Platinum drug reprogramming of protein phosphorylation

Luyu Qi[1,2,#], Yinzhu Hou[1,2,#], Qun Luo[1,2],Yan Xu[2], Xingkai Liu[1], Wanchen Yu[3], Bobo Xin,[4] Xiangjun Li[2], Yanyan Zhang[1], Feifei Jia[1], Tiantian Fang[1], Shijun Wang[3,*], Peter J. Sadler[5,*], Yao Zhao[1,*], Fuyi Wang[1,2,3,*]

[1] Beijing National Laboratory for Molecular Sciences; CAS Research/Education Center for Excellence in Molecular Sciences; National Centre for Mass Spectrometry in Beijing; CAS Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing, 100190, People's Republic of China
[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[3] Basic Medical College, Shandong University of Traditional Chinese Medicine, Jinan 250355, P. R. China
[4] NGS Department, Sangon Biotech (Shanghai), Shanghai 201611, P. R, China
[5] Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom
[#]These authors contributed equally to this work.
*Correspondence: fuyi.wang@iccas.ac.cn (F.Y.W.), yaozhao@iccas.ac.cn (Y.Z.), p.j.sadler@warwick.ac.uk (P.J.S.), wsj@sdutcm.edu.cn (S.J.W.)

ABSTRACT

Cisplatin is a DNA-targeting chemotherapeutic. We have utilized a forward chemical genetics strategy to map 7585 cisplatin-damaged genes (CDGs) with a fold-enrichment of >12 from A549 human lung cancer cells. The highly associated signalling pathways of the CDGs include sperm motility, molecular mechanism of cancer, and protein kinase A signalling. Among the CDGs, there are 1330 enzyme, 747 transcription regulators and 486 transporter genes. Importantly, cisplatin targets 306 protein kinase genes, accounting for 59% of putative protein kinase genes in the human genome, and 92 protein phosphatase genes which account for 67.6% of all protein phosphatases in the human genome. This suggests that cisplatin can reprogram protein phosphorylation genome-wide, evidenced by cisplatin-induced reduction in expression of 7 protein kinase genes in the sperm motility signalling pathway, and by CRISPR/dCas9-mediated imaging, which showed that cisplatination on the *PTPRN2* gene recruits HMGB1, but repels Smad3, a transcription factor. Silencing NCCIT testicular cancer cell *SPAG9*, which expresses JIP-4 in testicular haploid germ cells to activate MAPK signalling, resulted in similar apoptosis-inducing activity to cisplatin, implicating *SPAG9* as a potential target for precise testicular cancer therapy.

INTRODUCTION

Cisplatin, *cis*-[$PtCl_2(NH_3)_2$], is a first-line chemotherapeutic drug, in particular by systemic administration in combination therapy for solid tumours[1,2]. After hydrolysis in cells, cisplatin binds strongly to a variety of biological molecules, e.g., proteins, thiol-containing small molecules, and nucleic acids[3,4]. However, the major target is believed to be nuclear DNA (nDNA)[3,5], although nDNA adduct formation has been linked to side effects and toxicity[6], and not the sole determinant of cisplatin-induced cytotoxicity[7] The formation of 1,2-intrastrand -GG-/-AG- (90%) adducts, 1,3-intrastrand -GNG- adducts and interstrand crosslinks unwinds and bends DNA helices, inhibiting DNA replication and transcription, and in turn inducing cell apoptosis and death[3].

The human high mobility group box 1 (HMGB1) protein is a non-specific DNA binding protein, and well recognized for its specific binding to cisplatin-crosslinked double–stranded DNA in both cell-free media[8,9] and in cells[9-11]. HMGB1 functions as a protector of cisplatin-induced DNA damage from nucleotide excision repair (NER)[5], and may serve as a negative transcription cofactor to block the binding of transcription factors such as Smad3 to DNA, as revealed by single cell imaging[11]. Box a domain of HMGB1 has been demonstrated to have a higher affinity than box b for 1,2-cisplatin-crosslinked DNA[9,12]. Based on such a high affinity of HMGB1a domain for cisplatin-bound DNA, Shu et al. developed a "cisplatin-seq" method to map cisplatin lesion sites with a single base resolution throughout the whole genome[12]. They revealed that cisplatin was more likely to attack mitochondrial DNA (mDNA) than nDNA, and that the cisplatin damage was enriched within promotors and regions harbouring transcription termination sites[12]. The aim of the research reported here was to map for the first time the exact genomic distribution of cisplatin-damaged gene loci.

Despite technical challenges arising from the extremely low incidence of cisplatin damage on DNA on a genome-wide scale[13], and impacts of cisplatin damage on the PCR-based sequencing, e.g., stalling PCR process and making base mismatch, there have been a number of innovative strategies that provide insights into the cisplatin damage sites at the genome-wide level[14]. Apart from the cisplatin-seq approach described above, Hu et al. developed a set of methods, termed "damage-seq" and "eXcision Repair-seq" (XR-seq), to map at single-nucleotide resolution, the location and repair of cisplatin lesions in genomic DNA adducts[15], respectively. Their mapping data demonstrated that cisplatin damage was uniform and dominated by the underlying sequence, but the repair of cisplatin lesions was heterogeneous throughout the whole genome and affected by multiple factors such as transcription and chromatin states. Therefore, they thought that the overall effect of cisplatin damage in the genome is not primarily decided by cisplatin damage but by the effectiveness of repair[15]. However, similar to the cisplatin-seq approach, the damage-seq method was not able to assign cisplatin lesions to specific gene loci. To address this issue, the same group integrated their "damage-seq", "XR-seq" and RNA-seq approaches to study comprehensively the genome-wide profiles of cisplatin lesions in mouse organs[16]. They revealed that the expression of 12 genes was commonly downregulated due to cisplatin damage across mouse kidney, liver, lung, and spleen, and also a large number of genes in a specific group of organs, e.g. lung or kidney, lung and liver, or liver, spleen and kidney[16]. However, the combined use of transcriptomics strategies did not produce a full map of cisplatin lesions throughout the whole genome.

Here, we have mapped cisplatin-damaged gene loci genome-wide by using forward chemical genetics strategy[17,18]. The biotin-streptavidin system is often applied to enrich and isolate target DNA and proteins at ultra trace levels in complicated biological samples[17-21]. However, to avoid the effect of biotin modification on the chemical properties and pharmacological functions of the drug cisplatin itself, we combined HMGB1 box a (HMGB1a)-based affinity microprobes and high throughput sequencing to screen gene lesion loci induced by cisplatin on a whole-genome scale (Fig. 1), and to correlate the gene lesions with the action of the drug. Combined with extensive Ingenuity Pathway Analysis (IPA) on significantly damaged genes by cisplatin (fold-enrichment >12 compared to the same genes captured from the control cells), our studies reveal that cisplatin damaged 445 genes out of the 628 prior therapeutic gene targets for cancers identified by CRISPR/Cas9-based screening[22]. Significantly, we show that cisplatin attacks 306 out of 518 human putative protein kinase genes[23]

and 92 out 136 protein phosphatases in human genome[24], suggesting that cisplatin's mode of pharmacological activity involves reprogramming protein phosphorylation pathways. These findings provide new insights into the mechanisms of action of this front-line anticancer drug and new targets for drug design.

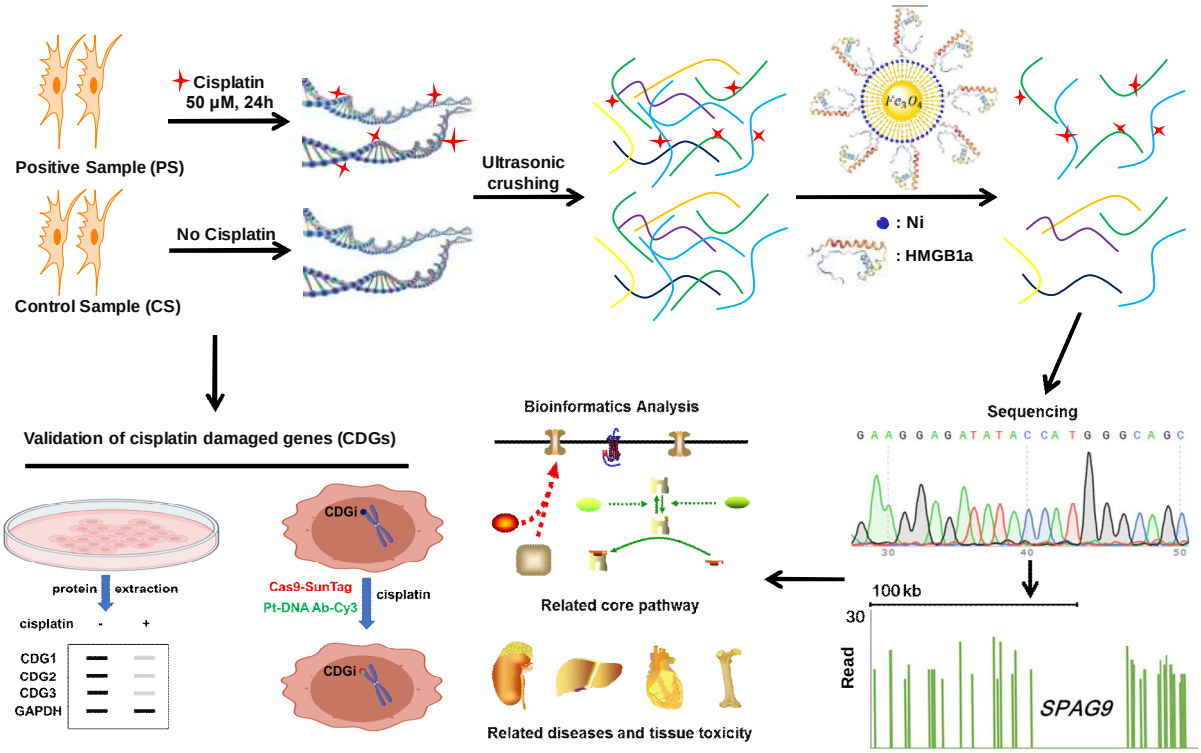

Fig. 1. Diagrammatic illustration of the workflow for capture and mapping of gene loci attacked by cisplatin. The procedure includes cell culture, extraction and fragmentation of genomic DNA, capture, sequencing and alignment of cisplatin-damaged DNA fragments, identification of cisplatin-damaged genes (CDGs) and bioinformatics analysis of CDGs, validation of CDGs by Western Blot assays and confocal fluorescence imaging.

RESULTS

Enrichment and sequencing of cisplatin-damaged DNA

We functionalized Ni-based magnetic microbeads with HMGB1a to assemble affinity microprobes for capturing and isolating 1,2-cisplatin-crosslinked DNA[3,9] from genomic DNA fragments derived from human A549 non-small cell lung cancer cells treated with 50 μM cisplatin for high throughput Illumina® sequencing (Fig. 1). Details of the expression and purification of HMGB1a, construction of affinity microprobes, cell harvesting, and DNA extraction and fragmentation are provided in Methods and Details.

Since HMGB1 binds to cisplatin-crosslinked DNA with high affinity, but can also bind to bent native DNA[3], we extracted and fragmented genomic DNA from A549 cells with (positive samples) and without (control samples) exposure to cisplatin for DNA sequencing. ICP-MS analysis for Pt showed that the platination level of genomic DNA extracted from cisplatin-treated cells reached 320 Pt atoms per 10^6 base pairs (Fig. 2A), consistent with our previous report[13]. Furthermore, gel-electrophoresis data demonstrated that the size of the DNA fragments gained from both the positive sample (PS) and the control sample (CS) ranged from 250 to 1000 bp, but the number of larger DNA fragments from positive samples was less than those from control samples (Fig. 2B). This is likely ascribed to platination, which could render genomic DNA more susceptible to sonication and more prone to aggregation[25]. Given that HMGB1a binds to both non-platinated and platinated DNA, it is not surprising that the same amount of the microprobes captured similar amounts of DNA fragments from positive and control samples (Extended Data Table 1). However, the subsequent sequencing revealed that there were significant differences between the DNA fragments enriched from the positive and control samples.

Next, we constructed DNA libraries based on the captured DNA fragments. Since cisplatin damage on DNA may stall DNA replication and/or cause base mismatch[14,26], NaCN was used to remove the Pt coordinated to DNA to eliminate the Pt-modification-induced bias for PCR amplification during construction of the sequencing libraries of cisplatin-damaged DNA[15]. To avoid the use of toxic NaCN, we tried to remove Pt bound to DNA bases by using excess of cysteine, since Pt(II) also has a high affinity for S-donor ligands like cysteine[27]. However, we did not observe any difference between the amplification rates of cysteine-treated and intact platinated DNA (data not shown). Then, considering the effect of PCR time on DNA replication[26], we extended the PCR expansion time of both positive (cells treated with cisplatin) and negative (cells without exposure to cisplatin) samples to 30 min. The results showed that the DNA sequences in DNA libraries of the positive samples and control samples have similar sizes (Fig. 2C). This means that possible Pt-modification-induced bias for PCR amplification of DNA[26] was circumvented by extension of the PCR expansion time during construction of the DNA libraries of extracted DNA from both the positive sample and control sample (details given in Methods and Details), which minimized the risk of false negative results from gene mapping. Indeed, the total read counts and total base counts of the positive samples did not show a pronounced difference from those of the control samples. However, because cisplatin prefers to crosslink -GG- sites on DNA, the GC base-pair counts of the positive samples accounted for a slightly higher ratio to the total base counts than those of the

control samples (Fig. 2D; Extended Data Table 1). Consequently, the mapping rates of total reads in proper base pairs of positive samples were slightly lower than those of control samples (Fig. 2E; Extended Data Table 1). Nevertheless, there are no differences in the total mapping rates (≈99%) and unique mapping rates (>91%) of genomic DNA fragments from both positive and control samples (Fig. 2E; Extended Data Table 1).

