medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

# Gene-Specific Common Deletion Regions and Detection of Gene Copy Number Deletion

**Running title**: Gene-specific common deletion regions

Yuan Tian, Ph.D. [1#], Rui Xing, Ph.D. [2#], Jing Zhou[1#], Juanli Qiao, Ph.D. [1], Zhaojun Liu, Ph.D. [1], Liankun Gu[1], Baozhen Zhang, Ph.D. [1], Youyong Lu, M.D. [2], Dajun Deng, M.D. [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

# Equal first authors

***Corresponding author**: 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)

**Word count**: 3945



**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**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 May 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.


## Abstract

A feasible method to detect somatic copy number deletion (SCND) of tumor suppressor genes is still absent up to date. To overcome the obstacle of SCND detection, we analyzed genomic coordinates of gene deletion fragments using the Catalogue Of Somatic Mutation In Cancer (COSMIC) datasets. Interstitial base-resolution deletion/fusion coordinates for *CDKN2A* were extracted from published articles and our whole genome sequencing (WGS) datasets. The results showed that estimated common deletion regions (CDRs) were observed in many tumor suppressor genes, such as *ATM*, *CDKN2A*, *FAT1*, *miR31HG*, *PTEN*, and *RB1* in the SNP array-based COSMIC datasets. A 5.1-kb base-resolution CDR could be identified in >90% cancer samples by sequencing. Using the true *CDKN2A* CDR as a PCR target, a multiplex quantitative PCR assay P16-Light was programmed to detect *CDKN2A* gene copy number with a lower detection limit of 20%. P16-Light was further confirmed with WGS as the gold standard among cancer tissue samples from 139 patients. In conclusion, CDRs are common in many tumor suppressor genes. The 5.1-kb *CDKN2A* CDR was found in >90% cancers containing *CDKN2A* deletion. The CDKN2A CDR was used a potential target for developing the P16-Light assay to detect *CDKN2A* SCND and amplification for routine clinical practices.


**Key words:** somatic copy number variation (SCNV); common deleted region (CDR); tumor suppressor gene; *CDKN2A*; whole genome sequencing (WGS)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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 copy number variations (SCNVs) of tumor related genes are landmarks of human cancers (1,2). Somatic copy number deletion (SCND) and amplification are two kinds of well-known SCNVs. However, current gene copy number detection methods, including microsattelite instability (MSI), loss/gain of heterozygosity (LOH/GOH), fluorescence-in-situ hybridization (FISH), whole genome sequencing (WGS) or whole exome sequencing (WES), are not sensitive enough or 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 SCND of tumor suppressor genes, including *CDKN2A*, are still rare owing to the lack of a feasible detection assay.

The frequency of *CDKN2A* SCND detected by single nucleotide polymorphism (SNP) microarray, WGS or WES was found to range from 30% to 60% in bladder cancer, melanoma, head and neck cancer, pleural mesothelioma, glioblastoma, and esophageal squamous cell cancer (ESCC) with an average frequency of 13% in pan-cancer datasets in The Cancer Genome Atlas (TCGA) (Figure S1A) (2-6). *CDKN2A* deep-deletion is associated with downregulation of the *CDKN2A* gene expression, while *CDKN2A* amplification is associated with upregulation of the *CDKN2A* gene expression in Pan-TCGA cancers (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 (7-11). Therefore, a convenient and sensitive assay to detect *CDKN2A* SCND is eagerly awaited.

In the present study, we characterized patterns of estimated genomic coordinates for SCNDs in a set of tumor suppressor genes using public Catalogue Of Somatic Mutations In Cancer (COSMIC) SCNV datasets and found common deletion regions (CDRs) in many frequently deleted genes. Then, we further defined a 5.1-kb base-resolution CDR within the *CDKN2A* gene using sequencing data for the first time. A sensitive P16-Light assay targeted to the *CDKN2A* CDR was established for clinical practices.

# Methods & Material

**COSMIC and TCGA SCNV datasets**

SNP6 array-based Estimated genomic coordinates of interstitial copy number deletion/fusion of the *CDKN2A* gene in cancer cell lines (n=273) with homozygous *CDKN2A* deletion and estimated genomic coordinates of deep-deleted fragments of *CDKN2A*, *PTEN*, *RB1*, and other frequently deleted genes in cancer tissues were downloaded from were downloaded from the Copy Number Analysis (CONA) datasets in the COSMIC project (Data file 1-11) (12).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

**Patients, tissues, and DNA preparation**

Frozen fresh GC and paired surgical margin (SM) tissue samples were collected from 156 patients in the WGS study (13). These samples were frozen in liquid nitrogen about 30 min after surgical dissection and then stored in a -80$^{o}$ C refreezor for 2-5 yrs. Clinicopathologic information were also obtained. The 2010 UICC tumor-node-metastasis (TNM) system was used to classify these GCs (14). Genomic DNA was extracted from these samples with a phenol/chloroform method coupled with RNase treatment. Concentrations of these DNA samples were determined with NanoVue Plus (Biochrom LTD, Cambridge, UK). DNA samples with $OD_{260nm}/OD_{280nm}$ ratios ranging from 1.7 to 1.9 were used for detection of gene copy number as described below.

**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 conservative sequences within the CDR in the *CDKN2A* (HGNC: 1787) and *GAPDH* (HGNC: 4141) gene sequences by Bacon Designer 8 software. Multiplex PCR assays were established according to the Applied Biosystems (ABI) TaqMan universal PCR master mix manual. The performance of these assays for the detection of *CDKN2A* copy numbers was compared with each other. Finally, a multiplex primer and probe combination targeted to *CDKN2A* intron-2 was selected (Table 1) and their components' concentrations were optimized. Each multiplex PCR assay was carried out in a total volume 20 µL 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 fewer cycles, this sample was considered as *CDKN2A* SCNV informative. The specificity of PCR reaction was monitored through running gel. Distilled water was used as no template control for each experiment.

**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

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

number for the *CDKN2A* gene were calculated using the *GAPDH* gene as the internal reference. When the *CDKN2A* copy number in the A549-diluted template was consistently lower than that in the RKO control template, and the difference was statistically significant (t-test, $p<0.05$), 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 SM sample, the sample was defined as somatic *CDKN2A* CDR deletion-positive or amplification-positive, respectively. For each experiment, the 100% A549, 100% RKO, and 20% A549 + 80% RKO DNA mix controls were analyzed.

**Quantitative detection of *CDKN2A/P14$^{ARF}$* exon-1β copy number by PCR assay (P14-qPCR)**

The copy number of *CDKN2A* exon-1β was detected uisng the primer and probe set (Table 1) as previously reported (15). When the relative copy number of *CDKN2A* exon-1β in a tissue sample was significantly lower or higher than that of the paired SM sample, the sample was defined as somatic *CDKN2A/P14$^{ARF}$* deletion-positive or amplification-positive, respectively.

**Call for *CDKN2A* interstitial deletion/fusion and calculate purity of cancer cells in the GC WGS datasets**

We used 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 (13). This method used soft-clipped and split reads to identify candidate breakpoints, and precise breakpoints were refined by local alignments. *CDKN2A* deletion information of 157 GC samples was obtained from WGS datasets. We also estimated copy number profiling over 10-kb windows with Patchwork 28 and calculated the ratio of standardized average depth between normal tissue and tumor tissue (log2R ratio). The purity and ploidy of each tumor were calculated using ABSOLUTE software (16).

