medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Characterization of a base-resolution common deletion region within *CDKN2A* gene in cancers and its biological and clinical significances

Running title: Somatic *CDKN2A* deletion in epithelial dysplasia

Juanli Qiao, Ph.D.[1#], Rui Xing, Ph.D.[2#], Zhiyuan Fan, M.D.[3#], Jing Zhou, Ms.[1], Yuan Tian, Ph.D., M.D.[1], Yu Qin, Ph.D.[3], Zhaojun Liu, Ph.D.[1], Liankun Gu, Ms.[1], Jiafu Ji, Ph.D., M.D.[4], Youyong Lu, M.D., M.S.[2], Sanford M. Dawsey, Ph.D.[5], Wenqiang Wei, Ph.D.[3]*, Dajun Deng, M.D., M.S.[1]*

[1] Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Etiology, Peking University Cancer Hospital and Institute, Beijing, 100142, China;

[2] Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Tumor Biology, Peking University Cancer Hospital and Institute, Beijing, 100142, China;

[3] National Cancer Center, Cancer Hospital, Chinese Academy of Medical Science, Beijing, 100021, China;

[4] Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Surgical Oncology, Peking University Cancer Hospital and Institute, Beijing, 100142, China

[5] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, 20892, USA.

Equal first authors

* Corresponding authors

Corresponding authors: Prof. Wenqiang Wei, National Cancer Center, Chinese Academy of Medical Science, Beijing, 100021, China (E-mail: weiwq@cicams.ac.cn); Dajun Deng, Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Etiology, Peking University Cancer Hospital and Institute, Fu-Cheng-Lu #52, Haidian District, Beijing, 100142, China, Phone:+8610-88196752 (E-mail: dengdajun@bjmu.edu.cn)

ABSTRACT

Background: Frequency of somatic copy number deletion of *CDKN2A* gene is upto 60% in human esophageal squamous cell cancer. However, it is unknown whether *CDKN2A* deletion could be a biomarker for esophageal squamous cell dysplasia (ESCdys) due to absence of a feasible detection method.

Methods: Information on base-resolution common deletion region (CDR) for *CDKN2A* were extracted from published articles and confirmed with whole genome sequencing (WGS). A quantitative PCR targeted to the CDR (P16-Light) was established and used to detect *CDKN2A* copy number in ESCdys biopsies from patients (*n*=205) enrolled in a multicentre follow-up study.

Results: A 5.1-kb CDR from the *CDKN2A/P16*^{INK4A} promoter to intron-2 was firstly characterized in 90% (83/92) of cancer cell lines and confirmed with WGS. The CDR covers *CDKN2A* exon-2 which is the essential coding exon for both P16^{INK4a} and P14^{ARF}. And *CDKN2A* exon-2 deletion markedly promoted the proliferation and invasion and inhibited the apoptosis of HEK293T cells. In the follow-up study, both somatic *CDKN2A* deletion and amplification are prevalent in mild/moderate (m/M) ESCdys. *CDKN2A* deletion was less common among 70 patients whose ESCdys regressed than among 135 patients whose ESCdys progressed or remained stable, and *CDKN2A* amplification was more common in the patients who regressed than in the patients whose m/M ESCdys persisted or progressed over a median of 37 months of follow-up (p<0.0001).

Conclusion: There is A 1.5-kb CDR within *CDKN2A* gene in many cancers. CDR deletion could inactivate both P16^{INK4a} and P14^{ARF} and associate wtih prognosis of ESCdys.

Key words: *CDKN2A*; common deleted region (CDR); esophageal squamous cell dysplasia (ESCdys); prospective study

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

INTRODUCTION

Somatic *CDKN2A* copy number deletion is a landmark of human cancer (Beroukhim et al. 2010). The frequency of *CDKN2A* deletion detected by single nucleotide polymorphism (SNP) microarray or whole genome sequencing (WGS) was found to be 30% to 60% in bladder cancer, melanoma, pleural mesothelioma, head and neck cancer, glioblastoma, and esophageal squamous cell cancer (ESCC), with an average frequency of 13% (1384/10967) in pan-cancer datasets in The Cancer Genome Atlas (TCGA) (Supplemental Figure S1A) (Mermel et al. 2011; Cerami et al. 2012; Gao et al. 2013; Song et al. 2014; The Cancer Genome Atlas Research Network 2017; Cui et al. 2020). *CDKN2A* deep-deletion is associated with downregulation of *CDKN2A* gene expression, while *CDKN2A* amplification is associated with upregulation of *CDKN2A* gene expression in Pan-TCGA cancers (Supplemental Figure S1B).

It is well known that genetic *CDKN2A* inactivation contributes to malignant transformation, cancer metastasis, and therapy sensitivity of cancers to drugs, including CDK4/6 inhibitors and their combination with PD-1 blockades (Deng et al. 2018; Jerby-Arnon et al. 2018; Zhang et al. 2018; Yu et al. 2019). However, current gene copy number detection methods, including fluorescence-in-situ hybridization (FISH) and WGS, are not sensitive enough or are too costly for routine clinical use. While the amplification of oncogenes (such as *EGFR*, c-*ERBB2,* c-*MYC*, and c-*MET*) are increasingly driving decision-making for precise cancer treatments, clinical applications of somatic copy deletions of tumor suppressor genes, including *CDKN2A*, remain rare because of the lack of a feasible detection assay (Patel et al. 2014).

RESULTS

Characterization of a *CDKN2A* common deletion region (CDR) in human cancers

It has been previously reported that a homozygous deletion of approximately 170 kilobase pairs (kb), including the *CDKN2A* locus, can be detected in human cancers by microsatellite analyses (Cairns et al. 1995). To characterize the base-resolution genomic coordinates of *CDKN2A* deletions in cancers, we extracted sequence information of interstitial *CDKN2A* deletions from available published articles (Supplemental Table S1). We found a 5.1-kb CDR (chr9: 21,970,277 - 21,975,386, hg19) that spanned from the $P16^{INK4a}$ promoter to intron-2 in 83 (90%) of 92 reported cancer cell lines or tissue samples containing interstitial *CDKN2A* deletions (Figure 1). This CDR sequence is exactly the same as the *CDKN2A* deletion fragment in the HCC193 lung cancer cell line (Sasaki et al. 2003). The CDR coordinates were also confirmed in our WGS datasets (sequencing depth, 36×) of 18/18 gastric carcinomas (GC) (Xing et al. 2019), in which interstitial *CDKN2A* deletions were identified (Figure 1; Supplemental Table S1).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