To identify cisplatin-damaged genes (CDGs), we compared the total reads of all peaks mapped to a gene in the positive sample with those of all peaks mapped to the same gene in the corresponding control sample. If the ratio, which refers to the fold-enrichment (FE_P) of a peak mapped to a gene, was ≥ 1.5, the peak was counted to the gene. The fold-enrichment (FE_G) of a gene was the sum of the FE_P of all peaks mapped to the gene (more details given in Methods and Details). Following this method, we identified 246,424 and 362,594 peaks with $FE_P \geq 1.5$ from replicates 1 and 2, respectively, allowing us to map totally 17729 CDGs with $FE_G \geq 1.5$ (Fig. 2F; Extended Data Table 2). The four CDGs with the highest FE_G mapped in this way are shown in Fig. 2G. Since HMGB1a binds to double-stranded DNA over 5 base-pairs[3], we calculated the occurrence frequency of pentanucleotide motifs in all peaks containing platination sites, and found that the top 5 most-frequently platinated pentanucleotide motifs are CTGGG, CCAGG, GCTGG, GGAGG and CAGGC (Fig. 2H). Remarkably, *PTPRN2,* the CDG with the highest FE_G (=830, Extended Data Table 2), contains 82 (A/T)GGGG repeats instead of the five pentanucleotide motifs described above. Nevertheless, both the five most-frequently platinated pentanucleotide motifs and the pentanucleotide repeating in *PTPRN2* contain at least a GG site, consistent with the well-established finding that cisplatin prefers to bind to GG sites of DNA[3,5].

The 17729 CDGs are unevenly distributed over all 23 pairs of chromosomes, showing a moderate dependence on the average G-density of the chromosomes, i.e., the higher the average G-density of a chromosome, the more the genes in it were attacked by cisplatin (Fig. 2I, J), again consistent with the known preference of cisplatin for crosslinking -GG- sites. Moreover, the results show that cisplatin mainly attacks the distal intergenic and intron regions (Fig. 2K). This can be attributed to the relatively larger size of these genomic regions compared with others. Given that the promoter sequences are generally short, their platination rate counted in the number of bases will be much higher than those counted in other regions as shown in Fig. 2K, in line with previous reports[12].

It is notable that due to the large difference in the length of genes, the fold-enrichment (EF_G) of a CDG calculated on the fold-enrichment of all peaks to the gene does not correspond simply to the base coverage of all identified peaks on the gene (Extended Data Table 2). For example, the base coverage of all 361 peaks identified to *PTPRN2,* which contain 255499 G-bases in its sequences and has the highest FE_G (830), was 23.4%. In contrast, for microRNA gene *MIR182* which totally has 29 G-bases, only one peak was identified with a EF_P (ratio of the read of a DNA fragment in a positive sample over the control) of 3, but the base coverage of this peak to the whole gene was 99.1% (Extended Data Table 2). With regard to this, the EF_G in this work is an indicator only of CDG-mapping confidence, but not of the degree of damage to a gene.

We utilized the Ingenuity Pathway Analysis (IPA) program to perform genome-wide annotation on CDGs which we mapped. We input the 17729 CDGs into the IPA data pool and matched 14012 protein encoding genes to the human genome (hg19), including 2263 enzymes, 1308 transcription regulators, 768 transporters, 616 kinases and 207 phosphatases (Fig. 2L), as well as 18 microRNA (miRNA) encoding genes of which the

majority have relative low fold-enrichment as their sequences are much shorter than genomic DNA (Extended Data Table 2). Notably, among the 616 cisplatin-damaged kinase genes, 419 are protein kinase genes (PKGs, Extended Data Table 3), accounting for 81% of putative protein kinase genes in the human genome[23]. Besides, there are 140 protein phosphatase genes (PPGs) among the 207 cisplatin-damaged phosphatase genes (Extended Data Table 4). Significantly, among the 14012 CDGs, 445 genes belong to the 628 prior therapeutic gene targets for cancers identified by the CRISPR/Cas9-based screening, 6% of which are targets of clinically-used drugs or candidates in preclinical development[22] (Extended Data Table 5). Furthermore, 67.5% (27) of the first group of 40 prior gene targets[22] are damaged by cisplatin, including 14 PKGs, e.g., the well-known drug targets *EGFR* and *MTOR* (Extended Data Table 5).

Limited by the data handling capacity of IPA, only 7585 CDGs with FE_G >12 could be analysed and annotated further. Among the 7585 CDGs are 259 prior therapeutic gene targets for cancers, 8% of which are clinical drug or preclinical candidate targets. Moreover, the CDGs include 306 protein kinase genes (CDPKGs) and 92 protein phosphatase genes (CDPPGs), of which the average percentage of G-bases is significantly higher than that of all the genes in the human genome (Fig. 2M; Extended Data Tables 3 and 4). This again implies that it is the higher G-density that makes PKGs and PPGs more vulnerable to cisplatin attacking.

Mitochondrial DNA (mDNA) has been shown to be more susceptible to cisplatin attack than nDNA[7,12]. However, here we detected only 14 cisplatin-damaged mitochondrial genes (CDMGs) with low EF_G ranging from 1.5 to 2.4 (Extended Data Table 6). This difference between the previous and present work is likely ascribable to the fact that once HMGB1 is transferred outside the nucleus, it prefers to bind to circular mitochondrial DNA[28,29], which enabled the HMGB1a probe to capture more native mitochondrial DNA from the control samples in the absence of competitive cisplatin-crosslinked DNA, lowering the EF_G ($=\sum EF_P$) of CDMGs for which EF_P is the ratio of the read of an mDNA fragment in a positive sample to that in the control sample. However, considering the much lower G-density in mDNA (Fig. 2M), the possibility should not be excluded that it is the affinity of HMGB1a for native mitochondrial DNA that leads to a false positive report of the susceptibility of mDNA to cisplatin attack when the HMGB1a probe is used to enrich cisplatin-damaged DNA without use of blank controls[12].

It is notable that among the 17729 CDGs with EF_G >1.5, there are 730 pseudogenes, and among the 7585 CDGs with EF_G >12, there are 150 pseudogenes (Extended Data Table 2). In general, pseudogenes are thought to be non-functional “junk genes”. However, increasing evidence argues that pseudogenes can act as key regulators at DNA, RNA or protein levels in various human diseases such as cancer[30,31]. Therefore, the biological implications of the cisplatin-damaged pseudogenes deserve further investigation.

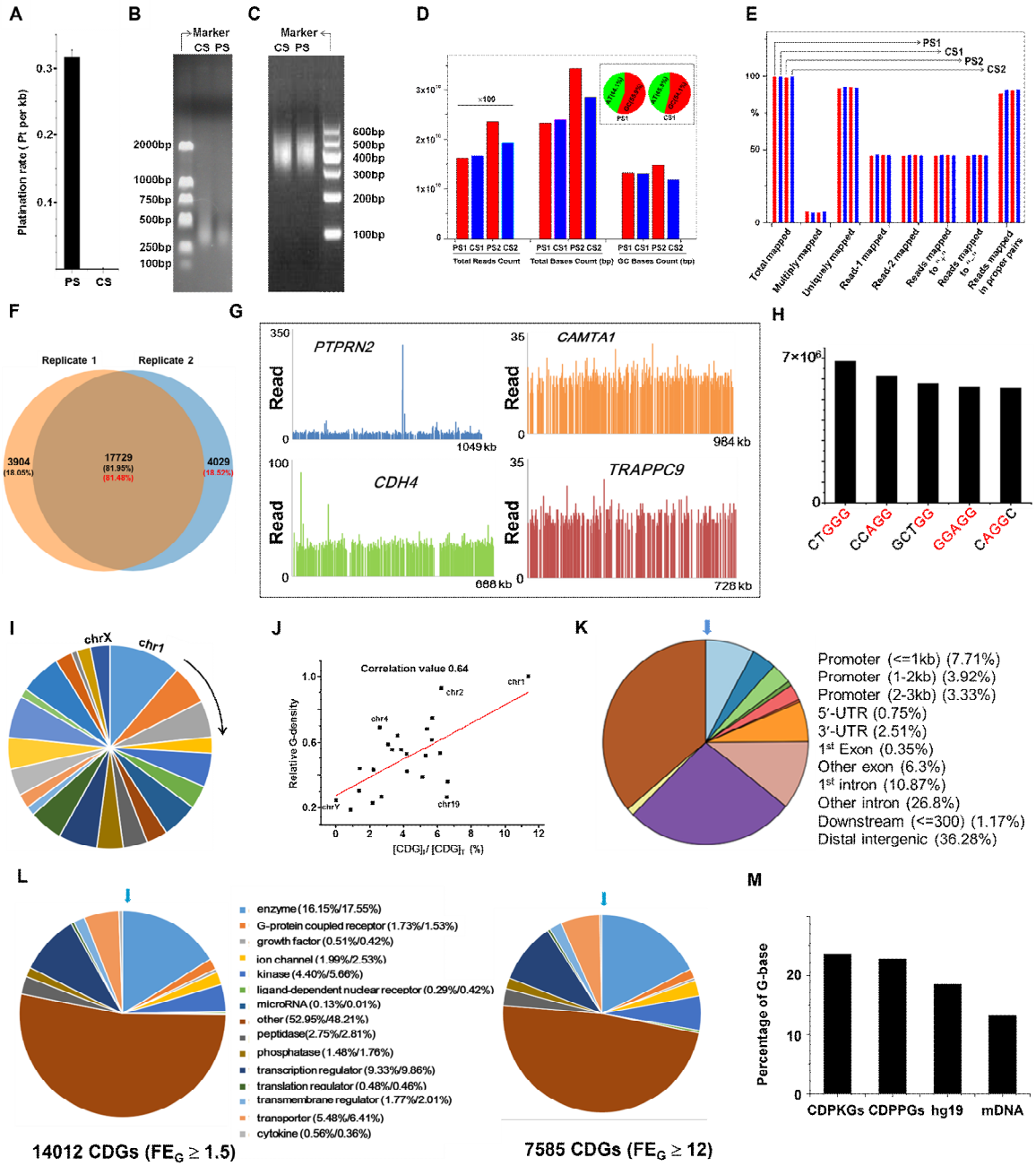

Fig. 2. Mapping of cisplatin damaged genes. **(A)** The platination level of DNA fragments extracted from positive sample (PS) and control sample (CS), determined by ICP-MS. **(B)** The size distribution of DNA fragments from positive sample (PS) and control sample (CS). **(C)** The size distribution of sequences in DNA libraries constructed from captured DNA fragments. **(D)** Total reads, total bases and GC bases mapped for positive samples and control samples. **(E)** Mapping rates counted for positive samples and control samples. **(F)** Venn Diagram of total cisplatin-damaged genes (CDGs) with FE_G >1.5 in two independent biological replicates. **(G)** Reads for each peak mapped to the four CDGs with the highest fold-enrichments. The x axis represents the location of a peak in the linear sequence of each gene, and the y axis represents the read of the peak in the gene. **(H)** The five pentanucleotide motifs that were the most frequently detected in CDGs. **(I)** Distribution of CDGs over chromosomes. **(J)** Correlation of the proportion of CDGs detected in a chromosome ($[CDG]_i$) to total

CDGs ($[CDG]_T$) in human genome with the relative G-density in the chromosome. **(K)** Distribution of cisplatin-damaged DNA fragments in genetic regions. **(L)** Biological function classification of 14012 cisplatin-damaged genes (CDGs) with FE_G >1.5 (left) and 7578 CDGs with FE_G >12 (right). **(M)** Average percent of G bases in CDPKGs and CDPPGs with FE_G >12, hg19 and mDNA.

Highly associated core signalling pathways of cisplatin-damaged genes

Further IPA annotation demonstrated that the 7584 CDGs are mostly associated with the sperm motility signalling pathway (P=10^{-22}), among 302 core signalling pathways (CSPs) to which they are closely related (Extended Data Fig. 1A; Extended Data Table 7). These CDGs associated core signalling pathways can be classified into over 40 categories but are mostly enriched in the ones that are closely related to the inhibition of cancer cell growth and toxicities of cisplatin (Fig. 3A). The cisplatin-inhibited signalling pathways predicted by IPA (with negative z-scores) in the categories, such as signal transduction, organismal growth and development, cellular growth, proliferation and development, growth factor signalling, cellular stress and injury, cancer, apoptosis, and cell cycle regulation are closely related to the inhibition of cell growth. Also, those categories containing inhibited signalling pathways such as neurotransmitters and other nervous system signalling, and cardiovascular signalling, are closely related to the side effects of cisplatin, such as neurotoxicity and cardiotoxicity[6,32].

