1 Detection of Somatic Copy Number Deletion of the *CDKN2A* Gene

2 by Quantitative Multiplex PCR for Clinical Practice

- 3 Running title: Gene-specific common deletion regions
- 4 Yuan Tian^{1,†}, Jing Zhou^{1,†}, Juanli Qiao¹, Zhaojun Liu¹, Liankun Gu¹, Baozhen Zhang¹, Youyong Lu²,
- 5 Rui Xing^{2,*}, Dajun Deng^{1,*}
- 6 ¹ Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Etiology,
- 7 Peking University Cancer Hospital and Institute, Beijing, 100142, China;
- 8 ² Key Laboratory of Carcinogenesis and Translational Research (MOE/Beijing), Division of Tumor
- 9 Biology, Peking University Cancer Hospital and Institute, Beijing, 100142, China

10

- 11 * To whom correspondence should be addressed. Tel: +8610-88196752; Fax: +8610-88122437;
- 12 Email: xingrui@bjmu.edu.cn or dengdajun@bjmu.edu.cn
- ¹³ [†]These authors contributed equally to this work.

14

1 Abstract

- 2 Background: While the amplification of oncogenes (such as EGFR, c-ERBB2, c-MYC, and c-MET)
- 3 are increasingly driving decision-making for precise cancer treatments, a feasible method to detect
- 4 somatic copy number deletion (SCND) of tumor suppressor genes is still absent to date.
- 5 Methods: The genomic coordinates of gene deletion fragments were analyzed using the Catalogue
- 6 Of Somatic Mutation In Cancer (COSMIC) datasets. Interstitial base-resolution deletion/fusion
- 7 coordinates for CDKN2A were extracted from published articles and our whole genome sequencing
- 8 (WGS) datasets. The copy number of the CDKN2A gene was measured with a multiplex quantitative
- 9 PCR assay P16-Light and confirmed with whole genome sequencing (WGS).
- 10 Results: Estimated common deletion regions (CDRs) were observed in many tumor suppressor
- 11 genes, such as ATM, CDKN2A, FAT1, miR31HG, PTEN, and RB1, in the SNP array-based COSMIC
- 12 datasets. A 5.1-kb base-resolution CDR could be identified in >90% of cancer samples with CDKN2A
- 13 deletion by sequencing. The CDKN2A CDR covers exon-2, which is essential for P16^{INK4A} and P14^{ARF}
- 14 synthesis. Using the true CDKN2A CDR as a PCR target, a multiplex quantitative PCR assay P16-
- 15 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 from139 patients.
- 18 **Conclusion:** CDRs are common in many tumor suppressor genes. The 5.1-kb *CDKN2A* CDR was
- 19 found in >90% of cancers containing CDKN2A deletion. The CDKN2A CDR was used as a potential
- 20 target for developing the P16-Light assay to detect *CDKN2A* SCND and amplification for routine
- 21 clinical practices.

22 Keywords

23 CDKN2A; somatic copy number deletion; whole genome sequencing; gastric carcinoma

1 Background

2 Somatic copy number variations (SCNVs) of tumor-related genes are landmarks of human cancers 3 [1,2]. Somatic copy number deletion (SCND) and amplification are two kinds of well-known SCNVs. 4 However, current gene copy number detection methods, including microsatellite instability (MSI), 5 loss/gain of heterozygosity (LOH/GOH), fluorescence-in situ hybridization (FISH), whole genome 6 sequencing (WGS) or whole exome sequencing (WES), are not sensitive enough or too costly for 7 routine clinical use. While the amplification of oncogenes (such as EGFR, c-ERBB2, c-MYC, and c-8 MET) is increasingly driving decision-making for precise cancer treatments, clinical applications of 9 SCND of tumor suppressor genes, including CDKN2A, are still rare owing to the lack of a feasible 10 detection assay. 11 The frequency of CDKN2A SCND detected by single nucleotide polymorphism (SNP) microarray, 12 WGS or WES was found to range from 30% to 60% in bladder cancer, melanoma, head and neck 13 cancer, pleural mesothelioma, glioblastoma, and esophageal squamous cell cancer (ESCC), with an 14 average frequency of 13% in pan-cancer datasets in The Cancer Genome Atlas (TCGA) (Figure S1A) 15 [2-6]. CDKN2A deep deletion is associated with downregulation of CDKN2A gene expression, while 16 CDKN2A amplification is associated with upregulation of CDKN2A gene expression in Pan-TCGA 17 cancers (Figure S1B). It is well known that genetic CDKN2A inactivation contributes to malignant 18 transformation, cancer metastasis, and therapeutic sensitivity of cancers to drugs, including CDK4/6 19 inhibitors and their combination with PD-1 blockade [7-11]. Therefore, a convenient and sensitive 20 assay to detect CDKN2A SCND is eagerly awaited. 21 In the present study, we characterized patterns of estimated genomic coordinates for SCNDs in a 22 set of tumor suppressor genes using the public Catalogue Of Somatic Mutations In Cancer (COSMIC) 23 SCNV datasets and found common deletion regions (CDRs) in many frequently deleted genes. Then, 24 we further defined a 5.1-kb base-resolution CDR within the CDKN2A gene using sequencing data for 25 the first time. A sensitive P16-Light assay targeting the CDKN2A CDR was established for clinical

