plates. After 48h transfection, we measured firefly luciferase activity on a bio-luminometer. All reading measurements were obtained from at least three replicates. Statistical significance was determined by two tailed Student's t test using GraphPad Prism 6.

2.4 *MLPH* transcript-specific plasmid construction

To clone transcript V1 coding sequence, we designed forward primer CCCAAGCTTCAAGAAGCAGAAATGGGGAAG (with HindIII cutting site) and reverse primer GCTCTAGACTGTCCCGTTAGGACTGGTG (with Xbal cutting site). This pair of primers were used to amplify full length coding sequence of *MLPH* transcript V1. By double digestion using Hind III and Xbal, the digested PCR product and pGL4 vector were ligated with T4 DNA ligase for subsequent bacterial transformation. To obtain full length transcript V4 coding sequence, we took advantage of previously amplified transcript V1 coding sequence and used Gibson Assembly assay to knock out exons 7–9. Specifically, we designed internal primer A: CCTAGACCCTGTGGGCTGTCCCTGGCCTCA and internal primer B: CCCACAGGGTCTAGGTGCTGGAGTGCGCACGGAGG. The primers A and B had 15 bp overlap at 5' end (underlined) to facilitate Gibson-based assembly. We used primer A and transcript V1 forward primer to amplify exons 1–6 while primer B and transcript V1 reverse primer to amplify exons 10–16. We performed Sanger sequencing to confirm that the final assembly of transcript V4 had excluded internal exons 7–9 (429bp).

2.5 Cell transfection and *MLPH* expression quantification

For transient transfection, 2.5 μg *MLPH* transcript V1 or V4 was used to transfect DU145 and 22Rv1 through Lipofectamine 3000 reagent (ThermoFisher Scientific, Waltham, MA, USA) following the manufacturer's protocol. After incubating at 37°C for 6 h, the medium was replaced with culture medium and continued growth for 48 h. For RT-qPCR assay, transfected cellular RNA was isolated through TRIzol reagent (Zymo Research, Irvine, CA, USA) and its quantity was determined by NanoDrop (ThermoFisher Scientific). 100ng of total RNA was reverse-transcribed to cDNA by superscript VILO cDNA synthesis kit (ThermoFisher Scientific). qPCR was performed in PikoReal real-time PCR system in 20 μL solution containing 10 μL 2× Gene Expression Master Mix (ThermoFisher Scientific), 1 μL each of appropriate primer and probe, and 1 μL cDNA. The primer and probe sequences were the same as described in eQTL analysis.

For Western blot analysis, cells were lysed in RIPA buffer after 48h transfection. Protein concentration was tested by Pierce BCA Protein Assay Kit. 20μg of total protein was applied to SDS-PAGE in 1x running buffer for 120 min in 80 volts and transferred to nitrocellulose membrane (Bio-Rad, Hercules, CA, USA). The membrane was incubated with MLPH antibody (LSBio, Seattle, WA, USA) overnight after blocked in 3% bovine serum albumin (BSA) in TBST (Tris-buffered saline, 0.1% Tween 20). After 30 minutes washing, the membrane was incubated with secondary anti-rabbit IgG (LSBio) for 1 hour. Protein was detected through an electrochemiluminescent system using a chemiluminescent detection kit (Thermofisher Scientific).

2.6 Cell proliferation and apoptosis assays

To determine effect of *MLPH* transcripts V1 and V4 on cell proliferation, we applied MTS cell proliferation assay in prostate cancer cell lines DU145 and 22Rv1. We seeded 100μL cells into 96-well plates at the density of $2.5*10^5$ /mL with three technical repeats. After 48 hours transfection, we pipetted 20μl of CellTiter 96 AQueous One Solution Reagent (Promega) into each well and incubated the mixture for 4h. OD values at 490nm was recorded using a microplate reader. To determine cell apoptosis after transfection, we applied Annexin V-APC detection kit (ThermoFishier Scientific). 48 hours after transfection, cells were detached through trypsinization and resuspended in 100μl 1×Binding Buffer. By adding 5 μl Annexin V-APC, we incubated cells for 15 minutes at room temperature and then washed cells in 1×Binding Buffer. Immediately before flow cytometry analysis, we added 5 μl Propidium Iodide Staining Solution and put the cell solution on ice in the dark. The experiment was repeated three times.

2.7 Cell migration and invasion assays

We utilized a wound healing assay with live-cell imaging to test cell migration in 96-well ImageLock tissue culture plate (Essen BioScience, Ann Arbor, MI, USA). Cells were first incubated in a standard cell incubator. When grown to confluence, cells were then transfected with 100 ng *MLPH* transcript V1 or V4 plasmid for 48h. The confluent monolayer was wounded by a WoundMaker tool (Essen Bioscience) to create consistent wound areas in each well. Culture media was used to wash culture plate twice to prevent cells from setting and reattaching. After adding 100 μl media, the assay plate was placed into IncuCyte (Essen BioScience). Cell migration was recorded every 3h for 3 days. Relative Wound Density (%) were calculated by the IncuCyte software.

We also tested cell invasion on Essen ImageLock plates using live imaging. Wells were coated with a thin layer of Matrigel (BD Bioscience, San Jose, CA, USA) overnight at 37°C. Prostate cancer cells were seeded into 96-well ImageLock tissue culture plate and incubated in a standard cell incubator. When grown to confluence, cells were transfected with 100 ng *MLPH* transcript 1 or transcript 4 for 48h, followed use of WoundMaker to create precise and reproducible wounds. Fifty microliters of 80% Matrigel:20% culture medium was added to each well and the assay plate was placed in a 37°C incubator for 30 min. After adding 100 μl additional culture media to each well, the assay plate was placed into IncuCyte for 3 days with repeat scanning every 3h. Relative Wound Density (%) was calculated by the IncuCyte software.

2.8 RNA sequencing

To determine effect of *MLPH* on signaling pathways, we transfected 22Rv1 cells with 20 μg plasmid containing *MLPH* transcripts V1 or V4 for 48 h. Total RNA was isolated using total RNA extraction kit (Zymo Research). The RNA concentration was determined by measuring the absorbance at 260 nm using a NanoDrop. High quality RNA samples (1 μg each) were sent to Novogene (Chula Vista, CA, USA) for RNA library preparation and sequencing. 150bp pair-end sequencing was performed in a HiSeq2500. DNASTAR Genomics Suite (Madison, WI) was used for RNA mapping and read count calculation. A greater than or equal to 8-fold change in expression was considered as differential gene expression. Online DAVID functional microarray analysis tool (https://david.ncifcrf.gov/summary.jsp) was used for pathway enrichment analysis. FDR adjustment was applied for multiple testing correction.