Notably, cisplatin-damaged protein kinase genes (CDPKGs) on average account for 24.0% of all CDGs involving in the 302 CSPs (insert in Extended Data Fig. 1A; Extended Data Table 7). This proportion is much higher than that for cisplatin-damaged kinase genes to all CDGs with FE_G >12 in the whole genome (5.7%, insert in Extended Data Fig. 1A). For example, there are 87 CDPKGs involved in the sperm motility signalling pathway (Extended Data Fig. 2A), accounting for 60% of all CDGs (146) in this pathway which is predicted to be greatly inhibited with an inhibition index of −8.2 (z-score) by cisplatin-induced gene lesions (Extended Data Table 7). Remarkably, the 87 CDPKGs in the sperm motility signalling pathway are members of the PK family, including *PKA*, *PTK* and *PKC* subfamilies (Extended Data Fig. 2B). We also noted that CDGs are highly associated as well with protein kinase A (PKA) signalling with P=2.5×10^{-14} (Extended Data Fig. 1A), and 209 CDGs account for 54.4% of all genes involved in this pathway, including 38 CDPKGs and 29 CDPPGs (Extended Data Fig. 2A). *PRKCA* is a key node in both sperm motility (Extended Data Fig. 2A) and PKA signalling pathways (Extended Data Fig. 2B and Extended Data Fig. 2C), and is associated with many diseases and functions and involved in phosphorylation of tyrosine[33] and threonine[34]. The high fold-enrichment (=180, Extended Data Table 1) indicates that cisplatin-induced lesions on *PRKCA* play a crucial role in the activity of this drug, meriting its further investigation as a potential target for cancer treatment.

Interestingly, bioinformatics analysis showed that 73 CDPKGs cover 96.0% of all CDGs involved in the IL-15 production signalling pathway (Extended Data Fig. 2D), and are predicted to inhibit IL-15 production with a z value of -8.72 (Extended Data Table 7). Furthermore, three other interleukin (IL) signalling pathways, IL-22 signalling, role of IL-17A in arthritis and IL-15 signalling, have 73.3%, 50% and 39.0% of CDPKG coverage (Extended Data Table 7), respectively, all much higher than the average CDPKG coverage (24.0%) on the CDGs of the 302 CSPs (the insert in Extended Data Fig. 1A). IL-15 production is activated by PTK-PKC phosphorylation signalling. IL-15 induces the activation of JAK kinases, and phosphorylates and activates transcription activators STAT3, STAT5 and STAT6, which may promote the expression of apoptosis inhibitor BCL2L1/BCL-x(L), thus hampering apoptosis[35]. On the other hand, gene expression regulated by IL-22 has been demonstrated to enhance antimicrobial defence and increase cellular mobility[36]. These imply that cisplatin damage to protein

kinase genes of the IL-related signalling pathways may play important roles in the activity of cisplatin, e.g., inducing apoptosis by inhibiting the expression of apoptosis inhibitors such as BCL2L1/BCL-x(L) (Extended Data Fig. 2D).

Another core signalling pathway with which the CDGs are highly associated (P=10^{-15}) is the molecular mechanism of cancer signalling (Extended Data Fig. 2E). We found that 213 CDGs are involved in this pathway, and cover 55.5% of all genes in it. However, due to the complexity of the molecular pathogenesis of cancer, IPA cannot predict how these gene lesions induced by cisplatin regulate the development of cancers (Extended Data Table 7). Further investigation into how CDGs influence each cancer signalling pathway and the roles which each (group) of CDG(s) play in specific pathways are needed.

Moreover, we found that CDGs are also highly associated (P=10^{-6}) with the non-small cell lung cancer (NSCLC) signalling pathway (Extended Data Fig. 2F, Extended Data Table 7). Forty-seven genes, including 13 PKGs, of this core signalling pathway are damaged by cisplatin, covering 64.4% of all genes in it. The z score of this pathway was -4.6, which means that it is predicted to be highly inhibited by cisplatin. Many patients with non-small-cell lung cancer have no response to the tyrosine kinase inhibitor gefitinib, which targets the epidermal growth factor receptor (EGFR)[37]. However, cisplatin can damage the *EGFR* gene in the NSCLC signalling pathway, which may play a role in the anticancer activity of cisplatin towards NSCLC. Given that AKT/PKB can promote survival of NSCLC[38], the damage to the *AKT* gene by cisplatin may contribute to the killing of the NSCLC cancer cells.

Apart from the signalling pathways described above, unexpectedly, a number of signalling pathways with which the cisplatin damaged genes are closely associated (P<10^{-10}) are gathered in the neural system. They include synaptic long-term depression (P=6.3×10^{-19}, Extended Data Fig. 2G), GREB signalling in neurons (P=3.2×10^{-18}, Extended Data Fig. 2H, axonal guidance signalling (P=6.3×10^{-18}), GNRH signalling (P=2.0×10^{-16}), synaptogenesis signalling pathway (P=3.2×10^{-16}), endocannabinoid neuronal synapse pathway (P=2.5×10^{-15}) and netrin signalling (P=2.5×10^{-13}) (Extended Data Table 7). Among them, the synaptogenesis signalling pathway is most inhibited by cisplatin damage on genes, with a z-score of −12.0. Due to the blood-brain barrier, cisplatin accumulates at only a low level in the human brain[39], thus we did not analyse the impact of gene damage on the neural system further. However, this finding will stimulate exploration of whether and how cisplatin damage to neural signalling-related genes is related to its neurotoxicity, such as peripheral sensory neuropathy and central nervous side effects, which can be associated with major toxic side effects of cisplatin in clinical use[6,32].

Additionally，IPA revealed that cisplatin-damaged genes are highly related to the role of NFAT in the cardiac hypertrophy signalling pathway with an association P value of 3.2×10^{-18} (Extended Data Fig. 2I). The 137 CDGs account for 64.9% of all 211 genes involved in this signalling pathway (Extended Data Table 7). The nuclear factor of activated T cells (NFAT) as a transcription factor plays an important role in cardiac hypertrophic signalling which is mediated by the calcium-activated phosphatase calcineurin[40]. This implies that calcineurin-NFAT signalling is dominated by calcium signalling, with which 116 CDGs are highly associated with an association P value of 3.2×10^{-11} (Extended Data Fig. 2J), covering 58.6% of all the 198 genes involved in this signalling pathway (Extended Data Table 7). The calcium signalling pathway is predicted to be strongly

inhibited by cisplatin damage to genes with a *z*-score of −8.5. The calcium channel genes *CACNA1A/B/C* and *CACNA2D3* are all severely damaged by cisplatin with a fold enrichment of 253, 179, 194 and 223, respectively (Extended Data Table 2 and 7), implicating gene damage by cisplatin in cardiac hypertrophy, one of the clinical side effects of cisplatin, *via* disturbance of calcium homeostasis.

As protein kinases and protein phosphatases play dominant roles in cancer cell proliferation and growth, 92 CDPPGs and 306 CDPKGs out of the 7584 CDGs were selected for further IPA annotation. As shown in Extended Data Fig. 3A and Extended Data Table 8, 522 core signalling pathways (CSPs) are closely related to these CDPPGs and CDPKGs with −log(p) > 1.3 and 217 CSPs with −log(p) > 10. Similar to the IPA annotation by all the CDGs, these signalling pathways can be classified into a number of categories that are closely related to the inhibition of cell growth and side effects of cisplatin, including apoptosis, cellular immune response, cytokine signalling, organismal growth and development, cellular growth, proliferation and development and cell cycle regulation. (Fig. 3B). There are also a series of categories that are correlated with the toxic side effects of cisplatin, such as neurotransmitters and other nervous system signalling, cardiovascular signalling, and neuronal system, corresponding to the neurotoxicity and cardiotoxicity.

Specifically, IL-15 production, sperm motility, and protein kinase A are the most notable signalling pathways related to the CDPKGs and CDPPGs, being similar to the IPA annotation of all the CDGs. Moreover, a series of new pathways related to phosphate metabolism were discovered, including D-myo-inositol-5-phosphate biosynthesis, 3-phosphoinositide biosynthesis and degradation (Extended Data Fig. 3A). As shown in Extended Data Table 8, 23 signal pathways are related to cancer. More importantly, there is a signal pathway related to “production of nitric oxide and reactive oxygen species in macrophages” which may be attributed to the disturbance of redox homeostasis in cancer cells by cisplatin[41].

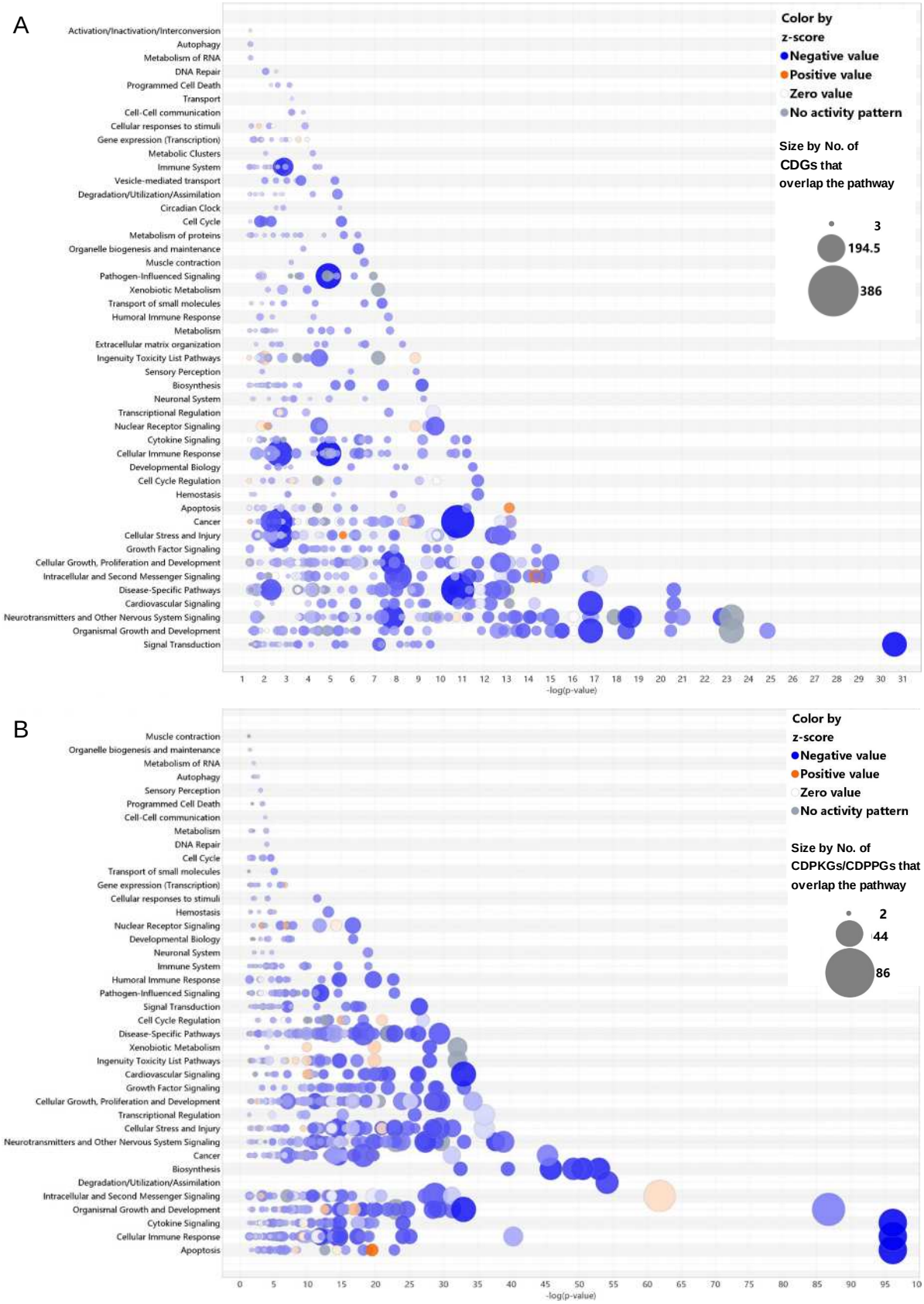

Fig. 3. Bioinformatics analysis of cisplatin-damaged genes. **(A)** Classified core signalling pathways with which the CDGs are closely associated. **(B)** Classified core signalling pathways with which the CDPKGs and CDPPGs are closely associated. The *P* values represent associativity; the smaller *P* values (larger –log(*P* value)) suggest that

the CDGs are more closely related to the pathway. The z-scores indicate the downregulation degree of cisplatin damage on the signalling pathways.