**Cell lines and cultures**

The *CDKN2A* allele homozygously 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 *CDKN2A* 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. A Goldeneye™ 20A STR Identifiler PCR Amplification kit was used to analyze the STR patterns.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

### Statistical analysis

Chi-square or Fisher's exact tests were used to compare the proportion of *CDKN2A* SCND or amplification between different groups of tissue samples. Student t-test was used to compare the proportion of the *CDKN2A* gene copy number between genomic DNA samples. All statistical tests were two-sided, and *p* value less than 0.05 was considered to be statistically significant.

## Results

### Prevalence of estimated CDRs within various tumor suppressor genes

It has been previously reported that homozygous deletion of approximately 170 kilobase pairs (kb), including the *CDKN2A* locus, can be detected in human cancers by MSI analyses (17). SCND inactivates the *CDKN2A* gene in 273 human cancer cell lines according to the COSMIC dataset (Data file 1). We found that an 8-kb estimated *CDKN2A* CDR could be detected among these cell lines through ording "start" coordinates of these breaking points (Figure S1). To investigate the prevalence of CDRs within tumor suppressor genes in human cancer tissues with a high deletion frequency (1,2), we further downloaded the estimated genomic coordinates for deletion fragments that overlap with these genes, respectively. We found that CDRs could be detected not only within the *CDKN2A* gene (Figure 1A; about 17-kb), but also within *ATM* (middle to downstream), *FAT1* (promoter to middle), *miR31HG* (promoter to exon-1), *PTEN* (promoter to exon-1), and *RB1* genes (promoter to intron-2) (Figure 2; about 158-kb, 23-kb, 33-kb, 5-kb, and 2442-kb, respectively) (Data file 2-7). No CDR could be observed within *CCSER1*, *FHIT*, *LRP1B*, and *WWOX* genes according to the SNP-array data (Data file 8-11).

### Characterization of a true *CDKN2A* CDR at the base-resolution in human cancers

It was reported that the error in *CDKN2A* breakpoint estimation based on SNP-array data is roughly 10-kb (18). To characterize true genomic coordinates of *CDKN2A* deletion fragments in cancers, we extracted base-resolution sequence information of interstitial *CDKN2A* deletions from available published articles and our sequencing data (Data file 12) (19-28). 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 1B, blue lines). This CDR sequence is the same as the *CDKN2A* deletion fragment in the HCC193 lung cancer cell line (25). The CDR coordinates were also confirmed in our WGS datasets (average sequencing depth, 36×) of 18 (100%) of 18 GCs (13), in which interstitial *CDKN2A* deletions/fusions were identified (Figure 1B, purple lines; Data file 12).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

It is well known that germline *CDKN2A* inactivation can lead to a high predisposition for melanoma and pancreatic cancer (29-31). Interestingly, we found that 14 (93.3%) of 15 *CDKN2A* allelic variants in the Online Mendelian Inheritance in Man (OMIM) database are located within the CDR sequence, especially in *CDKN2A* exon-2 (Figure S3) (32,33).

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

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

The current clinical method FISH for detecting SCND is composed of a set of biotin-labeled probes that should cover at least 50-kb DNA sequence. Thus, FISH is not a suitable method for detecting the copy number deletion of the 5.1-kb *CDKN2A* CDR. To provide a convenient way for routine clinical use, we designed and experimentally evaluated a set of multiplex quantitative PCR assays and finally optimized the *CDKN2A* CDR-specific quantitative multiplex PCR assay called P16-Light for detecting the copy number of a 129-bp amplicon within the *CDKN2A* intron-2 (Figure 3A), which covers 86% (94/110) of known *CDKN2A* deletion fragments (Figure 1B, violet line).

The copy number of the *GAPDH* gene was used as the internal reference. Genomic DNA from human A549 cells (with homozygous deletion of *CDKN2A* alleles) and RKO cells (with 2 wild-type *CDKN2A* alleles) were used as *CDKN2A* CDR deletion-positive and deletion-negative controls, respectively. The amplification efficiency of two amplicons in *GAPDH* and *CDKN2A* were very similar (Figure 3B). No template inhibition was abserved when the amount of template DNA ranged from 10 to 0.63 ng (Figure 3C). 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 (10 ng/reaction) when the A549 DNA was spiked in at different proportions for the P16-Light analyses (Figure 3D). Furthermore, there was a high reproducibility when DNA with homozygous deletion of *CDKN2A* was present in ≥20% of the cells verified in ten experimental repeats performed on different days (Figure 3E). Thus, when the proportion of *CDKN2A* copy number was significantly decreased (or increased) in a sample relative to the paired normal control (t-test, $p<0.05$) in the P16-Light analyses, the sample would be defined as *CDKN2A* SCND-positive (or amplification-positive).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

**Comparison of P16-Light with WGS datasets**

As we described above, information on interstitial copy number deletion/fusion of the *CDKN2A* gene was extracted from WGS datasets for 156 of 168 GC patients enrolled in a GC genome study (13) and a total of 18 *CDKN2A* deletion/fusion coordinates at the base-resolution were detected in 17 (10.8%) GCs (Data file 12 and 13). To compare the performance of P16-Light with WGS, we analyzed the status of SCNVs, including SCND and amplification, of the *CDKN2A* gene in 156 of these GCs with enough amounts of genomic DNA samples with P16-Light using the paired surgical margin tissue (SM) as the diploid reference (Data file 13). *CDKN2A* SCND and amplification were detected in 40 (25.6%) and 34 (21.8%) of these GCs, respectively. The P16-Light analysis was confirmed by the WGS results: the frequency of *CDKN2A* SCND (or amplification) by P16-Light was significantly higher (or lower) in 17 GCs containing interstitial *CDKN2A* deletion/fusion than those 139 GCs without (chi-square test, $p<0.028$; Figure 4A). These results also indicate that there is a significantly higher sensitivity for detecting *CDKN2A* SCND by the quantitative P16-Light assay than the hemi-quantitative WGS.

Moreover, it is well known that the propertion of cancer cells in tissue samples (ie. sample purity) may affect the detecton values of various genome data. To study whether cancer cell proportion disturbs detection of *CDKN2A* SCNVs, we calculated cancer cell propertion in above GC samples using WGS data (Data file 13). We found that the difference in sample purity between GC subgroups with the different status of *CDKN2A* SCNVs was not statistically significant (t-test, $p=0.075$; Figure 4B), although the propertion was slightly higher in the GCs with *CDKN2A* SCND than thoses without. No correlation was observed between the proportion of cancer cells and relative copy number of the *CDKN2A* gene among these GCs (Figure 4C).

**Comparison of P16-Light with P14-qPCR assay**

The P14-qPCR assay was previously established for detecting the copy number of *CDKN2A/P14*$^{ARF}$ exon-1β (15). Two amplicons in the P16-Light and P14-qPCR assays cover 98% (108/110) of known *CDKN2A* deletion fragments (Figure 1B, violet and green lines). Therefore, we further compared the performance of P16-Light, P14-qPCR, and their combination using GCs and paired SM samples from patients who were recently included in the cross-sectional cohort in our association study (34). GC samples ($n=139$) with enough amount of genomic DNA were used in P14-qPCR analysis (Data file 14). The SCND-positive rate for *P14*$^{ARF}$ was similar to that for the *CDKN2A* CDR (31.7% vs 36.7%) (Table 2). *CDKN2A* SCND was found only in 19 GCs by both assays. While *CDKN2A* CDR SCND was significantly associated with distant metastasis of GC

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

(odd ratio=4.09, $p$<0.001), no association was observed between *P14*$^{ARF}$ SCND and GC metastasis. Using merged *CDKN2A* SCND data (*CDKN2A* CDR SCND-positive and/or *P14*$^{ARF}$ SCND-positive), only a weaker association was observed. These results suggest that individual P16-Light alone may be good enough for detecting *CDKN2A* SCND in tissue samples.