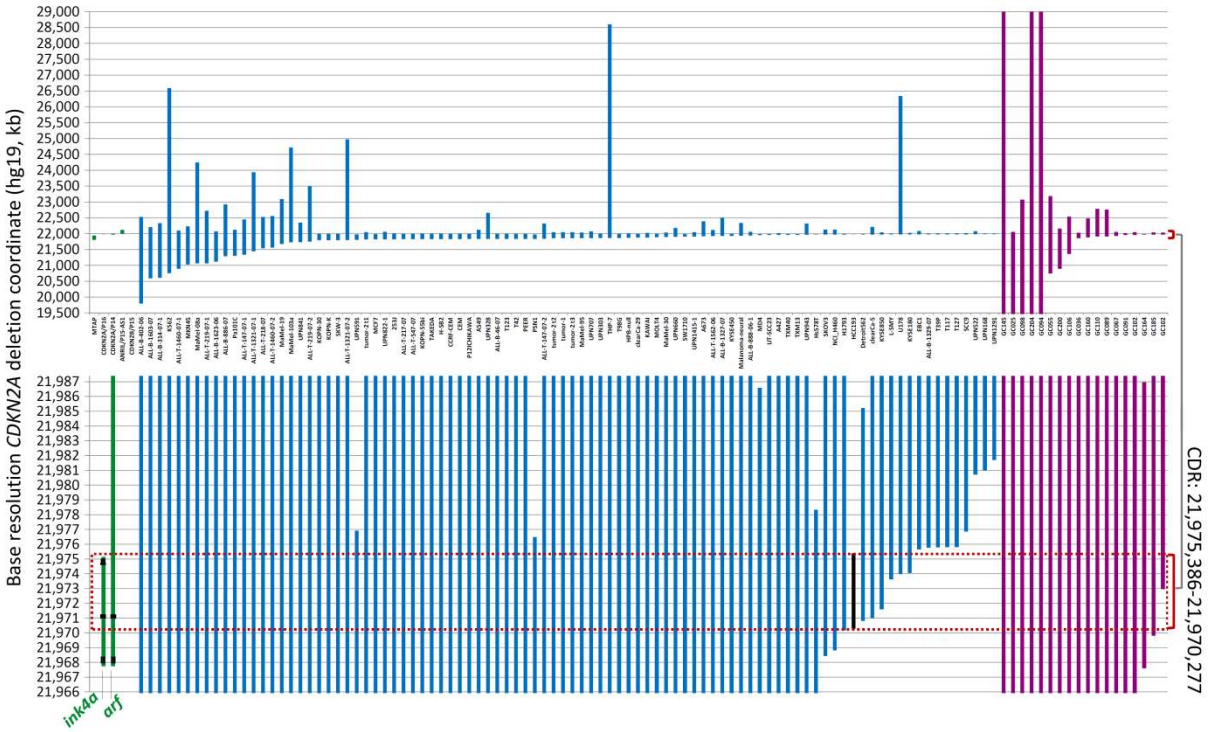

Figure 1. Base resolution of the genomic coordinates of *CDKN2A* deletion in human cancers. The top chart displays the coordinates of whole deleted fragments by Sanger sequencing of chromosome 9p21 in *CDKN2A*-deleted cancer cell lines and tissue samples (*n*=92, deep-blue lines), and in gastric carcinomas (GC, *n*=18, deep-red lines) by whole genome sequencing (WGS). The sample ID is labeled under each column. The bottom chart displays an amplified view of the *CDKN2A* gene, where the 5.1-kb common deletion region (CDR) occurred from the $P16^{INK2a}$ promoter to intron-2 (highlighted with a red dashed line rectangle), which is exactly the same region as the deleted *CDKN2A* fragment in the HCC193 lung cancer cell line (highlighted with a black line). The detailed deletion coordinates for each sample are listed in Supplemental Table S1. Each line represents a *CDKN2A* deletion fragment. The locations of $P16^{INK4a}$ and $P14^{ARF}$ (green lines) and exon-1α/2/3 (black dots) are also labeled as landmarks.

It is well known that germline *CDKN2A* inactivation can lead to a high predisposition for melanoma and pancreatic cancer (Hussussian et al. 1994; Freedberg et al. 2008; Harinck et al. 2012). Interestingly, we found that 14 of 15 *CDKN2A* allelic variants in the Online Mendelian Inheritance in Man (OMIM) database are located within the CDR sequence (Supplemental Figure S2) (Hamosh et al. 2005; Amberger et al. 2009). These phenomena indicate that both inherited and somatic defects of the *CDKN2A CDR* region are cancer drivers.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

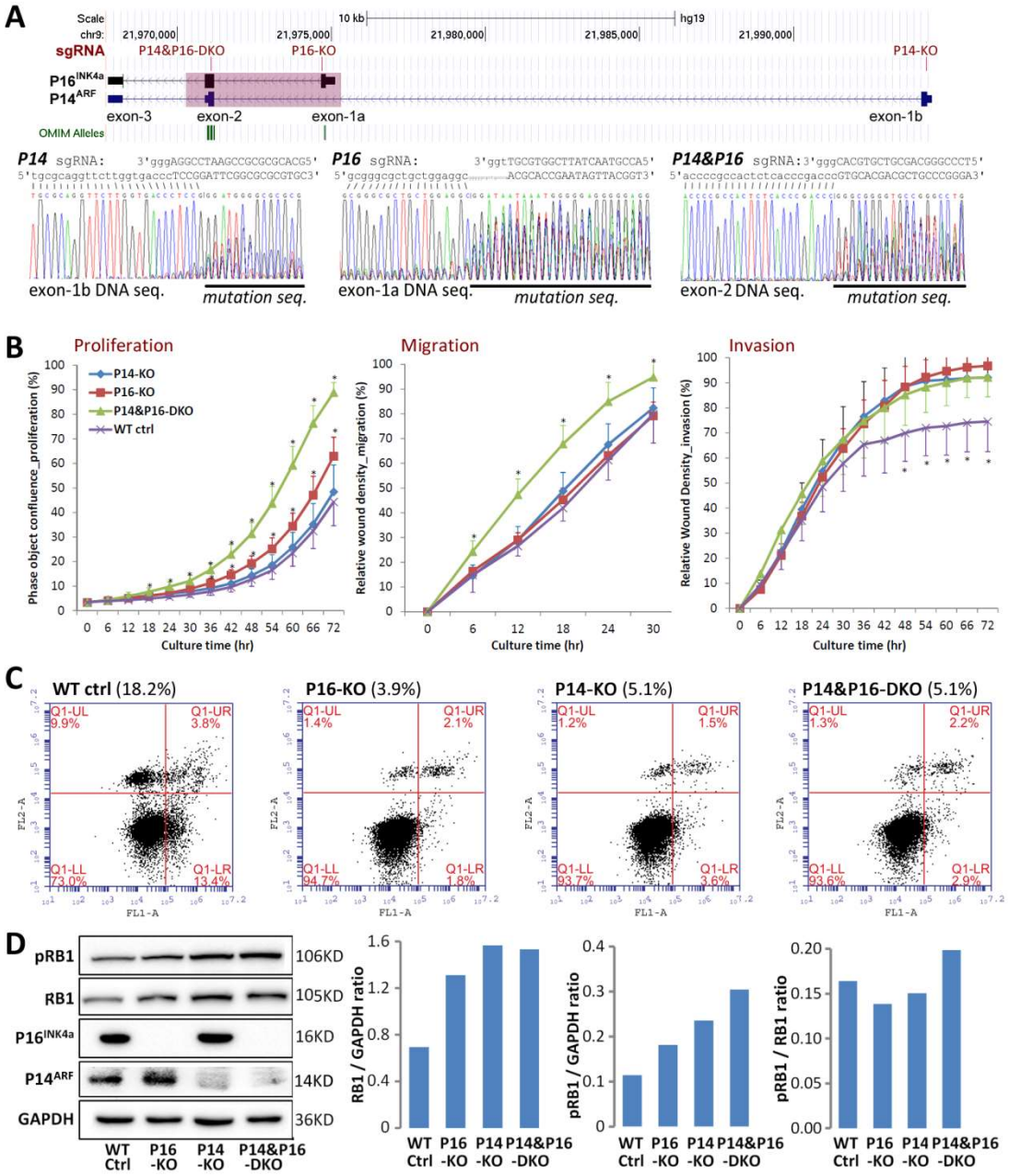

Figure 2. Comparison of behavioral alterations of HEK293T cells with various *CDKN2A* knockout (KO) genotypes. (**A**) KO of *CDKN2A* exon-1b, exon-1a, and exon-2 in cells by CRISPR/Cas9 and corresponding single guide RNA (sgRNA). Locations of sgRNAs and exons are labeled, and the 5.1-kb common deletion region is highlighted with a pink shadow (top chart). (**B**) The proliferation, migration, and invasion of pooled clones with various *CDKN2A* inactivation genotypes was assessed in long-term dynamic IncuCyte analyses. Each point represents the average value of 9 or 12 wells. The SD value is also displayed. *: *P*<0.01 compared with the *CDKN2A* wild-type control cells. (**C**) The proportions of apoptotic and dead cells with various *CDKN2A* inactivation genotypes were detected in FACS analyses using annexin V-isothiocyanate (FITC, FL1-A) and propidine iodide (PI, FL2-A) staining. The values in parentheses indicate the total proportions of early and late apoptotic cells with various *CDKN2A* KO genotypes. (**D**) The amounts of total RB1 and phosphorylated RB1 (pRB1) proteins in cells with various *CDKN2A* KO genotypes was determined by Western blot analyses.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Both *P16*^{INK4a} and *P14*^{ARF} mRNAs are transcribed from the human *CDKN2A* gene at chromosome 9p21, but with different transcription start sites (Stone et al. 1995); they share the same exon-2 but have different translation reading frames. Because *CDKN2A* exon-2 is the essential exon for coding P16^{INK4a} and P14^{ARF} proteins and because it located within the CDR, our above findings indicate that *P16*^{INK4a} and *P14*^{ARF} are coinactivated in 87% (96/110) of human cancer cell lines and tissues containing *CDKN2A CDR* deletion (Figure 1).

The *P16*^{INK4a} and *P14*^{ARF} coinactivation promotes cell proliferation and migration and inhibits apoptosis.