2.9 Clinical association analysis for *MLPH* expression

To estimate clinical relevance of *MLPH*, we examined the association of the gene expression with prostate cancer and clinicopathological features using RNA profiling data from TCGA. Normalized expression levels of *MLPH* splicing transcripts were downloaded from TCGA SpliceSeq (https://bioinformatics.mdanderson.org/TCGASpliceSeq/index.jsp). *MLPH* gene expression level in 52 normal and 497 prostate cancer tissues were included. Samples were stratified into two groups based on the tissue types (normal vs tumor) for differential analysis or median *MLPH* expression in tumor tissues for survival analysis. Graphpad Prism (version 6) was used to perform statistical analyses.

3 RESULTS

3.1 Allele-dependent regulation of *MLPH* transcript V4

Our previous study has shown significant eQTL signals at prostate cancer risk locus of 2q37.3 [8]. For all eQTL signals in the linkage equilibrium block, the association of rs13426236 with *MLPH* expression was among the most significant (p= 8.49E-09) (Figure 1A). The SNP rs13426236, which demonstrated allele-dependent protein binding difference, was selected from our previous study [24]. Full-length gene has a total of 16 exons (transcript V1, NM_024101), while alternative splicing generates three additional protein-coding transcripts. One of the alternative transcripts is transcript V4 (NM_001281474), which excludes exons 7–9. Additionally, transcript V4 has a unique transcription start site, which is significantly different from the other three transcripts (Figure 1B). Interestingly, about half of the genomic region covering the entire *MLPH* gene resides in a super enhancer region [25]. Furthermore, we checked *MLPH* RNA expression in the Genotype-Tissue Expression (GTEx) database and found that the gene had the highest expression level in prostate tissue among 53 tissues tested (Figure 1C). In addition to prostate tissue (eQTL p=8.30E-06), the GTEx data also showed significant association between rs13426236 and *MLPH* mRNA expression in thyroid tissue (p=1.50E-08) and esophagus-mucosa (p=7.40E-09).

To determine potential effect of rs13426236 genotypes on *MLPH* alternative splicing, we performed splicing variant-specific qPCRs in 87 benign prostate tissues. This analysis revealed a significant association of the allele G with increased expression of *MLPH*

transcript V4 (P= 7.60E−5). The same analyses of the other three transcripts did not show any statistical association (Figure 2A–B). To determine if rs13426236 regulated the target *MLPH* in an allele-dependent manner, we replaced the pGL4.28 minimal promoter with either *MLPH* transcript V1 or V4 promoter. Luciferase assays showed that allele G had significantly higher enhancer activity than allele A in regulating *MLPH* transcript V4 promoter, either in cell line DU145 (P =0.003) or in 22Rv1 (p=0.0005) (Figure 2C–D). However, alleles A and G had no significant difference when using transcript V1 promoter in the two cell lines. This result suggests that *MLPH* transcript 4 is a direct target of the SNP rs13426236.

3.2 *MLPH* transcript V4 promotes cell growth and anti-apoptosis

To evaluate the effect of *MLPH* overexpression on cell growth characteristics, we transfected pGL4 vector containing the full coding region of either transcript V1 or V4 into two prostate cancer cell lines. We performed qPCR and Western blot assays to determine transfection efficiency. Both cell lines (DU145 and 22Rv1) showed low level baseline expression of transcript V1 and V4. Transfection of the two variant transcripts significantly increased their expression at mRNA and protein levels (Figure 3A–B). After confirming successful transfection, we first performed MTS assays to determine the effect of these two transcripts on cell viability. We seeded an equal number of cells for each group of control, transcript V1 and transcript V4. Compared to baseline controls, cell viability was significantly higher in transcript V4-overexpressing cells than in controls (p=0.01) (Figure 3C). Although transcript 1-overexpressing cells also showed an increased cell viability, it did not show statistical significance. We then performed flow cytometry to evaluate effect of the two transcripts on cell death. We found that apoptotic cells were significantly reduced in transcript V4-overexpressing cells than in the baseline controls and in transcript V1-overexpressing cells ($p \leq 0.005$) (Figure 4). Those results suggest an increased cell viability and a decreased apoptosis by overexpressing the transcript V4.

3.3 *MLPH* transcript V4 increases cell migration and invasion

To evaluate effect of *MLPH* transcripts on cell migration and invasion, we performed wound healing assays using an IncuCyte live imaging system. Compared to control-transfected cells, relative wound healing was significantly higher in both transcript V4 and V1-overexprssing DU145 cells across a 72-hour culture period (Figure 5A, C). Furthermore, the transcript V4-overexpressing cells showed higher relative wound density than transcript V1 overexpressing cells. We then performed Matrigel invasion assays to test the effect of transcript variant overexpression on cell invasiveness. Relative invasion through the Matrigel matrix was significantly higher in transcript V4-overexpressing cells than in controls and transcript 1-overexpressing cells (Figure 5B, D). These results indicate that both transcripts V4 and V1 have significant effect on cell migration and invasion. However, this effect is stronger in transcript V4 than transcript V1. Additionally, since the effect of transcript V4 on cell migration was similar regardless of Matrigel, the effect is most likely related to migration, rather than a direct effect on invasiveness.

3.4 *MLPH* transcript V4 causes over-representation of metal-binding and glycoprotein-related categories

To evaluate genes that were significantly affected by overexpressing transcripts V4 and V1, we performed RNA-seq profiling analysis in 22Rv1 cell line. Three RNA-seq samples included baseline (empty vector) control, transcript V1-transfected cells and transcript V4-transfected cells. Overall, we received over 49.4 million raw reads (45.3–55.9 million) per sample and 94.1% (93.2–94.7%) reads were mapped to human genes (hg19). We applied log2-transformed RPKM as gene expression values for data analysis. When compared to the baseline RNA profile, we identified 390 genes and 314 genes showing ≥8-fold change of expression in transcript V4-overexpressing cells and in transcript V1-overexpressing cells, respectively. Among those, 252 genes were common, meaning that these genes demonstrated significant difference of expression in transfected cells (both V4 and V1) compared to vector control (Figure 6). Interestingly, all of the 252 shared genes were up-regulated by *MLPH* variant over-expression. Noticeably, the two splicing variants induced over 56, 36 and 29-fold increases in expression of the genes *METTL15*, *SEL1L* and *GNA13*, respectively. Enrichment analysis from these up-regulated genes demonstrated over-representation in metal-binding and glycoprotein-related categories. Supplementary Tables list all statistics regarding the RNA-seq, *MLPH*-induced genes and fold changes, and over-representation categories by enrichment analysis.

