

RESEARCH ARTICLE

# *KLF3* and *PAX6* are candidate driver genes in late-stage, MSI-hypermuted endometrioid endometrial carcinomas

Meghan L. Rudd<sup>1#a</sup>, Nancy F. Hansen<sup>1b</sup>, Xiaolu Zhang<sup>1#b</sup>, Mary Ellen Urick<sup>1</sup>, Suiyuan Zhang<sup>2</sup>, Maria J. Merino<sup>3</sup>, National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program<sup>4¶</sup>, James C. Mullikin<sup>1,4</sup>, Lawrence C. Brody<sup>5</sup>, Daphne W. Bell<sup>1\*</sup>

**1** Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **3** Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **4** NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland, United States of America, **5** Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America

<sup>#a</sup> Current address: NIH Training Center, Workforce Support and Development Division, National Institutes of Health, Bethesda, Maryland, United States of America

<sup>#b</sup> Current address: Department of Neurosciences, University of Toledo College of Medicine and Life Sciences, Toledo, Ohio, United States of America

<sup>¶</sup> Membership of the National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program is provided in the Acknowledgments.

\* [belldaph@mail.nih.gov](mailto:belldaph@mail.nih.gov)



**OPEN ACCESS**

**Citation:** Rudd ML, Hansen NF, Zhang X, Urick ME, Zhang S, Merino MJ, et al. (2022) *KLF3* and *PAX6* are candidate driver genes in late-stage, MSI-hypermuted endometrioid endometrial carcinomas. PLoS ONE 17(1): e0251286. <https://doi.org/10.1371/journal.pone.0251286>

**Editor:** Obul Reddy Bandapalli, German Cancer Research Center (DKFZ), GERMANY

**Received:** May 4, 2021

**Accepted:** November 5, 2021

**Published:** January 26, 2022

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All relevant data are within the manuscript, the [Supporting Information](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001153.v1.p1) files, dbGaP Study Accession phs001153.v1.p1 (URL: [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001153.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001153.v1.p1)), the open access MC3 mutation annotation format (MAF) file (mc3.v0.2.8.PUBLIC.maf.gz) (URL: <https://gdc.cancer.gov/about-data/publications/mc3-2017>) for The Cancer Genome Atlas pan-cancer UCEC study, publicly available clinicopathologic information for TCGA-UCEC samples as published by Berger et al., Cancer Cell.

## Abstract

Endometrioid endometrial carcinomas (EECs) are the most common histological subtype of uterine cancer. Late-stage disease is an adverse prognosticator for EEC. The purpose of this study was to analyze EEC exome mutation data to identify late-stage-specific statistically significantly mutated genes (SMGs), which represent candidate driver genes potentially associated with disease progression. We exome sequenced 15 late-stage (stage III or IV) non-ultramutated EECs and paired non-tumor DNAs; somatic variants were called using Strelka, Shimmer, SomaticSniper and MuTect. Additionally, somatic mutation calls were extracted from The Cancer Genome Atlas (TCGA) data for 66 late-stage and 270 early-stage (stage I or II) non-ultramutated EECs. MutSigCV (v1.4) was used to annotate SMGs in the two late-stage cohorts and to derive p-values for all mutated genes in the early-stage cohort. To test whether late-stage SMGs are statistically significantly mutated in early-stage tumors, q-values for late-stage SMGs were re-calculated from the MutSigCV (v1.4) early-stage p-values, adjusting for the number of late-stage SMGs tested. We identified 14 SMGs in the combined late-stage EEC cohorts. When the 14 late-stage SMGs were examined in the TCGA early-stage data, only *Krüppel-like factor 3 (KLF3)* and *Paired box 6 (PAX6)* failed to reach significance as early-stage SMGs, despite the inclusion of enough early-stage cases to ensure adequate statistical power. Within TCGA, nonsynonymous mutations in *KLF3* and *PAX6* were, respectively, exclusive or nearly exclusive to the microsatellite

2018 Apr 9;33(4):690-705.e9. doi: [10.1016/j.ccell.2018.03.014](https://doi.org/10.1016/j.ccell.2018.03.014), and publicly available data for the TCGA-UCEC study within the cBioPortal for Cancer Genomics (URL: [https://www.cbioportal.org/study/summary?id=ucec\\_tcga\\_pan\\_can\\_atlas\\_2018](https://www.cbioportal.org/study/summary?id=ucec_tcga_pan_can_atlas_2018)).

Data contained within dbGaP Study Accession phs001153.v1.p1 (URL: [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001153.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001153.v1.p1)) are available via dbGAP Authorized/Controlled Access with Data Use Limitations of "Use of the data must be related to endometrial cancer" and "collaboration required." The Human Protein Atlas results for KLF3 and PAX6 protein expression (by antibody staining) in the normal glandular epithelium of the endometrium are publicly available at the following URLs (<https://www.proteinatlas.org/ENSG00000109787-KLF3/tissue/endometrium>, and <https://www.proteinatlas.org/ENSG0000007372-PAX6/tissue/endometrium>).

**Funding:** This work was funded by the Intramural Program of the National Human Genome Research Institute at the National Institutes of Health, grant number ZIA HG200338, Principal Investigator DWB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

instability (MSI)-hypermethylated molecular subgroup and were dominated by insertions-deletions at homopolymer tracts. In conclusion, our findings are hypothesis-generating and suggest that *KLF3* and *PAX6*, which encode transcription factors, are MSI target genes and late-stage-specific SMGs in EEC.

## Introduction

Defects in mismatch repair can result in DNA strand slippage and the appearance of microsatellite instability (MSI) [1]. MSI is common in endometrial carcinoma (EC) in which it occurs in ~30% of sporadic tumors. In this context, MSI generally results from *MLH1* hypermethylation and is associated with a hypermutated genome [2–4]. MSI/hypermethylated ECs are one of four distinct molecular subgroups of EC, defined by The Cancer Genome Atlas (TCGA) [2]. The three remaining subgroups are referred to as POLE/ultramutated, copy number-low/microsatellite stable (MSS), and copy number-high (serous-like) [2]. Each molecular subgroup has distinct clinical outcomes [2] (and reviewed in [5]) and the prognostic utility of this molecular classification is an area of active exploration.

Endometrial carcinoma (EC) exacts a significant toll on women's health. It resulted in 89,929 deaths globally in 2018 [6], and is projected to cause 12,940 deaths within the United States in 2021 [7]. Importantly, EC incidence is increasing annually in the US and many other countries [8]. This phenomenon is likely partly due to increasing rates of obesity [9], a well-recognized epidemiological risk factor for endometrioid endometrial carcinomas (EECs) that make up 75%–80% of all newly diagnosed endometrial tumors. EECs most often present as low-grade, early-stage (stage I or II) tumors, that are confined within the uterus [10]. Five-year survival rates for patients with low-grade, early-stage disease are high because surgery is often curative for this patient population, due to the limited extent of disease [10]. In contrast, patients with late-stage EEC have relatively poor outcomes [11], despite more aggressive treatment approaches of surgery with adjuvant chemotherapy or radiotherapy [12–14]. Thus, increasing tumor stage is an adverse prognosticator for EEC that is used in the clinical setting, as are high tumor grade (Grade 3; G3), and extent of lymphovascular space invasion [15]. The prognostic utility of molecular classification, according to *POLE*, microsatellite instability (MSI), and *TP53*/p53 status, is an area of active exploration originating from The Cancer Genome Atlas (TCGA) discovery that EECs can be subclassified into four molecular subgroups associated with distinct clinical outcomes [2] (and reviewed in [5]).

TCGA's initial comprehensive molecular characterization of primary endometrial carcinomas included exome sequencing of 200 EECs [2]; an expanded analysis that included 188 additional EECs was subsequently reported [16]. These studies confirmed prior findings that EEC exhibits high frequencies of somatic alterations resulting in activation of the PI3-kinase pathway, the RAS-RAF-MEK-ERK pathway, and the WNT/ $\beta$ -catenin pathway, frequent mutations in *ARID1A* (BAF250A) tumor suppressor, and mismatch repair defects resulting in MSI [2,16–18]. Moreover, many additional "significantly mutated genes" (SMGs), which represent candidate pathogenic driver genes, were annotated in EECs by TCGA using statistical approaches [2].