26 practice.

27 Materials and methods

28 COSMIC and TCGA SCNV datasets

29 SNP6 array-based estimated genomic coordinates of interstitial copy number deletion/fusion of the

- 30 CDKN2A gene in cancer cell lines (n=273) with homozygous CDKN2A deletion and estimated
- 31 genomic coordinates of deep-deleted fragments of CDKN2A, PTEN, RB1, and other frequently
- 32 deleted genes in cancer tissues were downloaded from the Copy Number Analysis (CONA) datasets
- 33 in the COSMIC project (Data file 1-11) [12].

34 Patients, tissues, and DNA preparation

35 Frozen fresh GC and paired surgical margin (SM) tissue samples were collected from 156 patients in

- 36 the WGS study [13]. These samples were frozen in liquid nitrogen approximately 30 min after surgical
- 37 dissection and then stored in a -80°C freezer for 2-5 yrs. Clinicopathological information was also

- 1 obtained. The 2010 UICC tumor-node-metastasis (TNM) system was used to classify these GCs [14].
- 2 Genomic DNA was extracted from these samples with a phenol/chloroform method coupled with
- 3 RNase treatment. Concentrations of these DNA samples were determined with NanoVue Plus
- 4 (Biochrom LTD, Cambridge, UK). DNA samples with OD_{260nm}/OD_{280nm} ratios ranging from 1.7 to 1.9
- 5 were used for the detection of gene copy number as described below.

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

7 Multiplex primer and probe combinations were designed based on the best multiplex primer probe 8 scores for conserved sequences within the CDR in the CDKN2A (HGNC: 1787) and GAPDH (HGNC: 9 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 10 11 performance of these assays for the detection of CDKN2A copy numbers was compared with each 12 other. Finally, a multiplex primer and probe combination targeting CDKN2A intron-2 was selected 13 (Table 1), and the concentrations of the components were optimized. Each multiplex PCR assay was 14 carried out in a total volume of 20 μ L that included 5-10 ng of input DNA, 10 μ M of forward and 15 reverse primers and probe for CDKN2A intron-2, $10 \,\mu$ M forward and reverse primers and probe for 16 GAPDH, and 10 µL of 2x TagMan Universal Master Mix II with uracil-N-glycosylase (Kit-4440038, ABI, 17 Lithuania). The PCRs were performed in triplicate in a MicroAmp Fast Optical 96-Well Reaction Plate 18 with a barcode (0.1 mL; ABI, China) with an ABI 7500 Fast Real-Time PCR System. The specific 19 conditions of the PCR were as follows: initial incubation for 10 min at 95°C, followed by 40 cycles of 20 95°C for 20 sec and 58°C for 60 sec. When the Ct value for GAPDH input for a sample was 34 or 21 fewer cycles, this sample was considered CDKN2A SCNV informative. The specificity of the PCR was 22 monitored through running the gel. Distilled water was used as a no-template control for each 23 experiment.

24 Definitions of CDKN2A CDR deletion positivity and amplification positivity

25 We used genomic DNA from A549 cells containing no CDKN2A allele to dilute genomic DNA from 26 RKO cells containing 2 wild-type CDKN2A alleles, and then we set the standard curve according to 27 the relative copy number of the CDKN2A gene at different dilutions. The Δ Ct value and relative copy 28 number for the CDKN2A gene were calculated using the GAPDH gene as the internal reference. 29 When the CDKN2A copy number in the A549-diluted template was consistently lower than that in the 30 RKO control template and the difference was statistically significant (t test, p < 0.05), it was judged that 31 the lowest dilution concentration was the detection limit of CDKN2A deletion (the difference in 32 CDKN2A copy number between the 100% RKO template and 80% RKO template spiked with 20% 33 A549 DNA). When the CDKN2A relative copy number in a tissue sample was significantly lower or 34 higher than that of the paired SM sample, the sample was defined as somatic CDKN2A CDR deletion-35 positive or amplification-positive, respectively. For each experiment, the 100% A549, 100% RKO, and