## Discussion

Somatic copy number deletion and amplication are two main kinds of SCNVs. The detection of copy number amplification of oncogenes is routinely used for precise cancer treatments. However, the detection of SCND of tumor suppressor genes is absent and its significance in clinical practices is not well studied. The reason should be the lack of feasible detection approches. Here, we report that there are CDRs in many tumor suppressor genes, such as *CDKN2A*, *miR31HG*, *PTEN*, and *RB1*, which are commonly inactivated by SCND in various human cancers (1,2). Notably, we characterized, for the first time, the 5.1-kb true CDR from the *CDKN2A*/*P16*$^{INK4A}$ promoter to intron-2 in >90% *CDKN2A*-deleted cancers. Using the *CDKN2A* CDR as a PCR target, we further established the feasible P16-Light assay to detect *CDKN2A* SCND and amplification in tissue samples. These findings indicate that CDRs are prevalent sequences in tumor suppressor genes, and characterization of the base-resolution genomic coordinates of CDRs could enable us to setup conventient methods for SCND detection of genes.

Interstitial deletion/fusion is the main type of *CDKN2A* SCND and the breaking/fusing coordinates for *CDKN2A* SCNDs in cancer genomes are diverse, which blocks the establishment of a feasible detection assay for *CDKN2A* SCND, although many efforts have been afforded (20). In the present study, we initially found the 8~17-kd estimated *CDKN2A* CDR in both monoclonal cancer cell lines and cell-heterogenous cancer tissues with *CDKN2A* copy number deletion according to the SNP-array datasets from COSMIC and TCGA projects (1,12). Then, we further characterized the 5.1-kb true CDR at the base-resolution within the *CDKN2A* gene in cancer genomes using DNA sequencing data (19-28), and confirmed the CDR using WGS datasets in all 18 GCs containing *CDKN2A* SCND (13). Because the true *CDKN2A* CDR was observed in more than 90% of *CDKN2A*-deleted cancer samples and the P16-Light assay is highly reproducible and convenient, the quantitative P16-Light assay should be considered as a viable assay for detecting *CDKN2A* SCNVs in clinical practices. This is supported by the result that *CDKN2A* SCND by P16-Light was significantly associated with GC metastasis, and further supported by the results of our prospective study, in which *CDKN2A* SCND was closely associated with hematogenous metastasis of GCs (34). In another long-term prospective study, we also found that *CDKN2A* SCND and amplification by P16-Light were significantly associated with

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

malignant transformation and complete regression of mild or moderate esophageal squamous cell dysplasia, respectively (Fan et al. submitted for publication). The results of these studies also suggest that the sensitivity of 20% for the P16-Light assay may be good enough for routine clinical use.

WGS is generally used as a kind of gold standard to study structural alterations of genomic DNAs, especially for insterstitial gene copy deletion/fusions. However, WGS is a cost assay and its accuracy depends on sequencing depth, WGS at sequencing depth 36× would be considered as the hemi-quantitative assay. In our calling *CDKN2A* SCND coordinate processes, it was found that 18 *CDKN2A* SCND coordinates were identified in 17 (10.8%) of 157 GCs, which was consistent with the frequency (11.4% =50/438) of homozygous deletion of *CDKN2A* in GCs in WES or WGS sequencing datasets (Data file 14) (35). The positive rate (25.6%) of *CDKN2A* SCND in 156 GCs with enough genomic DNA samples in the P16-Light analysis was more than twice of WGS. And the results of P16-Light analyses were significantly correlated with those of WGS. These phenomena suggest that P16-Light is a much more sensitive, convenient, and less expensive assay than WGS.

P14-qPCR is a method used to detect the copy number of *CDKN2A/P14*$^{ARF}$ exon-1 (15). Although the combination of P16-Light with P14-qPCR may detect both SCNDs overlapped with the *CDKN2A* CDR and not overlapped with the *CDKN2A* CDR, the results of our comparison analysis among 139 GC patients showed that detecting *CDKN2A* SCND by individual P16-Light may be good enough for clinical practices because combination with P14-qPCR could not improve the performance of P16-Light. However, for other genes such as *RB1* and *PTEN*, whether a qPCR array needs to be employed for detecting their SCNVs should be studied case by case.

Generally, IHC is a popular method used to detect expression changes of protein-coding genes. For example, P16$^{INK4A}$ overexpression in cervical mucosa samples is currently used for rapidly HPV infection screening. We compared the status of P16$^{INK4A}$ expression by IHC between GCs with *CDKN2A* SCND (n=4) and without *CDKN2A* SCND (n=12) and did not find any difference in the P16$^{INK4A}$ positive-staining rate between these GCs (3/4 vs. 9/12). The expression level of *CDKN2A/P16*$^{INK4A}$ is not only affected by SCNVs, but also regulated by the methylation status of CpG island, histone modifications, and high-risk HPV infection (36,37). These factors may partially account for the insistence between IHC and P16-Light.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

humans (38,39). The exact mechanisms leading to the difference among species is still unclear. Here, we reported that approximately 87% of genetic *P16*$^{INK4a}$ inactivation by *CDKN2A* SCND 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. Th report supports this explanation that knocking out both *p16*$^{ink4a}$ and *p19*$^{arf}$ leads to more cancer development than individual inactivation in mice (40). This also may account that *P14*$^{ARF}$ exon-1β deletion was not associated with GC metastasis whereas *CDKN2A* CDR deletion was significantly associated with GC metastasis as described above.

In conclusion, we have found estimated CDRs in many tumor suppressor genes in the cancer genome. And there is a 5.1-kb CDR region within the *CDKN2A* gene, and most *CDKN2A* deletions lead to *P16*$^{INK4a}$ and *P14*$^{ARF}$ inactivation in human cancers. Using the *CDKN2A* CDR as a target sequence, we developed a convenient quantitative multiplex PCR assay, the P16-Light, to detect *CDKN2A* SCNVs in clinical practice, suggesting that the strategy to detect *CDKN2A* SCNVs may be suitable for the establishment of SCNV detection methods for other tumor suppressor genes.

## Abbreviations

CDR, common deletion region;

DHPLC, denaturing high performance liquid chromatography

GC, gastric carcinoma

LOH, loss of heterozygosity

MSI, microsatellite instability

OMIM, the Online Mendelian Inheritance in Man;

SCND, somatic copy number deletion

SM, surgical margin

TCGA, The Cancer Genome Atlas

WGS, whole genome sequencing

**Ethical approval and consent to participate:** The Institution Review Board of Peking University Cancer Hospital & Institute approved this study and was carried out in accordance with the principles outlined in the Declaration of Helsinki. Informed consent was obtained from each patient prior to their inclusion in the study.

**Consent for publication**: Not applicable.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

**Availability of supporting data**: Detailed information on the base-resolution interstitial common deletion region in the *CDKN2A* gene in 110 cancer samples and other data were provided in Supplementary Data files 1-14.