To study whether *P16*^{INK4a} and *P14*^{ARF} coinactivation plays a larger role in cancer development than individual *P16*^{INK4a} or *P14*^{ARF} inactivation alone, we knocked out *P16*^{INK4a}-specific exon-1a (P16-KO), *P14*^{ARF}-specific exon-1b (P14-KO), and *P16*^{INK4a} and *P14*^{ARF}-shared exon-2 (P14&P16-DKO) with CRISPR/Cas9 in *P16*^{INK4a} and *P14*^{ARF}-expressing human non-tumor embryo kidney HEK293T cells (Figure 2A). Two KO subclones for each genotype were obtained and pooled for the following experiments.

As expected, the proliferation and migration of P14&P16-DKO HEK293T cells were highest among the cells with different genotypes, and the invasion of P14-KO, P16-KO, and P14&P16-DKO cells were similarly increased, as shown by long-term dynamic IncuCyte analyses (Figure 2B). The proportions of these HEK293T cells that were undergoing apoptosis or death were similarly decreased (Figure 2C).

P16^{INK4a} and P14^{ARF} proteins play crucial roles in cell senescence, apoptosis, and cell cycle arrest, preventing cell replicative stress via P16^{INK4a}-CDK4-RB1 and P14^{ARF}-MDM2-P21^{CIP1}-CDK2-RB1 pathways (Serrano et al. 1993 and 1996; Kamijo et al. 1997; Chen et al. 2009). The amount of phosphorylated RB1 (pRB1) protein was highest in the P14&P16-DKO clones by Western blot analysis (relative to both GAPDH and RB1; Figure 2D). These results indicate that P16^{INK4a} and P14^{ARF} coinactivation leads to a more dramatic effect on cell proliferation than individual inactivation. This may account for the phenomena that most *CDKN2A* genetic diseases (mainly cancers) in the OMIM database are related to exon-2 variations (12/15=80%; Supplemental Figure S2).

Establishment of a quantitative PCR assay (P16-Light) to detect somatic *CDKN2A CDR* deletion

To study the application potential of somatic copy number variations (SCNVs) of the *CDKN2A* gene, we designed and experimentally evaluated a set of multiplex quantitative PCR assays and finally established a *CDKN2A CDR*-specific quantitative multiplex PCR assay for detecting the copy number of a 129-bp amplicon within *CDKN2A/P16*^{INK4a} intron-2 (P16-Light, Figure 3A), which covers 86% (94/110) of known *CDKN2A*

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

deletion fragments (Figure 1). Using genomic DNA from human A549 cells (with a homozygous *CDKN2A* deletion) and RKO cells (with 2 wild-type *CDKN2A* alleles) as *CDKN2A CDR* deletion-positive and deletion-negative controls, respectively, the proportions of *CDKN2A CDR* copy number were linearly correlated with the ratios (0 - 100%) of RKO cell DNA and A549 cell DNA in the input mixtures when the A549 DNA was spiked in at different proportions for the P16-Light analyses (Figure 3B). Furthermore, there was a high reproducibility when DNA with homozygous deletion of *CDKN2A* was present in ≥20% of the cells; results were verified in ten experimental repeats that were performed on different days (Figure 3C). Thus, when the proportion of *CDKN2A* copy number is significantly decreased or increased in a tumor sample relative to the paired blood DNA sample ($p<0.05$) in the P16-Light analyses, the tumor is defined as somatic *CDKN2A* deletion-positive or amplification-positive.

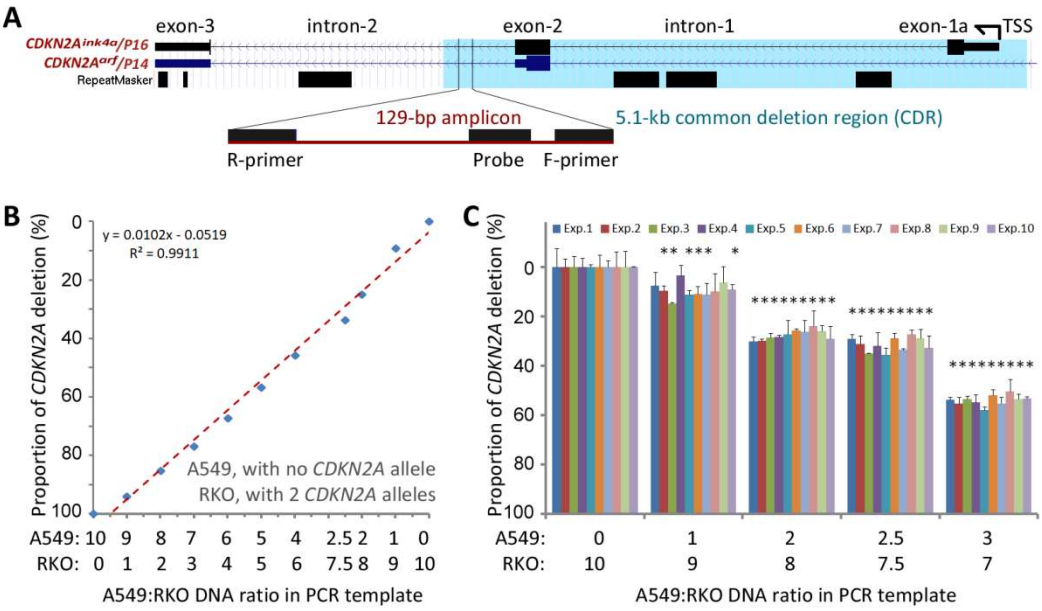

Figure 3. Detection of the copy number of *CDKN2A* intron-2 with quantitative gene-specific multiplex PCR (P16-Light). (**A**) The location of the 129-bp amplicon within the common deletion region (CDR) and its host genes. (**B**) The linear relationship between the proportion of *CDKN2A* copy number deletion and ratios of RKO cells (with two wild-type *CDKN2A* alleles) spiked with A549 cells (with a homozygous *CDKN2A* deletion). (**C**) Stability of the proportion of the *CDKN2A* copy number deletion by P16-Light during ten experiments over different days. The RKO cell DNA templates were spiked with 0, 10%, 20%, 25%, and 30% A549 cell DNA. Each column represents the average proportion of *CDKN2A* copy number deletion in triplicate. Exp. 1 - 10: the results of 10 repeated experiments performed on different days. *: compared to the 100% RKO control in chi-square test, $p<0.05$.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Somatic *CDKN2A CDR* deletion blocks regression of esophageal squamous cell dysplasia (ESCdys)

Recently, a panel of SCNV biomarkers has been reported to assess the progression potential of Barrett's esophagus to esophageal adenocarcinoma (Killcoyne et al. 2020). *P16*^{INK4a} is frequently inactivated by DNA methylation in both this cancer and this precancer. Our previous work, along with others, demonstrates that *P16*^{INK4a} methylation is significantly associated with malignant transformation of low-grade gastric dysplasia, oral epithelial dysplasia and Barrett's esophagus (Sun et al. 2004; Hall et al. 2008; Jin et al. 2009; Liu et al. 2015). This encouraged us to study whether somatic *CDKN2A* deletion could also be used to assess the prognosis of ESCdys, from which most ESCCs develop.