- 36 20% A549 + 80% RKO DNA mix controls were analyzed.
- 37

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

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

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

7 datasets

- 8 We used Meerkat 23 to predict somatic SVs and their breakpoints in WGS datasets (accession
- 9 numbers, EGAD00001004811 with 36× of sequencing depth) for gastric adenocarcinoma samples
- 10 from 168 patients using the suggested parameters [13]. This method used soft-clipped and split reads
- 11 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
- 13 number profiling over 10-kb windows with Patchwork 28 and calculated the ratio of standardized
- 14 average depth between normal tissue and tumor tissue (log2R ratio). The purity and ploidy of each
- 15 tumor were calculated using ABSOLUTE software [16].

16 Cell lines and cultures

- 17 The CDKN2A allele homozygously deleted cell line A549 (kindly provided by Dr. Zhiqian Zhang of
- 18 Peking University Cancer Hospital and Institute) was grown in RPMI-1640 medium, and the RKO cell
- 19 line containing two wild-type *CDKN2A* alleles was purchased from American Type Culture Collection
- 20 and grown in DMEM. The medium was supplemented with 10% (v/v) fetal bovine serum (FBS). These
- 21 cell lines were tested and authenticated by Beijing JianLian Genes Technology Co., Ltd. before they
- 22 were used in this study. A Goldeneye[™] 20A STR Identifier PCR Amplification kit was used to analyze
- 23 the STR patterns.

24 Statistical analysis

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

29 **Results**

30 Prevalence of estimated CDRs within various tumor suppressor genes

- 31 It has been previously reported that homozygous deletion of approximately 170 kilobase pairs (kb),
- 32 including the CDKN2A locus, can be detected in human cancers by MSI analyses [17]. SCND
- 33 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 by
- 35 ordering "start" genomic coordinates of these breaking points (Figure S2). To investigate the

- 1 prevalence of CDRs within tumor suppressor genes in human cancer tissues with a high deletion
- 2 frequency [1,2], we further downloaded the estimated genomic coordinates for deletion fragments that
- 3 overlapped with these genes. We found that CDRs could be detected not only within the CDKN2A
- 4 gene (Figure 1A; approximately 17 kb) but also within the ATM (middle to downstream), FAT1
- 5 (promoter to middle), miR31HG (promoter to exon-1), PTEN (promoter to exon-1), and RB1 genes
- 6 (promoter to intron-2) (Figure 2; approximately 158 kb, 23 kb, 33 kb, 5 kb, and 2442 kb, respectively)
- 7 (Data files 2-7). No CDR could be observed within CCSER1, FHIT, LRP1B, and WWOX genes
- 8 according to the SNP-array data (Data files 8-11).

9 Characterization of a true CDKN2A CDR at base resolution in human cancers

10 It was reported that the error in *CDKN2A* breakpoint estimation based on SNP-array data is

- 11 approximately 10 kb [18]. To characterize the true genomic coordinates of CDKN2A deletion
- 12 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
- 14 5.1-kb CDR (chr9: 21,970,277 21,975,386, hg19) that spanned from the *P16^{INK4A}* promoter to intron-
- 15 2 in 83 (90%) of 92 reported cancer cell lines or tissue samples containing interstitial CDKN2A
- 16 deletions (Figure 1B, blue lines). This CDR sequence is the same as the CDKN2A deletion fragment
- 17 in the HCC193 lung cancer cell line [25]. The CDR coordinates were also confirmed in our WGS
- 18 datasets (average sequencing depth, 36×) of 18 (100%) of 18 GCs [13], in which interstitial CDKN2A
- 19 deletions/fusions were identified (Figure 1B, purple lines; Data file 12).
- 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 *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 coinactivated in 87% (96/110) of human cancer cell lines and tissues containing *CDKN2A* CDR deletion (Figure 1B).