**Competing Insterests**: DD, YT, JZ, and ZL are the creators for the pending patent "A quantitative method for detection of human CDKN2A gene copy number using a primer set and their applications" (PCT/CN2019/087172; WO2020228009). All of other authors have nothing to disclose.

**Funding**: This work was supported by the Beijing Natural Science Foundation (grant number 7181002 to DD); Beijing Capital's Funds for Health Improvement and Research (grant number 2018-1-1021 to DD.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author contributions**: DD designed the study, performed bioinformatics analyses, characterized base resolution *CDR* coordinates in *CDKN2A* and other genes, and wrote the manuscript draft. JZ, YT, ZJL, and LKG developed the P16-Light assay and analyzed *CDKN2A* SCNVs in samples. RX and YYL provided the *CDKN2A CDR* coordinates from whole genome sequencing datasets. All authors approved the final manuscript.

**Acknowledgements**: We thank Dr. Sanford Dawsey at NCI, NIH, Bethesda, Maryland for critical comments for the manuscript. We also thank Miss Gina Mckeown in New York, USA for English language editing.

**Authors' information**: Professor Dajun Deng graduated from Peking Union Medical College in 1986 and employed by Peking University Cancer Hospital (Beijing Institute for Cancer Research) from then on. His current research activities are focused on genetic and epigenetic *CDKN2A* inactivation in cancer development. He published approximately 150 scientific papers, including Mol Cancer, Nucl Acids Res, Genome Biol, Proc Nat Acad Sci USA, Clin Cancer Res, and EBioMed. His ORCiD is 0000-0001-5161-5943.

## References

1. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, et al. The landscape of somatic copy-number alteration across human cancers [eng]. Nature 2010 Feb;463 7283:899-905 as doi: 10.1038/nature08822.
2. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data [eng]. Cancer Discov 2012 May;2 5:401-4 as doi: 10.1158/2159-8290.CD-12-0095.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

3. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers [eng]. Genome Biol 2011;12 4:R41. Epub 2011/04/28 as doi: 10.1186/gb-2011-12-4-r41.

4. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal [eng]. Sci Signal 2013 Apr;6 269:pl1. Epub 2013/04/02 as doi: 10.1126/scisignal.2004088.

5. Network CGAR, University AWGA, Agency BC, Hospital BaWs, Institute B, University B, University CWR, et al. Integrated genomic characterization of oesophageal carcinoma [eng]. Nature 2017 01 12;541 7636:169-75. Epub 20170104 as doi: 10.1038/nature20805.

6. Cui Y, Chen H, Xi R, Cui H, Zhao Y, Xu E, Yan T, et al. Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma [eng]. Cell Res 2020 10;30 10:902-13. Epub 20200512 as doi: 10.1038/s41422-020-0333-6.

7. Serrano M, Lee H, Chin L, Cordon-Cardo C, Beach D, DePinho R. Role of the INK4a locus in tumor suppression and cell mortality. [eng]. Cell 1996 Apr;85 1:27-37.

8. Deng J, Wang ES, Jenkins RW, Li S, Dries R, Yates K, Chhabra S, et al. CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-cell Activation [eng]. Cancer Discov 2018 02;8 2:216-33. Epub 2017/11/03 as doi: 10.1158/2159-8290.CD-17-0915.

9. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, Leeson R, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade [eng]. Cell 2018 11;175 4:984-97.e24 as doi: 10.1016/j.cell.2018.09.006.

10. Zhang J, Bu X, Wang H, Zhu Y, Geng Y, Nihira NT, Tan Y, et al. Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance [eng]. Nature 2018 01;553 7686:91-5. Epub 2017/11/16 as doi: 10.1038/nature25015.

11. Yu J, Yan J, Guo Q, Chi Z, Tang B, Zheng B, Yin T, et al. Genetic Aberrations in the CDK4 Pathway Are Associated with Innate Resistance to PD-1 Blockade in Chinese Patients with Non-Cutaneous Melanoma [eng]. Clin Cancer Res 2019 Nov;25 21:6511-23. Epub 2019/08/02 as doi: 10.1158/1078-0432.CCR-19-0475.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

12. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer [eng]. Nucl Acids Res 2019 Jan; 47 D1:D941–7. Epub 2019/01/08 as doi: 10.1093/nar/gky1015.

13. Xing R, Zhou Y, Yu J, Yu Y, Nie Y, Luo W, Yang C, et al. Whole-genome sequencing reveals novel tandem-duplication hotspots and a prognostic mutational signature in gastric cancer [eng]. Nat Commun 2019 05;10 1:2037. Epub 2019/05/02 as doi: 10.1038/s41467-019-09644-6.

14. Sobin L, Gospodarowicz M, Wittekind C. (2009) TNM Classification of Malignant Tumours, 7th edn. International Union Against Cancer (UICC). Wiley Press, New York, USA.

15. Berggren P, Kumar R, Sakano S, Hemminki K, Wada T, Steineck G, Adolfsson J, et al. Detecting homozygous deletions in the CDKN2A(p16INK4a)/ARF(p14ARF) gene in urinary bladder cancer using real-time quantitative PCR [eng]. Clin Cancer Res 2003 01;9： 235-42.

16. Carter SL, Cibulskis K, Helman E, Mckenna A, Shen H, Zack T, Laird PW. et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012 05;30 5:413–21. as doi: 10.1038/nbt.2203

17. Cairns P, Polascik TJ, Eby Y, Tokino K, Califano J, Merlo A, Mao L, et al. Frequency of homozygous deletion at p16/CDKN2 in primary human tumours [eng]. Nat Genet 1995 Oct;11 2:210-2 as doi: 10.1038/ng1095-210.

18. Novara F, Beri S, Bernardo ME, Bellazzi R, Malovini A, Ciccone R, Cometa AM, et al. Different molecular mechanisms causing 9p21 deletions in acute lymphoblastic leukemia of childhood [eng]. Hum Genet 2009 Oct;126 4:511-20. Epub 20090530 as doi: 10.1007/s00439-009-0689-7.

19. Xie H, Rachakonda PS, Heidenreich B, Nagore E, Sucker A, Hemminki K, Schadendorf D, et al. Mapping of deletion breakpoints at the CDKN2A locus in melanoma: detection of MTAP-ANRIL fusion transcripts [eng]. Oncotarget 2016 Mar 29;7 13:16490-504 as doi: 10.18632/oncotarget.7503.

20. Patel A, Schwab R, Liu YT, Bafna V. Amplification and thrifty single-molecule sequencing of recurrent somatic structural variations [eng]. Genome Res 2014 Feb;24 2:318-28. Epub 20131204 as doi: 10.1101/gr.161497.113.

21. Norris AL, Kamiyama H, Makohon-Moore A, Pallavajjala A, Morsberger LA, Lee K, Batista D, et al. Transflip mutations produce deletions in pancreatic cancer [eng]. Genes Chromosomes Cancer 2015 Aug;54 8:472-81. Epub 20150529 as doi: 10.1002/gcc.22258.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

22. Guney S, Bertrand P, Jardin F, Ruminy P, Kerckaert JP, Tilly H, Bastard C. Molecular characterization of 9p21 deletions shows a minimal common deleted region removing CDKN2A exon 1 and CDKN2B exon 2 in diffuse large B-cell lymphomas [eng]. Genes Chromosomes Cancer 2011 Sep;50 9:715-25. Epub 20110602 as doi: 10.1002/gcc.20893.