Association of cisplatin-damaged genes with diseases and tissue toxicities

Next, we analysed the association of CDGs with diseases and functions, and discovered that the most associated disease of the 7578 CDGs is cancer with an association *P* value of 1×10^{-312} (Extended Data Fig. 1B and Extended Data Table 9). The four cancer-related diseases or functions, i.e., cancer ($P=8.8\times10^{-154}$), cancer of cells ($P=2.27\times10^{-74}$), cell transformation ($P=9.5\times10^{-9}$) and neoplasia of cells ($P=5.4\times10^{-81}$), are significantly inhibited by cisplatin-induced damage to genes with *z*-scores of −3.15, −2.10, −3.56 and −2.48, respectively (Extended Data Fig. 1C, Extended Data Table 10). The top five types of cancers to which CDGs are highly related are head and neck tumours, endocrine cancer, thyroid gland tumour, gastrointestinal adenocarcinoma and skin tumours (Extended Data Table 10). Of the 92 CDPPGs and 306 CDPKGs selected out of the 7584 CDGs for IPA annotation, cancer is still one of the most related disease and function, with an association *P* value of 5.68×10^{-86} (Extended Data Fig. 3B, Extended Data Table 11).

It is notable that the sperm associated antigen 9 (*SPAG9*) gene is involved in 147 of 361 cancers or functions with which the CDGs are highly associated (Extended Data Table 10). *SPAG9* abundantly expresses C-jun-amino-terminal kinase-interacting protein 4 (JIP-4) in testicular haploid germ cells, and is essential for the development, migration and invasion of cancer[42]. JIP-4 as a scaffolding protein knits mitogen-activated protein kinases (MAPKs) and their transcription factors to activate specific signalling pathways[43]. These imply that cisplatin damage on *SPAG9* (FE_G = 61) may play a crucial role in the anticancer activity of cisplatin.

We further analysed the diseases and functions to which *SPAG9* gene is related. As shown in Extended Data Fig. 4A, *SPAG9* (JIP-4) activates JUN/MAP kinase signalling[43,44], and is associated with various types of cancers, e.g. anaplastic thyroid carcinoma[45], hairy cell leukaemia[46] and intestinal gastric adenocarcinoma[47]. It also plays a role in several important biological processes, such as fertility[48], differentiation of skeletal muscle cells[49], and differentiation of neurons[50]. Moreover, the *SPAG9* gene is regulated by ligand-dependent nuclear receptor ESR1[51], transmembrane receptor CAV1[52], transcription factor ZNE217[53] and transporter BSG[54], and regulates mainly ERK, JNK and p38 MAPK signalling[55] (Extended Data Fig. 4B). JIP-4 functions as a scaffold protein that structurally organizes MAPKs and mediates c-Jun-terminal kinase signalling. It has been demonstrated that *SPAG9* is associated with tumour growth, migration, and invasion in renal cell carcinoma[56]. Sinha et al. discovered that down regulation of *SPAG9* by siRNA silencing reduces growth and invasion potential of triple-negative breast cancer cells[57]. A survival analysis based on an online database (http://gepia2.cancer-pku.cn/#survival) shows that patients who have higher expression level of *SPAG9* have a lower overall survival rate than the patients with lower expression level of *SPAG9* (Extended Data Fig. 4C). Given that *SPAG9* specifically expresses JIP-4 in high abundance in testicular haploid germ cells, the cisplatin damage on it may account for its high cure rate in testicular cancer, which also deserves further investigation.

Tissue toxicity is a major factor which limits the clinical application of cisplatin. However, the molecular basis for its toxicity remains unclear. We performed IPA annotation on the associated toxicity of cisplatin damage to genomic DNA, and found that the most related tissue toxicity is liver hyperplasia or hyperproliferation with an association *P* value of 6.80×10^{-108}, and then cardiac arteriopathy, heart failure, cardiac arrhythmia and

nephrosis with *P* values of 4.83×10^{-15}, 6.59×10^{-8}, 1.76×10^{-7} and 1.12×10^{-6}, respectively (Extended Data Fig. 1C and Extended Data Table 12). Since liver hyperplasia or hyperproliferation are the most related to liver tumours, liver carcinoma and liver cancer (Extended Data Table 12) rather than hepatotoxicity, and that the most frequently reported tissue toxicity of cisplatin in the clinic is nephrotoxicity[6], it seems likely that the most relevant toxicity to all CDGs is nephrotoxicity (Extended Data Fig. 1D). However, further investigation is needed to correlate the gene lesions induced by cisplatin with the phenotypes of clinical nephrotoxicity. Of the 92 CDPPGs and 306 CDPKGs selected for IPA annotation, the most related tissue toxicity is hepatocellular carcinoma and liver hyperplasia/hyperproliferation with association *P* values for both of 3.34×10^{-31}. This is consistent with the hepatotoxicity of cisplatin[6,32]. Side effects on the heart (heart failure, cardiac enlargement, cardiac necrosis/cell death, cardiac dysfunction, cardiac damage, cardiac dilation) and the kidney (renal inflammation, renal nephritis, renal damage, kidney failure, renal dysplasia, renal hypoplasia) are also revealed to be associated with the protein kinase and phosphatase genes damaged by cisplatin (Extended Data Fig. 3C, Extended Data Table 13).

Surprisingly, we detected that apoptosis-related genes *CASP9* and *BCL2*, which encode apoptosis factor caspase-9 and apoptosis regulator BCL2[6,58], were damaged by cisplatin with FE_G values of 32 and 29, respectively (Extended Data Table 2). This indicates that cisplatin-induced lesions on these genes may contribute to acquired cisplatin-resistance of cells, an important clinical problem.

Western Blot assays

To verify the genome-wide mapping results of CDGs, we performed classical Western Blot assays to measure the gene expression level of 20 functionally important genes with and without cisplatin exposure (Extended Data Table 14). These included 18 genes involved in the sperm motility signalling pathway, of which 13 are protein kinases. Two other genes, *SPAG9* and *AKT3*, are the upstream mediator of the MAPK signalling pathway and middle transducer of the PI3K signalling pathway, respectively (Extended Data Fig. 4A-B).

As shown in Fig. 4A, the expression levels of all selected genes decreased after cisplatin treatment. The changes in expression levels of 12 genes, including 10 protein kinase genes and *SPAG9*, were significant ($P < 0.05$) or very significant ($P < 0.01$) subject to cisplatin treatment (Fig. 4B). This indicates that these genes are indeed damaged by cisplatin, strongly supporting the reliability and credibility of the gene mapping results described above. However, there is no simple correlation between the fold-enrichment of a specific CDG and base coverage (%), and fold-change in the expression level of the gene (Fig. 4C), most probably due to complexity of the gene expression machinery that consists of DNA replication, transcription, translation and repair. This sophisticated system finely controls gene expression such that any alteration in this system can influence the expression of a specific gene in addition to the direct damage imposed by cisplatin on the gene. Another possible reason is that the expression level of a selected gene is too low to be characterized, e.g., PRKG1 in NCCIT cells and *EPHB1* in HeLa cells (Extended Data Fig. 5A). Additionally, the location of cisplatin induced lesions on genes is also an important factor in determining their inhibition potency.

Interestingly, the full-length protein product (JIP-4) of *SPAG9* significantly decreased in abundance by 2.0-fold, but the two subunits visualized in band b of Fig. 4A-B showed pronounced increases in intensity in NCCIT

cells upon the treatment of cisplatin. This implies that the cisplatin lesion more likely occurs in the non-encoding region and/or the linkage region of *SPAG9* than in the promotors of SPAG9 and the encoding region of its four subunits (Extended Data Fig. 5B). Indeed, when we further analysed the cisplatin-damaged sites on *SPAG9* (Extended Data Table 15), we found that (i) no cisplatin lesions were identified on the four promotor sequences; and (ii) cisplatin damaged the non-encoding regions, i.e., intron and intergenic regions, much more than the encoding regions of the gene, i.e., exons (Extended Data Fig. 5C). These make it possible that although the full-length JIP-4 protein decreased in abundance, one or two subunits of JIP4 increased in abundance. However, the increase in abundance of subunits shown in band b (Fig. 4A) does not compensate the loss of full-length JIP-4. Furthermore, we applied siRNA to silence the *SPAG9* gene (Fig. 4D), and found that silencing of *SPAG9* induced a similar ratio of early- and late-stage apoptosis in NCCIT cells as 12 μM cisplatin (Fig. 4E), though its potency to inhibit the growth of the NCCIT cell line was weaker than cisplatin (Fig. 4F). This confirmed that *SPAG9* plays an important role in the action of cisplatin as the upstream mediator of the MAPK pathway (Extended Data Fig. 4).

The expression of *MERTK* was the most strongly inhibited (FC = 32) due to cisplatin damage (Fig. 4A-B). This receptor tyrosine kinase, MERTK, phosphorylates Akt1-Y26, mediating Akt activation and survival signalling, which in turn drives oncogenesis and therapeutic resistance[59]. Thus, cisplatin damage to *MERTK* may be implicated in mechanism of action of cisplatin, also supporting that *MERTK* is a potential therapeutic target[59,60].

The expression of *MERTK*-related gene *AKT3* was inhibited 3.5-fold by cisplatin (Fig. 4A-B). *AKT3* encodes a member of the AKT serine/threonine protein kinase family. As a middle-stream regulator of IGF1 signalling, AKT3 plays a key role in signalling and regulating cell survival in insulin signalling, angiogenesis and tumour formation. Cisplatin-induced damage to this gene could trigger a series of downstream changes, inducing apoptosis and death of cancer cells[61].

Validation of cisplatin-damaged genes by fluorescence imaging

To further validate the mapping results of cisplatin-damaged gene loci, we utilized CRISPR/dCas9-mediated fluorescence imaging[62] to visualize the *PTPRN2* gene, which was damaged by cisplatin with the highest fold-enrichment (*vide supra*), accompanied with immunostaining of cisplatin-DNA adducts[63] at the single cell level (more details in Method and Details). The repeating peptide array termed SunTag was genetically fused to the endonuclease-deficient Cas9 (dCas9) protein[64] and recruited multi-copies of sfGFP-scFv complex to illuminate *PTPRN2* gene in GIST882 (human gastrointestinal stromal tumour) and HEK293T (human embryonic kidney) cells directed by *PTPRN2*-specific guide RNA (Extended Data Fig. 6A). The superposition of *PTPRN2*-sfGFP signal and the platinated DNA antibody signal in GIST882 cells showed remarkable co-localization of *PTPRN2* and platinated DNA, evidence for cisplatin-induced damage on the gene.

To further verify the cisplatin-induced damage on *PTPRN2*, we genetically fused HMGB1 or Smad3, with blue fluorescent protein (BFP), and then *in situ* located these proteins and *PTPRN2* gene in living HEK293T cells by confocal fluorescent imaging. Smad3 is a transcription factor binding to GGC(GC)/(CG) and AGAC motifs of genomic DNA[65,66]. As shown in Fig. 5A, in the absence of cisplatin, HMGB1-BFP hardly co-localized with the sfGFP-labelled *PTPRN2*. However, upon the exposure to cisplatin, HMGB1-BFP strikingly co-localized with

PTPRN2, indicating the recognition and complexation of cisplatin-damaged *PTPRN2* gene by HMGB1. In contrast, the transcription factor Smad3 significantly co-localized with *PTPRN2* in the absence of cisplatin, but hardly bound to sfGFP-labelled *PTPRN2* in the presence of cisplatin in HEK293T cells (Fig. 5B). The statistical results (Fig. 5C – F) confirmed the phenomenon that cisplatin-bound genomic DNA facilitates the binding of HMGB1 to nDNA, but blocks the binding of Smad3 to nDNA, though this effect did not show simple and obvious concentration- and time-relevance (Fig. 5C – F). In other words, these results suggested that cisplatin damage on *PTPRN2* recruited HMGB1, but hampered the binding of Smad3 to this gene, most likely inhibiting expression of the damaged genes[67]. Taken together, the CRISPR/dCas9-mediated fluorescence imaging further verified the cisplatin-induced damage on *PTPRN2* at the single cell level.