3.5 Up-regulation of *MLPH* transcript V4 in tumor tissues is common and predicts recurrence-free survival

To associate expression of splicing variants with clinical outcomes, we first compared expression differences between normal tissues (N=52) and tumor tissues (N=496) from the TCGA prostate adenocarcinoma dataset. The average read count (normalized) in 52 normal prostate tissues is 1005 for transcript V1 and 56 for transcript V4 with V1/V4 ratio being 19. Compared to normal controls, tumor tissues demonstrated significant upregulation in transcript V4 (P=2.49E-6) and V1 (P=2.17E-7) (Figure 7A). We then tested the transcript abundance for their potential association with disease outcomes. Although there was no association with overall survival, we observed a significant association between transcript V4 and recurrence-free survival. Kaplan-Meier analysis showed that higher expression of the transcript V4 was correlated with poor biochemical recurrence-free survival (p=0.028, HR=1.63, 95%CI=1.05–2.42) (Figure 7B). However, transcript V1 did not show such an association (Figure 7C).

4 DISCUSSION

Previous GWAS has identified a significant association of 2q37.3 locus with increased risk to prostate cancer [6,26]. Further eQTL analysis has revealed *MLPH* as one candidate gene since SNPs at this locus are associated with the gene expression in prostate tissues [8,24]. In this study, we performed detail functional analysis of the SNP-gene association, elucidated the genetic control of *MLPH* expression and determined functional role of the gene in cell growth and migration. In particular, this study focused on transcript level analysis, which not only characterized differential genetic regulation of *MLPH* splicing variants but also associated these splicing variants with clinical outcomes. From cell lines-based functional

assays to clinical correlative analysis, this study provided strong evidence showing that splicing variant V4 of *MLPH* protein-coding transcripts is a candidate responsible for the increased risk to prostate cancer.

MLPH is an essential member of the melanosome trafficking complex, which can work as a Rab effector protein involved in intracellular melanosome transport [27] and acting as a link between Rab27a and myosin Va [28]. A study has shown that Rab family and their effector proteins often have abnormal expression in tumors, which may drive tumor aggressiveness [29]. Mutations altering GTP/GDP-binding of Rabs involvement with effectors may reduce the efficiency and specificity in membrane traffic that are involved in disease development such as cancer [30]. Interestingly, *MLPH* has multiple splicing transcript variants. Based on reference gene database, transcript V1 is the longest protein-coding transcript while V4 is the shortest protein-coding transcript. From InterPro protein domain search, the V1 shows only one Rab effector domain while V4 has two Rab effectors. We hypothesize that the extra Rab effector in transcript V4 may contribute to the enhanced activity of *MLPH* gene on cell proliferation and migration. Although our study has provided some evidence to support the hypothesis, more detail analysis is needed to fully elucidate functional role of transcript V4.

Genetic control and functional role of *MLPH* in prostate cancer is not clear. By integrating ChIP-seq and microarray expression profiling with GWAS risk SNPs, a study identified a SNP rs11891426 in an intron of *MLPH* that T→G change attenuated the transcriptional activity of the ARBS in an AR reporter gene assay [31]. The study also reported that the expression of *MLPH* in primary prostate tumors was significantly lower in those with the risk allele G compared to the wildtype allele T, suggesting tumor suppressive role of the gene. This result, however, seems contradictory to the observation in our study. Instead of gene-level analysis, we performed transcript-level analysis. We show significant up-regulation of *MLPH* transcripts in prostate cancer tissues from TCGA dataset. We also demonstrate that risk allele G of rs13426236 increases *MLPH* transcript V4 expression. Our cell lines-based analyses further reveal that the transcript V4 enhances anti-apoptotic effect, promotes cell growth and invasion. Importantly, the increased transcript V4 is associated with PSA-based biochemical progression. All these results support oncogenic role of *MLPH* transcript V4 (if not other transcripts) in prostate cancer.

Oncogenic function of *MLPH* (at least transcript V4) also has significant effect on genes critical for prostate cancer. We found that *METTL15* was significantly increased in *MLPH*-transfected cells. A genome-wide CRISPR screen has identified *METTL15* (*HNRNPL*) as a prostate cancer dependency regulating RNA splicing. *METTL15* has ability to directly regulates the alternative splicing of a set of RNAs, including those encoding *AR*, the key lineage-specific prostate cancer oncogene [32]. In another study, bayesian network modelling of microarray and mass spectrometry data have identified an N-terminal *SEL1L* sequence as a putative serum biomarker of prostate cancer [33], which is supported by our finding that the SEL1L is up-regulated in prostate cancer cells transfected with *MLPH*. Our study also showed significant increase of *GNA13* expression in *MLPH*-overexpression cells. It is believed that *GNA13* is an important mediator of prostate cancer cell invasion [34]. Knockdown of GNA13 in highly invasive PC3 cells has revealed that these cells depend on

GNA13 expression for their invasion, migration, and Rho activation. These results suggest *MLPH* as a potential key driver for prostate cancer initiation and progression.

It is worth mentioning that exact role of rs13426236 in regulating *MLPH* expression requires further investigation. The regulatory SNP is over 51kb from its target promoter region. A chromosome conformation capture (3C) test may be needed to demonstrate allele-dependent long-range chromatin interaction in prostate-originated cell lines. Nevertheless, our functional analysis shows that rs13426236 is a regulatory SNP that specifically controls *MLPH* transcript V4 expression. Variant allele G of rs13426236 increases transcript V4 expression. The splicing variant V4 has shown an enhanced ability to promote cell growth and anti-apoptosis, and to increase cell migration and invasion. Up-regulation of the transcript V4 in tumor tissues is common and predicts biochemical recurrence-free survival. Therefore, *MLPH* transcript V4 plays an oncogenic role in prostate cancer. Further study is needed to evaluate *MLPH* and its transcript variants as potential biomarker for prostate cancer detection and outcome prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by National Institute of Health (R01 CA157881) to LW and China Scholar Counsel Fellowship awarded to FKX. Authors declare no conflicts of interest.