23. Florl AR, Schulz WA. Peculiar structure and location of 9p21 homozygous deletion breakpoints in human cancer cells [eng]. Genes Chromosomes Cancer 2003 Jun;37 2:141-8 as doi: 10.1002/gcc.10192.

24. Kitagawa Y, Inoue K, Sasaki S, Hayashi Y, Matsuo Y, Lieber MR, Mizoguchi H, et al. Prevalent involvement of illegitimate V(D)J recombination in chromosome 9p21 deletions in lymphoid leukemia [eng]. J Biol Chem 2002 Nov 29;277 48:46289-97. Epub 20020912 as doi: 10.1074/jbc.M208353200.

25. Sasaki S, Kitagawa Y, Sekido Y, Minna JD, Kuwano H, Yokota J, Kohno T. Molecular processes of chromosome 9p21 deletions in human cancers [eng]. Oncogene 2003 Jun;22 24:3792-8 as doi: 10.1038/sj.onc.1206589.

26. Cayuela JM, Gardie B, Sigaux F. Disruption of the multiple tumor suppressor gene MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic leukemias [eng]. Blood 1997 Nov 01;90 9:3720-6.

27. Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bièche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF [eng]. Cancer Res 2007 Apr 15;67 8:3963-9 as doi: 10.1158/0008-5472.CAN-06-2004.

28. Raschke S, Balz V, Efferth T, Schulz WA, Florl AR. Homozygous deletions of CDKN2A caused by alternative mechanisms in various human cancer cell lines [eng]. Genes Chromosomes Cancer 2005 Jan;42 1:58-67 as doi: 10.1002/gcc.20119.

29. Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH, et al. Germline p16 mutations in familial melanoma [eng]. Nat Genet 1994 Sep;8 1:15-21 as doi: 10.1038/ng0994-15.

30. Freedberg DE, Rigas SH, Russak J, Gai W, Kaplow M, Osman I, Turner F, et al. Frequent p16-independent inactivation of p14ARF in human melanoma [eng]. J Natl Cancer Inst 2008 Jun;100 11:784-95. Epub 2008/05/27 as doi: 10.1093/jnci/djn157.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

31. Harinck F, Kluijt I, van der Stoep N, Oldenburg RA, Wagner A, Aalfs CM, Sijmons RH, et al. Indication for CDKN2A-mutation analysis in familial pancreatic cancer families without melanomas [eng]. J Med Genet 2012 Jun;49 6:362-5. Epub 2012/05/25 as doi: 10.1136/jmedgenet-2011-100563.

32. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders [eng]. Nucleic Acids Res 2005 Jan;33 Database issue:D514-7 as doi: 10.1093/nar/gki033.

33. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM) [eng]. Nucleic Acids Res 2009 Jan;37 Database issue:D793-6. Epub 2008/10/08 as doi: 10.1093/nar/gkn665.

34. Qiao JL, Tian Y, Cheng XJ, Liu ZJ, Zhou J, Gu LK, Zhang BZ, et al. CDKN2A deletion leading to hematogenous metastasis of human gastric carcinoma. Frontiers in Oncology 2021 Dec 06; DOI: 10.3389/fonc.2021.801219.

35. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, Seoane JA, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas [eng]. Cancer Cell 2018 04 09;33 4:721-35.e8. Epub 20180402 as doi: 10.1016/j.ccell.2018.03.010.

36. Peters G. Tumor suppression for ARFicionados: the relative contributions of p16INK4a and p14ARF in melanoma [eng]. J Natl Cancer Inst 2008 Jun 04;100 11:757-9. Epub 20080527 as doi: 10.1093/jnci/djn156.

37. Cui CH, Gan Y, Gu LK, Wilson J, Liu ZJ, Zhang BZ, Deng DJ. P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis [eng]. Genome Biol 2015 Nov 23;16:252. doi: 10.1186/s13059-015-0819-6.

38. Li Q, Wang XH, Lu ZM, Zhang BZ, Guan ZP, Liu ZJ, Zhong QM, et al. Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9 at the p16 locus [eng]. PLoS One 2010 Oct 29;5(10):e13732. doi: 10.1371/journal.pone.0013732.

39. Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, Blasco MA, et al. The Ink4/Arf locus is a barrier for iPS cell reprogramming [eng]. Nature 2009 Aug 27;460 7259:1136-9. Epub 20090809 as doi: 10.1038/nature08290.

40. Sharpless NE, Ramsey MR, Balasubramanian P, Castrillon DH, DePinho RA. The differential impact of p16(INK4a) or p19(ARF) deficiency on cell growth and tumorigenesis [eng]. Oncogene 2004 Jan;23 2:379-85 as doi: 10.1038/sj.onc.1207074.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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 Legends

**Figure 1.** Genomic coordinates of interstitial *CDKN2A* deletion/fusion in human cancer genomes. (**A**) Estimated coordinates of *CDKN2A* deep-deletion in cancer tissues according to the COSMIC data. (**B**) True coordinates at the base-resolution of *CDKN2A* deletion in cancer cell lines (*n*=92, blue lines) and gastric cancer (*n*=18, purple lines) according to sequencing data. Two top charts display the coordinates of most deletion fragments. The sample ID is labeled under each column. Two bottom charts display the amplified view of these deletion fragments, where the 17-kb and 5.1-kb common deletion regions (CDRs) are highlighted with a red dashed line rectangle. The 5.1-kb true CDR from the $P16^{INK2a}$ promoter to intron-2 is exactly the same region as the deleted *CDKN2A* fragment in the HCC193 lung cancer cell line (highlighted with a black line). Each line represents a *CDKN2A* deletion fragment. The locations of $P16^{INK4a}$ and $P14^{ARF}$ (grey shadow) and exon-1α/1β/2/3 (black dots) are also labeled as landmarks. The position of amplicons for P16-Light and P14-qPCR are respectively illustrated with violet and green lines. The detailed deletion coordinates for each sample are listed in Data file 2 and Data file 12.

**Figure 2.** The estimated coordinates of deep-deletion fragments overlaped with tumor suppressor genes *ATM*, *FAT1*, *RB1*, *PTEN*, and *miR31HG* according to the COSMIC data. Common deletion region (CDR) for each gene is highlighted with a red line rectangle. The detailed deletion coordinates for each sample are listed in Data file 3 and Data file 11.

**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 amplication efficiency of two amplicons for *GAPDH* and *CDKN2A* genes in the template titration assays using standard DNA samples from RKO cells (with two wild-type *CDKN2A* alleles) and A549 cells (with a homozygous *CDKN2A* deletion). (**C**) Effects of the amount of template DNA on the efficiency of PCR amplification for amplicons in *CDKN2A* and *GAPDH* genes (Left chart) and detection of the relative *CDKN2A* gene copy number (Right chart). (**D**) The linear relationship between the proportion of *CDKN2A* copy number deletion and ratios of RKO cells spiked with A549 cells. (**E**) 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

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

experiments performed on different days. *: compared to the 100% RKO control in chi-square test, $p<0.05$.