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

31 The current clinical method FISH for detecting SCND is composed of a set of biotin-labeled probes 32 that should cover at least 50-kb DNA sequences. Thus, FISH is not a suitable method for detecting 33 the copy number deletion of the 5.1-kb CDKN2A CDR. To provide a convenient method for routine 34 clinical use, we designed and experimentally evaluated a set of multiplex quantitative PCR assays 35 and finally optimized the CDKN2A CDR-specific guantitative multiplex PCR assay called P16-Light for 36 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). 37 38 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

- 1 CDKN2A alleles) were used as CDKN2A CDR deletion-positive and deletion-negative controls,
- 2 respectively. The amplification efficiencies of the two amplicons in GAPDH and CDKN2A were very
- 3 similar (Figure 3B). No template inhibition was observed when the amount of template DNA ranged
- 4 from 10 to 0.63 ng (Figure 3C). The proportions of CDKN2A CDR copy number were linearly
- 5 correlated with the ratios (0 100%) of RKO cell DNA and A549 cell DNA in the input mixtures (10
- 6 ng/reaction) when the A549 DNA was spiked in at different proportions for the P16-Light analyses
- 7 (Figure 3D). Furthermore, there was a high reproducibility when DNA with homozygous deletion of
- 8 CDKN2A was present in ≥20% of the cells verified in ten experimental repeats performed on different
- 9 days (Figure 3E). Thus, when the proportion of CDKN2A copy number was significantly decreased (or
- 10 increased) in a sample relative to the paired normal control (t test, *p*<0.05) in the P16-Light analyses,
- 11 the sample was defined as *CDKN2A* SCND-positive (or amplification-positive).

12 Comparison of P16-Light with WGS datasets

13 As we described above, information on interstitial copy number deletion/fusion of the CDKN2A gene 14 was extracted from WGS datasets for 156 of 168 GC patients enrolled in a GC genome study [13], 15 and a total of 18 CDKN2A deletion/fusion coordinates at the base resolution were detected in 17 16 (10.8%) GCs (Data files 12 and 13). To compare the performance of P16-Light with WGS, we 17 analyzed the status of SCNVs, including SCND and amplification, of the CDKN2A gene in 156 of 18 these GCs with enough genomic DNA samples with P16-Light using the paired SM sample as the 19 diploid reference (Data file 13). CDKN2A SCND and amplification were detected in 40 (25.6%) and 34 20 (21.8%) of these GCs, respectively. The P16-Light analysis was confirmed by the WGS results: the 21 frequency of CDKN2A SCND (or amplification) by P16-Light was significantly higher (or lower) in 17 22 GCs containing interstitial CDKN2A deletion/fusion than in 139 GCs without interstitial CDKN2A 23 deletion/fusion (chi-square test, p<0.028; Figure 4A). These results also indicate that there is a 24 significantly higher sensitivity for detecting CDKN2A SCND by the quantitative P16-Light assay than 25 the hemi-quantitative WGS. 26 Moreover, it is well known that the proportion of cancer cells in tissue samples (i.e., sample purity)

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

34 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-1 β [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 GC and paired SM samples from patients who were recently included in the cross-sectional cohort in our association study [34]. GC

- 1 samples (*n*=139) with enough genomic DNA were used in P14-qPCR analysis (Data file 14). The
- 2 SCND-positive rate for P14^{ARF} was similar to that for the CDKN2A CDR (31.7% vs. 36.7%) (Table 2).
- 3 CDKN2A SCND was found only in 19 GCs by both assays. While CDKN2A CDR SCND by P16-Light
- 4 was significantly associated with distant metastasis of GC (odds ratio=4.09, *p*<0.001), no association
- 5 was observed between GC metastasis and *P14*^{ARF} SCND by P14-qPCR. Using merged *CDKN2A*
- 6 SCND data (*CDKN2A* CDR SCND-positive and/or *P14*^{ARF} SCND-positive), only a weaker association
- 7 was observed. These results suggest that individual P16-Light alone may be good enough for
- 8 detecting *CDKN2A* SCND in tissue samples.