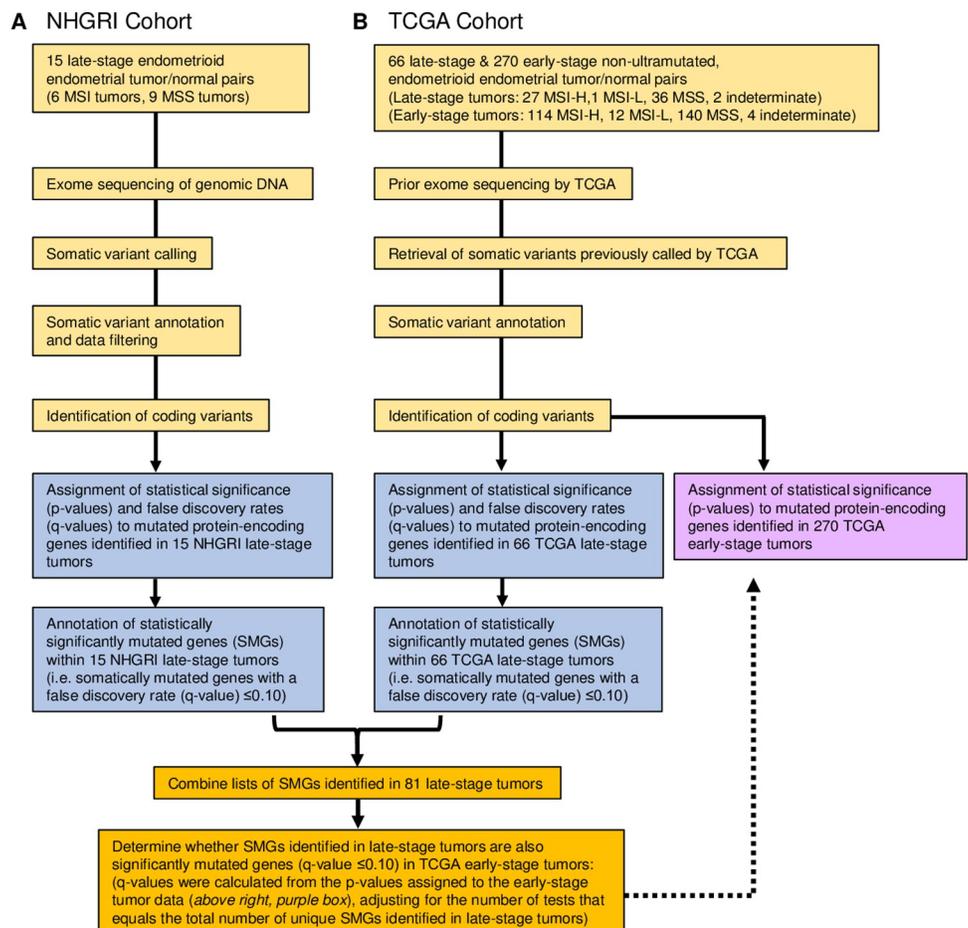
Given the dynamic nature of tumor genomes during disease initiation and progression, it is conceivable that the repertoire of pathogenic driver genes may differ in late-stage compared to early-stage EEC. However, the annotation of SMGs in primary EEC exomes by TCGA was performed in a stage-agnostic manner [2,16]. An improved understanding of the molecular etiology of late-stage EEC may provide novel insights into disease pathogenesis and progression. The aim of this study was to delineate SMGs in late-stage EEC exomes, and to determine whether these

genes are also significantly mutated in early-stage disease. To this end, we exome sequenced 15 “in-house” late-stage EECs (National Human Genome Research Institute (NHGRI) cohort) and reanalyzed somatic mutation calls from 66 late-stage and 270 early-stage non-ultramutated EECs within TCGA (Fig 1). Collectively, we identified 14 SMGs in 81 late-stage tumors. Krüppel-like factor 3 (*KLF3*) and Paired box 6 (*PAX6*), which encode transcription factors, were SMGs in late-stage tumors, but were not statistically significantly mutated in early-stage tumors. All *KLF3* mutations, and almost all *PAX6* mutations, were in the MSI-hypermutated EEC subgroup; within this subgroup, *KLF3* and *PAX6* mutations were more frequent in late-stage than early-stage tumors. The mutation spectrum of both genes included recurrent insertions-deletions (indels) at homopolymer tracts, consistent with strand slippage resulting from mismatch repair defects and suggesting that *PAX6* and *KLF3* are likely MSI target genes.

## Materials and methods

### Ethics statement

The NHGRI cohort of de-identified, fresh-frozen endometrioid endometrial tumors and matched non-tumor (normal) samples were obtained from the Cooperative Human Tissue



**Fig 1.** Flow diagram summarizing the approaches used in the step-wise analysis of somatic mutation data for (A) 15 late-stage endometrioid endometrial cancers (EECs) in the NHGRI tumor cohort, and for (B) 66 late-stage and 270 early-stage non-ultramutated EECs in The Cancer Genome Atlas (TCGA) tumor cohort.

<https://doi.org/10.1371/journal.pone.0251286.g001>

Network (CHTN). The National Institutes of Health Office of Human Subjects Research Protections determined that research using these specimens was exempt from IRB review. Because the specimens were obtained from CHTN as de-identified specimens with an agreement that we will never request re-identification, we do not have information on whether consent was written or oral.

### NHGRI clinical specimens

For 15 cases in the NHGRI cohort, de-identified, fresh-frozen endometrioid endometrial tumors and matched non-tumor (normal) samples were obtained from the Cooperative Human Tissue Network (CHTN) (**Table A in S1 Table**). The National Institutes of Health Office of Human Subjects Research Protections determined that this research was not human subject research, per the Common Rule (45 CFR 46). For each tumor sample, an H&E stained section was reviewed by an experienced gynecologic pathologist to identify regions containing  $\geq 70\%$  neoplastic cellularity; accompanying surgical pathology reports were retrospectively evaluated by the same gynecologic pathologist to annotate tumor stage using the International Federation of Gynecology and Obstetrics (FIGO) 2009 classification (**Table A in S1 Table**).

### Genomic DNA preparation and next-generation sequencing

Genomic DNA extraction, identity testing and MSI analysis of tumor and normal samples in the NHGRI cohort were performed as previously described [19]. DNA was purified by phenol-chloroform extraction prior to library preparation. DNA libraries were prepared using the SeqCap EZ Exome + UTR capture kit (Roche) and sequenced with the Illumina HiSeq 2000 platform (Illumina). A flow diagram summarizing the approaches and methods used to generate and analyze the NHGRI exomes is provided in **S1 Fig**.

### Alignment and variant calling

Short sequence reads from NHGRI cohort exomes were aligned to the Hg19 human reference sequence using NovoAlign version 2.08.02 (University of California at Santa Cruz). Four somatic mutation detection algorithms, Strelka [20], Shimmer [21], SomaticSniper [22], and MuTect [23], were used to call potential somatic variants. Insertions and deletions (indels) were identified by Shimmer and Strelka, while single nucleotide variants (SNVs) were identified by all four somatic algorithms. Strelka workflow version 1.0.14 (<https://doi.org/10.1093/bioinformatics/bts271>) was run with default parameters. Shimmer version 0.2 (<https://github.com/nhansen/shimmer>) was run with `--min_som_reads = 6` and `--minqual = 20` [21]. SomaticSniper version 1.0.5 was run with options `-Q 40 -G -L`, followed by the "standard somatic detection filters" described in Larsen et al [22]. MuTect version 1.1.5 was run with default parameters, and data were then filtered to include only calls designated as "KEEP" in the program's output [23]. Following analysis with each algorithm, a VarSifter-formatted file was generated containing the somatic variant allele frequencies observed in each tumor and matched normal sample for every called variant [24]. ANNOVAR (downloaded on August 12, 2014) was used to annotate all variants using the UCSC "known genes" gene structures [25].

### Variant filtering

Coding, splicing, and non-coding (intronic, 3' or 5' untranslated region (UTR), and 1kb upstream of the transcription start or downstream of the transcription end site) somatic variant calls in the NHGRI cohort were displayed using VarSifter [24]. We prioritized mutations for the NHGRI tumors using criteria similar to those that have been shown to yield accurate

mutation datasets in past studies [26–31]. A minimum of 14 reads covering a site in the tumor and 8 in the normal were required for mutation calling [26,27]; potential germline variants (those with a variant allele frequency (VAF) of greater than 3% in matched normal samples) were excluded. Coding and splice-site single nucleotide variants (SNVs) were annotated against dbSNP Build 135 and nonpathogenic single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) greater than 5% were excluded. Indel variants that were present in dbSNP Build 135 were excluded without further evaluation of MAF. SNVs called by all four algorithms and indels called by either Strelka or Shimmer were retained and further annotated against GENCODE hg19 using Oncotator (v1.5.3.0) (<http://www.broadinstitute.org/oncotator>) [32]; noncoding variants, those with a variant classification of UTR, Flank, lincRNA, RNA, Intron, or De novo start were excluded.

### TCGA data analysis

A subset of TCGA Uterine Corpus Endometrial Carcinoma (UCEC) somatic mutation data (TCGA UCEC PanCancer Atlas [16]) was extracted from the MC3 Public MAF file (mc3.v0.2.8.PUBLIC.maf.gz, <https://gdc.cancer.gov/about-data/publications/mc3-2017>) [33]. Briefly, the MC3 Public MAF file was filtered to include somatic variants from 336 EECs from the MSI-hypermutated ( $n = 141$ ), copy number-low/MSS ( $n = 140$ ) or copy number-high ( $n = 55$ ) molecular subgroups; variants from EECs within the ultramutated-POLE molecular subgroup or those without a molecular subgroup assignment were excluded (Table B in S1 Table). The TCGA mutation dataset used in our manuscript had been previously filtered to retain only the highest quality calls using both coverage and population frequency information [33]. Molecular subtype annotation for each sample was obtained from the cBioPortal for Cancer Genomics [34,35]. Variants with a PASS, WGA, or Native\_WGA\_mix designation as described by [33] were retained and further filtered to include SNVs called by MuTect and Indels called by Indelocator [16]. The final set of selected variants was annotated against GENCODE hg19 using Oncotator (v1.5.3.0) (<http://www.broadinstitute.org/oncotator>) [32]; noncoding variants, those with a variant classification of UTR, Flank, lincRNA, RNA, Intron, or De novo start were excluded. Additional clinicopathologic information for each tumor, including histology, stage, and grade, was obtained from Berger et al [16], and the cBioPortal for Cancer Genomics (URL: <https://www.cbioportal.org/>) [34,35] (Table B in S1 Table). Early-stage tumors were defined herein as stage I or II; late-stage tumors were defined as stage III or IV. A flow diagram summarizing the approaches and methods used to analyze the TCGA mutation calls is provided in S2 Fig.