Abbreviations

MLPH	melanophilin
SNP	single nucleotide polymorphism
eQTL	expression Quantitative Trait Loci
TCGA	The Cancer Genome Atlas
GTEx	Genotype-tissue expression
GWAS	Genome-wide association studies
TF	Transcription factor
EMSA	Electrophoretic mobility shift assay
SNPs-seq	Single-nucleotide polymorphisms sequencing
RPKM	Read per kilobase per million sequences

References

1. Zareba P, Flavin R, Isikbay M, et al. Perineural Invasion and Risk of Lethal Prostate Cancer. Cancer Epidemiol Biomarkers Prev. 2017;26(5):719–726. [PubMed: 28062398]

2. Carroll PR, Parsons JK, Andriole G, et al. Prostate cancer early detection, version 1.2014. Featured updates to the NCCN Guidelines. J Natl Compr Canc Netw. 2014;12(9):1211–1219; quiz 1219. [PubMed: 25190691]

3. Hjelmborg JB, Scheike T, Holst K, et al. The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev. 2014;23(11):2303–2310. [PubMed: 24812039]

4. Narod S Genetic epidemiology of prostate cancer. Biochim Biophys Acta. 1999;1423(1):F1–13. [PubMed: 9989206]

5. Benafif S, Kote-Jarai Z, Eeles RA, Consortium P. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiol Biomarkers Prev. 2018;27(8):845–857. [PubMed: 29348298]

6. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50(7):928–936. [PubMed: 29892016]

7. Chen H, Yu H, Wang J, et al. Systematic enrichment analysis of potentially functional regions for 103 prostate cancer risk-associated loci. Prostate. 2015;75(12):1264–1276. [PubMed: 26015065]

8. Thibodeau SN, French AJ, McDonnell SK, et al. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. Nat Commun. 2015;6:8653. [PubMed: 26611117]

9. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. [PubMed: 22955616]

10. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–330. [PubMed: 25693563]

11. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045–1048. [PubMed: 20944595]

12. Cowper-Sal lari R, Zhang X, Wright JB, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nature genetics. 2012;44(11):1191–1198. [PubMed: 23001124]

13. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012;22(9):1790–1797. [PubMed: 22955989]

14. Wang Y, Lieberman R, Pan J, et al. miR-375 induces docetaxel resistance in prostate cancer by targeting SEC23A and YAP1. Mol Cancer. 2016;15(1):70. [PubMed: 27832783]

15. Guo L, Du Y, Qu S, Wang J. rVarBase: an updated database for regulatory features of human variants. Nucleic Acids Res. 2016;44(D1):D888–893. [PubMed: 26503253]

16. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930–934. [PubMed: 22064851]

17. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. Nucleic Acids Res. 2012;40(18):e139. [PubMed: 22684628]

18. Khurana E, Fu Y, Colonna V, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science (New York, NY. 2013;342(6154):1235587.

19. Smith AJ, Palmen J, Putt W, Talmud PJ, Humphries SE, Drenos F. Application of statistical and functional methodologies for the investigation of genetic determinants of coronary heart disease biomarkers: lipoprotein lipase genotype and plasma triglycerides as an exemplar. Hum Mol Genet. 2010;19(20):3936–3947. [PubMed: 20650961]

20. Ipe J, Swart M, Burgess KS, Skaar TC. High-Throughput Assays to Assess the Functional Impact of Genetic Variants: A Road Towards Genomic-Driven Medicine. Clin Transl Sci. 2017;10(2):67–77. [PubMed: 28213901]

21. Yoshimi K, Kaneko T, Voigt B, Mashimo T. Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform. Nat Commun. 2014;5:4240. [PubMed: 24967838]

22. Jin HJ, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. Oncotarget. 2016;7(34):54616–54626. [PubMed: 27409348]
23. Gao P, Xia JH, Sipeky C, et al. Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. Cell. 2018;174(3):576–589 e518. [PubMed: 30033361]
24. Zhang P, Xia JH, Zhu J, et al. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. Nat Commun. 2018;9(1):2022. [PubMed: 29789573]
25. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Res. 2016;44(D1):D164–171. [PubMed: 26438538]
26. Kote-Jarai Z, Olama AA, Giles GG, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat Genet. 2011;43(8):785–791. [PubMed: 21743467]
27. Van Gele M, Dynoodt P, Lambert J. Griscelli syndrome: a model system to study vesicular trafficking. Pigment Cell Melanoma Res. 2009;22(3):268–282. [PubMed: 19243575]
28. Strom M, Hume AN, Tarafder AK, Barkagianni E, Seabra MC. A family of Rab27-binding proteins. Melanophilin links Rab27a and myosin Va function in melanosome transport. J Biol Chem. 2002;277(28):25423–25430. [PubMed: 11980908]
29. Gundry C, Marco S, Rainero E, et al. Phosphorylation of Rab-coupling protein by LMTK3 controls Rab14-dependent EphA2 trafficking to promote cell:cell repulsion. Nat Commun. 2017;8:14646. [PubMed: 28294115]
30. Tzeng HT, Wang YC. Rab-mediated vesicle trafficking in cancer. J Biomed Sci. 2016;23(1):70. [PubMed: 27716280]
31. Bu H, Narisu N, Schlick B, et al. Putative Prostate Cancer Risk SNP in an Androgen Receptor-Binding Site of the Melanophilin Gene Illustrates Enrichment of Risk SNPs in Androgen Receptor Target Sites. Hum Mutat. 2016;37(1):52–64. [PubMed: 26411452]
32. Fei T, Chen Y, Xiao T, et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. Proc Natl Acad Sci U S A. 2017;114(26):E5207–E5215. [PubMed: 28611215]
33. Deng X, Geng H, Ali HH. Cross-platform analysis of cancer biomarkers: a Bayesian network approach to incorporating mass spectrometry and microarray data. Cancer Inform. 2007;3:183–202. [PubMed: 19455243]
34. Rasheed SA, Teo CR, Beillard EJ, Voorhoeve PM, Casey PJ. MicroRNA-182 and microRNA-200a control G-protein subunit alpha-13 (GNA13) expression and cell invasion synergistically in prostate cancer cells. J Biol Chem. 2013;288(11):7986–7995. [PubMed: 23329838]