9 Discussion

10 Somatic copy number deletion and amplification are two main kinds of SCNVs. The detection of copy 11 number amplification of oncogenes is routinely used for precise cancer treatments. However, the 12 detection of SCND of tumor suppressor genes is absent, and its significance in clinical practice is not 13 well studied. The reason should be the lack of feasible detection approaches. Here, we report that 14 there are CDRs in many tumor suppressor genes, such as CDKN2A, miR31HG, PTEN, and RB1, 15 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% of 16 17 cancers containing CDKN2A deletion. Using the CDKN2A CDR as a PCR target, we further 18 established a feasible P16-Light assay to detect CDKN2A SCND and amplification. These findings 19 indicate that CDRs are prevalent sequences in tumor suppressor genes, and characterization of the 20 base-resolution genomic coordinates of CDRs could enable us to establish convenient methods for 21 SCND detection of genes. 22 Interstitial deletion/fusion is the main type of CDKN2A SCND, and the breaking/fusing 23 coordinates for CDKN2A SCNDs in cancer genomes are diverse, which blocks the establishment of a 24 feasible detection assay for CDKN2A SCND, although many efforts have been made [20]. In the 25 present study, we initially found the 8~17-kd estimated CDKN2A CDR in both monoclonal cancer cell 26 lines and cell-heterogeneous cancer tissues with CDKN2A copy number deletion according to the 27 SNP-array datasets from COSMIC and TCGA projects [1,12]. Then, we further characterized the 5.1-28 kb true CDR at the base resolution within the CDKN2A gene in cancer genomes using DNA 29 sequencing data [19-28] and confirmed the CDR using WGS datasets in all 18 GCs containing 30 CDKN2A SCND [13]. Because the true CDKN2A CDR was observed in more than 90% of CDKN2A-31 deleted cancer samples and the P16-Light assay is highly reproducible and convenient, the 32 quantitative P16-Light assay should be considered a viable assay for detecting CDKN2A SCNVs in 33 clinical practice. This is supported by the result that CDKN2A SCND by P16-Light was significantly 34 associated with GC metastasis and further supported by the results of our prospective study, in which 35 CDKN2A SCND was closely associated with hematogenous metastasis of GCs [34]. In another long-36 term prospective study, we also found that CDKN2A SCND and amplification by P16-Light were 37 significantly associated with malignant transformation and complete regression of mild or moderate 38 esophageal squamous cell dysplasia, respectively [Fan et al. submitted for publication]. The results of

1 these studies also suggest that the sensitivity of 20% for the P16-Light assay may be good enough for

2 routine clinical use.

3 WGS is generally used as a kind of gold standard to study structural alterations of genomic 4 DNAs, especially for interstitial gene copy deletion/fusions. However, WGS is a cost assay, and its 5 accuracy depends on sequencing depth. WGS at sequencing depth 36× would be considered a hemi-6 quantitative assay. In our calling of CDKN2A SCND coordinate processes, it was found that 18 7 CDKN2A SCND coordinates were identified in 17 (10.8%) of 157 GCs, which was consistent with the 8 frequency (11.4% =50/438) of homozygous deletion of CDKN2A in GCs in WES or WGS sequencing 9 datasets (Data file 14) [35]. The positive rate (25.6%) of CDKN2A SCND in 156 GCs with enough 10 genomic DNA samples in the P16-Light analysis was more than twice that of WGS. The results of 11 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. 12 P14-qPCR is a method used to detect the copy number of CDKN2A/P14^{ARF} exon-1 β [15]. 13 Although the combination of P16-Light with P14-gPCR may detect both SCNDs overlapping with the 14 15 CDKN2A CDR and not overlapping with the CDKN2A CDR, the results of our comparison analysis 16 among 139 GC patients showed that detecting CDKN2A SCND by individual P16-Light may be good 17 enough for clinical practice because combination with P14-qPCR could not improve the performance 18 of P16-Light. However, for other genes, such as RB1 and PTEN, whether a gPCR array needs to be 19 employed for detecting SCNVs should be studied case by case. 20 Generally, IHC is a popular method used to detect expression changes in protein-coding genes. For example, P16^{INK4A} overexpression in cervical mucosa samples is currently used for rapid HPV 21 infection screening. We compared the status of P16^{INK4A} expression by IHC between GCs with 22 CDKN2A SCND (n=4) and GCs without CDKN2A SCND (n=12) and did not find any difference in the 23 P16^{INK4A} positive-staining rate between these GCs (3/4 vs. 9/12). The expression level of 24 CDKN2A/P16^{INK4A} is not only affected by SCNVs but also regulated by the methylation status of CpG 25 islands, histone modifications, and high-risk HPV infection [36,37]. These factors may partially 26 27 account for the inconsistency between IHC and P16-Light. The driver function of the CDKN2A gene in cancer development is enigmatic. P16^{ink4a} inactivation 28 contributes less than P19^{arf} (the murine counterpart of human P14^{ARF}) inactivation to cancer 29 development in mice, while *P16^{INK4A}* inactivation contributes more than *P14^{ARF}* inactivation to cancer 30 development in humans [38,39]. The exact mechanisms leading to the difference among species are 31 still unclear. Here, we reported that approximately 87% of genetic P16^{INK4A} inactivation by CDKN2A 32 SCND is accompanied by P14^{ARF} inactivation in human cancer cell lines or tissues. This may account 33 34 for the species-related functional difference in the CDKN2A gene. The report supports this explanation that knocking out both $p16^{ink4a}$ and $p19^{art}$ leads to more cancer development than 35 individual inactivation in mice [40]. This also may account for the fact that $P14^{ARF}$ exon-1 β deletion 36 37 was not associated with GC metastasis, whereas CDKN2A CDR deletion was significantly associated 38 with GC metastasis, as described above. 39 In conclusion, we have found estimated CDRs in many tumor suppressor genes in the cancer