### Power analysis

MutSigCV's statistical power to detect SMGs was estimated using the binomial model described in [36]. Briefly, the probability of obtaining a  $p$ -value  $\leq 0.1/14$  (for 14 tests) was calculated assuming a background mutation rate of  $p_0 = 1 - (1 - \mu f_g)^{3/4L}$ , where  $\mu$  is the background mutation rate, and  $f_g = 3.9$  and  $L = 1500$  are the 90<sup>th</sup> percentile gene-specific mutation rate factor and gene length, respectively. We also assumed a signal mutation rate of  $p_1 = p_0 + r(1 - m)$ , where  $r$  is the frequency of non-silent mutations in tumor samples and  $m = 0.1$  is the mis-detection rate. Power estimates were performed and plotted for a range of mutation rates and frequencies (S3 Fig) using an R script available at <https://github.com/nhansen/LateStageEECs>.

### Annotation of SMGs

SMGs were annotated using MutSigCV (v1.4). Briefly, MutSigCV (v1.4) was run on the NIH high-performance computing Biowulf cluster (<http://hpc.nih.gov>) using the coverage,

covariate, and mutation type dictionary files provided by the Broad Institute. Filtered somatic variants for each data set were annotated against GENCODE hg19 using Oncotator (<http://www.broadinstitute.org/oncotator>) [32], noncoding variants were excluded in accordance with a published approach [37], and the resulting coding mutation annotation format (maf) files were uploaded to the Biowulf cluster. Somatically mutated genes with a false discovery rate (q-value)  $\leq 0.10$  were defined as SMGs in accordance with a published approach [36].

### Determining whether late-stage SMGs are statistically significantly mutated in early-stage tumors

MutSigCV (v1.4) was run as described above on the set of filtered somatic variants from the 270 early-stage EECs to obtain p-values for all mutated genes. For all genes annotated as SMGs in late-stage tumors, q-values were re-calculated from the MutSigCV (v1.4) p-values assigned to the early-stage data, adjusting for 14 tests (reflecting the total number of SMGs identified in late-stage tumors).

### *In silico* prediction of functional consequences for somatic variants

MutationAssessor [38], PROVEAN (Protein Variation Effect Analyzer) [39], SIFT (Sorting Intolerant From Tolerant) [40], and PolyPhen-2 (Polymorphism Phenotyping v2) [41], were used to predict the effects of missense mutations on protein function. For each algorithm, the following descriptors were considered as impacting protein function: “high” (MutationAssessor), “deleterious” (PROVEAN), “damaging” (SIFT), and “probably-damaging” (PolyPhen-2). Agreement across at least three of the four prediction methods was required to assign an overall determination of “functional impact” for a missense mutation.

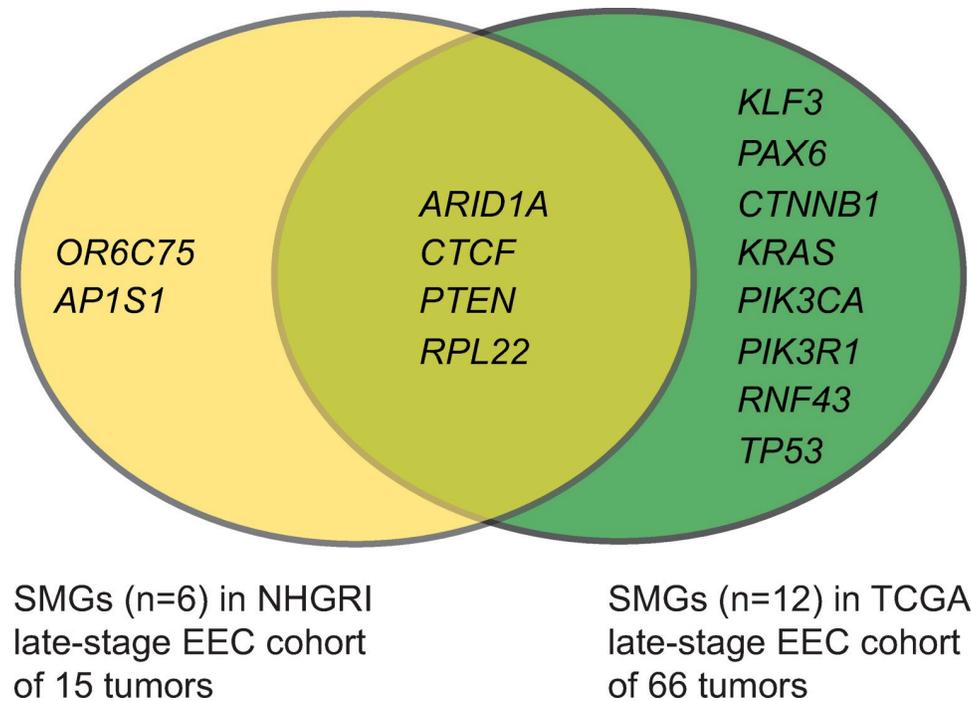
### Survival analyses

We utilized the cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) to query the relationship between SMG mutation status and survival (overall-, disease-free-, progression-free-, and disease-specific-survival) stratifying cases by stage (all stages, early-stage, late-stage) and molecular subgroup (MSI-hypermutated, CN-low, CN-high, all non-ultramutated), and applying a Bonferroni correction to account for multiple testing.

## Results

### Identification of SMGs among late-stage EECs

For the NHGRI late-stage cohort (n = 15), the average depth of coverage within regions targeted by the capture kit for tumor and normal samples was 67.2x and 65.5x, respectively; 90.87% of targeted bases for each tumor/normal pair had sufficient coverage for variant calling (Table C in S1 Table). Using a combination of somatic variant calling algorithms and stringent filtering parameters, we identified 2,879 high-confidence coding and splice-site somatic variants (consisting of 2,214 nonsynonymous (1,405 SNVs, 809 indels), 92 splice-site, and 573 synonymous variants) (Table D in S1 Table). Combined, the 2,306 nonsynonymous and splice-site variants affected 1,968 protein-coding genes and averaged 153.7 variants per tumor (range 9–542 per tumor) (Table D and Table E in S1 Table). For the TCGA late-stage cohort (n = 66), we extracted a total of 28,996 somatic coding and splice-site variants distributed among 10,504 protein-encoding genes (Table F and Table G in S1 Table). Using MutSigCV (v1.4), we identified a total of 14 unique late-stage SMGs (Fig 2), representing 6 SMGs (q-value  $\leq 0.1$ ) in the NHGRI (Table 1) and 12 SMGs in the TCGA late-stage EEC cohorts (Table 2).



**Fig 2. Statistically significantly mutated genes (SMGs) in late-stage EEC cohorts.** Venn diagram showing gene names for SMGs identified by MutSigCV (v1.4) analysis of 15 NHGRI late-stage EECs and 66 TCGA late-stage EECs. Late-stage tumors were defined as stage III or stage IV tumors. SMGs are defined as genes mutated in a tumor cohort at statistically significantly ( $q\text{-value} \leq 0.10$ ) higher rates than the background mutation rate.

<https://doi.org/10.1371/journal.pone.0251286.g002>

### ***KLF3* and *PAX6* are SMGs in late-stage but not early-stage EEC**

To test whether each of the 14 late-stage SMGs are also statistically significantly mutated in the TCGA early-stage EECs ( $n = 270$ ), we first estimated MutSigCV's power to detect genes as significantly mutated in the early-stage cohort. Estimating power using a binomial model as described in [42], we determined that the data from 270 tumors, when tested on 14 genes, yields >95% power to detect genes as significantly mutated across a wide range of background

**Table 1. SMGs ( $q \leq 0.10$ ) identified within the NHGRI cohort of 15 late-stage EEC exomes.**

Gene symbol	Gene name	Number and frequency (%) of NHGRI late-stage tumors with non-silent mutation(s)**	p-value	q-value
<i>PTEN</i>	Phosphatase and tensin homolog	13 (86.7%)	2.11E-15	3.98E-11
<i>ARID1A</i>	AT-rich interaction domain 1A	11 (73.3%)	1.27E-10	1.20E-06
<i>RPL22</i>	Ribosomal protein L22	4 (26.7%)	1.44E-07	9.07E-04
<i>OR6C75</i>	Olfactory receptor family 6 subfamily C member 75	4 (26.7%)	7.84E-06	3.70E-02
<i>CTCF</i>	CCCTC-binding factor	5 (33.3%)	1.26E-05	4.74E-02
<i>AP1S1</i>	Adaptor related protein complex 1 subunit sigma 1	3 (20.0%)	2.13E-05	6.70E-02

\*\* Non-silent mutations consist of nonsynonymous and splice junction mutations.