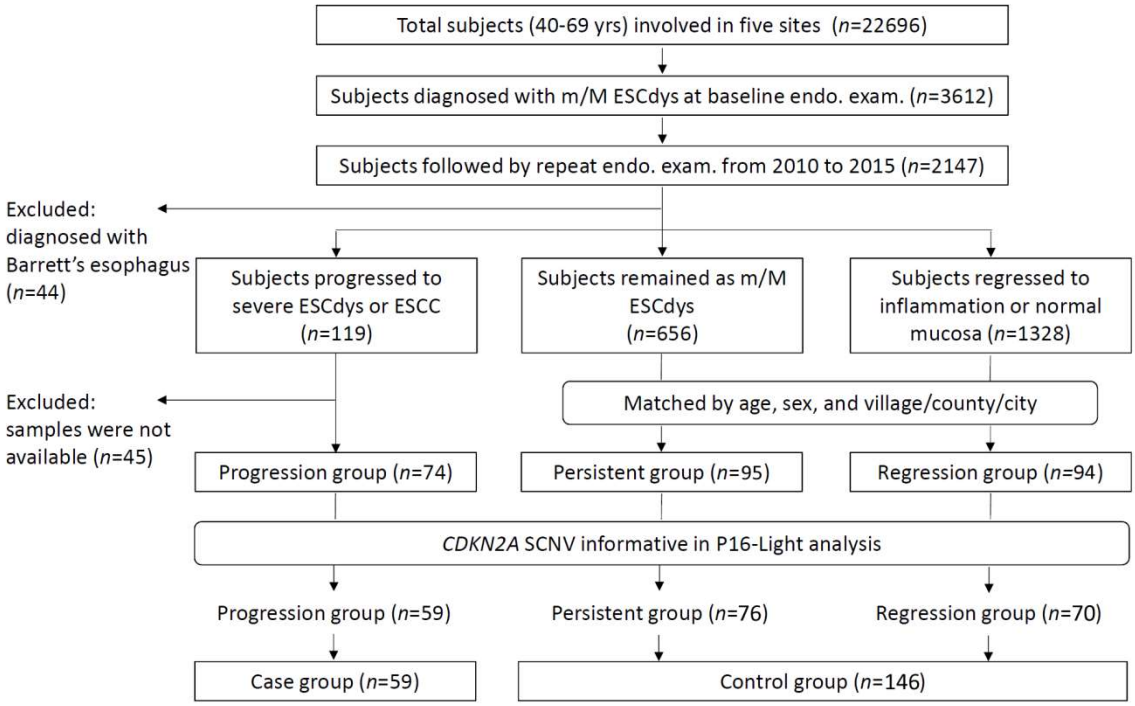

Figure 4. Participant flow diagram

We are performing a long-term prospective multicentre endoscopic screening program among residents aged 40-69 yrs in populations in rural areas of China in which residents had a high risk of ESCC.[32] In this program, patients with severe ESCdys or ESCC are sent for endoscopic or surgical therapy, patients with mild or moderate (m/M) ESCdys are scheduled for repeat endoscopy within 3 years, and patients with inflammation (esophagitis) or normal mucosa are excluded from the follow-up cohort, according to the National Esophageal Cancer Screening guideline (Wang et al. 2020). For the current evaluation, we divided patients with baseline m/M ESCdys into three groups based on the results of their first repeat endoscopy (a median of 37.4 months after baseline): patients (*n*=74) whose worst squamous diagnosis progressed to

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

severe ESCdys or ESCC; patients (*n*=95) who had persistent m/M ESCdys; and patients (*n*=94) who regressed to inflammation or normal (Supplemental Figure S3). Information on *CDKN2A* SCNVs was obtained by P16-Light analysis in the baseline m/M ESCdys lesions from 78.0% (205/263) patients (Figure 4, Supplemental Table S2). Interestingly, both somatic *CDKN2A* deletion and amplification were prevalent in these ESCdys lesions (33.7% for deletion and 27.3% for amplification, relative to the gene copy number in the paired blood DNA samples). More *CDKN2A* amplification was detected in patients with low annual income than those with high income, and in patients drinking tea than those who did not (Supplemental Table S3). Because no significant difference in somatic *CDKN2A* deletion or amplification was observed between the ESCdys patients in the progression and persistent groups (*p*=0.1934; Supplemental Table S4), we merged these two groups together in further frequency comparisons. No significant difference was observed between the merged progression & persistent group and the regression group in baseline grades of ESCdys, location of ESCdys within the esophagus, sex or age of the patients, or the patients' city or county of residence (Table 1).

Table 1. Comparison of the prevalence of somatic *CDKN2A* gene copy number variations (SCNVs) in baseline esophageal squamous cell dysplasia biopsy samples from patients in the progression & persistent group and patients in the regression group

		Progression &Persistent group				Regression group				Chi-square test	
		n (%)	Cases of CDKN2A SCNVs (%)			n (%)	Cases of CDKN2A SCNVs (%)			Chi-square	p-value
			Deleted	Diploid	Amplified		Deleted	Diploid	Amplified		
Baseline	Mild ESCdys	94(69.6)	34(36.2)	40(42.6)	20(21.3)	49(70.0)	7(14.3)	23(46.9)	19((38.8)	8.871	0.0029
	Mod ESCdys	41(30.4)	22(53.7)	12(29.3)	7(17.1)	21(30.0)	6(28.6)	5(23.8)	10(47.6)	6.090	0.0136
Location	Upper	20(14.8)	10(50.0)	5(25.5)	5(25.0)	18(25.7)	2(11.1)	8(44.4)	8(44.4)	4.776	0.0290
	Middle	68(50.4)	28(41.2)	29(42.6)	11(16.2)	38(54.3)	9(23.7)	13(34.2)	16(42.1)	7.653	0.0057
	Lower	47(34.8)	18(38.3)	18(38.3)	11(23.4)	14(20.0)	2(14.3)	7(50.0)	5(35.7)	2.389	0.1222
Sex	Male	79((58.5)	32(40.5)	29(36.7)	18(22.8)	44(62.9)	7(15.9)	18(40.9)	19(43.2)	9.188	0.0024
	Female	56(41.5)	24(42.9)	23(41.1)	9(16.1)	26(37.1)	6(23.1)	10(38.5)	10(38.5)	5.382	0.0204
Age	<60yrs	58(43.0)	24(41.4)	22(37.9)	12(20.7)	34(48.6)	4(11.8)	15(44.1)	15(44.1)	9.980	0.0016
	≥60yrs	77(57.0)	32(41.6)	30(39.0)	15(19.5)	36(51.4)	9(25.0)	13(36.1)	14(38.9)	5.172	0.0230
County	Cixian	34(25.2)	22(64.7)	10(29.4)	2(5.9)	32(45.7)	6(18.8)	17(53.1)	9(28.1)	14.395	0.0002
	Yanting	70(51.9)	26(37.1)	28(40.0)	16(22.9)	25(35.7)	5(20.0)	9(36.0)	11(44.0)	4.389	0.0362
	LZ&YZ&FC*	31(23.0)	8(25.8)	14(45.2)	9(29.0)	13(18.6)	2(15.4)	2(15.4)	9(69.2)	3.802	0.0512
(Total)		135	56(41.5)	52(38.5)	27(20.0)	70	13(18.6)	28(40.0)	29(41.4)	14.888	0.0001

*Linzhou, Yangzhong and Feicheng

Notably, the positive rate of somatic *CDKN2A* deletion was much lower in the baseline ESCdys lesions from the 70 patients in the regression group than in the baseline ESCdys lesions from the 135 patients in

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

the progression & persistent group, and the positive rate of somatic *CDKN2A* amplification was much higher in the baseline ESCdys lesions in the regression group than in the baseline ESCdys lesions of the progression & persistent group (p<0.0001). These significant differences also remained in most strata of baseline grades of ESCdys, location of ESCdys within the esophagus, sex and age of the patients, and the patients' city or county of residence (Table 1).