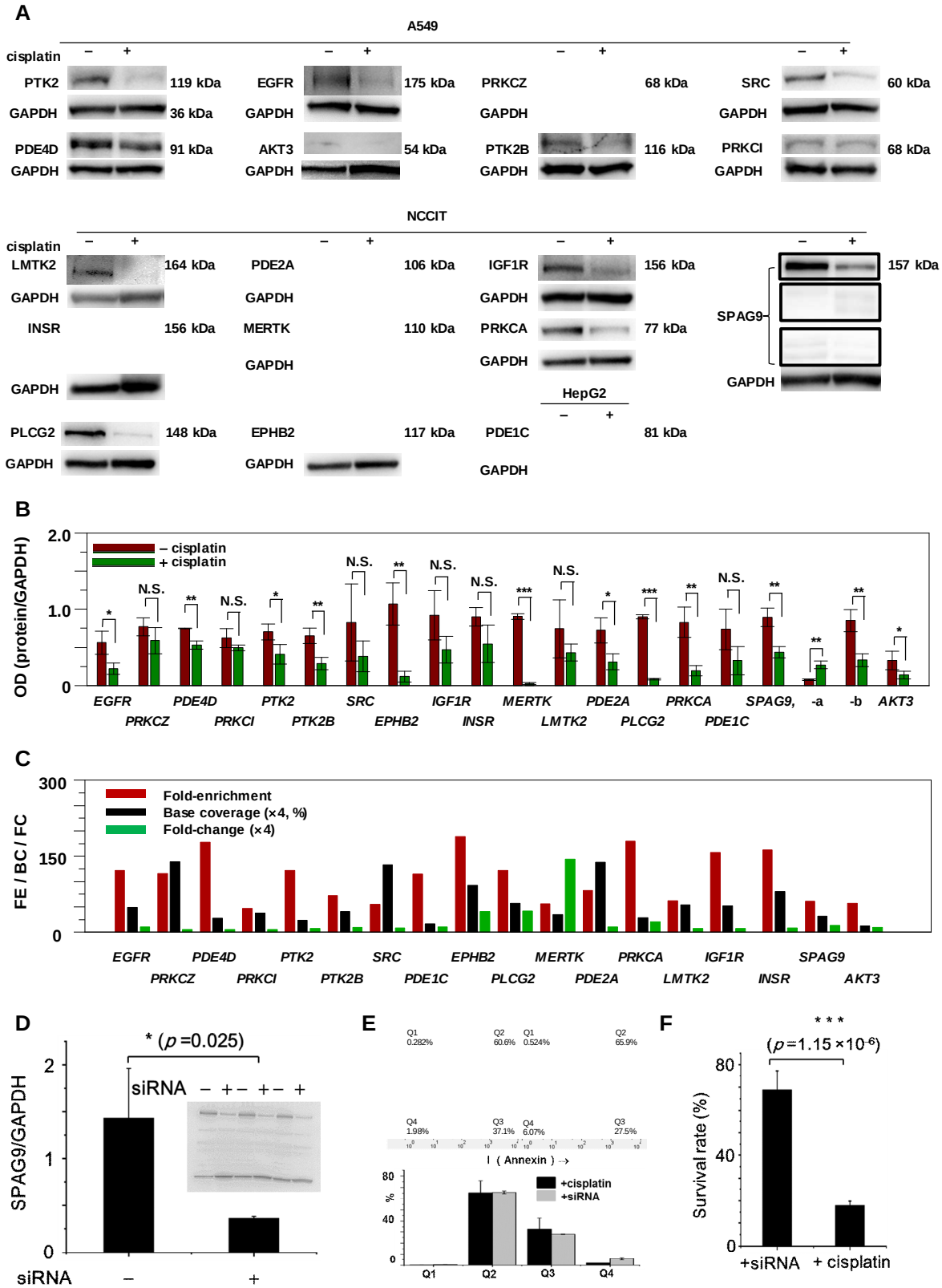

Fig. 4. Western Blot assays and siRNA silencing. **(A)** Western Blotting images of selected gene encoding proteins expressed in A549 and NCCIT cells with (+) and without (−) cisplatin treatment (12 μM). GAPDH was used as the internal reference. **(B)** Optical density (OD) ratio of a gene-encoded protein to GAPDH with (+) and

without (-) cisplatin treatment. Two-tailed unpaired Student's t-test (n = 3), N.S. refers to not significant; * indicates $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. P_{EGFR}=0.0244, P_{PRKCZ}=0.2057, P_{PDE4D}=0.0024, P_{PRKCI}=0.1615, P_{PTK2}=0.0366, P_{PTK2B}=0.0088, P_{SRC}=0.2343, P_{EPHB2}=0.0044, P_{IGF1R}=0.1000, P_{INSR}=0.0893, P_{MERTK}=2.71×10^{-6}, P_{LMTK2}=0.2453, P_{PDE2A}=0.0216, P_{PLCG2}=1.16×10^{-6}, P_{PRKCA}=0.0062, P_{PDE1C}=0.0925,$P_{SPAG9\text{-a}}$=0.0049, $P_{SPAG9\text{-b}}$=0.0030, $P_{SPAG9\text{-c}}$=0.0061, P_{AKT3}=0.0126. **(C)** The fold-enrichment (FE_G) and base coverage of 18 mapped cisplatin-damaged genes (CDGs) selected for Western Blotting, and the fold-change in the expression of the CGDs. **(D)** Western Blotting data for siRNA silencing of SPAG9. **(E)** Flow cytometric quantification of viability (Q1), early-stage apoptosis (Q2), late-stage apoptosis (Q3) and necrosis (Q4) for NCCIT cells subject to cisplatin treatment (12 μM) or siRNA silencing of *SPAG9*. **(F)** Survival rate (%) of NCCIT cells subject to cisplatin treatment (12 μM) or siRNA silencing of *SPAG9*.

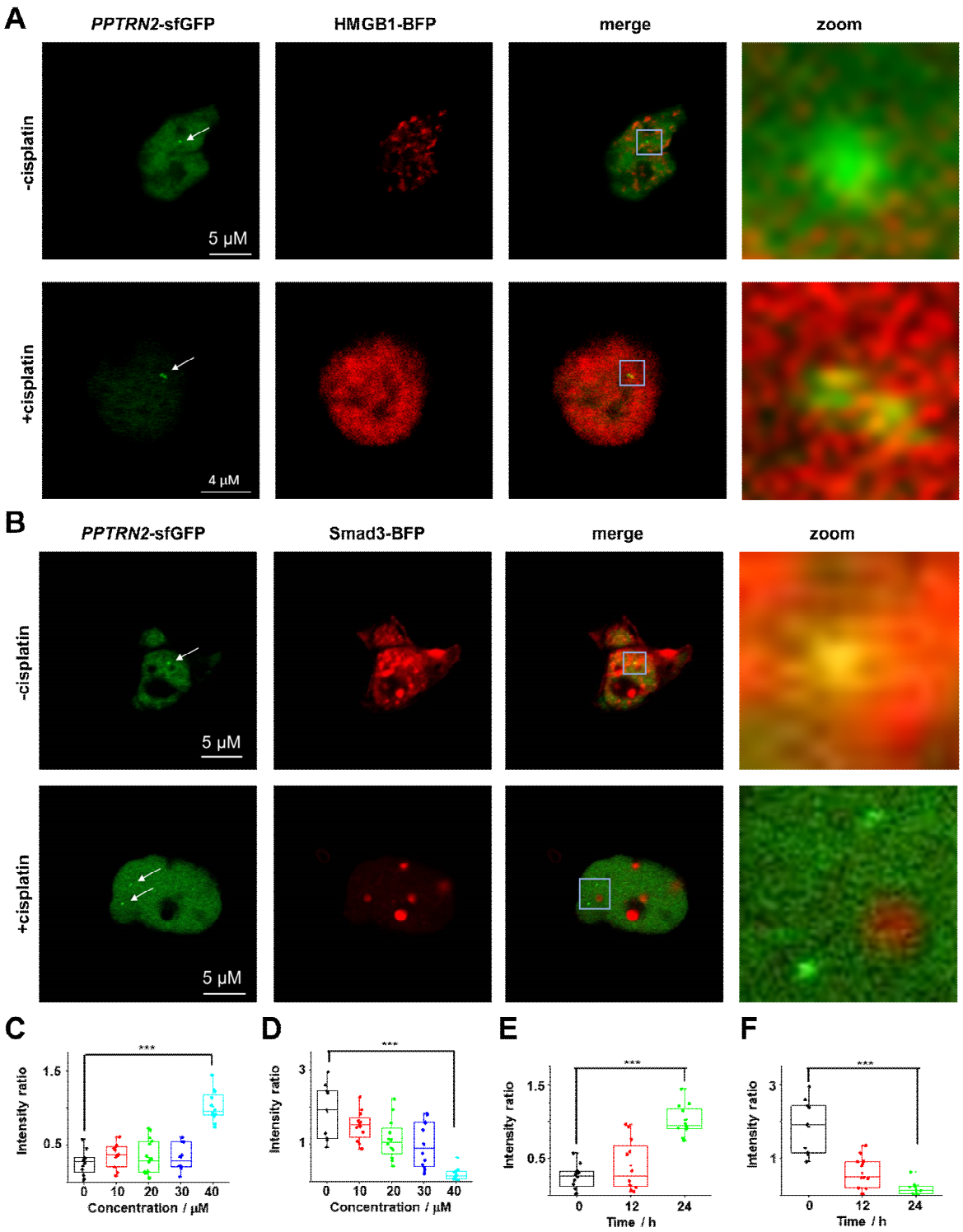

Fig. 5. Validation of cisplatin damage on *PTPRN2* by confocal fluorescence imaging of HEK293 cells without (–) or with (+) cisplatin treatment. **(A)** Co-localization of CRISPR/dCas9 mediated sfGFP labelled *PTPRN2* (green, λ_{ex} = 488 nm, λ_{em} = 500 – 550 nm) and BFP fused HMGB1 (red, λ_{ex} = 405 nm, λ_{em} = 410 – 460 nm) in the absence (top panel) and the presence (bottom panel) of cisplatin (12 μM) in HEK293T cells. The zoomed-in view indicates that without cisplatin treatment, HMGB1 did not co-localize with *PTPRN2* gene, while co-localizing remarkably with *PTPRN2* signal subject to cisplatin treatment. **(B)** Co-localization of CRISPR/dCas9 mediated

sfGFP labelled *PTPRN2* (green) and BFP fused Smad3 (red, λ_{ex} = 405 nm, λ_{em} = 410 – 460 nm) in the absence (top panel) and the presence (bottom panel) of cisplatin (12 μM) in HEK293T cells. The zoomed-in view indicates that without cisplatin treatment, Smad3 significantly co-localized with *PTPRN2* gene, whereas hardly co-localized with the *PTPRN2* signal in the presence of cisplatin. (**C - F**) The statistic results for the colocalization of PTPRN2-sfGFP and HMGB1-BFP (**C** and **E**) or Smad3-BFP (**D** and **F**) in HEK293T cells treated with different concentration of cisplatin for 24 h (**C** and **D**) or with 40 μM of cisplatin for various times (**E** and **F**). The details for calculating the intensity ratio (colocalization ratio) are shown in Extended Data Fig. 6B.

DISCUSSION

Cisplatin is a genetically toxic anticancer drug, but its gene targets or precise gene loci are largely unknown, limiting the understanding of the cellular impact of DNA damage induced by the drug. In 2016, two methods, termed "cisplatin-seq"[12] and "damage-seq"[15], respectively, were developed back-to-back to map or locate the cisplatin damaged sites in genomic DNA. Both the methods achieved single-base (nucleotide) resolution mapping of cisplatination sites on specific genic regions such as promotors, transcription termination sequences and intergenic regions[12,15]. However, the precise gene loci damaged by cisplatin remain unclear.

To address this issue, transcriptomics has been used in combination with the "damage-seq" method to map the cisplatin damaged gene loci across mouse organs[16]. However, due to the tissue-specificity of transcriptomics, only genes which are transcribed in specific tissues of mouse, e.g., liver, lung and spleen, could be mapped for cisplatin lesions. In recent years, a number of studies have applied chemical genetic strategies, in particular chemical proteomic approaches to uncover gene targets of cisplatin, and have produced significant advances in dissecting its mechanism of action. However, similar to the transcriptomics approach, these studies do not fully reveal cisplatin-damaged gene loci throughout the whole genome because they focused on discovering gene targets of cisplatin *via* measuring the difference between gene expressions in a specific type of cancer cells or tumour tissues upon cisplatin exposure. For example, Jimenez and co-workers found that UGGT1, COL6A1 and MAP4 can be used as potential biomarkers for determining cisplatin response, by proteomic analysis of non-small cell lung cancer cells[68]. Wilmes et al. have used integrated omics techniques to study differentially-expressed proteins and related pathways for cisplatin nephrotoxicity[69].

In this work we have used the principle of forward chemical genetics[17,18], combining affinity separation and high throughput sequencing to map cisplatin-damaged gene loci. By integration with bioinformatics analysis, our data give rise to two important and innovative findings. One is that among the 14012 identified cisplatin-damaged protein-encoding genes, 445 genes are included in the 628 prior gene targets for cancer treatment reported very recently[22]. Another is that cisplatin attacks 306 protein kinase genes with high fold-enrichment (>12); these genes cover 59% of consensual protein kinase genes in the human genome[23]. These cisplatin-damaged protein kinase genes are involved in the majority of the 302 core signalling pathways, which are closely related to the anticancer activity and toxic side effects of cisplatin. Moreover, we discovered that cisplatin also induced damage on 140 protein phosphatase genes. It is clear that cisplatin genome-wide disturbs the signalling of protein phosphorylation of cells.

Protein kinases accompanied by phosphatases regulate protein phosphorylation, the most common and important protein post-translational modification, involved in many natural biochemical pathways and cellular processes. Since abnormal protein phosphorylation is usually and closely related to cancer development, protein kinases have been major drug targets for cancer therapy[70,71]. During the first and second wave of intensive studies of molecularly-targeted drugs, a number of small molecule protein kinase inhibitors (PKIs), e.g. gefitinib, imatinib and dasatinib[72], have been developed and approved for clinical treatment of various cancers. Given our findings that protein kinase genes are strongly susceptible to cisplatin attack, and that the cisplatin-damaged protein kinase genes are highly associated with 260 core signalling pathways, this work unambiguously demonstrates that cisplatin acts as a broad-spectrum protein kinase inhibitor by inducing

lesions in protein kinase genes. However, on the one hand, cisplatin irreversibly blocks protein phosphorylation by damaging protein kinase genes, and on the other hand, cisplatin-induced lesions on protein kinase genes appear to be relatively non-specific. These results suggest that a strategy for the design of next-generation platinum anticancer drugs and/or drug delivery systems might involve minimizing non-specific attack on protein kinase genes of healthy cells[2,73], and use of small molecule inhibitors for specific protein kinases which play crucial roles in the development of particular cancers. Hence, molecular targeting of protein kinase genes and protein kinases should provide a valuable approach to precise cancer therapy.