40 genome. There is a 5.1-kb CDR region within the CDKN2A gene, and most CDKN2A deletions lead to

- 1 P16^{INK4A} and P14^{ARF} inactivation in human cancers. Using the CDKN2A CDR as a target sequence,
- 2 we developed a convenient quantitative multiplex PCR assay, P16-Light, to detect *CDKN2A* SCNVs
- 3 in clinical practice, suggesting that the strategy to detect *CDKN2A* SCNVs may be suitable for the
- 4 establishment of SCNV detection methods for other tumor suppressor genes.

5 Supplementary material

6 Supplementary Figures 1-3 and Data files 1-14 can be found online.

7 Funding

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

12 Acknowledgments

- 13 We thank Dr. Sanford Dawsey at NCI, NIH, Bethesda, Maryland for critical comments on the
- 14 manuscript. We also thank Miss Gina Mckeown in New York, USA for English language editing.
- 15 *Conflict of Interest Statement:* DD, YT, JZ, and ZL are the creators for the pending patent "A
- 16 quantitative method for detection of human CDKN2A gene copy number using a primer set and their
- 17 applications" (PCT/CN2019/087172; WO2020228009). All of the other authors have nothing to
- 18 disclose.

19 **Declarations**

20 Ethical Approval

- 21 This study was approved by the Institution Review Board of Peking University Cancer Hospital &
- 22 Institute and carried out in accordance with the principles outlined in the Declaration of Helsinki.
- 23 Informed consent was obtained from each patients prior to their inclusion in the study.

24 Authors' contributions

- 25 Yuan Tian: Methodology, Writing- Original draft preparation; Jing Zhou: Methodology, Formal analysis,
- 26 Writing- Original draft preparation; Juanli Qiao: Investigation; Zhaojun Liu: Data curation; Liankun Gu:
- 27 Investigation; Baozhen Zhang: Investigation; Youyong Lu: Supervision, Resources; Rui Xing:
- 28 Conceptualization, Resources, Data curation, Validation, Writing- Original draft preparation; Dajun
- 29 Deng: Conceptualization, Supervision, Funding acquisition, Methodology, Data curation, Visualization,
- 30 Writing- Original draft preparation.

31 Availability of data and materials

32 The original contributions presented in the study are included in the article as Supplemental Files 1-14.

33

1 References

Beroukhim, R. *et al.* (2010) The landscape of somatic copy-number alteration across human
 cancers. *Nature*, **463**, 899-905.

Cerami, E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring
multidimensional cancer genomics data. *Cancer Discov.*, 2, 401-404.

Mermel, C.H. *et al.* (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets
of focal somatic copy-number alteration in human cancers. *Genome Biol.*, **12**, R41. doi: 10.1186/gb2011-12-4-r41.

9 4. Gao, J. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using 10 the cBioPortal. *Sci. Signal.*, **6**, pl1. doi: 10.1126/scisignal.2004088.

Network CGAR. *et al.* (2017) Integrated genomic characterization of oesophageal carcinoma.
 Nature, **541**, 169-175.

13 6. Cui,Y. *et al.* (2020) Whole-genome sequencing of 508 patients identifies key molecular features 14 associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res.*, **30**, 902-913.

Serrano, M. *et al.* (1996) Role of the INK4a locus in tumor suppression and cell mortality. *Cell*,
 85, 27-37.

Beng J. *et al.* (2018) CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-cell
 Activation. *Cancer Discov.*, **8**, 216-233.

Jerby-Arnon, L. *et al.* (2018) A Cancer Cell Program Promotes T Cell Exclusion and Resistance to
 Checkpoint Blockade. *Cell*, **175**, 984-997.e24.