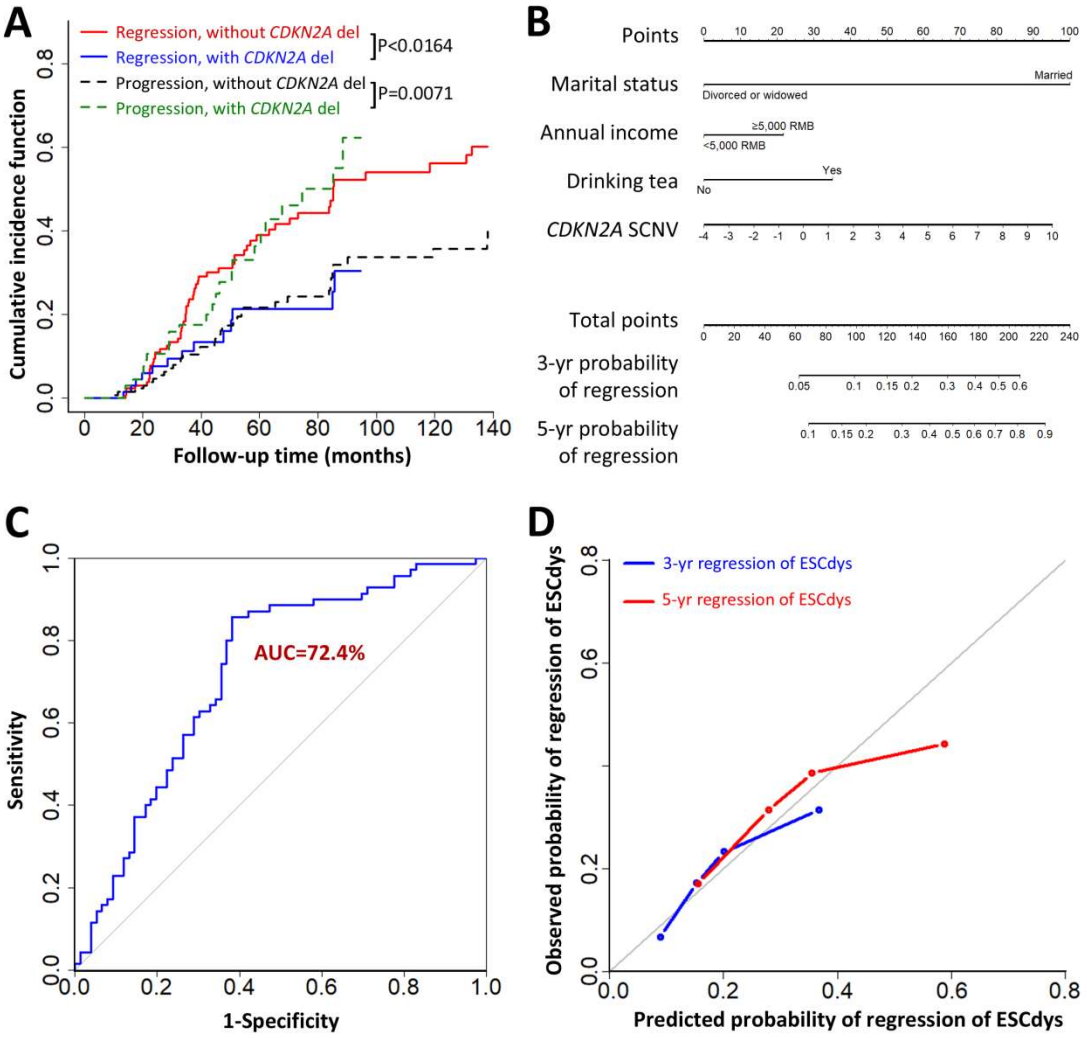

Figure 5. Comparison of cumulative incidence function (CIF) for somatic *CDKN2A* copy number deletion in human esophageal mucosa biopsy samples from patients in the follow-up study. (**A**) Cumulative incidence of regression and progression of esophageal squamous cell dysplasia (ESCdys) with and without *CDKN2A* deletion using the competing risk analysis. The progression of ESCdys was used as a competing risk. The p-values in the Fine-Gray univariate analysis are labled. (**B**) Competing-risk nomogram including *CDKN2A* SCNV, the marrial status, annual income, and tea consumption to predict regression of ESCdys following start of follow-up in the presence of the competing risk progression of ESCdys. (**C**) The receiver operating characteristic (ROC) curve for regression of ESCdys within 3 years, calculated using the prediction model constructed with *CDKN2A* SCNV, the marrial status, annual income, and drinking tea. (**D**) The calibration curves at 3 and 5 years between predicted and actual observed probability of regression of ESCdys

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 2. The results of univariate and multivariate Fine-Gray regression analyses*

Variable	Univariate analysis		Multivariate analysis	
	HR (95% CI)**	p-value	HR (95% CI)	p-value
Age (yrs)	0.99 (0.95-1.03)	0.4900		
Gender (Male vs. Female)	0.96 (0.60-1.55)	0.8800		
Marriage (Married vs. Divorced or widowed)	7.56 (1.12-51.20)	0.0380	7.28 (1.03-51.46)	0.0470
Education		0.1800		
No education	Reference	0.9900		
Primary education	0.56 (0.30-1.04)	0.0670		
Higher educations	0.67 (0.33-1.36)	0.2700		
Annual income per capita (≥5000 vs. <5000)	1.92 (1.16-3.17)	0.0110	1.67 (0.968-2.89)	0.0650
Drinking alcohol (Yes vs. No)	0.84 (0.47-1.51)	0.5700		
Smoking (Yes vs. No)	0.95 (0.58-1.56)	0.8400		
Drinking tea (Yes vs. No)	2.78 (1.35-5.74)	0.0057	2.28 (1.061-4.90)	0.0350
Water source (Tap water vs. Others)	0.85 (0.32-2.28)	0.7500		
Taking fruit (Yes vs. No)	1.21 (0.58-2.53)	0.6100		
Taking pickled food (Yes vs. No)	1.15 (0.64-2.07)	0.6400		
Taking fried food (Yes vs. No)	1.00 (0.55-1.79)	0.9900		
Taking hot food (Yes vs. No)	0.71 (0.43-1.17)	0.1800		
Taking mildew food (Yes vs. No)	0.28 (0.04-1.80)	0.1800		
Digestive diseases (Yes vs. No)	1.32 (0.75-2.34)	0.3300		
Family cancer history (Yes vs. No)	0.91 (0.55-1.49)	0.7000		
BMI (≥24 kg/m^2 vs. <24 kg/m^2)	1.54 (0.98-2.44)	0.0630		
CDKN2A SCNVs	1.14 (1.05-1.24)	0.0018	1.12 (1.02-1.23)	0.0140

*Association of all variables with regression of ESCdys is analyzed using a two-sided Gray's test.

**HR, hazard ratio; CI, confidence interval

The cumulative incidence of regression of ESCdys with *CDKN2A* deletion was significantly lower than those without (Fine-Gray univariate analysis: $p<0.0164$) while the cumulative incidence of progression of ESCdys with *CDKN2A* deletion was significantly higher than those without ($p=0.0071$; Figure 5A). In the multivariate Fine-Gray analysis, *CDKN2A* deletion was still a significant independent regression predictor

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

(*p*=0.0140; Table 2). In additon, the marital status, annual income, and tea consumption were also significantly correlated to regression of ESCdys. Thus, we used these significant factors to construct a monogram for predicting regression of m/M ESCdys (Figure 5B). The area under the receiver operating characteristic (ROC) curve (AUC) to predict regression of m/M ESCdys within 3 years was 72.4% (95% confidence interval: 64.0%–80.7%) (Figure 5C). The calibration curves at 3 and 5 years showed good agreement between the estimations with the nomogram and actual observations (Figure 5D). In addition, a significant majority of the patients with baseline somatic *CDKN2A* deletions still had somatic *CDKN2A* deletions in the follow-up biopsy samples (*n*=170) (Supplemental Figure S4).

Taken together, these findings indicate that somatic *CDKN2A* deletions are usually persistent, and they may block regression and promote progression of m/M ESCdys.

DISCUSSION

ESCC is one of the main causes of cancer death in China (Bray et al. 2018). Most ESCCs progress from ESCdys lesions. While a panel of gene SCNVs has been used to predict the malignant transformation of Barrett's esophagus within the lower esophagus (Killcoyne et al. 2020), a clinical biomarker is not yet available to predict the prognosis of ESCdys into ESCC. Somatic *CDKN2A* copy number deletion is a frequent event in both ESCdys and ESCC (Song et al. 2014; Liu et al. 2017; Cui et al. 2020). In the present study, we characterized a 5.1-kb CDR sequence at base resolution within the *CDKN2A* gene in various human cancers, and confirmed this finding using WGS datasets for all of 18 gastric cancers (Xing et al. 2019). We also established a convenient quantitative PCR assay, P16-Light, to detect *CDKN2A* SCNVs, and found that both somatic *CDKN2A* copy number deletion and amplification were prevalent in ESCdys lesions. We also discovered that *CDKN2A* SCNVs were significantly and consistently associated with the prognosis of ESCdys years before regression and progression. Using *CDKN2A* SCNVs and other 3 risk factors to construct a monogram to predict regression of ESCdys within 3 follow-up years, the AUC was 72.4% (95% confidence interval: 64.0%-80.7%).

It is well known that amplifications of oncogenes drive cancer development and progression. Here, we found that the amplification of the tumor suppressor gene *CDKN2A* is positively associated with the regression of ESCdys. To best of our knowledge, this is the first report that the amplification of a tumor suppressor gene can decrease cancer risk in patients with precancer. Also, the fact that *CDKN2A* deletion is associated with progression/persistence of ESCdys suggests that *CDKN2A* inactivation may play a crucial role in the initiation stage during malignant transformation of normal squamous cells in the esophageal mucosa

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

and maintenance of ESCdys. Among subjects under a normal environment, two *CDKN2A* alleles are functionally enough for diploid cells to physiologically maintain. However, among patients with ESCdys in high cancer risk areas, it is unknown whether two *CDKN2A* alleles are functionally enough because environmental factors may cause epigenetic inactivation of the *CDKN2A/P16*^{INK4a} gene. *CDKN2A* amplification is consistently associated with the upregulation of gene transcription in cancer cells. It is worth studying whether *CDKN2A* amplification may favor the recovery of function of this gene in precancer cells, and subsequently contribute to regression of ESCdys lesions.