Ninety-two protein phosphatases are damaged by cisplatin with high fold-enrichment (> 12). It is notable that the protein tyrosine phosphatase receptor, *PTPRN2*, gained the highest fold-enrichment among 7578 CDGs we mapped, and other PTPR members, *PTPRD/T/M/J/F/G* were detected with fold-enrichment of > 100 as well. These results mean that cisplatin may disturb protein phosphorylation signalling not only by attacking protein kinase genes, but also by inducing lesions on protein phosphatase genes. This may explain why intracellular response to cisplatin activates rather than inhibits Ataxia Telangiectasia mutated (ATM)/ATM and Rad3-related (ATR)/DNA-protein kinase (PK) dependent pathways[74,75] while it induces damage on protein kinase genes as revealed in this work. Taken together, cisplatin-induced damage on protein kinase and phosphatase genes genome-wide, imply that cisplatin exerts its pharmacological functions by reprogramming protein phosphorylation in cancer cells.

Importantly, we found significantly reduced expression of *SPAG9* by cisplatin-induced lesions, notably contributing to cisplatin-induced apoptosis as verified by siRNA silencing. Because *SPAG9* exclusively expresses the upstream mediator JIP-4 of p38 MAPK signalling in the testis[76,77], our findings suggest that *SPAG9* likely accounts for the unique and high cure rate of cisplatin for testicular cancer, and is a potential target for precise therapy of testicular cancer.

REFERENCES

1 Ramalingam, S. S., Owonikoko, T. K. & Khuri, F. R. Lung cancer: new biological insights and recent therapeutic advances. *CA. Cancer J. Clin.* **61**, 91-112, (2011).

2 Johnstone, T. C., Suntharalingam, K. & Lippard, S. J. The next generation of platinum drugs: targeted Pt(II) agents, nanoparticle delivery, and Pt(IV) prodrugs. *Chem. Rev.* **116**, 3436-3486, (2016).

3 Jamieson, E. R. & Lippard, S. J. Structure, recognition, and processing of cisplatin-DNA adducts. *Chem. Rev.* **99**, 2467-2498, (1999).

4 Mezencev, R. Interactions of cisplatin with non-DNA targets and their influence on anticancer activity and drug toxicity: the complex world of the platinum complex. *Current Cancer Drug Targets* **14**, 794-816, (2014).

5 Jung, Y. & Lippard, S. J. Direct cellular responses to platinum-induced DNA damage. *Chem. Rev.* **107**, 1387-1407, (2007).

6 Qi, L. *et al.* Advances in toxicological research of the anticancer drug cisplatin. *Chem. Res. Toxicol.* **32**, 1469-1486, (2019).

7 Yang, Z. *et al.* Cisplatin Preferentially Binds Mitochondrial DNA and Voltage-Dependent Anion Channel Protein in the Mitochondrial Membrane of Head and Neck Squamous Cell Carcinoma: Possible Role in Apoptosis. *Clin. Cancer Res.* **12**, 5817-5825, (2006).

8 Zeng, W. *et al.* Proteomic strategy for identification of proteins responding to cisplatin-damaged DNA. *Anal. Chem.* **91**, 6035-6042, (2019).

9 Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708-712, (1999).

10 Jia, F. F. *et al.* In Situ Visualization of Proteins in Single Cells by Time-of-Flight-Secondary Ion Mass Spectrometry Coupled with Genetically Encoded Chemical Tags. *Anal. Chem.* **92**, 15517-15525, (2020).

11 Lin, Y. *et al.* Single cell imaging reveals cisplatin regulating interactions between transcription (co)factors and DNA. *Chem. Sci.* **12**, 5419-5429, (2021).

12 Shu, X. T., Xiong, X. S., Song, J. H., He, C. & Yi, C. Q. Base-Resolution Analysis of Cisplatin-DNA Adducts at the Genome Scale. *Angew. Chem. Int. Ed.* **55**, 14244-14247, (2016).

13 Guo, W. *et al.* Transferrin serves as a mediator to deliver organometallic ruthenium(II) anticancer complexes into cells. *Inorg. Chem.* **52**, 5328-5338, (2013).

14 Mingard, C., Wu, J., McKeague, M. & Sturla, S. J. Next-generation DNA damage sequencing. *Chem. Soc. Rev.* **49**, 7354-7377, (2020).

15 Hu, J. C., Lieb, J. D., Sancar, A. & Adar, S. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11507-11512, (2016).

16 Yimit, A., Adebali, O., Sancar, A. & Jiang, Y. Differential damage and repair of DNA-adducts induced by anti-cancer drug cisplatin across mouse organs. *Nat. Commun.* **10**, 309, (2019).

17 Walsh, D. P. & Chang, Y.-T. Chemical genetics. *Chem. Rev.* **106**, 2476-2530, (2006).

18 Johnson, E. O. *et al.* Large-scale chemical-genetics yields new M. tuberculosis inhibitor classes. *Nature* **571**, 72-78, (2019).

19 Liu, X., Bing, T. & Shangguan, D. Microbead-Based Platform for Multiplex Detection of DNA and Protein. *Acs Applied Materials & Interfaces* **9**, 9462-9469, (2017).

20 Wang, J. *et al.* Direct Quantification of MicroRNA at Low Picomolar Level in Sera of Glioma Patients Using a Competitive Hybridization Followed by Amplified Voltammetric Detection. *Anal. Chem.* **84**, 6400-6406, (2012).

21 Meng, X., Sun, P., Xu, H. & Wang, Z. Folic acid-functionalized magnetic nanoprobes via a PAMAM dendrimer/SA-biotin mediated cascade-amplifying system for the efficient enrichment of circulating tumor cells. *Biomaterials Science* **8**, 6395-6403, (2020).

22 Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511-516, (2019).
23 Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934, (2002).
24 Sacco, F., Perfetto, L., Castagnoli, L. & Cesareni, G. The human phosphatase interactome: An intricate family portrait. *FEBS Lett.* **586**, 2732-2739, (2012).
25 Onoa, G. B., Cervantes, G., Moreno, V. & Prieto, M. J. Study of the interaction of DNA with cisplatin and other Pd(II) and Pt(II) complexes by atomic force microscopy. *Nucleic Acids Res.* **26**, 1473-1480, (1998).
26 Marini, V. *et al.* Conformation, protein recognition and repair of DNA interstrand and intrastrand cross-links of antitumor trans-[PtCl2(NH3)(thiazole)]. *Nucleic Acids Res.* **33**, 5819-5828, (2005).
27 Reedijk, J. Why does cisplatin reach guanine-N7 with competing S-donor ligands available in the cell? *Chem. Rev.* **99**, 2499-2510, (1999).
28 Ito, H. *et al.* HMGB1 facilitates repair of mitochondrial DNA damage and extends the lifespan of mutant ataxin-1 knock-in mice. *EMBO Molecular Medicine* **7**, 78-101, (2015).
29 Liu, Y. *et al.* Hypoxia induced HMGB1 and mitochondrial DNA interactions mediate tumor growth in hepatocellular carcinoma through Toll-like receptor 9. *J. Hepatol.* **63**, 114-121, (2015).
30 Lou, W., Ding, B. & Fu, P. Pseudogene-Derived lncRNAs and Their miRNA Sponging Mechanism in Human Cancer. *Frontiers in Cell and Developmental Biology* **8**, (2020).
31 Cheetham, S. W., Faulkner, G. J. & Dinger, M. E. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics* **21**, 191-201, (2020).
32 Chiorazzi, A., Semperboni, S. & Marmiroli, P. Current view in platinum drug mechanisms of peripheral neurotoxicity. *Toxics* **3**, 304-321, (2015).
33 Saito, Y., Hojo, Y., Tanimoto, T., Abe, J. & Berk, B. C. Protein kinase C-alpha and protein kinase C-epsilon are required for Grb2-associated binder-1 tyrosine phosphorylation in response to platelet-derived growth factor. *J. Biol. Chem.* **277**, 23216-23222, (2002).
34 Chen, Y. M. *et al.* Microtubule affinity-regulating kinase 2 functions downstream of the PAR-3/PAR-6/atypical PKC complex in regulating hippocampal neuronal polarity. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8534-8539, (2006).
35 Masuda, A., Matsuguchi, T., Yamaki, K., Hayakawa, T. & Yoshikai, Y. Interleukin-15 prevents mouse mast cell apoptosis through STAT6-mediated Bcl-x(L) expression. *J. Biol. Chem.* **276**, 26107-26113, (2001).
36 Wolk, K. *et al.* IL-22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis. *Eur. J. Immunol.* **36**, 1309-1323, (2006).
37 Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129-2139, (2004).
38 Brognard, J., Clark, A. S., Ni, Y. C. & Dennis, P. A. Akt/protein kinase B is constitutively active in non-small cell lung cancer cells and promotes cellular survival and resistance to chemotherapy and radiation. *Cancer Res.* **61**, 3986-3997, (2001).
39 Heydorn, K., Rietz, B. & Krarup-Hansen, A. Distribution of platinum in patients treated with cisplatin determined by radiochemical neutron activation analysis. *J. Trace Elem. Exp. Med.* **11**, 37-43, (1998).
40 Kumarapeli, A. R. K. *et al.* Alpha B-crystallin suppresses pressure overload cardiac hypertrophy. *Circ. Res.* **103**, 1473-1482, (2008).
41 Zhang, J. *et al.* Systematic identification of anticancer drug targets reveals a nucleus-to-mitochondria ROS-sensing pathway. *Cell* **186**, 2361-2379.e2325, (2023).

42 Yan, Q. *et al.* SPAG9 is involved in hepatocarcinoma cell migration and invasion via modulation of ELK1 expression. *OncoTargets and therapy* **9**, 1067-1075, (2016).
43 Kelkar, N., Standen, C. L. & Davis, R. J. Role of the JIP4 scaffold protein in the regulation of mitogen-activated protein kinase signaling pathways. *Mol. Cell. Biol.* **25**, 2733-2743, (2005).
44 Lee, C. M., Onesime, D., Reddy, C. D., Dhanasekaran, N. & Reddy, E. P. JLP: A scaffolding protein that tethers JNK/p38MAPK signaling modules and transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14189-14194, (2002).
45 Wagle, N. *et al.* Response and acquired resistance to everolimus in anaplastic thyroid cancer. *N. Engl. J. Med.* **371**, 1426-1433, (2014).
46 Abaan, O. D. *et al.* The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer Res.* **73**, 4372-4382, (2013).
47 Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573-582, (2014).
48 Iwanaga, A. *et al.* Ablation of the scaffold protein JLP causes reduced fertility in male mice. *Transgenic Res.* **17**, 1045-1058, (2008).
49 Lu, M. & Krauss, R. S. N-cadherin ligation, but not Sonic hedgehog binding, initiates Cdo-dependent p38 alpha/beta MAPK signaling in skeletal myoblasts. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4212-4217, (2010).
50 Oh, J.-E. *et al.* Cdo promotes neuronal differentiation via activation of the p38 mitogen-activated protein kinase pathway. *FASEB J.* **23**, 2088-2099, (2009).
51 Notas, G. *et al.* Whole transcriptome analysis of the ER alpha synthetic fragment P-295-T-311 (ER alpha 17p) identifies specific ER alpha-isoform (ER alpha, ER alpha 36)-dependent and -independent actions in breast cancer cells. *Molecular Oncology* **7**, 595-610, (2013).
52 Mercier, I. *et al.* Genetic ablation of caveolin-1 drives estrogen-hypersensitivity and the development of DCIS-like mammary lesions. *Am. J. Pathol.* **174**, 1172-1190, (2009).
53 Krig, S. R. *et al.* Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. *J. Biol. Chem.* **282**, 9703-9712, (2007).
54 Supper, V., Hartl, I., Boulegue, C., Ohradanova-Repic, A. & Stockinger, H. Dynamic interaction- and phospho-proteomics reveal Lck as a major signaling hub of CD147 in T cells. *J. Immunol.* **198**, 2468-2478, (2017).
55 Wang, H. M. *et al.* Scaffold protein JLP is critical for CD40 signaling in B lymphocytes. *J. Biol. Chem.* **290**, 5256-5266, (2015).
56 Garg, M. *et al.* Sperm-Associated Antigen 9 Is Associated With Tumor Growth, Migration, and Invasion in Renal Cell Carcinoma. *Cancer Res.* **68**, 8240-8248, (2008).
57 Sinha, A. *et al.* Down regulation of SPAG9 reduces growth and invasive potential of triple-negative breast cancer cells: possible implications in targeted therapy. *J. Exp. Clin. Cancer Res.* **32**, 69-79, (2013).
58 Dasari, S. & Tchounwou, P. B. Cisplatin in cancer therapy: Molecular mechanisms of action. *European Journal of Pharmacology* **740**, 364-378, (2014).
59 Jiang, Y. *et al.* MERTK mediated novel site Akt phosphorylation alleviates SAV1 suppression. *Nat. Commun.* **10**, 1515, (2019).
60 Graham, D. K., DeRyckere, D., Davies, K. D. & Earp, H. S. The TAM family: phosphatidylserine-sensing receptor tyrosine kinases gone awry in cancer. *Nat. Rev. Cancer* **14**, 769-785, (2014).
61 Khan, P. *et al.* Aspirin enhances cisplatin sensitivity of resistant non-small cell lung carcinoma stem-like cells by targeting mTOR-Akt axis to repress migration. *Sci. Rep.* **9**, 16913, (2019).
62 Chen, B. H. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479-1491, (2013).
63 Hu, J., Lieb, J. D., Sancar, A. & Adar, S. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11507-11512, (2016).