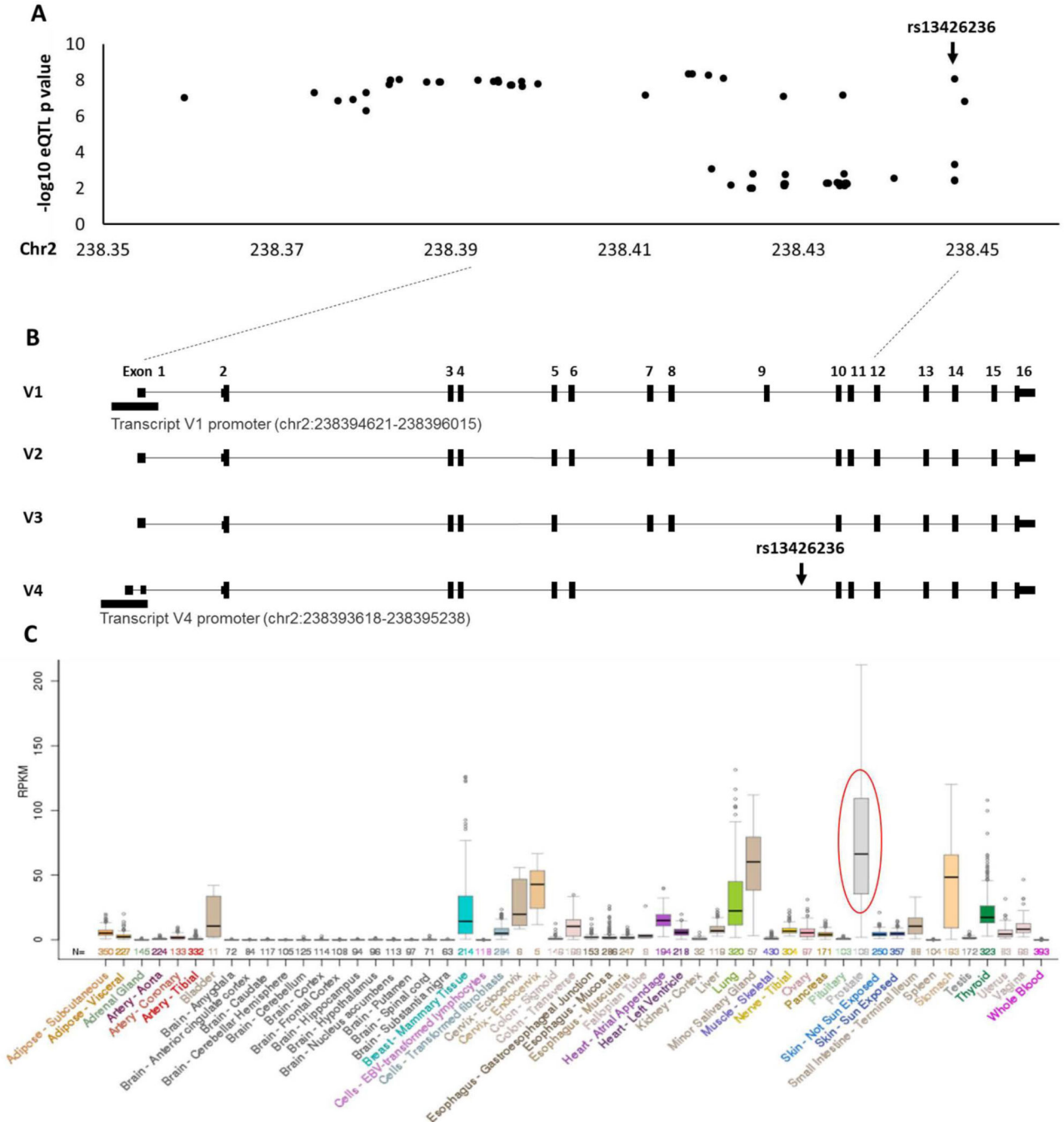

Figure 1. Functional SNP rs13426236 and its target gene *MLPH*.

A. eQTL analysis shows that rs13426236 has one of the most significant association signals in the linkage disequilibrium (LD) block. Each black dot represents a SNP with its height being -log10 eQTL p value. **B.** Four major *MLPH* protein-coding transcripts in reference gene database. Transcript V4 has unique transcription start site. The SNP rs13426236 is in intron 9 of transcript 1 (V1). **C.** Genotype-Tissue Expression (GTEx) dataset shows the highest expression of *MLPH* in prostate tissue (circled) among 53 tested human tissues.

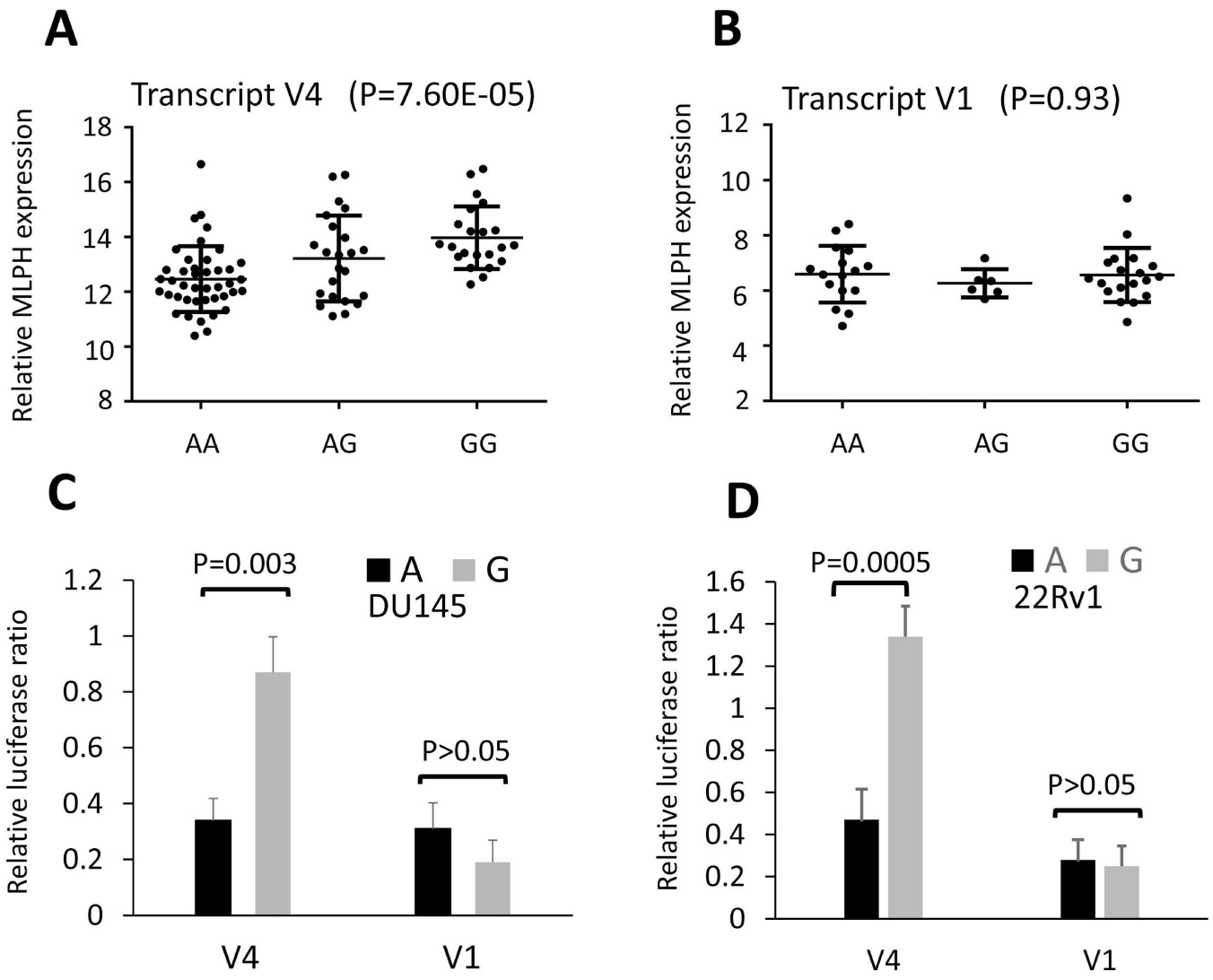

Figure 2. Splicing-specific eQTL analysis and luciferase assay.