21 10. Zhang, J. *et al.* (2018) Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer
 22 immune surveillance. *Nature*, **553**, 91-95.

11. Yu,J. *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.

12. Tate, J.G. *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucl. Acids Res.*,
47, D941–D947.

Xing, R. *et al.* (2019) Whole-genome sequencing reveals novel tandem-duplication hotspots and
a prognostic mutational signature in gastric cancer. *Nat. Commun.*, **10**, 2037. doi: 10.1038/s41467019-09644-6.

Sobin,L. *et al.* (2009) TNM Classification of Malignant Tumours, 7th edn. International Union
 Against Cancer (UICC). Wiley Press, New York, USA.

33 15. Berggren, P. *et al.* (2003) Detecting homozygous deletions in the

CDKN2A(p16INK4a)/ARF(p14ARF) gene in urinary bladder cancer using real-time quantitative PCR.

35 Clin. Cancer Res., **9**, 235-242.

36 16. Carter, S.L. *et al.* (2012) Absolute quantification of somatic DNA alterations in human cancer.

37 Nat. Biotechnol., **30**, 413–421.

- 1 17. Cairns, P. *et al.* (1995) Frequency of homozygous deletion at p16/CDKN2 in primary human
- 2 tumours. Nat. Genet., **11**, 210-212.
- 18. Novara, F. *et al.* (2009) Different molecular mechanisms causing 9p21 deletions in acute
 lymphoblastic leukemia of childhood. *Hum. Genet.*, **126**, 511-520.
- Sie, H. *et al.* (2016) Mapping of deletion breakpoints at the CDKN2A locus in melanoma:
 detection of MTAP-ANRIL fusion transcripts. *Oncotarget*, 7, 16490-16504.
- Patel, A. *et al.* (2014) Amplification and thrifty single-molecule sequencing of recurrent somatic
 structural variations. *Genome Res.*, 24, 318-328.
- 9 21. Norris, A.L. *et al.* (2015) Transflip mutations produce deletions in pancreatic cancer. Genes 10 Chromosomes *Cancer* **54**, 472-481.
- 12 22. Guney, S. *et al.* (2011) 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.
 Genes Chromosomes Cancer, **50**, 715-725.
- 14 23. Florl,A.R. *et al.* (2003) Peculiar structure and location of 9p21 homozygous deletion breakpoints
 15 in human cancer cells. *Genes Chromosomes Cancer*, **37**, 141-148.
- 16 24. Kitagawa, Y. *et al.* (2002) Prevalent involvement of illegitimate V(D)J recombination in
- 17 chromosome 9p21 deletions in lymphoid leukemia. J. Biol. Chem., 277, 46289-46297.
- 25. Sasaki, S. *et al.* (2003) Molecular processes of chromosome 9p21 deletions in human cancers.
 Oncogene, **22**, 3792-3798.
- 20 26. Cayuela, J.M. et al. (1997) Disruption of the multiple tumor suppressor gene
- 21 MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic
- 22 leukemias. *Blood*, **90**, 3720-3726.
- 23 27. Pasmant, E. et al. (2007) 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. *Cancer Res.*, **67**, 3963-3969.
- 28. Raschke,S. *et al.* (2005) Homozygous deletions of CDKN2A caused by alternative mechanisms in
 various human cancer cell lines. *Genes Chromosomes Cancer*, **42**, 58-67.
- 28 29. Hussussian, C.J. *et al.* (1994) Germline p16 mutations in familial melanoma. *Nat. Genet.*, **8**, 1521.
- 30. Freedberg, D.E. *et al.* (2008) Frequent p16-independent inactivation of p14ARF in human
 31 melanoma. *J. Natl Cancer Inst.*, **100**, 784-795.
- 31. Harinck, F. *et al.* (2012) Indication for CDKN2A-mutation analysis in familial pancreatic cancer
 families without melanomas. *J. Med. Genet.*, **49**, 362-365.
- 34 32. Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of 35 human genes and genetic disorders. *Nucl. Acids Res.*, **33**, D514-D517.
- 33. Amberger, J. *et al.* (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucl. Acids Res.*, **37**, D793-D796.