The driver function of the *CDKN2A* gene in cancer development is enigmatic. *P16*^{ink4a} inactivation contributes less than *P19*^{arf} (the murine counterpart of human *P14*^{ARF}) inactivation to cancer development in mice while *P16*^{INK4a} inactivation contributes more than *P14*^{ARF} inactivation to cancer development in humans (Peter et al. 2008; Li et al. 2009). The exact mechanisms leading to the difference among species is not clear. Here, we reported that aproximately 87% of genetic *P16*^{INK4a} inactivation is accompanied by *P14*^{ARF} inactivation in human cancer cell lines or tissues. This may account for the species-related functional difference of the *CDKN2A* gene. This explanation is also supported by the report that knocking out both *p16*^{ink4a} and *p19*^{arf} leads to more cancer development than individual inactivation in mice (Sharpless et al. 2004).

In conclusion, we have, for the first time, found that there is a 5.1-kb CDR region within the *CDKN2A* gene, and that most *CDKN2A* deletions lead to *P16*^{INK4a} and *P14*^{ARF} coinactivation in human cancers. Using the CDR as a target sequence, we developed a convenient quantitative multiplex PCR assay, the P16-Light to detect *CDKN2A* SCNVs in clinical practice. Both *CDKN2A* deletion and amplification were prevalent in ESCdys lesions and were significantly associated with progression or persistence and regression, respectively, of mild or moderate ESCdys in a population-based follow-up study. *CDKN2A* SCNVs are potential biomarkers for predicting the regression of precancer in esophageal squamous cells. In addition, we observed a similar CDR region within other tumor suppressor genes such as *ATM*, *FAT1*, *miR31HG*, *PTEN*, and *RB1* in the SNP array-based TCGA SCNV datasets (Supplemental Figure S5), suggesting that our strategy to detect *CDKN2A* SCNVs may be suitable for the establishment of detection methods for other genes.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

METHODS

Specimens and DNA preparation

Since 1999, the National Cancer Center of the Chinese Academy of Medical Science has conducted a multicentre prospective ESCC endoscopic screening project among 22696 residents aged 40-69 yrs in Cixian County (Hebei Province), Yanting County (Sichuan Province), Linzhou City (Henan Province) and Yangzhong City (Jiangsu Province) and Feicheng City (Shandong Province), China, rural areas in which residents have a high risk of ESCC (Project Registration no. NCT02094105, ChiCTR-EOC-17010553). In baseline exams, mild and moderate (m/M) esophageal squamous cell dysplasia (ESCdys) was pathologically diagnosed in 3612 patients. From 2010 to 2015, a repeat endoscopic screening was offered to these eligible subjects to re-evaluate the status of baseline ESCdys lesions and to look for the development of new lesions, and 2147 (59.4%) underwent these repeat endoscope exams. As in the baseline exams, endoscopic biopsy specimens were taken from all lesions visible in the esophagus after mucosal iodine staining. Peripheral blood samples were also taken from each patient before this repeat endoscopic examination. The biopsies were fixed immediately in buffered formalin and then embedded, sectioned, and stained with hematoxylin and eosin. The microscopic slides were read by a panel of three senior pathologists as we previously described (Wei et al. 2015). The presence or absence of ESCC, ESCdys (mild, moderate, severe), inflammation (mild, moderate, severe), and other pathological changes were recorded for each biopsy, and a global diagnosis was made in each case representing the most advanced lesion.

Subjects (*n*=74) with a diagnosis of m/M ESCdys at the baseline who progressed to severe ESCdys or ESCC during the follow-up were included in the progression group. Subjects remaining as mild or moderate ESCdys were included in the persistent group. Subjects who regressed to inflammation or normal mucosa were included in the regression group (Supplemental Figure S6). m/M ESCdys samples in the persistent group (*n*=95) and the regression group (*n*=94) were selected from the tissue block archive, matched to the progressing subjects by age, sex, and village/county/city.

All baseline biopsy samples of mild or moderate ESCdys in the progression group were used, if sections were available from archived paraffin blocks, for P16-Light analysis. Baseline m/M ESCdys biopsy samples were collected from 263 patients and used for P16-Light analysis. Finally, information on the status of *CDKN2A* SCNVs was obtained for 205 patients with sufficient DNA extracted from their baseline biopsies to complete the P16-Light analysis (Supplemental Figure S3). The peripheral blood samples were selected for all these cases from the National Cohort of Esophageal Cancer (NCEC) biobank. Information including

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

demographic characteristics and risk factors by questionnaire were also collected for each case. Genomic DNA was extracted from these samples with a phenol/chloroform method in a standard fashion. These studies were approved by the Institutional Review Board of the China Cancer Foundation and the National Cancer Center, Cancer Hospital of the Chinese Academy of Medical Sciences and Peking Union Medical College (Approval No. 16-171/1250), and all of the patients provided written informed consent.

Optimized quantitative multiplex PCR assay (P16-Light) to detect *CDKN2A* copy number

A number of multiplex primer and probe combinations were designed based on the best multiplex primer probe scores for the CDR in the *CDKN2A* gene and *GAPDH* sequences by Bacon Designer 8 software. Multiplex PCR assays were established according to the Applied Biosystems (ABI) TaqMan universal PCR master mix manual. The performances of these assays for the detection of *CDKN2A* copy numbers were compared with each other. Finally, a multiplex primer and probe combination was selected (Supplemental Table S5) and their components' concentrations were optimized. Briefly, each multiplex PCR assay was carried out in a total 20 μL volume that included 5-10 ng of input DNA, 10 μM of forward and reverse primers and probe for *CDKN2A* intron-2, 10 μM forward and reverse primers and probe for *GAPDH*, and 10 μL of 2 x TaqMan Universal Master Mix II with uracil-N-glycosylase (Kit-4440038, ABI, Lithuania). The PCRs were performed in triplicate in a MicroAmp Fast Optical 96-Well Reaction Plate with a barcode (0.1 mL; ABI, China) with an ABI 7500 Fast Real-Time PCR System. The specific conditions of the PCR were as follows: initial incubation for 10 min at 95°C, followed by 40 cycles of 95°C for 20 sec and 58°C for 60 sec. When Ct value for *GAPDH* input for a sample was 34 or less cycles, this sample was considered as *CDKN2A* SCNV informative.

Definitions of *CDKN2A CDR* deletion-positive and amplification-positive

We used the genomic DNA from A549 cells containing no *CDKN2A* allele to dilute genomic DNA from RKO cells containing 2 wild-type *CDKN2A* alleles, and then we set the standard curve according to the relative copy number of the *CDKN2A* gene at different dilution concentrations. The ΔCt value and relative copy number for the *CDKN2A* gene were calculated using the *GAPDH* as the reference. When the *CDKN2A* copy number in the A549-diluted template was consistently lower than the copy number in the RKO control template, and the difference was statistically significant ($p<0.05$), it was judged that the lowest dilution concentration was the detection limit of *CDKN2A* deletion (the difference in *CDKN2A* copy number between the 100% RKO template and 80% RKO template spiked with 20% A549 DNA). When the *CDKN2A* relative copy number in a tissue sample was significantly lower or higher than that of the paired blood sample, the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

sample was defined as somatic *CDKN2A CDR* deletion-positive or amplification-positive, respectively. The 100% A549, 100% RKO, and 20% A549 + 80% RKO DNA mix controls were analyzed for each experiment.

Call for *CDKN2A* interstitial deletion in the GC WGS datasets

We used Meerkat (http://compbio.med.harvard.edu/Meerkat/) 23 to predict somatic SVs and their breakpoints in WGS datasets (accession numbers, EGAD00001004811 with 36× of sequencing depth) for gastric adenocarcinoma samples from 168 patients using the suggested parameters (Xing et al. 2020). This method used soft-clipped and split reads to identify candidate breakpoints, and precise breakpoints were refined by local alignments. *CDKN2A* deletion information was obtained from WGS datasets for 157 GC samples.

Cell lines and cultures

The human cell line HEK293T (kindly provided by Professor Yasuhito Yuasa of Tokyo Medical and Dental University) and the *P16* allele homogygously deleted cell line A549 (kindly provided by Dr. Zhiqian Zhang of Peking University Cancer Hospital and Institute) were grown in RPMI-1640 medium, and the RKO cell line containing two wild type *P16* alleles was purchased from American Type Culture Collection and grown in DMEM media. The medium was supplemented with 10% (v/v) fetal bovine serum (FBS). These cell lines were tested and authenticated by Beijing JianLian Genes Technology Co., Ltd. before they were used in this study. STR patterns were analyzed using a Goldeneye™ 20A STR Identifiler PCR Amplification kit.