<https://doi.org/10.1371/journal.pone.0251286.t001>

Table 2. SMGs ( $q \leq 0.10$ ) identified among 66 late-stage TCGA EECs.

Gene symbol	Gene name	$\xi$ Number and frequency (%) of TCGA late-stage tumors with non-silent mutation(s)	p-value	q-value
<i>ARID1A</i>	AT-rich interaction domain 1A	29 (43.9%)	0	0
<i>PIK3R1</i>	Phosphoinositide-3-kinase regulatory subunit 1	21 (31.8%)	5.55E-16	5.24E-12
<i>PTEN</i>	Phosphatase and tensin homolog	46 (69.7%)	2.00E-15	1.26E-11
<i>PIK3CA</i>	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	25 (37.9%)	7.27E-14	3.43E-10
<i>TP53</i>	Tumor protein p53	19 (28.8%)	3.07E-13	1.16E-09
<i>KRAS</i>	KRAS proto-oncogene, GTPase	18 (27.3%)	7.22E-11	2.27E-07
<i>RPL22</i>	Ribosomal protein L22	6 (9.1%)	4.89E-08	1.08E-04
<i>CTCF</i>	CCCTC-binding factor	12 (18.2%)	4.99E-08	1.08E-04
<i>CTNNB1</i>	Catenin beta 1	18 (27.3%)	5.16E-08	1.08E-04
<i>PAX6</i>	Paired box 6	7 (10.6%)	1.04E-06	1.96E-03
<i>RNF43</i>	Ring finger protein 43	9 (13.6%)	4.61E-06	7.91E-03
<i>KLF3</i>	Kruppel like factor 3	7 (10.6%)	6.34E-06	9.96E-03

$\xi$  Data were extracted from previously published TCGA data.

\*\* Non-silent mutations consist of nonsynonymous and splice junction mutations.

<https://doi.org/10.1371/journal.pone.0251286.t002>

mutation rates when at least 10% of the 270 tumors are mutated in that gene (S3 Fig). Next, we obtained somatic variants for the cohort of non-ultramutated TCGA early-stage EECs; there were 162,763 somatic coding- and splice-site variants affecting 17,435 protein-encoding genes (Table H and Table I in S1 Table). To determine whether any of the 14 late-stage SMGs were significantly mutated in this dataset, p-values for all somatically mutated genes in early-stage tumors were calculated and used to determine q-values adjusting for 14 tests (reflecting the 14 late-stage SMGs queried) using the Benjamini-Hochberg procedure [43] (Table 3). Results showed that 12 of 14 late-stage SMGs were statistically significantly mutated (q-value < 0.1) in early-stage EECs whereas two late-stage SMGs, *KLF3* and *PAX6* were not (Table 3). Somatic mutations were more frequent among late-stage tumors than early-stage tumors for both *KLF3* (10.6% (7 of 66) late-stage vs 4.8% (13 of 270) early-stage) and *PAX6* (10.6% (7 of 66) late-stage vs 1.9% (5 of 270) early-stage) (Table 4).

We constructed Q-Q plots to verify that our q-values, calculated using the Benjamini-Hochberg procedure on MutSigCV's p-values, are the result of real statistical significance and not stratification of our dataset (S4 Fig). The Q-Q plots show significant deviation from ideal behavior due to MutSigCV's testing model [44], and the limited number of tumors analyzed.

### ***KLF3* and *PAX6* mutations occur in MSI-hypermuted EEC and are predicted to affect protein function**

For the TCGA cohorts, we evaluated the distribution of *KLF3* and *PAX6* mutations across the MSI-hypermuted (n = 141 cases), CN-low (n = 140 cases), and CN-high (n = 55 cases)

**Table 3. PAX6 and KLF3 are the only late-stage EEC SMGs (q-value  $\leq 0.1$ ) that are not statistically significantly mutated in early-stage EEC.**

Gene symbol for 14 SMGs (q-value $\leq 0.1$ ), identified in late-stage EEC cohorts	MutSigCV (1.4) p-value in early-stage EEC TCGA cohort	MutSigCV (1.4) q-value in early-stage EEC TCGA cohort corrected for 14 tests
<i>AP1S1</i>	0.00E+00	0.00E+00
<i>ARID1A</i>	0.00E+00	0.00E+00
<i>CTNNB1</i>	0.00E+00	0.00E+00
<i>PIK3CA</i>	0.00E+00	0.00E+00
<i>PTEN</i>	0.00E+00	0.00E+00
<i>RNF43</i>	0.00E+00	0.00E+00
<i>RPL22</i>	0.00E+00	0.00E+00
<i>TP53</i>	1.11E-15	1.94E-15
<i>CTCF</i>	2.78E-15	4.05E-15
<i>PIK3R1</i>	2.89E-15	4.05E-15
<i>KRAS</i>	6.88E-15	8.76E-15
<i>OR6C75</i>	1.89E-03	2.21E-03
<i>PAX6</i>	1.29E-01	1.39E-01
<i>KLF3</i>	9.99E-01	9.99E-01

<https://doi.org/10.1371/journal.pone.0251286.t003>

molecular subgroups (Table 4). *KLF3* mutations occurred exclusively in the MSI-hypermethylated subgroup at an overall frequency of 14.2% (20 of 141 cases), which was significantly higher than the occurrence of *KLF3* mutations among the combined CN-high and CN-low subgroups (0 of 195 cases) (p-value < 0.0001 2-tailed Fisher's exact test). Within the MSI-hypermethylated subgroup, *KLF3* was mutated in 25.9% (7 of 27) of late-stage tumors versus 11.4% (13 of 114) of early-stage tumors. There were no statistically significant differences in *KLF3* mutation frequency according to tumor grade; mutations were present in 14.2% of grade 1 (4 of 28), 8.1% of grade 2 (3 of 37), and 13.2% of grade 3 (11 of 83) MSI tumors (Table J in S1 Table).

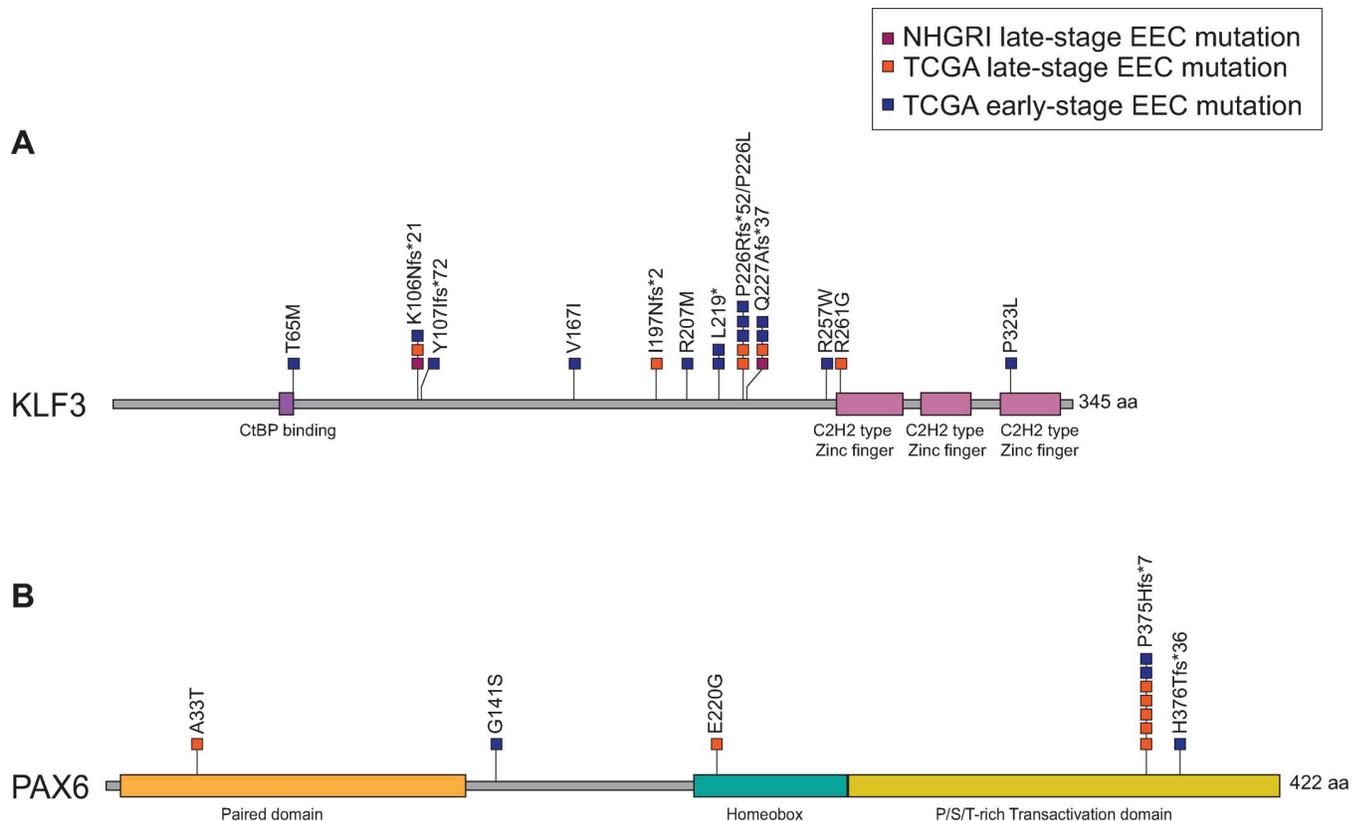
All but one (11 of 12) of *PAX6* mutations were in the MSI subgroup; the *PAX6*<sup>X306\_splice</sup> mutation was present in a CN-low tumor (Table 4). The higher frequency of *PAX6* mutations