A-B. RT-qPCR shows significant association of rs13426236 genotypes with *MLPH* transcript V4 (**A**) but not transcript V1 (**B**). Allele G of the rs13426236 is associated with an increased expression of transcript V4. **C-D.** Luciferase assay shows that allele G-containing fragment has stronger enhancer effect on transcript V4 promoter than allele A-containing fragment. However, both allele G and allele A have no significant effect on transcript V1 promoter. **C.** DU145 cell line. **D.** 22Rv1 cell line.

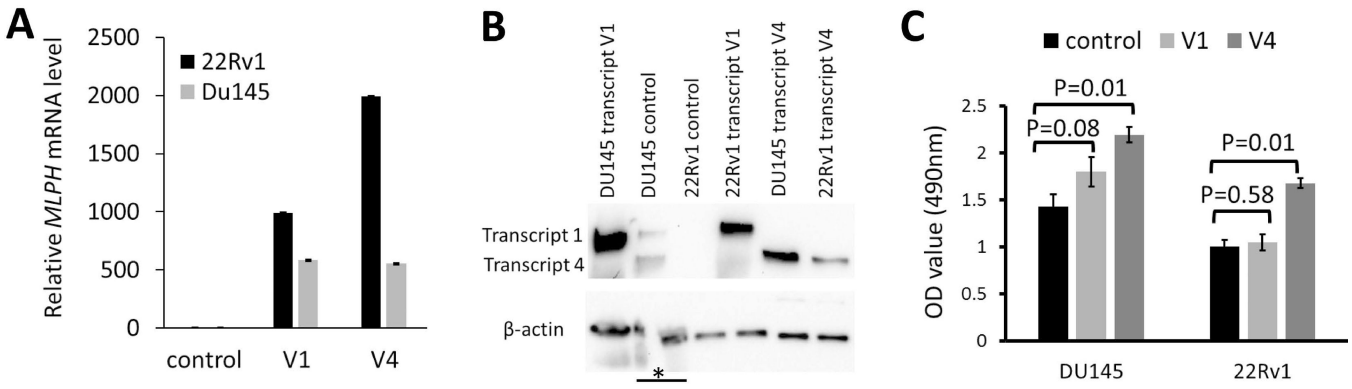

Figure 3. Cell viability test by transfection of *MLPH* transcript V1 and V4.

A. RT-qPCR shows a significant increase of *MLPH* RNA transcripts after 48h transfection. **B.** Western blot analysis shows a significant increase of *MLPH* protein after 48h transfection. Asterisk (*) indicates that a single beta-actin band was split in sample DU145 control due to broken gel during membrane transfer. **C.** Cell viability test shows significantly higher live cell population in transcript V4-transfected cells than vector control. However, transcript V1-transfected cells did not show significant increase in live cell population when compared to vector control.

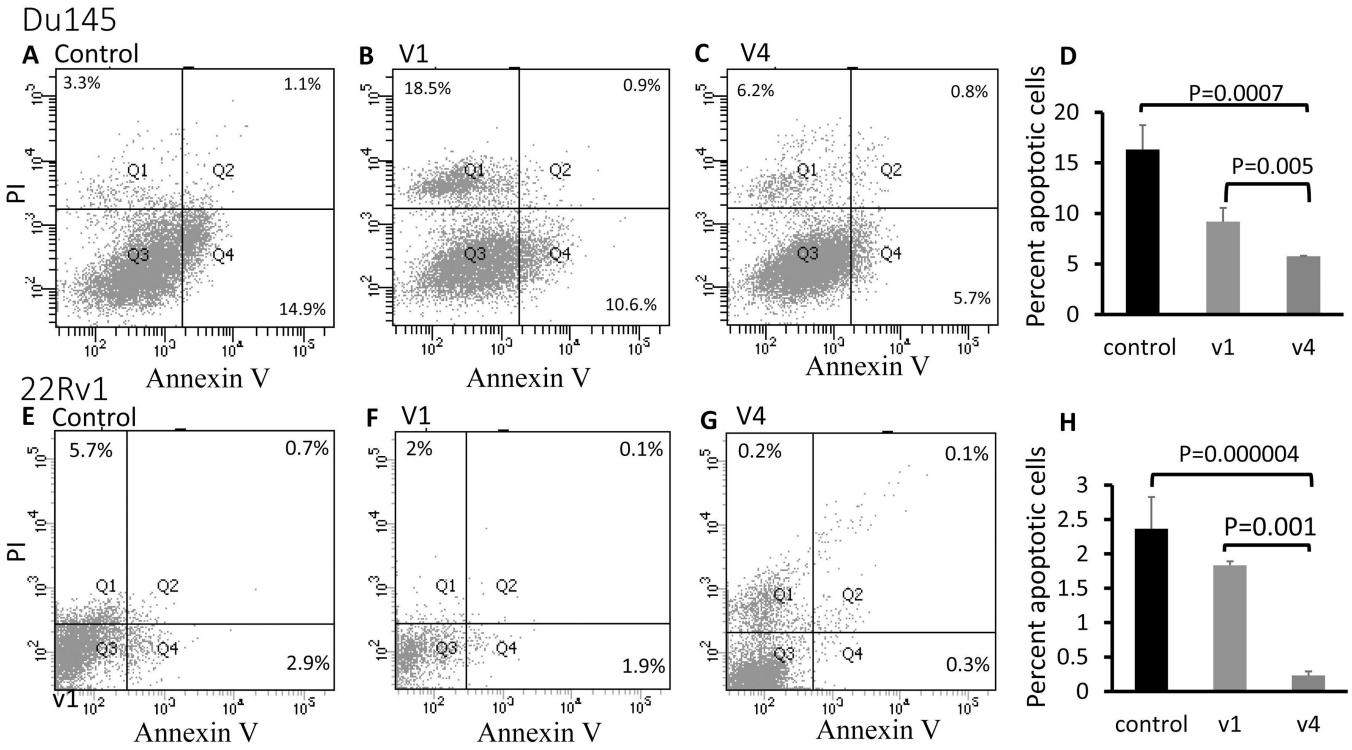

Figure 4. Different effect of *MLPH* transcripts on programed cell death.