Cell proliferation, migration, and invasion assays using IncuCyte

Cells were seeded into 96-well plates (2,000 cells/well, 10 wells/group) and cultured for at least 96 hr to determine the proliferation curves. The cells were photographed every 6 hr using a long-term dynamic observation platform (IncuCyte, Essen, MI, USA). The cell confluence was analyzed using IncuCyte ZOOM software (Essen, Ann Arbor, MI, USA). For continuous observation of cell migration and invasion, the cells were seeded into 96-well plates at a density of 25,000 cells/well and then were cultured for 24 hr. After a wound scratch was established, the cells were washed three times with PBS. For the invasion test, after PBS washed, 50 µL Matrigel (BD Bioscience, San Jose, CA) diluted with RPMI 1640 Medium at ratio of 1:8 was added, and the cells were cultured for 30 min at 37 °C. The cells were then regularly cultured and photographed every 6 hr for at least 96 hr. The relative wound width was calculated using the same software.

Knockout of *CDKN2A* exon-1a, exon-1b, and exon-2 with CRISPR/Cas9

A single gRNA approach was used to knock out target sequences in the *CDKN2A* gene using the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

CRISPR/Cas9 system. The sgRNAs were designed using online software found at the website (http://crispr.mit.edu), and they were synthesized by Thermo Scientific, Inc., Rockford, IL, USA (Supplemental Table S5). The sgRNAs were cloned into the Lenti-CRISPR-V2 vector expressing Cas9 (Plasmid #52961, Addgene, Inc.) at the *BsmBI* restriction site. Then, the lentiviral plasmid expressing gRNA and Cas9 was introduced into HEK293FT cells. The viral supernatants were collected and filtered through a 0.45 μm PVDF filter (Millipore, USA) 72 hr after transfection, and then the viruses were used to infect HEK293T cells. Three days later, the infected cells were subjected to puromycin selection for one week, and genomic DNA from the surviving cells was isolated for PCR amplification and sequencing using the primers listed in Table S5. Then, the cells were seeded into 96-well plates to select the monoclonal cells. Cells transfected with an empty Lenti-CRISPR-V2 control vector were used as a wild-type (WT) control.

Western blot

Cells were collected and lysed to obtain protein lysate. The resulting proteins were electrophoresed through a 10% SDS-PAGE gel and then were transferred onto a PVDF membrane. After blocking with 5% fat-free milk overnight at 4 °C, the membrane was incubated with primary antibodies (anti-P16, Abcam, ab81278 UK; anti-P14, Abcam, ab185620 UK; anti-RB1, Abcam, ab181616 UK; anti-Phospho-RB1 (Ser807/811), Cell Signaling Technology, #8516, USA; anti-GAPDH, Protein Tech, 60004-1, China) for 1 hr at room temperature. The membrane was then washed 3 times with PBST (PBS with 0.1% Tween 20). After washing, the membrane was incubated with the corresponding horseradish peroxidase-conjugated goat anti-goat or anti-mouse IgG at room temperature for 1 hr. The signals were visualized using an Immobilon Western Chemiluminescent HRP Substrate kit (WBKLS0500, Millipore, Billerica, USA).

Cell apoptosis and death analyses

Cells were seeded in six-well plates (2×10^5/well). After 48 hr, the cells were treated with trypsin and washed twice with cold PBS. They were labeled with annexin V-FITC and propidium iodide (PI) according to the manufacturer's protocol (Dojindo, Japan). Then, the cells were analyzed with a BD Accuri C6 flow cytometer (BD Biosciences, USA). The percentages of cells in early apoptosis (annexin V-positive, PI-negative) and late apoptosis or necrosis (annexin V- and PI-positive) were calculated using BD Accuri C6 Software.

Statistical analysis

Chi-square or Fisher's exact tests were used to compare the proportion of somatic *CDKN2A* CDR deletion or amplification between different groups of tissue samples. Student t-test was used to compare

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

differences of the density or confluence of cells with different *CDKN2A* genotypes and the proportion of *CDKN2A* gene copy number between genomic DNA samples. Univariate competing risk analysis was used to identify risk factors for regression of ESCdys, and cumulative regression rate was calculated using the cumulative incidence function method with progression of ESCdys as a competing risk. Factors with $p<0.05$ in the univariate analysis were incorporated into the multivariate analysis, and used to construct a competing-risk nomogram. The discrimination of the nomogram was evaluated by AUC of the ROC curve. The calibration, which compares estimations with actual observations, was graphically assessed with a calibration curve. All statistical tests were two-sided, and *p* value less than 0.05 was considered statistically significant. All statistical analyses were performed using SPSS software (version 16.0) and R software (version 4.0.5).

Finding

This work was supported by the Beijing Natural Science Foundation (grant number 7181002 to D.J.D.); Capital's Funds for Health Improvement and Research (grant number 2018-1-1021 to D.J.D.); and by the National Key R&D Program of China (grant number 2016YFC0901404 to W.Q.W.).

Competing Insterest statement: The authors have nothing to disclose.

Acknowledgements: We sincerely thank Dr. Guohui Song from Cixian County Cancer Institute, Dr. Changqing Hao from Linzhou Cancer Hospital, Dr. Zhaolai Hua from Yanzhong Cancer Hospital, Dr. Jun Li from Yanting Cancer Hospital, and Dr. Yanyan Li from Feicheng People's Hospital. Thanks to all cooperating demonstration centers and their staff whose hard work in follow-up made this study possible. We also than Miss Gina Mckeown in New York, USA for English language editing.

Author contributions: D.J.D. and W.Q.W. designed the study. D.J.D. performed bioinformatics analyses, characterized base resolution *CDKN2A CDR* coordinates, and wrote the manuscript draft. W.Q.W, Z.Y.F., and Y.Q. organized the follow-up study. J.Z., Y.T., Z.J.L., and L.K.G. developed the P16-Light assay and analyzed *CDKN2A* SCNVs in samples. R.X., Y.Y.L., J.F.J., provided the *CDKN2A CDR* coordinates from whole genome sequencing datasets. J.L.Q. performed gene KO experiments. S.M.D., critically reviewed the manuscript. All authors approved the final manuscript.

Data access: Detailed information on the base-resolution interstitial common deletion region in *CDKN2A* gene in 110 cancer samples was listed in Supplemental Table S1. Deidentified individual participant data that underlie the reported results was attached in the manuscript as Supplemental Table S2. Individual

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

participant data will not be shared.

REFERENCES

Amberger J, Bocchini CA, Scott AF, Hamosh A. 2009. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**: D793-796.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**: 899-905.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**: 394-424.

Cairns P, Polascik TJ, Eby Y, Tokino K, Califano J, Merlo A, Mao L, Herath J, Jenkins R, Westra W et al. 1995. Frequency of homozygous deletion at p16/CDKN2 in primary human tumours. *Nat Genet* **11**: 210-212.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401-404.

Chen Z, Carracedo A, Lin HK, Koutcher JA, Behrendt N, Egia A, Alimonti A, Carver BS, Gerald W, Teruya-Feldstein J et al. 2009. Differential p53-independent outcomes of p19(Arf) loss in oncogenesis. *Sci Signal* **2**: ra44.

Cui Y, Chen H, Xi R, Cui H, Zhao Y, Xu E, Yan T, Lu X, Huang F, Kong P et al. 2020. Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res* **30**: 902-913.

Deng J, Wang ES, Jenkins RW, Li S, Dries R, Yates K, Chhabra S, Huang W, Liu H, Aref AR et al. 2018. CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-cell Activation. *Cancer Discov* **8**: 216-233.

Freedberg DE, Rigas SH, Russak J, Gai W, Kaplow M, Osman I, Turner F, Randerson-Moor JA, Houghton A, Busam K et al. 2008. Frequent p16-independent inactivation of p14ARF in human melanoma. *J Natl Cancer Inst* **100**: 784-795.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**: pl1.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Hall GL, Shaw RJ, Field EA, Rogers SN, Sutton DN, Woolgar JA, Lowe D, Liloglou T, Field JK, Risk JM. 2008. p16 Promoter methylation is a potential predictor of malignant transformation in oral epithelial dysplasia. *Cancer Epidemiol Biomarkers Prev* **17**: 2174-2179.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**: D514-517.