**Table 4. Frequency of non-silent *KLF3* and *PAX6* mutations in non-ultramutated EECs, according to molecular subgroup.**

ξTumor stage and molecular subgroup	ξ <i>KLF3</i> mutation frequency	ξ <i>PAX6</i> mutation frequency
<b>All stages of EEC (n = 336)</b>	<b>5.9% (20 of 336)</b>	<b>3.6% (12 of 336)</b>
MSI subgroup (n = 141)	14.2% (20 of 141)	7.8% (11 of 141)
CN low subgroup (n = 140)	0% (0 of 140)	0.7% (1 of 140)
CN high subgroup (n = 55)	0% (0 of 55)	0% (0 of 55)
<b>Late-stage EECs (n = 66)</b>	<b>10.6% (7 of 66)</b>	<b>10.6% (7 of 66)</b>
MSI subgroup (n = 27)	25.9% (7 of 27)	25.9% (7 of 27)
CN low subgroup (n = 21)	0% (0 of 21)	0% (0 of 21)
CN high subgroup (n = 18)	0% (0 of 18)	0% (0 of 18)
<b>Early-stage EECs (n = 270)</b>	<b>4.8% (13 of 270)</b>	<b>1.9% (5 of 270)</b>
MSI subgroup (n = 114)	11.4% (13 of 114)	3.5% (4 of 114)
CN low subgroup (n = 119)	0% (0 of 119)	0.8% (1 of 119)
CN high subgroup (n = 37)	0% (0 of 37)	0% (0 of 37)

ξ Data were extracted from previously published TCGA data [16].

<https://doi.org/10.1371/journal.pone.0251286.t004>



**Fig 3. Spectrum of *KLF3* and *PAX6* somatic mutations in late-stage and early-stage non-ultramutated EECs.** Lollipop plots showing the positions of somatic mutations in (A) *KLF3* and (B) *PAX6* relative to protein domains. Mutations in late-stage EEC (orange and red) cohorts, and the TCGA early-stage (blue) EEC cohort are distinguished. Abbreviations used: aa (amino acids); CtBP (C-terminal Binding Protein); C2H2 (Cysteine-Cysteine-Histidine-Histidine); fs (frameshift); P/S/T (Proline/Serine/Threonine).

<https://doi.org/10.1371/journal.pone.0251286.g003>

in the MSI-hypermutated subgroup compared to other subgroups was statistically significant ( $p$ -value = 0.0004, 2-tailed Fisher's exact test). Within the MSI-hypermutated subgroup, *PAX6* was mutated in 7.8% (11 of 141) of tumors; mutations in late-stage tumors were more frequent compared to early-stage tumors (25.9% (7 of 27) versus 3.5% (4 of 114)). There was no significant difference in the frequency of *PAX6* mutations between tumors of differing grade; *PAX6* mutations were present in 3.6% of grade 1 (1 of 28), 13.5% of grade 2 (5 of 37) and 7.9% of grade 3 (6 of 76) MSI-hypermutated tumors (Table J in S1 Table). We observed no statistically significant differences in *KLF3* or *PAX6* mutation frequencies between *POLE/POLD1*-mutated and *POLE/POLD1*-wildtype cases within the MSI-hypermutated subgroup (Table K in S1 Table).

A majority of *KLF3* and *PAX6* mutations were indels within homopolymer tracts, resulting in frameshifts; the *KLF3*<sup>K106Nfs\*21</sup>, *KLF3*<sup>P226Rfs\*52</sup>, *KLF3*<sup>Q227Afs\*37</sup>, and *PAX6*<sup>P375Hfs\*7</sup> frameshift mutations were recurrent (Fig 3). Six of 21 (28.6%) *KLF3* mutations and 3 of 11 (27.3%) *PAX6* mutations were missense mutations; *KLF3*<sup>R257W</sup>, *KLF3*<sup>R261G</sup> and *PAX6*<sup>A33T</sup> were predicted to affect protein function by 3 of 4 *in silico* algorithms (Table L in S1 Table).

### Survival analysis

We utilized the cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) to query the relationship between patient survival and somatic mutation status of all 14 late-stage SMGs

identified herein, applying a Bonferroni correction to account for multiple testing (456 tests). With respect to *KLF3* and *PAX6* in the MSI-hypermutated subgroup, no significant differences in overall survival (OS), progression-free survival (PFS), disease-free survival (DFS) or disease-specific survival (DSS) were observed between mutated and non-mutated tumors when all stages were combined or when early- and late-stage tumors were considered separately (**Table M in S1 Table**). For the remaining 12 SMGs, there were no statistically significant differences in survival for any stage or molecular subgrouping (**Table N through Table V in S1 Table**).

## Discussion

The mutational landscape of EEC was reported by TCGA in an initial 2013 study and a subsequent “pan-gyn” study which included the 2013 EEC cohort and additional cases. Both studies performed *in silico* annotation of SMGs, which represent candidate driver genes, in a stage-agnostic manner. However, cancer genomes are dynamic and the mutational repertoire of tumors can evolve during progression and metastasis [45]. Recent comparisons of primary and metastatic endometrial cancer genomes have demonstrated divergence in their mutational landscapes [46–48]. But exome-wide comparisons of late-stage and early-stage primary tumors are lacking. Here, our stage-specific analysis of TCGA mutation data for non-ultramutated EECs showed that *KLF3* and *PAX6* are SMGs in late-stage (III/IV) but not early-stage (I/II) disease, raising the possibility that *KLF3* and *PAX6* mutations undergo positive selection during tumor progression.

*KLF3* encodes a zinc finger transcription factor with roles in adipogenesis, erythroid maturation, B-cell differentiation, and cardiovascular development (reviewed in [49]). In the Human Protein Atlas, *KLF3* expression was detected at “medium” levels in the normal in the glandular epithelium of the endometrium (<https://www.proteinatlas.org/ENSG00000109787-KLF3/tissue/endometrium>), by immunohistochemistry. The encoded protein includes an N-terminal CtBP-binding motif, three C-terminal Cys2His2 zinc finger domains, and a primary phosphorylation site at serine-249 that is important for DNA binding and enhancing transcriptional repression [49]. In our analysis of NHGRI EEC exomes and TCGA mutation data, the majority of *KLF3* mutations, including three mutation hotspots, were frameshift mutations that occur N-terminal to the zinc finger domains and to serine-249. Because frameshift mutations often generate a downstream premature stop codon, they may result in the production of a truncated protein or the transcript may be subjected to nonsense-mediated decay resulting in haploinsufficiency [50]. Based on the positional rules for nonsense-mediated decay [51], it is likely that the *KLF3* frameshift mutations among the EECs in this study result in nonsense-mediated decay and haploinsufficiency because the associated premature stop codons are located more than 50–55 nucleotides upstream of the final exon-exon junction [51]. In addition, *in silico* analyses predicted deleterious effects for the *KLF3*<sup>R257W</sup> and *KLF3*<sup>R261G</sup> missense mutants that occur in EEC; *KLF3*<sup>R257W</sup> also occurs somatically in 2 colorectal cancers (1 MSI-high/CIMP (CpG island methylator phenotype)-low; 1 CIN (chromosome instability)-subgroup) [52,53].

The fact that *KLF3* mutations in EEC occur predominantly at homopolymer tracts, were restricted to the MSI-hypermutated EEC subgroup, and are more frequently mutated in late-stage than early-stage MSI-hypermutated tumors (25.9% versus 11.4%, respectively), indicate that *KLF3* is an MSI target gene that may be involved in the etiology and progression of a subset of hypermutated EECs. Consistent with the idea that *KLF3* is an MSI target gene, frameshift mutations at codons 106 and 227, which are recurrent in MSI-EECs, are also recurrent in the colorectal MSI-colorectal and MSI-stomach TCGA molecular subgroups [35,54,55].

Studies in other tumor types have reported *KLF3* alterations as adverse prognosticators. For example, decreased *KLF3* expression in colorectal and cervical cancers is associated with lymph node positivity and poorer outcomes [56,57]. Conflicting data exist regarding the occurrence and effects of reduced *KLF3* levels in lung cancer. However, one study reported lower levels of *KLF3* mRNA and protein expression in lung adenocarcinomas compared with adjacent normal tissues and more frequent loss of *KLF3* expression in late- versus early-stage disease [58]. Although we found *KLF3* is a late-stage-specific SMG in EEC, there was no significant association between *KLF3* mutation status and survival for EEC patients, possibly reflecting tissue-specific differences in *KLF3* association with outcome, and/or outcome differences between mutation and reduced expression of *KLF3*.