64 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635-646, (2014).
65 Mullen, A. C. *et al.* Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-beta Signaling. *Cell* **147**, 565-576, (2011).
66 Martin-Malpartida, P. *et al.* Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. *Nat. Commun.* **8**, 2070, (2017).
67 Lin, Y. *et al.* Single Cell Imaging Revealed Cisplatin Regulating Interactions between Transcription (Co)factors and DNA. *Chem. Sci.* **12**, 5333-5690, (2021).
68 Bottger, F. *et al.* Proteome analysis of non-small cell lung cancer cell line secretomes and patient sputum reveals biofluid biomarker candidates for cisplatin response prediction. *J. Proteomics* **196**, 106-119, (2019).
69 Wilmes, A. *et al.* Mechanism of cisplatin proximal tubule toxicity revealed by integrating transcriptomics, proteomics, metabolomics and biokinetics. *Toxicology in Vitro* **30**, 117-127, (2015).
70 Katsogiannou, M., Andrieu, C. & Rocchi, P. Heat shock protein 27 phosphorylation state is associated with cancer progression. *Frontiers in Genetics* **5**, 346, (2014).
71 Cohen, P. Protein kinases - the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309-315, (2002).
72 Baselga, J. Targeting tyrosine kinases in cancer: The second wave. *Science* **312**, 1175-1178, (2006).
73 Wang, X. & Guo, Z. Targeting and delivery of platinum-based anticancer drugs. *Chem. Soc. Rev.* **42**, 202-224, (2013).
74 Reinhardt, H. C., Aslanian, A. S., Lees, J. A. & Yaffe, M. B. p53-Deficient Cells Rely on ATM- and ATR-Mediated Checkpoint Signaling through the p38MAPK/MK2 Pathway for Survival after DNA Damage. *Cancer Cell* **11**, 175-189, (2007).
75 Pines, A. *et al.* Global Phosphoproteome Profiling Reveals Unanticipated Networks Responsive to Cisplatin Treatment of Embryonic Stem Cells. *Mol. Cell. Biol.* **31**, 4964-4977, (2011).
76 Jagadish, N., Rana, R., Mishra, D., Kumar, M. & Suri, A. Sperm associated antigen 9 (SPAG9): a new member of c-Jun NH2 -terminal kinase (JNK) interacting protein exclusively expressed in testis. *The Keio journal of medicine* **54**, 66-71, (2005).
77 Suri, A. *et al.* Cancer testis antigens A new paradigm for cancer therapy. *Oncoimmunology* **1**, 1194-1196, (2012).
78 Kumar, D., Mariappan, G., Husain, A., Monga, J. & Kumar, S. Design, synthesis and cytotoxic evaluation of novel imidazolone fused quinazolinone derivatives. *Arabian Journal of Chemistry* **10**, 344-350, (2017).
79 Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**, 669-680, (2009).
80 Kraemer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530, (2014).

Materials and Methods

Escherichia coil

E. coli strain BL21 with HMGB1a-encoding plasmid (DE3) (Tian Gen Biotech, China) was cultivated in Luria-Bertani (LB) medium with 50 μg/mL kanamycin. A single colony picked from a fresh plate was used to inoculate 2YT medium (containing Bacto™ Peptone (16 mg/mL), Bacto™ Yeast Extract (10 mg/mL, Becton, Dickinson and company, USA), sodium chloride (5 mg/mL, Beijing Chemical Works, China) and kanamycin (50 μg/mL, Solarbio, China)) overnight at 37 °C and 180 rpm. Then the cells were transferred to a large volume of fresh 2YT medium at 37 °C and 180 rpm. When the OD_{600} value of the cells reached 0.5 – 1.0, IPTG (Sigma-Aldrich, USA) was added at a final concentration of 1 mM to induce gene expression. After incubation at 37 °C and 180 rpm for further 12 h, the cells were harvested and washed once with PBS (Solarbio, China), and frozen at -80°C for at least one hour for better sonication.

Cell lines

The A549 human non-small cell lung cancer cell line, HeLa cervical cancer cell line, HepG2 hepatoma cell line (National Infrastructure of Cell Line Resource, Beijing, China) and NCCIT testicular cancer cell line (Fukesai Biological Technology Co., Ltd, Nanjing, China) were individually grown in DMEM medium (Gibco) supplemented with 10% FBS (Gibco) and 1% penicillin and streptomycin (GE Healthcare) at 37 °C. When the cell density reached confluency, the culture medium was replaced by fresh DMEM in the absence or presence of cisplatin at various concentrations (50 μM for A549 and HepG2, 12 μM for NCCIT, 25 μM for HeLa) for 24 h. Then the cells were harvested and washed twice by PBS. Whole cell protein or genomic DNA were extracted from different cells with different extraction kits.

Expression and purification of HMGB1a

The cDNA of human HMGB1 domain A (HMGB1a, Extended Data Chart 1.) was synthesized by Proteintech (Rosemont, USA) and cloned into pET-28a plasmid. The resulting plasmid encoding a his_6-tag was transfected into BL21 (DE3) *E. coli* cells (Tian Gen Biotech, China). The cells were grown in a small volume of 2YT medium containing Bacto™ Peptone (16 mg/mL), Bacto™ Yeast Extract (10 mg/mL, Becton, Dickinson and company, USA), sodium chloride (5 mg/mL, Beijing Chemical Works, China) and kanamycin (50 μg/mL, Solarbio, China) overnight at 37 °C and 180 rpm. Then the cells were transferred to a large volume of fresh 2YT medium at 37 °C and 180 rpm. When the OD_{600} value of the cells reached 0.5 – 1.0, IPTG (Sigma-Aldrich, USA) was added at a final concentration of 1 mM to induce gene expression. After incubation at 37 °C and 180 rpm for further 12 h, the cells were harvested and washed once with PBS (Solarbio, China), and frozen at -80°C for at least one hour for better sonication. The deep frozen cells were resuspended in 10 mL cold lysis buffer (PBS and 5 mM imidazole, pH 7.4), and lysed by sonication for 20 min at 0 °C, followed by centrifugation for 30 min at 4 °C and 12000 rpm. The supernatant was purified by His Trap™ FF crude column (GE Healthcare) which was eluted first by 50 mM imidazole (Innochem), then by 300 mM imidazole in phosphate buffer. The eluted HMGB1a containing a his_6-tag was characterized by SDS-PAGE and mass spectrometry (Extended Data Fig. 7) .

Assembly and characterization of HMGB1a-functionalized affinity microprobes

To construct affinity microprobes for cisplatin-crosslinked DNA, 2.1 mg purified HMGB1a was incubated with 7.5 mg nickel-modified magnetic beads (MAg25K/IDA Ni, Enriching Biotechnology, Shanghai, China) suspended

in 1 mL PBS, and rotated for 1 h at 4 °C. The beads were then separated by magnetic force and washed twice with PBS for 10 min. As shown in Extended Data Fig. 8A, the absorbance of the supernatant over the wavelength range 250 – 300 nm significantly decreased due to loading of HMGB1a onto the magnetic beads. This indicated successful functionalization of the magnetic beads by HMGB1a.

Cell culture and genomic DNA extraction

The A549 human non-small cell lung cancer cells obtained from the Centre for Cell Resource of Peking Union Medical College Hospital were cultured in DMEM medium (Gibco) supplemented with 10% FBS (Gibco) and 1% penicillin and streptomycin (GE Healthcare) at 37 °C. When the cell density reached confluency, the culture medium was replaced by fresh DMEM containing 50 μM cisplatin (Beijing Ouhe Technology, China), and the cells were incubated at 37 °C for 24 h. Then the cells (~10^7) were harvested, and genomic DNA was extracted by whole genome extraction kit (Tian Gen Biotech). The DNA extracts were crushed into fragments between 250 to 750 bp with an ultrasonic breaker (Scientz-IID ultrasonic homogenizer, SCIENTZ, Ningbo, China). At the same time, a similar amount of cells without exposure to cisplatin were harvested for genomic DNA extract and fragmentation. The DNA fragments from the cells treated with cisplatin were considered as a positive sample and those from the cells without cisplatin treatment as a control sample for capture and sequencing of cisplatin damaged genes. An inductively coupled plasma-mass spectrometer (ICP-MS, Agilent 7700) was used to determine the Pt concentrations of the DNA fragments obtained from positive and control samples.

Capture of cisplatin-bound DNA fragments

The same amounts (74 μg) of DNA fragments gained from cells treated and untreated with cisplatin were individually re-dissolved in 10 mM Tris-HCl (pH 7.4, Beijing Biodee Biotechnology), and added into the HMGB1a-functionlized magnetic beads suspended in PBS, followed by rotation for 1 h at room temperature. The beads were then separated by magnetic force and washed twice with 1 mL 10 mM Tris-HCl. The captured HMGB1a-DNA complexes were eluted with 1 mL of buffer containing 20 mM sodium phosphate, 500 mM NaCl and 400 mM imidazole (pH 7.4) *via* rotation at room temperature for 40 min. As shown in Extended Data Fig. 8B, C, the absorbance of the supernatant over the wavelength range 250 – 300 nm significantly decreased subject to incubation with HMGB1a-functionlized magnetic beads, providing evidence for the capture of DNA fragments by the HMGB1a-based affinity microprobes.

Each eluted HMGB1a-DNA complex sample was centrifuged and lyophilized, and then re-dissolved in 500 μL deionized water, followed by addition of 500 μL mixture of phenol (Solarbio) – chloroform (Beijing Chemical Works) – isopropanol (Concord Technology, Tianjin) (25: 24: 1). The resulting mixture was shaken firstly for 10 s, and kept at room temperature for 2 min, then centrifuged at 12000 rpm and 4 °C for 7 min. The supernatant was mixed with 500 μL of a mixture of chloroform-isopropanol (24: 1), shaken again for 10 s, kept at room temperature for 2 min, and then centrifuged at 12000 rpm and 4 °C for 7 min. The isolated supernatant was mixed with 1500 μL pre-chilled absolute ethanol (Concord Technology, Tianjin) and 50 μL sodium acetate solution (3 M, Beijing Chemical Works), kept at −20 °C overnight, and then centrifuged at 12000 rpm and 4 °C for 7 min. The supernatant was carefully removed, and the residues were left at room temperature for drying. Finally, the purified DNA fragments were re-dissolved in an appropriate amount of TE solution (Solarbio) for sequencing.

Western Blot assays

The A549 human non-small cell lung cancer cell line, HeLa cervical cancer cell line, HepG2 hepatoma cell line (National Infrastructure of Cell Line Resource, Beijing, China) and NCCIT testicular cancer cell line (Fukesai Biological Technology Co., Ltd, Nanjing, China) were individually grown in the presence and absence of cisplatin at various concentrations (50 μM for A549 and HepG2, 12 μM for NCCIT, 25 μM for HeLa) for 24 h. The cells were then harvested for extracting whole cell proteins using a total protein extraction kit (BestBio, China). The concentration of each protein extract was determined by standard BCA assay (Beyotime, China). To determine the change in expression level of a specific protein due to gene damage induced by cisplatin, the same amount of whole cell proteins extracted from a positive sample and a control sample were individually boiled with the gel-loading buffer (4×LDS, Genscript, Nanjing, China), and loaded onto a 4 – 12% gradient SDS-PAGE gel (Genscript) for electrophoresis. The separated proteins were transferred to a PVDF membrane (Millipore, 0.2 μm). The membrane was blocked in 5% nonfat dry milk (Biofroxx) dissolved in 0.1% TBST, of which 1 L contained 2.4 g Tris base (Innochem), 8.8 g NaCl and 0.1% (v/v) Tween 20 (Beijing Topbio Science&Technology), at 4 °C overnight, and then incubated with respective primary antibodies (Abcam) in appropriate dilutions at room temperature for 2 h. Thereafter, the membranes were washed three times with 5% nonfat dry milk in 0.1% TBST for 30 min each time and incubated with horseradish peroxidase-conjugated secondary antibody (Abcam) at room temperature for 1 h. After washing twice with 5% nonfat dry milk in 0.1% TBST and once with 0.1% TBST, the protein bands were visualized by enhanced chemiluminescence (Beyotime, China), and the optical densities were determined using an image analyser (Image J).