Flow cytometry analysis shows that overexpression of transcript V4 confers anti-apoptotic activity. V4-transfected cell lines have significantly less apoptotic cells than V1-transfected cell lines and vector controls. **A-D:** Flow cytometry analysis for DU145. **E-H:** Flow cytometry analysis for 22Rv1.

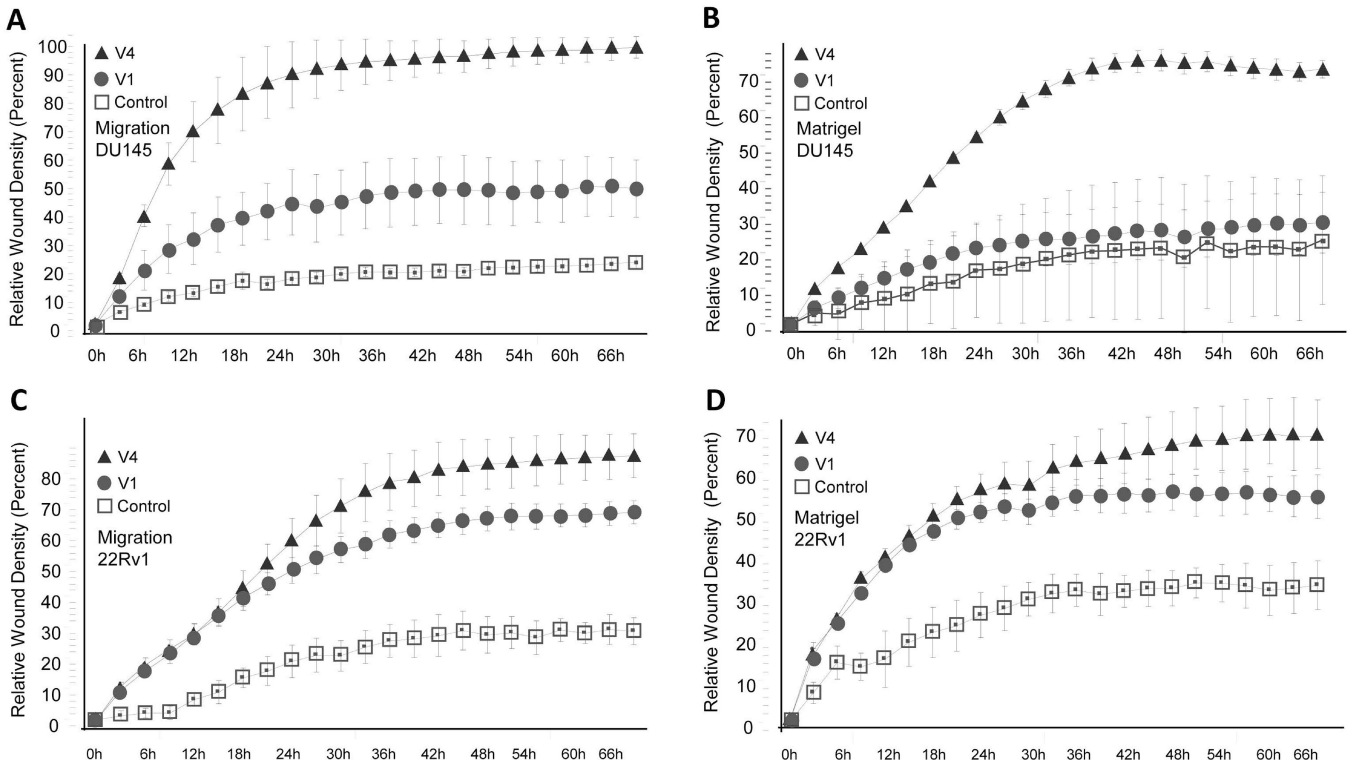

Figure 5. Different effect of *MLPH* transcripts on migration and invasion.

DU145 and 22Rv1 cells were transfected with *MLPH* transcripts V1 and V4. Cell migration and invasion via wound-healing assay with or without Matrigel were measured by IncuCyte Live Cell Analysis system during 72hr period after transfection. Compared to controls and transcript V1-transfected cells, the transcript V4-transfected cells show higher migration rate (A-DU145, C-22Rv1) and more invasive growth (B-DU145, D-22Rv1).

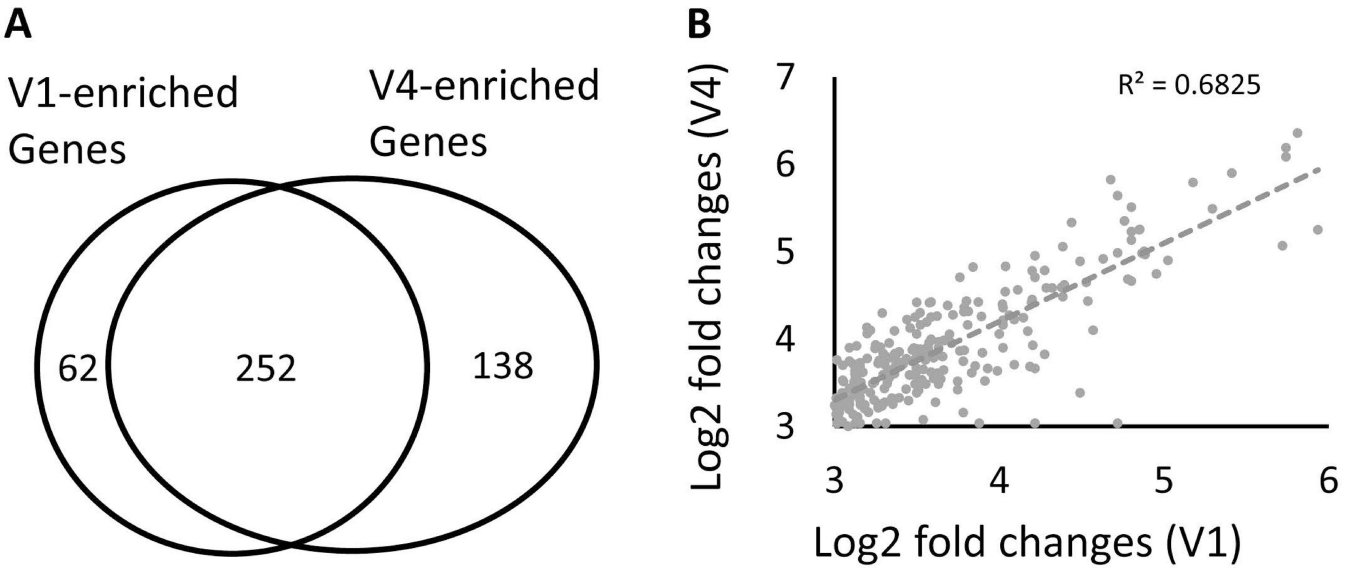

Figure 6. RNA profiling analysis of *MLPH* variants-induced genes.