The second late-stage-specific SMG identified in our study was *PAX6*. *PAX6* encodes a highly conserved paired box transcription factor that includes paired box and homeobox DNA-binding domains and a C-terminal transactivation domain (TAD); the final 40 residues of the TAD influence homeobox-DNA binding [59]. In the Human Protein Atlas, *PAX6* expression was undetectable by immunohistochemical analysis of the normal glandular epithelium of the endometrium (<https://www.proteinatlas.org/ENSG00000007372-PAX6/tissue/endometrium>). *PAX6* has important roles in the development of several tissue types, including the eye (reviewed in [60]). Inherited and *de novo* nonsense and frameshift mutations in *PAX6* cause the autosomal dominant eye disorder aniridia 1, whereas germline missense mutations are associated with attenuated ocular phenotypes [61]. Dysregulation of *PAX6* expression has been implicated in a variety of human cancers, resulting in tumor suppressive or oncogenic phenotypes depending on the cellular context [62–74]. A recent study reported a potential role for epigenetic silencing of *PAX6* in EC progression based on hypermethylation of *PAX6* in primary EC versus endometrial hyperplasia, and in metastatic EC versus primary EC [75]. Our analysis of TCGA mutation data found that *PAX6* mutations almost exclusively occur in MSI-hypermethylated tumors. This observation, coupled with the fact that *PAX6* mutations were more frequent among late-stage than early-stage MSI-hypermethylated tumors (25.9% versus 3.5%, respectively), raise the possibility that, like *KLF3* mutations, *PAX6* mutations may be pathogenic drivers of tumor progression in the context of MSI-hypermethylated EECs.

Most *PAX6* mutations in TCGA MSI-hypermethylated EECs were the recurrent *PAX6*<sup>P375Hfs\*7</sup> frameshift mutation in the transactivation domain [2,16]. We predict that *PAX6*<sup>P375Hfs\*7</sup> and an adjacent *PAX6*<sup>H376Tfs\*36</sup> frameshift mutation encode truncated proteins with reduced transactivation capacity, because the associated premature stop codons are located within 50 nucleotides of the penultimate exon-exon junction [51] and are located proximal to a synthetic nonsense mutation (*PAX6*<sup>Q422X</sup>) that exhibits reduced transactivation capacity *in vitro* [76]. Moreover, the fact that the *PAX6*<sup>P375Q</sup> aniridia-associated missense mutation results in attenuated DNA binding affinity *in vitro* [76], raises the possibility that the recurrent *PAX6*<sup>P375Hfs\*7</sup> mutant also may have attenuated DNA binding. Similar to *KLF3* frameshift mutations, the *PAX6*<sup>P375Hfs\*7</sup> and *PAX6*<sup>H376Tfs\*36</sup> frameshift mutations in EEC both arise within a (C)<sub>7</sub> homopolymer tract indicating that *PAX6* is an MSI target gene. Consistent with this idea is the fact that *PAX6* frameshift mutations originating at codon 375 and/or codon 376 are also recurrent in MSI-stomach cancer and MSI-colorectal carcinoma [34,35,77].

Compared to frameshift mutations, *PAX6* missense mutations are relatively rare in the non-ultramethylated TCGA cohort, occurring in three cases. The *PAX6*<sup>A33T</sup> EC-mutant occurs in the N-terminal paired box domain at a residue highly conserved across paired domains in Pax family members and other proteins and is predicted to impact function [78]. A different substitution at this residue (*PAX6*<sup>A33P</sup>) exhibits altered transactivation activity *in vitro* and is a germline variant associated with partial aniridia [78,79]. The other two *PAX6* missense mutations in EC (*PAX6*<sup>E220G</sup> and *PAX6*<sup>G141S</sup>) were not uniformly predicted to be functionally

significant in our analysis and, to our knowledge, are not pathogenic variants for ocular phenotypes.

In conclusion, our findings indicate that *KLF3* and *PAX6* are candidate driver genes in a subset of late-stage hypermutated EECs and are MSI target genes. Despite sufficient power, neither *KLF3* nor *PAX6* were detected as candidate driver genes in early-stage EECs. To our knowledge, this is the first study to annotate *KLF3* and *PAX6* as late stage-specific SMGs in EEC. Our findings warrant future studies to independently validate the enrichment of *PAX6* and *KLF3* mutations in late-stage, MSI-hypermutated EECs, to determine expression levels of *KLF3* and *PAX6* proteins in endometrial tumors, and to determine the functional effects of recurrent frameshift mutations in these genes particularly in regard to phenotypic properties associated with tumor progression.

## Supporting information

**S1 Fig. Flow diagram summarizing the approaches and methods used in the step-wise generation and analysis of NHGRI somatic mutation data for 15 late-stage endometrioid endometrial cancers (EECs) in the NHGRI tumor cohort.**

(PPTX)

**S2 Fig. Flow diagram summarizing the approaches and methods used in the step-wise analysis of TCGA somatic mutation calls for 336 non-ultramutated endometrioid endometrial cancers (EECs) within the Uterine Corpus Endometrial Carcinoma (UCEC) cohort.**

(PPTX)

**S3 Fig. Power to detect significantly mutated genes (SMGs) in early-stage tumors.** Curves show statistical power for different percentages ( $r$ ) of tumors that are somatically mutated. Calculations were performed as described in the text, assuming 270 tumors and 14 gene tests completed.

(PPTX)

**S4 Fig.** Q-Q plots for MutSigCV's p-values for differing mutation rates in background genes in (A) the 15 late-stage tumors sequenced and analyzed at NHGRI and (B) the set of 66 late-stage tumors from the TCGA project. Deviation from uniform p-value distribution here is a result of MutSigCV's assigned p-value and probably due to the limited number of tumors analyzed.

(PPTX)

**S1 Table. Contains supporting Tables A through V.**

(XLSX)

## Acknowledgments

This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). ^Membership of the National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program is as follows: Beatrice B. Barnabas<sup>1</sup>, Sean Black<sup>1</sup>, Gerard G. Bouffard<sup>1</sup>, Shelise Y. Brooks<sup>1</sup>, Holly Coleman<sup>1</sup>, Lyudmila Dekhtyar<sup>1</sup>, Joel Han<sup>1</sup>, Shi-ling Ho<sup>1</sup>, Juyun Kim<sup>1</sup>, Richelle Legaspi<sup>1</sup>, Quino L. Maduro<sup>1</sup>, Catherine A. Masiello<sup>1</sup>, Jennifer C. McDowell<sup>1</sup>, Casandra Montemayor<sup>1</sup>, James C. Mullikin<sup>1</sup>, Morgan Park<sup>1</sup>, Nancy L. Riebow<sup>1</sup>, Karen Schandler<sup>1</sup>, Brian Schmidt<sup>1</sup>, Christina Sison<sup>1</sup>, Sirintorn Stantripop<sup>1</sup>, James W. Thomas<sup>1</sup>, Pamela J. Thomas<sup>1</sup>, Meghana Vemulapalli<sup>1</sup>, and Alice C. Young<sup>1</sup>. The lead author for this group is Dr. James C. Mullikin (Email: [mullikin@mail.nih.gov](mailto:mullikin@mail.nih.gov)).

<sup>1</sup> NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland, United States of America.

## Author Contributions

**Conceptualization:** Meghan L. Rudd, Nancy F. Hansen, Mary Ellen Urick, Lawrence C. Brody, Daphne W. Bell.

**Data curation:** Nancy F. Hansen.

**Formal analysis:** Meghan L. Rudd, Nancy F. Hansen, Xiaolu Zhang, Suiyuan Zhang, Daphne W. Bell.

**Funding acquisition:** Daphne W. Bell.

**Investigation:** Meghan L. Rudd, Nancy F. Hansen, Xiaolu Zhang, Suiyuan Zhang, Daphne W. Bell.

**Methodology:** Meghan L. Rudd, Nancy F. Hansen, Xiaolu Zhang, Suiyuan Zhang, Lawrence C. Brody.

**Project administration:** Meghan L. Rudd, Nancy F. Hansen, Daphne W. Bell.

**Resources:** Nancy F. Hansen.

**Supervision:** James C. Mullikin, Daphne W. Bell.

**Validation:** Meghan L. Rudd, Nancy F. Hansen, Xiaolu Zhang, Suiyuan Zhang, Maria J. Merino.

**Visualization:** Meghan L. Rudd, Nancy F. Hansen, Daphne W. Bell.

**Writing – original draft:** Meghan L. Rudd, Daphne W. Bell.

**Writing – review & editing:** Meghan L. Rudd, Nancy F. Hansen, Xiaolu Zhang, Mary Ellen Urick, Suiyuan Zhang, Maria J. Merino, James C. Mullikin, Lawrence C. Brody, Daphne W. Bell.