- 1 34. Qiao, J.L. *et al.* (2021) CDKN2A deletion leading to hematogenous metastasis of human gastric
- 2 carcinoma. *Frontiers in Oncology*, **11**, 801219. doi: 10.3389/fonc.2021.801219.
- 3 35. Liu, Y. *et al.* (2018) Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas.
 4 *Cancer Cell*, **33**, 721-735.e8.
- 5 36. Peters, G. (2008) Tumor suppression for ARFicionados: the relative contributions of p16INK4a 6 and p14ARF in melanoma. *J. Natl. Cancer Inst.*, **100**, 757-759.
- 7 37. Cui, C.H. *et al.* (2015) P16-specific DNA methylation by engineered zinc finger methyltransferase
- 8 inactivates gene transcription and promotes cancer metastasis. *Genome Biol.*, **16**, 252. doi:
- 9 10.1186/s13059-015-0819-6.
- 10 38. Li,Q. *et al*. (2010) Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9 at 11 the p16 locus. *PLoS One*, **5**, e13732. doi: 10.1371/journal.pone.0013732.
- 12 39. Li,H. *et al.* (2009) The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature*, **460**, 11361139.
- 14 40. Sharpless, N.E. *et al.* (2004) The differential impact of p16(INK4a) or p19(ARF) deficiency on cell
- 15 growth and tumorigenesis. *Oncogene* **23**, 379-385.
- 16
- 17

1 Tables

2

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	

3

4

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

2 P16-Light and P14-qPCR

		n	<i>CDKN2A</i> CDR SCND- positive by P16-Light		CDKN2A	P14 ^{ARF} SCND	CDKN2A CDR or P14		^{RF} CDKN2A CDR & P14 ^{ARF}	
					positive by P14-qPCR		SCND-positive		SCND-positive	
			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
	Noncardiac	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	1-11	46	11	23.9°	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
	Т3	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 ^د	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; ^bOdds 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

1 Figure Legends

2 Figure 1. Genomic coordinates of interstitial CDKN2A deletion/fusion in human cancer genomes. (A) 3 Estimated coordinates of CDKN2A deep deletion in cancer tissues according to the COSMIC data. (B) 4 True coordinates at the base resolution of CDKN2A deletion in cancer cell lines (n=92, blue lines) and 5 gastric cancer (n=18, purple lines) according to sequencing data. The two top charts display the 6 coordinates of most deletion fragments. The sample ID is labeled under each column. The two bottom 7 charts display the amplified view of these deletion fragments, where the 17-kb and 5.1-kb common 8 deletion regions (CDRs) are highlighted with a red dashed line rectangle. The 5.1-kb true CDR from the P16^{INK4A} promoter to intron-2 is exactly the same region as the deleted CDKN2A fragment in the 9 10 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} (gray shadow) and exon-1 α /1 β /2/3 (black dots) are 11 also labeled as landmarks. The positions of amplicons for P16-Light and P14-gPCR are illustrated 12 13 with violet and green lines, respectively. The detailed deletion coordinates for each sample are listed 14 in Data file 2 and Data file 12.

Figure 2. The estimated coordinates of deep-deletion fragments overlapped with tumor suppressor genes *ATM*, *FAT1*, *RB1*, *PTEN*, and *miR31HG* according to the COSMIC data. The 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.

19 Figure 3. Detection of the copy number of CDKN2A intron-2 with quantitative gene-specific multiplex 20 PCR (P16-Light). (A) The location of the 129-bp amplicon within the common deletion region (CDR) 21 and its host genes. (B) The amplification efficiency of two amplicons for the GAPDH and CDKN2A 22 genes in the template titration assays using standard DNA samples from RKO cells (with two wildtype CDKN2A alleles) and A549 cells (with a homozygous CDKN2A deletion). (C) Effects of the 23 24 amount of template DNA on the efficiency of PCR amplification for amplicons in the CDKN2A and 25 GAPDH genes (left chart) and detection of the relative CDKN2A gene copy number (right chart). (D) 26 The linear relationship between the proportion of CDKN2A copy number deletion and ratios of RKO 27 cells spiked with A549 cells. (E) Stability of the proportion of the CDKN2A copy number deletion by 28 P16-Light during ten experiments over different days. The RKO cell DNA templates were spiked with 29 0, 10%, 20%, 25%, and 30% A549 cell DNA. Each column represents the average proportion of 30 CDKN2A copy number deletions in triplicate. Exp. 1 - 10: the results of 10 repeated experiments 31 performed on different days. *P<0.05.

Figure 4. Comparisons of somatic copy number variations (SCNVs) of the *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.

- 1 The average proportion of cancer cells in each GC group is labeled. (C) Correlation analysis between
- 2 the proportion and relative copy number of the *CDKN2A* gene in GCs.





[Fig.2]





[Fig.4]