**Figure 4.** Comparisons of somatic copy number variations (SCNVs) of *CDKN2A* gene in gastric carcinoma samples (GCs) from 156 patients in the P16-Light and WGS (30×) analyses. (**A**) The states of *CDKN2A* SCNVs by P16-Light (relative to paired surgical margin reference) in GC groups with and without *CDKN2A* deletion/fusion in the WGS analysis. (**B**) Comparison of the proportion of cancer cells (or sample purity; by WGS) in GC groups with various *CDKN2A* SCNVs by P16-Light. The average proportion of cancer cells in each GC group is labeled. (**C**) Correlation analysis between the proportion and relative copy number of the *CDKN2A* gene in GCs.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

1 **Table 1**. Oligo sequences

| Gene/STS | Assay | Oligo | Sequence (5'-3') | PCR product size |
|---|---|---|---|---|
| *CDKN2A* | P16-Light | F-primer | caggtctgtttcctcatttg | 129-bp |
| | P16-Light | R-primer | ggtcagattagttgagttgtg | |
| | P16-Light | Probe | FAM-ctggctggaccaacctcagg-BHQ1 | |
| | P14-qPCR | F-primer | ggaggcggcgagaacat | 92-bp |
| | P14-qPCR | R-primer | tgaaccacgaaaaccctcact | |
| | P14-qPCR | Probe | VIC-tgcgcaggttcttggtgaccctcc-TAMRA | |
| *GAPDH* | P16-Light | F-primer | gctcacatattctggaggag | 135-bp |
| | P16-Light | R-primer | ggtcattgatggcaacaata | |
| | P16-Light | Probe | Cy5-tgccttcttgcctcttgtctctt-BHQ2 | |
| | P14-qPCR | F-primer | ccactaggcgctcactgttct | 97-bp |
| | P14-qPCR | R-primer | gcgaactcacccgttgact | |
| | P14-qPCR | Probe | FAM-ctccctccgcgcagccgagc-TAMRA | |

2

3

medRxiv preprint doi: https://doi.org/10.1101/2021.09.16.21263412; this version posted May 12, 2022. 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.

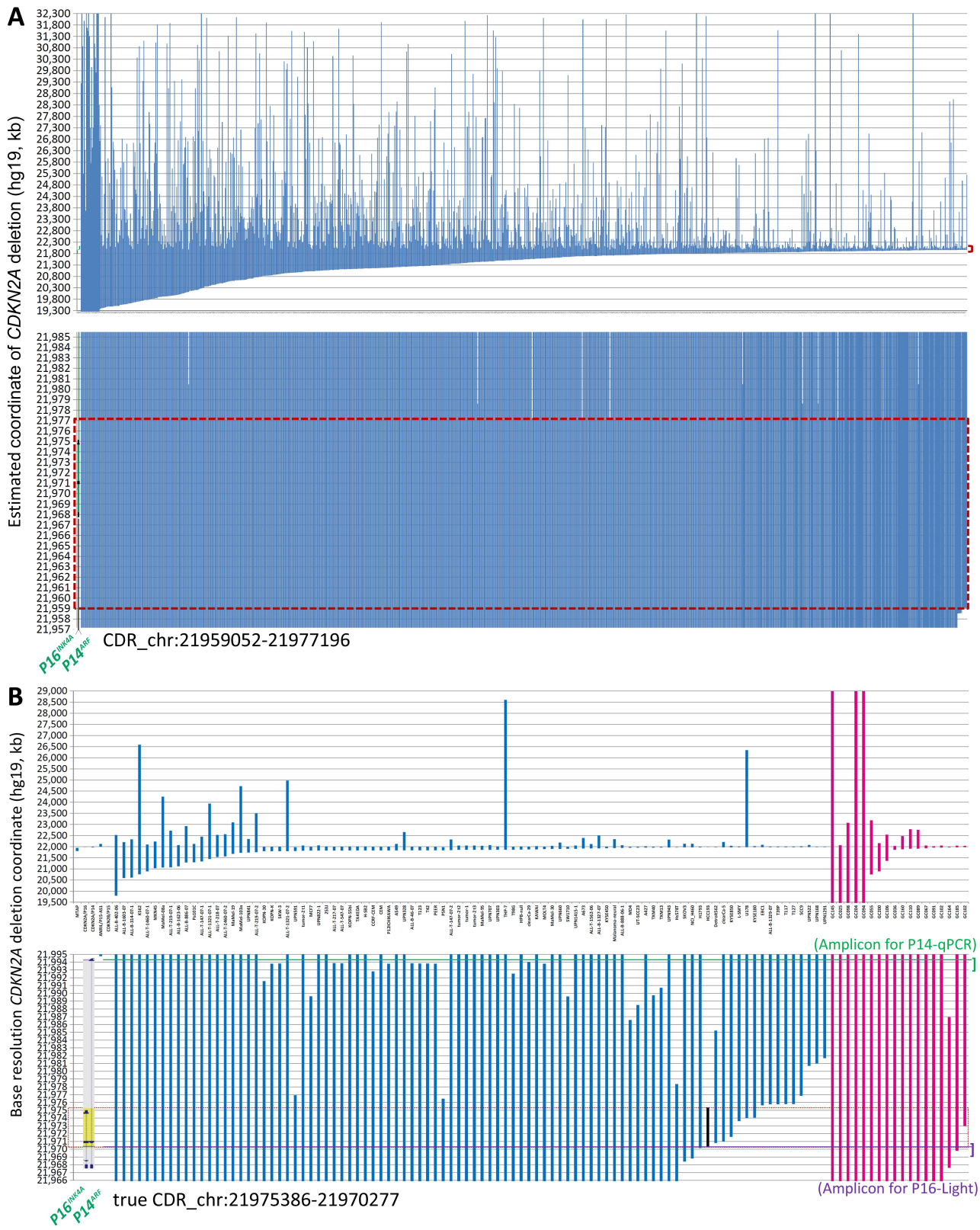
1 **Table 2.** Association between clinicopathological characteristics and *CDKN2A* SCND