## References

1. Eshleman JR, Markowitz S. Microsatellite instability in inherited and sporadic neoplasms. *Curr Opin Oncol*. 1995; 7(1):83–9. PMID: [7696368](https://pubmed.ncbi.nlm.nih.gov/7696368/)
2. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013; 497(7447):67–73. <https://doi.org/10.1038/nature12113> PMID: [23636398](https://pubmed.ncbi.nlm.nih.gov/23636398/)
3. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, Kovatich AJ, et al. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum Mol Genet*. 1999; 8(4):661–6. <https://doi.org/10.1093/hmg/8.4.661> PMID: [10072435](https://pubmed.ncbi.nlm.nih.gov/10072435/)
4. Esteller M, Levine R, Baylin SB, Ellenson LH, Herman JG. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene*. 1998; 17(18):2413–7. <https://doi.org/10.1038/sj.onc.1202178> PMID: [9811473](https://pubmed.ncbi.nlm.nih.gov/9811473/)
5. Urick ME, Bell DW. Clinical actionability of molecular targets in endometrial cancer. *Nat Rev Cancer*. 2019; 19(9):510–21. <https://doi.org/10.1038/s41568-019-0177-x> PMID: [31388127](https://pubmed.ncbi.nlm.nih.gov/31388127/)
6. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018; 68(6):394–424. <https://doi.org/10.3322/caac.21492> PMID: [30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)
7. American Cancer Society. Cancer facts & figures 2020. 2020:1–70.
8. Lortet-Tieulent J, Ferlay J, Bray F, Jemal A. International patterns and trends in endometrial cancer incidence, 1978–2013. *J Natl Cancer Inst*. 2018; 110(4):354–61. <https://doi.org/10.1093/jnci/djx214> PMID: [29045681](https://pubmed.ncbi.nlm.nih.gov/29045681/)

9. Smrz SA, Calo C, Fisher JL, Salani R. An ecological evaluation of the increasing incidence of endometrial cancer and the obesity epidemic. *Am J Obstet Gynecol*. 2020. <https://doi.org/10.1016/j.ajog.2020.10.042> PMID: 33127429
10. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, et al. Carcinoma of the corpus uteri. FIGO 26th annual report on the results of treatment in gynecological cancer. *Int J Gynaecol Obstet*. 2006; 95 Suppl 1:S105–43. [https://doi.org/10.1016/S0020-7292\(06\)60031-3](https://doi.org/10.1016/S0020-7292(06)60031-3) PMID: 17161155
11. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020; 70(1):7–30. <https://doi.org/10.3322/caac.21590> PMID: 31912902
12. Gadducci A, Cosio S, Genazzani AR. Old and new perspectives in the pharmacological treatment of advanced or recurrent endometrial cancer: Hormonal therapy, chemotherapy and molecularly targeted therapies. *Crit Rev Oncol Hematol*. 2006; 58(3):242–56. <https://doi.org/10.1016/j.critrevonc.2005.11.002> PMID: 16436330
13. Bakkum-Gamez JN, Gonzalez-Bosquet J, Laack NN, Mariani A, Dowdy SC. Current issues in the management of endometrial cancer. *Mayo Clin Proc*. 2008; 83(1):97–112. <https://doi.org/10.4065/83.1.97> PMID: 18174012
14. Fleming GF. Systemic chemotherapy for uterine carcinoma: Metastatic and adjuvant. *J Clin Oncol*. 2007; 25(20):2983–90. <https://doi.org/10.1200/JCO.2007.10.8431> PMID: 17617530
15. Singh N, Hirschowitz L, Zaino R, Alvarado-Cabrero I, Duggan MA, Ali-Fehmi R, et al. Pathologic prognostic factors in endometrial carcinoma (other than tumor type and grade). *Int J Gynecol Pathol*. 2019; 38 Suppl 1:S93–S113. <https://doi.org/10.1097/PGP.0000000000000524> PMID: 30550486
16. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell*. 2018; 33(4):690–705 e9. <https://doi.org/10.1016/j.ccell.2018.03.014> PMID: 29622464
17. Byron SA, Gartside M, Powell MA, Wellens CL, Gao F, Mutch DG, et al. Fgfr2 point mutations in 466 endometrioid endometrial tumors: Relationship with MSI, KRAS, PIK3CA, CTNNB1 mutations and clinicopathological features. *PLoS One*. 2012; 7(2):e30801. <https://doi.org/10.1371/journal.pone.0030801> PMID: 22383975
18. McMeekin DS, Trichter DL, Cohn DE, Mutch DG, Lankes HA, Geller MA, et al. Clinicopathologic significance of mismatch repair defects in endometrial cancer: An NRG oncology/gynecologic oncology group study. *J Clin Oncol*. 2016; 34(25):3062–8. <https://doi.org/10.1200/JCO.2016.67.8722> PMID: 27325856
19. Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, et al. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat Genet*. 2012; 44(12):1310–5. <https://doi.org/10.1038/ng.2455> PMID: 23104009
20. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK, Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012; 28(14):1811–7. <https://doi.org/10.1093/bioinformatics/bts271> PMID: 22581179
21. Hansen NF, Gartner JJ, Mei L, Samuels Y, Mullikin JC, Shimmer: Detection of genetic alterations in tumors using next-generation sequence data. *Bioinformatics*. 2013; 29(12):1498–503. <https://doi.org/10.1093/bioinformatics/btt183> PMID: 23620360
22. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012; 28(3):311–7. <https://doi.org/10.1093/bioinformatics/btr665> PMID: 22155872
23. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013; 31(3):213–9. <https://doi.org/10.1038/nbt.2514> PMID: 23396013
24. Teer JK, Green ED, Mullikin JC, Biesecker LG. Varsifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics*. 2012; 28(4):599–600. <https://doi.org/10.1093/bioinformatics/btr711> PMID: 22210868
25. Wang K, Li M, Hakonarson H. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38(16):e164. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685
26. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012; 44(6):685–9. <https://doi.org/10.1038/ng.2279> PMID: 22610119
27. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun*. 2015; 6:6744. <https://doi.org/10.1038/ncomms7744> PMID: 25855536

28. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet.* 2013; 45(5):478–86. <https://doi.org/10.1038/ng.2591> PMID: 23525077
29. Li M, Zhang Z, Li X, Ye J, Wu X, Tan Z, et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. *Nat Genet.* 46(8):872–6. <https://doi.org/10.1038/ng.3030> PMID: 24997986
30. Ho AS, Kannan K, Roy DM, G. ML, Ganly I, Katabi N, et al. The mutational landscape of adenoid cystic carcinoma. *Nat Genet.* 2013; 45(7):791–8. <https://doi.org/10.1038/ng.2643> PMID: 23685749
31. Dalin MG, Desrichard A, Katabi N, Makarov V, Walsh LA, Lee KW, et al. Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast cancer. *Clin Cancer Res.* 2016; 22(18):4623–33. <https://doi.org/10.1158/1078-0432.CCR-16-0637> PMID: 27103403
32. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: Cancer variant annotation tool. *Hum Mutat.* 2015; 36(4):E2423–9. <https://doi.org/10.1002/humu.22771> PMID: 25703262
33. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Systems.* 2018; 6(3):271–81.e7. <https://doi.org/10.1016/j.cels.2018.03.002> PMID: 29596782
34. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2(5):401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
35. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6(269):p11. <https://doi.org/10.1126/scisignal.2004088> PMID: 23550210
36. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013; 499(7457):214–8. <https://doi.org/10.1038/nature12213> PMID: 23770567
37. Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. *Proc Natl Acad Sci U S A.* 2016; 113(50):14330–5. <https://doi.org/10.1073/pnas.1616440113> PMID: 27911828
38. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* 2011; 39(17):e118. <https://doi.org/10.1093/nar/gkr407> PMID: 21727090
39. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012; 7(10):e46688. <https://doi.org/10.1371/journal.pone.0046688> PMID: 23056405
40. Ng PC, Henikoff S. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003; 31(13):3812–4. <https://doi.org/10.1093/nar/gkg509> PMID: 12824425
41. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4):248–9. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
42. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014; 505(7484):495–501. <https://doi.org/10.1038/nature12912> PMID: 24390350
43. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc.* 1995; 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
44. Jiang L, Zheng J, Kwan JHS, Dai S, Li C, Li MJ, et al. Witer: A powerful method for estimation of cancer-driver genes using a weighted iterative regression modelling background mutation counts. *Nucleic Acids Res.* 2019; 47(16):e96. <https://doi.org/10.1093/nar/gkz566> PMID: 31287869
45. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature.* 2009; 458(7239):719–24. <https://doi.org/10.1038/nature07943> PMID: 19360079
46. Gibson WJ, Hoivik EA, Halle MK, Taylor-Weiner A, Cherniack AD, Berg A, et al. The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat Genet.* 2016; 48(8):848–55. <https://doi.org/10.1038/ng.3602> PMID: 27348297
47. Soumerai TE, Donoghue MTA, Bandlamudi C, Srinivasan P, Chang MT, Zamarin D, et al. Clinical utility of prospective molecular characterization in advanced endometrial cancer. *Clin Cancer Res.* 2018; 24(23):5939–47. <https://doi.org/10.1158/1078-0432.CCR-18-0412> PMID: 30068706
48. Ashley CW, Da Cruz Paula A, Kumar R, Mandelker D, Pei X, Riaz N, et al. Analysis of mutational signatures in primary and metastatic endometrial cancer reveals distinct patterns of DNA repair defects and