RNA sequencing was performed in a prostate cancer cell line (22Rv1) transfected with either *MLPH* transcript V4 or V1. **A**. Vann diagram shows numbers of genes with ≥8-fold changes in either transcript V1 or V4 transfected cells. **B.** Scatter plot shows significant correlation in expression levels of 252 shared genes that were induced by overexpressing V1 and V4 transcripts.

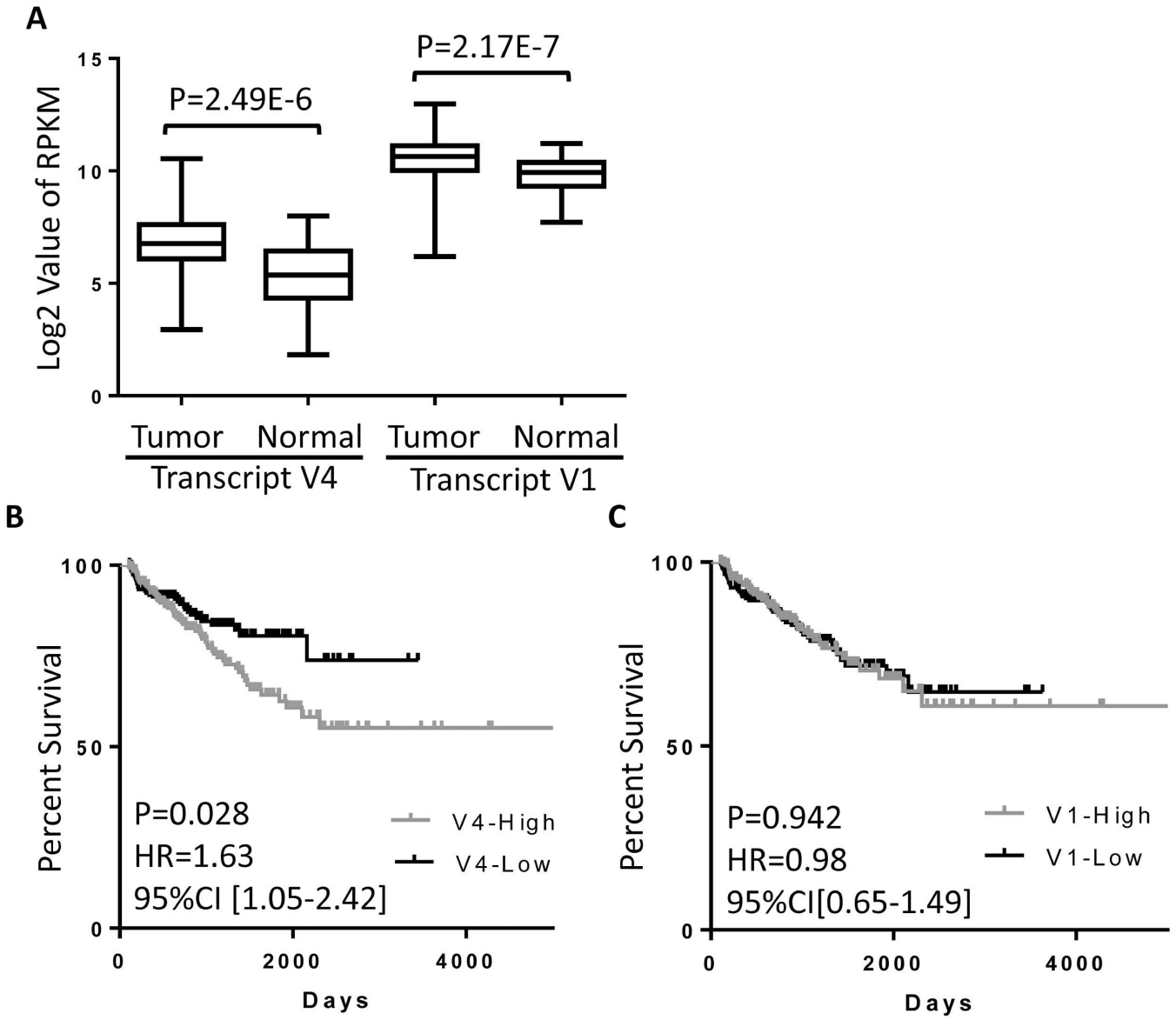

Figure 7. Association of *MLPH* expression levels with clinical outcomes.

A. TCGA prostate cancer dataset shows significant up-regulation of *MLPH* transcript V4 and V1 in tumor tissues. **B.** Kaplan-Meier analysis shows that higher expression of *MLPH* transcript V4 is associated with poor biochemical recurrence-free survival. **C.** *MLPH* transcript V1 is not associated with biochemical recurrence-free survival.

Table 1.

Primers and probes for quantification of MLPH transcripts

	Forward Primer (5′-->3′)	Reverse Primer (5′-->3′)	Note
Transcript V1	AATGCCTGCTGACCTACCTG (exon 9)	TCCTGTTGTACTGGACGGG (exon 11)	315bp
Transcript V2	GAGAGTCAGGGTCTAGGTGC (exon 8–10 junction)	GTCACTGCCACTCTGTCCTC (exon 11)	320bp
Transcript V3	CTGAGAGTCAGGGTCTAGGT (exon 8–10 junction)	TGTCTGAAACCTCCGGGT (exon 10–12 junction)	207bp
Transcript V4	GGACAGCCCACAGGGTCTA (exon 6–10 junction)	GTCTGTTGGTTTGATGGGCAG (exon 11)	230bp
V1-V4 common probe	AGAGCCCAACAGGGACAAATCAGT (exon 10)		FAM labelled
GADPH	CACCAGGGCTGTTTTAACTC	GAAGATGGTGATGGGATTTC	178bp
GADPH Probe	GTCAAGGCTGAGAACGGGAAGCT		HEX labelled