Since these chosen genes are expressed at different levels in different types of cells, we selected four human cancer cell lines, A549 (non-small cell lung cancer cell line), NCCIT (testicular cancer cell line), HepG2 (hepatocellular cancer cell line) and HeLa (cervical cancer cell line) to perform Western Blot assays. Among the four cell lines, NCCIT is the most sensitive to cisplatin with an IC_{50} value of 5.8 μM (24 h treatment). The IC_{50} values for HeLa, A549 and HepG2 cells were ca. 8.2 μM (24 h), 10.1 μM (48 h), 22.4 μM (24 h)[78], respectively. Thus, we applied different doses of cisplatin to treat the cell lines, i.e. 50 μM for A549 and HepG2, 25 μM for HeLa and 12 μM for NCCIT.

For siRNA silencing of *SPAG9* experiments, the NCCIT testicular cancer cell line were transfected with siRNA against *SPAG9* (sequence 5'-3': rGrGrArGrCrArGrArUrUrUrArCrUrArGrGrArATT, Suzhou Biosyn Biotechnology Co., Ltd.), or not, for 48 h. The cells were then harvested for extracting whole cell proteins using a total protein extraction kit (BestBio, China). The concentration of each protein extract was determined by standard BCA assay (Beyotime, China). To determine the change in expression level of JIP-4 due to gene silencing by siRNA, the same amount of whole cell proteins extracted from a positive sample and a control sample were individually analysed by using the same method for Western Blot assays described above.

CCK-8 assay

The survival rates of *SPAG9*-selicenced and cisplatin-treated NCCIT cells were measured using the CCK-8 (Cell Counting Kit-8) assay (MedChemExpress, USA). Briefly, the NCCIT cells (1×10^3 cells/per wells) were plated at a density of 2000 cells per well in 100 μL of media into 96-well plates. After 24 h for adherence, the cells were exposed to 12 μM cisplatin for 24 h or transfected with the siRNA against *SPAG9* and grown for 48 h. Then, 10

μL CCK-8 was added into the medium and the cells were grown for further 4 h. Finally, the optical density (OD) for each well was measured using a microplate reader (SpectraMax M5 Molecular Devices Corporation) at the wavelength of 450 nm. The survival rate (SR, %) of cells was calculated based on the equation (1):

$$\mathrm{SR}(\%) = \frac{\mathrm{OD}_{\mathrm{sample}} - \mathrm{OD}_{\mathrm{blank}}}{\mathrm{OD}_{\mathrm{control}} - \mathrm{OD}_{\mathrm{blank}}} \times 100\% \qquad (1)$$

All reported values of SR were averages of three independent experiments and expressed as mean ± SD (standard deviation).

Flow cytometric analysis

The apoptosis rate of NCCIT cells subject to cisplatin treatment and *SPAG9* silencing was determined by Annexin V / propidium iodide (PI) double staining assay following the Apoptosis Detection Kit I protocols (BD Pharmingen, USA). The NCCIT cells (1×10^3 cells/well) were plated into a six-well plate. Twenty-four hours later, NCCIT cells were transfected with siRNA against *SPAG9* for 48 h. Another group of cells were treated with 12 μM cisplatin for 24 h. Then, the supernatant was removed, and cells were transferred to tubes and centrifuged at 1000 rpm for 3 min. Thereafter, the cells were re-suspended in 1× Annexin V binding buffer, and incubated with 5 μL Annexin-V and 5 μL PI conjugate for 15 min in the dark prior to FACS analysis. The FACS assays were performed on a Calibur flow cytometer (BD，Franklin Lakes, New Jersey, US), of which the FL1 channel was used to record the intensity of annexin V-FITC staining and FL2 channel to record the intensity of PI staining. The data were quantified by Sell Quest software (BD, Franklin Lakes, New Jersey, US).

Laser scanning confocal microscopy fluorescence imaging

To achieve the co-localization imaging, human embryonic kidney cells (HEK293T) and human gastrointestinal stromal tumour cells (GIST882) (National Infrastructure of Cell Line Resource, Beijing, China) were seeded in confocal dish (ThermoFisher Scientific) and grew in DMEM in 5% CO_2 in a 37 °C incubator. When the cell density reached about 70%, the SunTag plasmids, including dCas9-SunTag10×-v4, scfv-GCN4-GFP-VP64 and sgPTPRN2 plasmids, which were kindly provided by Professor Yu Zhang at National Institute of Biological Sciences as gifts (the sgPTPRN2 plasmid was further cloned by GenScript), were simultaneously transfected with Lipo3000 transfection reagent (ThermoFisher Scientific). After 4 h, the cells were incubated in fresh DMEM containing 40 μM cisplatin for 24 h, and then treated with cold methanol-acetic acid fixative for 10 min, washed three times with phosphate-buffered saline (PBS), permeabilized with 0.1% Triton-X100/PBS on the decolorizing shaker for 30 min and then blocked with 5% BSA/PBS solution at 37 °C for 1 h. After that, the cells were incubated with the rabbit polyclonal primary antibody to HA tag (Abcam, ab9110) and the rat monoclonal primary antibody against cisplatin modified DNA (Abcam, ab103261) at 37 °C for 1 h, washed 5 times with buffer (0.2% BSA/PBS), then incubated with the donkey anti-rabbit secondary antibody conjugated with Alexa Fluor®488 (Abcam, ab150073) to ab9110 and the goat anti-rat secondary antibody conjugated with Alexa Fluor®555 (Cell signalling technology #4417) to ab103261 at 37 °C for 1 h, followed by washing three times with buffer prior to LSCM imaging. The excitation and emission wavelengths were 488 nm and 500 – 550 nm, 559 nm and 570 – 670 nm for Alexa Fluor®488 and Alexa Fluor®555, respectively.

To localize *PTPRN2* and HMGB1-BFP/Smad3-BFP, HEK293T cells were seeded in 4 confocal dishes (ThermoFisher Scientific) and grown in DMEM in 5% CO2 in a 37 °C incubator. When the cell density reached about 70%, the SunTag plasmids and HMGB1-BFP (2 dishes) or Smad3-BFP plasmid (2 dishes) were simultaneously transfected with Lipo3000 transfection reagent (ThermoFisher Scientific). One dish of cells transfected with HMGB1-BFP, or Smad3-BFP was incubated for 28 h, and then imaged by LSCM. Another dish of cells transfected with HMGB1-BFP, or Smad3-BFP was incubated for 4 h, followed by incubation with 40 μM cisplatin at 37 °C for 24 h, and then imaged by LSCM. The excitation and emission wavelengths were 405 nm and 425 nm – 475nm, 488 nm and 500 – 600 nm for BFP and sfGFP, respectively.

Quantification of protein and DNA concentration

The concentrations of purified protein HMGB1a and extracted genomic DNA were determined by NanoDrop OneC (Thermo Fisher Scientific Inc.). The concentrations of whole cell proteins from different cells were determined using a BCA kit (Beyotime Biotech Inc.).

Sequencing and gene mapping

Sequencing of captured and purified DNA fragments was performed by Shanghai Sangon Biotech Ltd. The DNA fragments were firstly characterized by 1% agarose gel electrophoresis for quality check, and then fragmented to about 500 bp by Covaris and recovered for DNA library construction. Library construction was performed following the instruction of NEB Next® Ultra™ DNA Library Prep Kit for Illumina® (https://international.neb.com/products/e7370-nebnext-ultra-dna-library-prep-kit-for-illumina#Protocols,%20Manuals%20&%20Usage). Since cisplatin binding to DNA fragments inhibits DNA replication and synthesis[3,5,26], the PCR expansion time was prolonged to 30 min. The constructed DNA libraries were characterized by a 2% agarose gel, and then each positive or control sample was sequenced by Illumina® providing at least 20 G of data. The sequencing data for each pair of positive/control sample were simultaneously matched with the human reference genome (hg38) to map genome-widely the DNA fragments captured.

Taking the protein kinase cGMP-dependent 1 (*PRKG1*) gene as an example (Extended Data Fig. 8D, E), we identified cisplatin damaged genes by comparing the total read of all peaks mapped to a gene in the positive sample with that of all peaks mapped to the gene in the corresponding control sample. The fold-enrichment (FE_P) of a peak was calculated using equation 2.

$$FE_p = \frac{R_{PS}}{R_{CS}} \qquad (2)$$

Where R_{PS} and R_{CS} are read (tag) counts for a peak in a pair of positive sample (PS) and control sample (CS), respectively. A peak was considered unreliable and then abandoned if it matched multiple genes. A peak with an fold-enrichment (FE_p) <1.5 was also taken as unreliable and discarded[79]. Thus, the fold-enrichment of a mapped gene is the sum of fold-enrichment of all peaks mapped to the gene, i.e.

$$FE_G = \sum_1^i (FE_P)_i \qquad (3)$$

Where i represents the number of peaks mapped for a gene in a positive sample/control sample pair. Using equations 2 and 3, the fold-enrichment of *PRKG1* was calculated to be 152, implying that this gene was heavily damaged by cisplatin.

Taking into account that cisplatin kills cells by attack on DNA, it might be expected that more cisplatin-damaged genes would be detected in the dead cells of positive samples. However, our sequencing data showed that there was no significant difference between the total numbers and fold-enrichments of cisplatin-damaged genes mapped from the DNA fragments collected from living cells alone and from living cells and dead cells together after 24 h of incubation with 50 μM cisplatin (data not shown). The software and databases used for gene screening and localization are listed in Extended Data Table 16.

Bioinformatics Analysis

Ingenuity Pathway Analysis (IPA, QIAGEN)[80] was used to analyse the core pathways, diseases, toxicities and functions associated with the cisplatin damage. IPA analysis maps each gene to the corresponding molecule in the Ingenuity Pathway Knowledge Base, which is available at the Ingenuity System's web site (http://www.ingenuity.com). We input the 17729 cisplatin damaged genes into the IPA data pool, and matched 14012 protein-encoding genes to the human genome. Limited by the data handling capacity of IPA, however, only 7578 genes with FE_G >12, could be analysed further.

Calculation of fold-change of protein expression

We carried out three technical repetitions of Western Blot assays. The optical densities of protein bands were determined using an image analyser (Image J). The fold-change (FC) of a protein was calculated using equation 4.

$$FC = \frac{OD_{PC}/OD_{GC}}{OD_{PN}/OD_{GN}} \qquad (4)$$

Where OD_{PC} and OD_{GC} are optical densities of a protein band and the GAPDH (the reference) band in a sample treated with cisplatin, and OD_{PN} and OD_{GN} are optical densities of the protein band and the GAPDH band in a sample without cisplatin treatment, respectively.

Statistical analysis

All data were analysed with Origin 8.0 and Microsoft offices excel 2010 software. Statistically significant differences were determined using unpaired t-tests, Student's t-tests, Fisher's exact test right-tailed, as applicable and according to test requirements. A two-sided P value < 0.05 was considered statistically significant. * indicates $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

DATA AVAILABILITY

The raw sequencing data were deposited on Mendeley Data, and available on https://data.mendeley.com/datasets/k9sm56gj5y/1, https://data.mendeley.com/datasets/2tv7ntcgh6/1, https://data.mendeley.com/datasets/4zj3kz34c5/1, https://data.mendeley.com/datasets/dmyfkn7pnz/1, https://data.mendeley.com/datasets/dw3v4sm9nb/1, https://data.mendeley.com/datasets/xb6s7vzcwm/1, https://data.mendeley.com/datasets/75dyzmvbrz/1, https://data.mendeley.com/datasets/s7jbxvwf4z/1, and https://data.mendeley.com/datasets/fv6f9rpg5m/1.

The flow cytometry data contained in this article are deposited on FlowRespository with a Repository ID of FR-FCM-Z45H on https://flowrepository.org/experiments/4273.

Acknowledgments

The authors thank the National Nature Science Foundation of China (No. 21575145, 22377130, and 21927804). Y.Z. was also supported by Beijing Natural Science Foundation, No. 2232034, National Key Research and Development Program of China, No. 2023YFF0713801. P.J.S. research on platinum is supported by Anglo American Platinum. The Medical Research Council, and the EPSRC (EP/P030572/1).

AUTHOR CONTRIBUTIONS

F.Y.W. and Y.Z. conceived the project. L.Y.Q., Y.Z.H, Q.L., Y.X. and T.T.F. carried out the expression of HMGB1a and construction of HMGB1a-functionilizaed microprobes. L.Y.Q., B.B. X., Y.Y.Z., F.F.J., X.J.L., P.J.S., Y.Z. and F.Y.W. analysed sequencing data. L.Y.Q., W.C.Y. and X.K.L. performed bioinformatics analysis and Western Blot assays. Y.Z.H. and Q.L. performed fluorescence imaging, L.Y.Q., F.Y.W. P.J.S., Y.Z. and S.J.W. wrote and revised the manuscript with inputs from all authors.

CONFLICT OF INTEREST

We have documented selected cisplatin-damaged genes as potential drug targets and biomarkers for cancer therapy in a Chinese patent application.