- shifts during tumor progression. *Gynecol Oncol.* 2019; 152(1):11–9. <https://doi.org/10.1016/j.ygyno.2018.10.032> PMID: 30415991
49. Pearson RC, Funnell AP, Crossley M. The mammalian zinc finger transcription factor kruppel-like factor 3 (KLF3/BKLF). *IUBMB Life.* 2011; 63(2):86–93. <https://doi.org/10.1002/iub.422> PMID: 21360637
  50. Lykke-Andersen S, Jensen TH. Nonsense-mediated mrna decay: An intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol.* 2015; 16(11):665–77. <https://doi.org/10.1038/nrm4063> PMID: 26397022
  51. Coban-Akdemir Z, White JJ, Song X, Jhangiani SN, Fatih JM, Gambin T, et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am J Hum Genet.* 2018; 103(2):171–87. <https://doi.org/10.1016/j.ajhg.2018.06.009> PMID: 30032986
  52. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep.* 2016; 15(4):857–65. <https://doi.org/10.1016/j.celrep.2016.03.075> PMID: 27149842
  53. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012; 487(7407):330–7. <https://doi.org/10.1038/nature11252> PMID: 22810696
  54. Chakravarty D, Srinivasan R, Ghosh S, Gopalan S, Rajwanshi A, Majumdar S. Estrogen receptor beta1 and the beta2/betacx isoforms in nonneoplastic endometrium and in endometrioid carcinoma. *Int J Gynecol Cancer.* 2007; 17(4):905–13. <https://doi.org/10.1111/j.1525-1438.2006.00851.x> PMID: 17309672
  55. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell.* 2018; 173(2):291–304 e6. <https://doi.org/10.1016/j.cell.2018.03.022> PMID: 29625048
  56. Lyng H, Brovig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, et al. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. *BMC Genomics.* 2006; 7:268. <https://doi.org/10.1186/1471-2164-7-268> PMID: 17054779
  57. Wang X, Jiang Z, Zhang Y, Wang X, Liu L, Fan Z. RNA sequencing analysis reveals protective role of kruppel-like factor 3 in colorectal cancer. *Oncotarget.* 2017; 8(13):21984–93. <https://doi.org/10.18632/oncotarget.15766> PMID: 28423541
  58. Sun W, Hu S, Zu Y, Deng Y. KLF3 is a crucial regulator of metastasis by controlling STAT3 expression in lung cancer. *Mol Carcinog.* 2019; 58(11):1933–45. <https://doi.org/10.1002/mc.23072> PMID: 31486564
  59. Shukla S, Mishra R. Predictions on impact of missense mutations on structure function relationship of PAX6 and its alternatively spliced isoform PAX6(5a). *Interdiscip Sci.* 2012; 4(1):54–73. <https://doi.org/10.1007/s12539-012-0114-0> PMID: 22392277
  60. Cvekl A, Callaerts P. PAX6: 25th anniversary and more to learn. *Exp Eye Res.* 2017; 156:10–21. <https://doi.org/10.1016/j.exer.2016.04.017> PMID: 27126352
  61. Lima Cunha D, Arno G, Corton M, Moosajee M. The spectrum of PAX6 mutations and genotype-phenotype correlations in the eye. *Genes (Basel).* 2019; 10(12):1050. <https://doi.org/10.3390/genes10121050> PMID: 31861090
  62. Li Y, Li Y, Liu Y, Xie P, Li F, Li G. Pax6, a novel target of microRNA-7, promotes cellular proliferation and invasion in human colorectal cancer cells. *Dig Dis Sci.* 2014; 59(3):598–606. <https://doi.org/10.1007/s10620-013-2929-x> PMID: 24185687
  63. Zong X, Yang H, Yu Y, Zou D, Ling Z, He X, et al. Possible role of PAX-6 in promoting breast cancer cell proliferation and tumorigenesis. *BMB Rep.* 2011; 44(9):595–600. <https://doi.org/10.5483/bmbrep.2011.44.9.595> PMID: 21944253
  64. Kiselev Y, Andersen S, Johannessen C, Fjukstad B, Standahl Olsen K, Stenvold H, et al. Transcription factor PAX6 as a novel prognostic factor and putative tumour suppressor in non-small cell lung cancer. *Sci Rep.* 2018; 8(1):5059. <https://doi.org/10.1038/s41598-018-23417-z> PMID: 29568088
  65. Mayes DA, Hu Y, Teng Y, Siegel E, Wu X, Panda K, et al. PAX6 suppresses the invasiveness of glioblastoma cells and the expression of the matrix metalloproteinase-2 gene. *Cancer Res.* 2006; 66(20):9809–17. <https://doi.org/10.1158/0008-5472.CAN-05-3877> PMID: 17047041
  66. Shyr CR, Tsai MY, Yeh S, Kang HY, Chang YC, Wong PL, et al. Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. *Prostate.* 2010; 70(2):190–9. <https://doi.org/10.1002/pros.21052> PMID: 19790232
  67. Zhou YH, Wu X, Tan F, Shi YX, Glass T, Liu TJ, et al. PAX6 suppresses growth of human glioblastoma cells. *J Neurooncol.* 2005; 71(3):223–9. <https://doi.org/10.1007/s11060-004-1720-4> PMID: 15735909
  68. Hegge B, Sjøttem E, Mikkola I. Generation of a PAX6 knockout glioblastoma cell line with changes in cell cycle distribution and sensitivity to oxidative stress. *BMC Cancer.* 2018; 18(1):496. <https://doi.org/10.1186/s12885-018-4394-6> PMID: 29716531

69. Huang BS, Luo QZ, Han Y, Li XB, Cao LJ, Wu LX. MicroRNA-223 promotes the growth and invasion of glioblastoma cells by targeting tumor suppressor PAX6. *Oncol Rep*. 2013; 30(5):2263–9. <https://doi.org/10.3892/or.2013.2683> PMID: 23970099
70. Maulbecker CC, Gruss P. The oncogenic potential of PAX genes. *EMBO J*. 1993; 12(6):2361–7. <https://doi.org/10.1002/j.1460-2075.1993.tb05890.x> PMID: 8099544
71. Ooki A, Dinalankara W, Marchionni L, Tsay JJ, Goparaju C, Maleki Z, et al. Epigenetically regulated PAX6 drives cancer cells toward a stem-like state via GLI-SOX2 signaling axis in lung adenocarcinoma. *Oncogene*. 2018; 37(45):5967–81. <https://doi.org/10.1038/s41388-018-0373-2> PMID: 29980786
72. Wu DM, Zhang T, Liu YB, Deng SH, Han R, Liu T, et al. The PAX6-ZEB2 axis promotes metastasis and cisplatin resistance in non-small cell lung cancer through PI3K/AKT signaling. *Cell Death Dis*. 2019; 10(5):349. <https://doi.org/10.1038/s41419-019-1591-4> PMID: 31024010
73. Jin M, Gao D, Wang R, Sik A, Liu K. Possible involvement of TGF- $\beta$ -SMAD epithelial-mesenchymal transition in pro-metastatic property of PAX6. *Oncol Rep*. 2020; 44(2):555–64. <https://doi.org/10.3892/or.2020.7644> PMID: 32627030
74. Urrutia G, Laurito S, Campoy E, Nasif D, Branham MT, Roque M. PAX6 promoter methylation correlates with MDA-MB-231 cell migration, and expression of MMP2 and MMP9. *Asian Pac J Cancer Prev* 2018; 19(10):2859–66. <https://doi.org/10.22034/APJCP.2018.19.10.2859> PMID: 30362313
75. Wu X, Miao J, Jiang J, Liu F. Analysis of methylation profiling data of hyperplasia and primary and metastatic endometrial cancers. *Eur J Obstet Gynecol Reprod Biol*. 2017; 217:161–6. <https://doi.org/10.1016/j.ejogrb.2017.08.036> PMID: 28910750
76. Singh S, Chao LY, Mishra R, Davies J, Saunders GF. Missense mutation at the c-terminus of PAX6 negatively modulates homeodomain function. *Hum Mol Genet*. 2001; 10(9):911–8. <https://doi.org/10.1093/hmg/10.9.911> PMID: 11309364
77. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell*. 2018; 33(4):721–35 e8. <https://doi.org/10.1016/j.ccell.2018.03.010> PMID: 29622466
78. Hanson I, Churchill A, Love J, Axton R, Moore T, Clarke M, et al. Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations. *Hum Mol Genet*. 1999; 8(2):165–72. <https://doi.org/10.1093/hmg/8.2.165> PMID: 9931324
79. Chauhan BK, Yang Y, Cveklova K, Cvekl A. Functional properties of natural human PAX6 and PAX6 (5a) mutants. *Invest Ophthalmol Vis Sci*. 2004; 45(2):385–92. <https://doi.org/10.1167/iovs.03-0968> PMID: 14744876