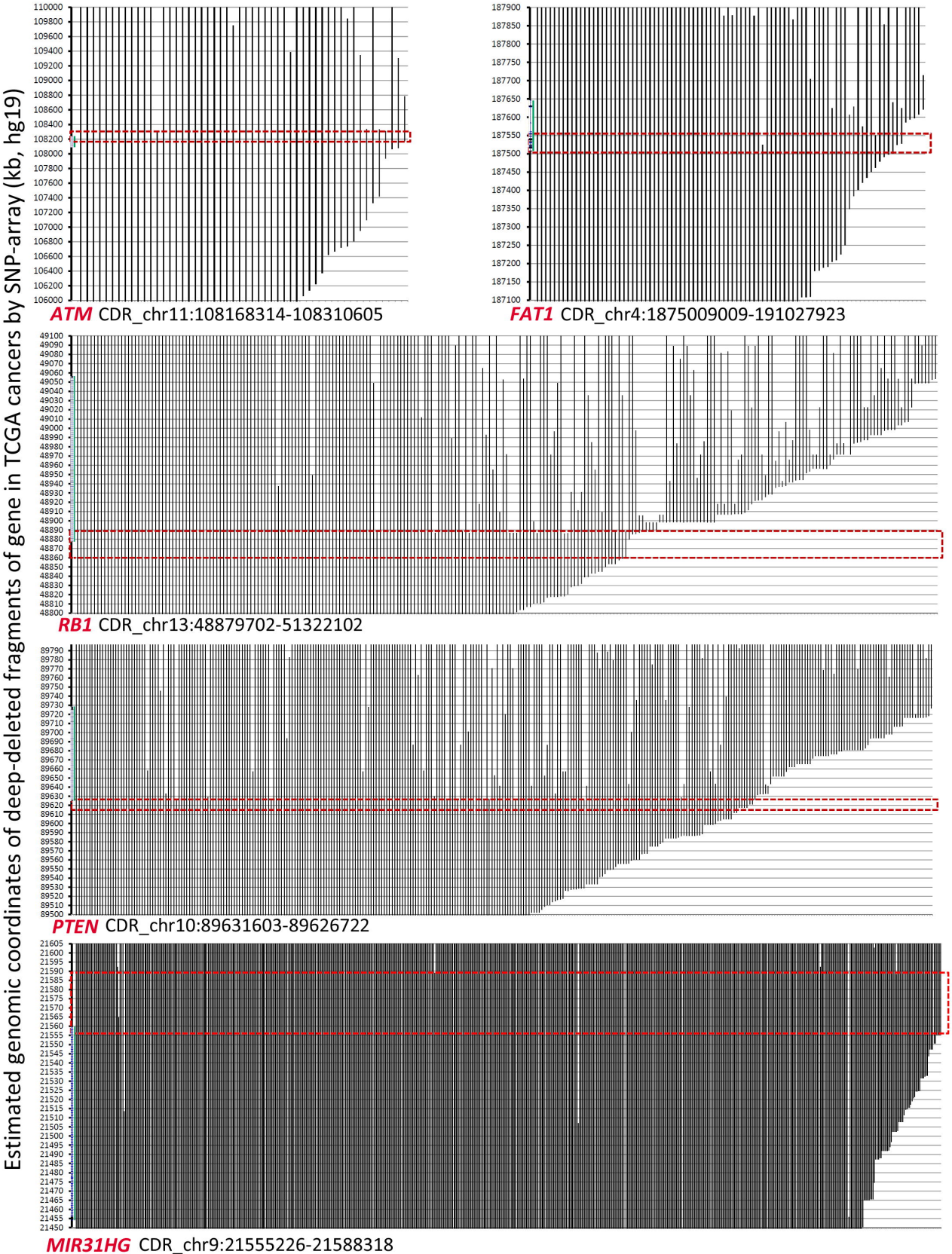
2 detected by P16-Light and P14-qPCR

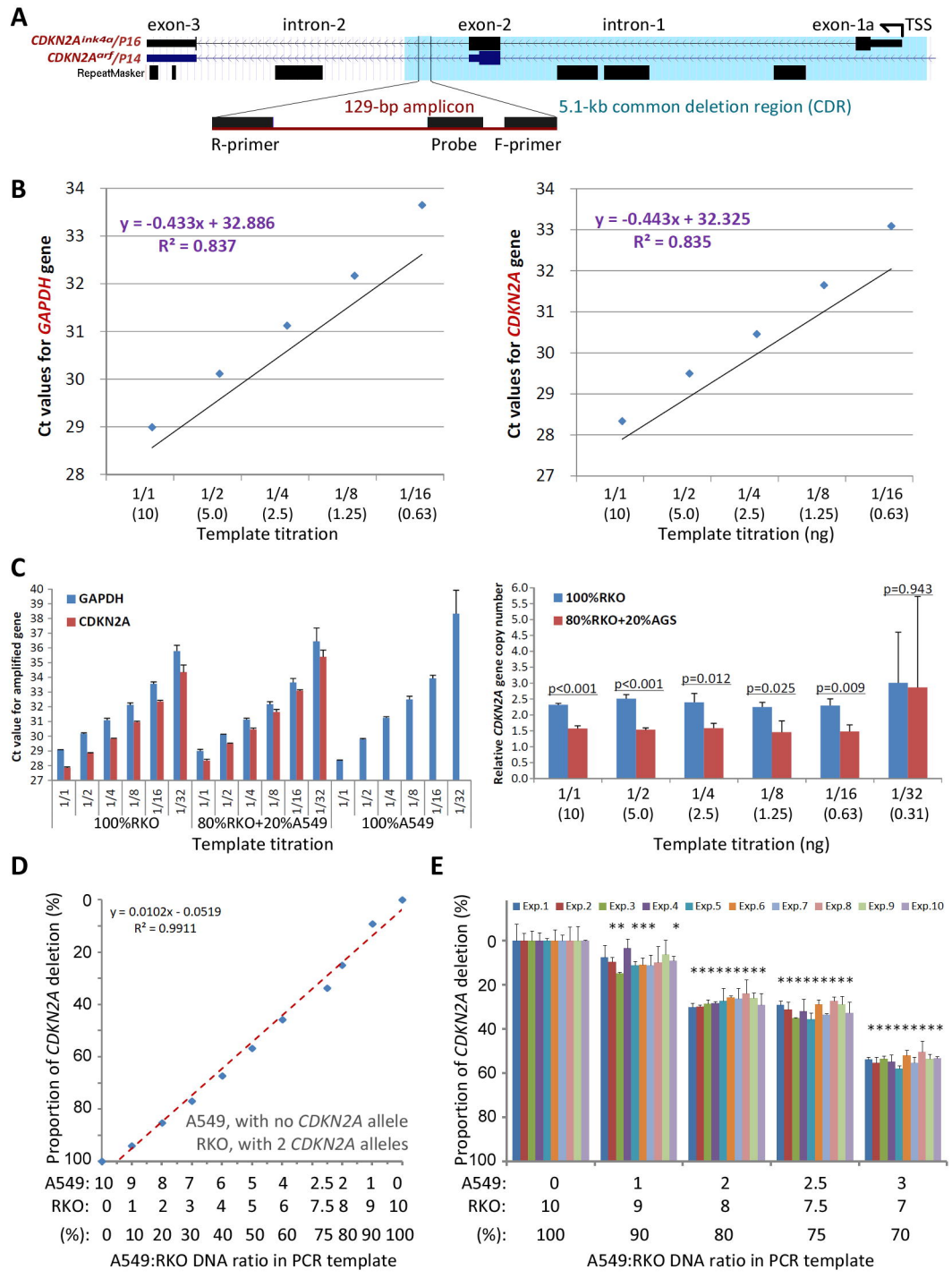
| | | | ***CDKN2A* CDR SCND-positive by P16-Light** | | ***CDKN2A P14$^{ARF}$* SCND-positive by P14-qPCR** | | ***CDKN2A* CDR or *P14$^{ARF}$* SCND-positive** | | ***CDKN2A* CDR & *P14$^{ARF}$* SCND-positive** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ***n*** | **Positive cases** | **Positive rate (%)** | **Positive cases** | **Positive rate (%)** | **Positive cases** | **Positive rate (%)** | **Positive cases** | **Positive rate (%)** |
| Age | <60 | 68 | 23 | 33.8 | 18 | 26.5 | 33 | 48.5 | 8 | 11.8 |
| | ≥60 | 71 | 28 | 39.4 | 26 | 36.6 | 43 | 60.6 | 11 | 15.5 |
| Sex | Male | 101 | 40 | 39.6 | 33 | 32.7 | 58 | 57.4 | 15 | 14.9 |
| | Female | 38 | 11 | 28.9 | 11 | 28.9 | 18 | 47.4 | 4 | 10.5 |
| Location[a] | Cardiac | 18 | 9 | 50.0 | 3 | 16.7 | 10 | 55.6 | 2 | 11.1 |
| | Non-cardiac | 121 | 42 | 34.7 | 41 | 33.9 | 66 | 54.5 | 17 | 14.0 |
| Different. | Poor | 99 | 33 | 33.3 | 30 | 30.3 | 51 | 51.5 | 12 | 12.1 |
| | Well/mod. | 37 | 16 | 43.2 | 14 | 37.8 | 23 | 62.2 | 7 | 18.9 |
| pTNM | I-II | 46 | 11 | **23.9**[a] | 16 | 34.8 | 23 | 50.0 | 4 | 8.7 |
| stage | III | 37 | 14 | **37.8** | 8 | 21.6 | 17 | 45.9 | 5 | 13.5 |
| | IV | 56 | 26 | **46.4** | 20 | 35.7 | 36 | 64.3 | 10 | 17.9 |
| Invasion | T1-2 | 27 | 11 | 40.7 | 13 | 48.1 | 19 | 70.4 | 5 | 18.5 |
| | T3 | 79 | 28 | 35.4 | 19 | 24.1 | 38 | 48.1 | 9 | 11.4 |
| | T4 | 33 | 12 | 36.4 | 12 | 36.4 | 19 | 57.6 | 5 | 15.2 |
| Lymph | Negative | 51 | 16 | 31.4 | 18 | 35.3 | 27 | 52.9 | 7 | 13.7 |
| metastasis | Positive | 88 | 35 | 39.8 | 26 | 29.5 | 49 | 55.7 | 12 | 13.6 |
| Distant | Negative | 107 | 31 | **29.0**[b] | 33 | 30.8 | 53 | **49.5**[c] | **11** | 10.3[d] |
| metastasis | Positive | 32 | 20 | **62.5** | 11 | 34.4 | 23 | **71.9** | **8** | 25.0 |
| (Total) | | 139 | 51 | 36.7 | 44 | 31.7 | 76 | 54.7 | 19 | 13.7 |