Harinck F, Kluijt I, van der Stoep N, Oldenburg RA, Wagner A, Aalfs CM, Sijmons RH, Poley JW, Kuipers EJ, Fockens P et al. 2012. Indication for CDKN2A-mutation analysis in familial pancreatic cancer families without melanomas. *J Med Genet* **49**: 362-365.

Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH, Tucker MA, Dracopoli NC. 1994. Germline p16 mutations in familial melanoma. *Nat Genet* **8**: 15-21.

Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, Leeson R, Kanodia A, Mei S, Lin JR et al. 2018. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**: 984-997.e924.

Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y, Olaru A, Paun B, Yang J, Kan T et al. 2009. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer Res* **69**: 4112-4115.

Kamijo T, Zindy F, Roussel MF, Quelle DE, Downing JR, Ashmun RA, Grosveld G, Sherr CJ. 1997. Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. *Cell* **91**: 649-659.

Killcoyne S, Gregson E, Wedge DC, Woodcock DJ, Eldridge MD, de la Rue R, Miremadi A, Abbas S, Blasko A, Kosmidou C et al. 2020. Genomic copy number predicts esophageal cancer years before transformation. *Nat Med* **26**: 1726-1732.

Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, Blasco MA, Serrano M. 2009. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**: 1136-1139.

Liu HW, Liu XW, Dong GY, Zhou J, Liu Y, Gao Y, Liu XY, Gu LK, Sun Z, Deng DJ. 2015. *P16* Methylation as an Early Predictor for Cancer Development From Oral Epithelial Dysplasia: A Double-blind Multicentre Prospective Study. *EBioMedicine* **2**: 6.

Liu X, Zhang M, Ying S, Zhang C, Lin R, Zheng J, Zhang G, Tian D, Guo Y, Du C et al. 2017. Genetic Alterations in Esophageal Tissues From Squamous Dysplasia to Carcinoma. *Gastroenterology* **153**: 166-177.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**: R41.

Network CGAR, University AWGA, Agency BC, Hospital BaWs, Institute B, University B, University CWR, Institute D-FC, University D, Centre GPC et al. 2017. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**: 169-175.

Patel A, Schwab R, Liu YT, Bafna V. 2014. Amplification and thrifty single-molecule sequencing of recurrent somatic structural variations. *Genome Res* **24**: 318-328.

Peters G. 2008. Tumor suppression for ARFicionados: the relative contributions of p16INK4a and p14ARF in melanoma. *J Natl Cancer Inst* **100**: 757-759.

Sasaki S, Kitagawa Y, Sekido Y, Minna JD, Kuwano H, Yokota J, Kohno T. 2003. Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* **22**: 3792-3798.

Serrano M, Hannon GJ, Beach D. 1993. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**: 704-707.

Serrano M, Lee H, Chin L, Cordon-Cardo C, Beach D, DePinho R. 1996. Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**: 27-37.

Sharpless NE, Ramsey MR, Balasubramanian P, Castrillon DH, DePinho RA. 2004. The differential impact of p16(INK4a) or p19(ARF) deficiency on cell growth and tumorigenesis. *Oncogene* **23**: 379-385.

Song Y, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang J, Xu L, Zhou Y et al. 2014. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**: 91-95.

Stone S, Jiang P, Dayananth P, Tavtigian SV, Katcher H, Parry D, Peters G, Kamb A. 1995. Complex structure and regulation of the P16 (MTS1) locus. *Cancer Res* **55**: 2988-2994.

Sun Y, Deng DJ, You WC, Bai H, Zhang L, Zhou J, Shen L, Ma JL, Xie YQ, Li JY. 2004. Methylation of p16 CpG islands associated with malignant transformation of gastric dysplasia in a population-based study. *Clinical Cancer Research* **10**: 5087-5093.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E et al. 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**: D941-D947.

Wang GQ and Wei WQ. 2020. Technology Scheme for Upper Gastrointestinal Cancer Early Detection and Treatment. People's Medical Publishing House: Beijing, China

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Wei WQ, Chen ZF, He YT, Feng H, Hou J, Lin DM, Li XQ, Guo CL, Li SS, Wang GQ et al. 2015. Long-Term Follow-Up of a Community Assignment, One-Time Endoscopic Screening Study of Esophageal Cancer in China. *J Clin Oncol* **33**: 1951-1957.

Xing R, Zhou Y, Yu J, Yu Y, Nie Y, Luo W, Yang C, Xiong T, Wu WKK, Li Z et al. 2019. Whole-genome sequencing reveals novel tandem-duplication hotspots and a prognostic mutational signature in gastric cancer. *Nat Commun* **10**: 2037.

Yu J, Yan J, Guo Q, Chi Z, Tang B, Zheng B, Yin T, Cheng Z, Wu X, Yu H et al. 2019. Genetic Aberrations in the CDK4 Pathway Are Associated with Innate Resistance to PD-1 Blockade in Chinese Patients with Non-Cutaneous Melanoma. *Clin Cancer Res* **25**: 6511-6523.

Zhang J, Bu X, Wang H, Zhu Y, Geng Y, Nihira NT, Tan Y, Ci Y, Wu F, Dai X et al. 2018. Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance. *Nature* **553**: 91-95.

SUPPLEMENTARY FIGURE LIST

Supplemental Figure S1. Prevalence of *CDKN2A* deep-deletion and the levels of gene expression in 10967 samples from cancer patients in Pan-TCGA studies. (**A**) Prevalence of *CDKN2A* deep deletion detected by Affymetrix SNP6.0 microarray. The number of total cancer cases and cases with *CDKN2A* deep deletion are listed for each kind of cancer. (**B**) The levels of $P16^{INK4a}$ mRNA determined by RNA sequencing in cancers with various *CDKN2A* genetic changes. The charts for patients (n=10953) in 32 Pan-TCGA studies were adapted from a graphic view at the cBioPortal Cancer Genomics website (www.cbioportal.org).

Supplemental Figure S2. Distribution pattern of the Online Mendelian Inheritance in Man (OMIM) allelic variants within the *CDKN2A* common deletion region (CDR, highlighted in blue shadow). 12 allelic variants are located in *CDKN2A* exon-2, 2 allelic variants are located in *CDKN2A* exon-1α, and 1 allelic variant is located in *CDKN2A* exon-1β. This chart was adapted from the UCSC Genome Browser on March 10, 2021.

Supplemental Figure S3. The prevalence of *CDKN2A* SCNVs in baseline follow-up esophageal mucosal biopsy samples from patients with baseline mild or moderate esophageal squamous cell dysplasia (ESCdys) and different follow-up experiences, including progression to severe ESCdys or ESCC (orange color), persistence as mild or moderate ESCdys (yellow color), or regression to inflammation or normal (green color). Deletion (blue), somatic *CDKN2A* deletion; Amplification (red), somatic *CDKN2A* amplification; Diploid (grey), no SCNV; not informative (white), the amount of genetic DNA from biopsies was not enough for P16-Light analysis

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted September 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Supplemental Figure S4. The proportion of various *CDKN2A* SCNVs in the followup esophageal mucosa biopsy samples from baseline ESCdys patients with and without *CDKN2A* SCNVs. No statistically significant difference of the proportion of *CDKN2A* SCNVs was observed between the baseline Diploid group and baseline Amplification group. The exact number of samples in each subgroup is labeled. The chi-square values and *p*-values are also listed between groups.

Supplemental Figure S5. Approximate locations of the estimated common deletion fragment within tumor suppressor genes *ATM*, *FAT1*, *RB1*, *PTEN*, and *miR31HG* in TCGA pan-cancers (by Affymetrix SNP6.0 microarray).

Supplemental Figure S6. Images of pathological lesions in esophageal mucosa biopsied from two representative patients with mild ESCdys at baseline and during the followup in the regressive, persistent, and progressive groups. H.E. staining; X100

SUPPLEMENTARY TABLE LIST

Supplemental Table S1. Base-resolution coordinates for *CDKN2A* deletions in 110 cancer samples

Supplemental Table S2. Clinicalpathological characteristics of all 205 patients with mild or moderate ESCdys at baseline and the status of *CDKN2A* SCNV in baseline ESCdys

Supplemental Table S3. Comparison of the prevalence of *CDKN2A* SCNVs between patients with various environmental and personal factors

Supplemental Table S4. Comparison of the prevalence of somatic *CDKN2A* gene copy number variations (SCNVs) in baseline esophageal squamous cell dysplasia biopsy samples from patients in the progression group and the persistent group in the multicenter followup study

Supplemental Table S5. Oligo sequences