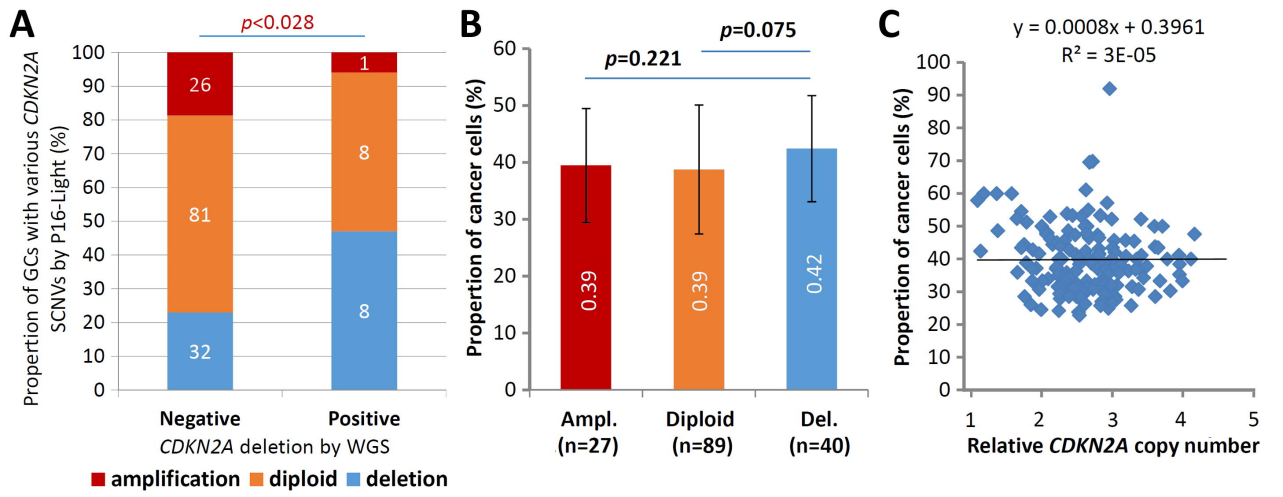
3 [a] Chi-square trend test, p<0.001; [b] Odds ratio (OR) =4.09, 95% confidence interval (CI) (4.66-10.19), *p*=0.001;

4 [C] OR =2.60, 95%CI (1.03-6.74), *p* <0.026; [d] OR =2.91, 95%CI (0.94-8.94), *p*=0.033


A
Estimated coordinate of CDKN2A deletion (hg19, kb)
CDR_chr:21959052-21977196
P16INK4A
P14ARF
B
Base resolution CDKN2A deletion coordinate (hg19, kb)
(Amplicon for P14-qPCR)
(Amplicon for P16-Light)
true CDR_chr:21975386-21970277
P16INK4A
P14ARF


Estimated genomic coordinates of deep-deleted fragments of gene in TCGA cancers by SNP-array (kb, hg19)
ATM CDR_chr11:108168314-108310605
FAT1 CDR_chr4:187500009-191027923
RB1 CDR_chr13:48879702-51322102
PTEN CDR_chr10:89631603-89626722
MIR31HG CDR_chr9:21555226-21588318


A
exon-3
intron-2
exon-2
intron-1
exon-1a
TSS
CDKN2Aink4a/P16
CDKN2Aarf/P14
RepeatMasker
129-bp amplicon
5.1-kb common deletion region (CDR)
R-primer
Probe
F-primer
B
y = -0.433x + 32.886
R² = 0.837
Ct values for GAPDH gene
1/1 (10)
1/2 (5.0)
1/4 (2.5)
1/8 (1.25)
1/16 (0.63)
Template titration
y = -0.443x + 32.325
R² = 0.835
Ct values for CDKN2A gene
Template titration (ng)
C
GAPDH
CDKN2A
Ct value for amplified gene
100%RKO
80%RKO+20%A549
100%A549
Template titration
100%RKO
80%RKO+20%AGS
Relative CDKN2A gene copy number
p<0.001
p<0.001
p=0.012
p=0.025
p=0.009
p=0.943
1/1 (10)
1/2 (5.0)
1/4 (2.5)
1/8 (1.25)
1/16 (0.63)
1/32 (0.31)
Template titration (ng)
D
y = 0.0102x - 0.0519
R² = 0.9911
Proportion of CDKN2A deletion (%)
A549, with no CDKN2A allele
RKO, with 2 CDKN2A alleles
A549: 10 9 8 7 6 5 4 2.5 2 1 0
RKO: 0 1 2 3 4 5 6 7.5 8 9 10
(%): 0 10 20 30 40 50 60 75 80 90 100
A549:RKO DNA ratio in PCR template
E
Exp.1 Exp.2 Exp.3 Exp.4 Exp.5 Exp.6 Exp.7 Exp.8 Exp.9 Exp.10
Proportion of CDKN2A deletion (%)
A549: 0 1 2 2.5 3
RKO: 10 9 8 7.5 7
(%): 100 90 80 75 70
A549:RKO DNA ratio in PCR template


A
p<0.028
Propertion of GCs with various CDKN2A SCNVs by P16-Light (%)
100
90
80
70
60
50
40
30
20
10
0
26
81
32
1
8
8
Negative
Positive
CDKN2A deletion by WGS
amplification
diploid
deletion
B
p=0.221
p=0.075
Propertion of cancer cells (%)
60
50
40
30
20
10
0
0.39
0.39
0.42
Ampl. (n=27)
Diploid (n=89)
Del. (n=40)
C
y = 0.0008x + 0.3961
R² = 3E-05
Propertion of cancer cells (%)
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
Relative CDKN2A copy number