Distinct and common features of numerical and structural chromosomal instability across different cancer types

Xiaoxiao Zhang[1,2] and Maik Kschischo[1*]

[1*]Department of Mathematics and Technology, University of Applied Sciences Koblenz, Joseph-Rovan-Allee 2, 53424 Remagen, Germany.
[2]Department of Informatics, Technical University of Munich, Munich, 81675, Germany.

*Corresponding author(s). E-mail(s): kschischo@rheinahrcampus.de;
Contributing authors: zhang@hs-koblenz.de;

Abstract

A large proportion of tumours is characterised by numerical or structural chromosomal instability (CIN), defined as an increased rate of gaining or losing whole chromosomes (W-CIN) or of accumulating structural aberrations (S-CIN). Both W-CIN and S-CIN are associated with tumourigenesis, cancer progression, treatment resistance and clinical outcome. Although S-CIN and W-CIN can co-occur, they are initiated by different molecular events. By analysing tumour genomic data from 32 cancer types, we show that the majority of tumours with high levels of W-CIN underwent whole genome doubling, whereas S-CIN levels are strongly associated to homologous recombination deficiency. Both CIN phenotypes are prognostic in several cancer types. Most drugs are less efficient in high CIN cell lines, but we report also compounds and drugs which could specifically target S-CIN or W-CIN. By analysing associations of gene expression with CIN at the pathway and single gene level, we complement gene signatures for CIN and report that the drug resistance gene CKS1B is strongly associated to both S-CIN and W-CIN. Finally, we identify a potential copy number dependent mechanism for the activation of the PI3K pathway in high CIN tumours.

1 Introduction

A large proportion of human tumours exhibits abnormal karyotypes with gains and losses of whole chromosomes or structural aberrations of parts of chromosomes [1–3]. In many cases, these karyotypic changes are the result of ongoing chromosomal instability (CIN), which is defined as an increased rate of chromosome changes. Accordingly, two major forms of CIN can be distinguished: Numerical or whole chromosome CIN (W-CIN) refers to the ongoing acquisition of gains and losses of whole chromosomes. Structural CIN (S-CIN) is characterised by an increased rate of smaller scale deletions, insertions, and amplifications and gross chromosomal rearrangements [1–4]. Although W-CIN can induce structural CIN and vice versa, both types of CIN arise through distinct molecular characteristics . Whilst W-CIN is caused by chromosome missegregation during mitosis, S-CIN is commonly attributed to errors in the repair of DNA double-strand breaks [5, 6]. Both types of CIN are intimately related to DNA replication stress [7, 8], which can induce CIN [9, 10], but also forms an immediate short term response to aneuploidy and CIN [11].

Aneuploidy and CIN have typically detrimental effects on cell fitness and proliferation [11–13]. Therefore, it was unclear why CIN is often associated with poor patient survival and more aggressive disease progression [1, 14, 15]. Stratification of breast cancer patient samples into groups of low, intermediate and high CIN revealed that patients with intermediate levels of CIN had the worst survival, whereas the low and high CIN groups had a better prognosis [16, 17]. These results hinted at mechanisms for tolerating CIN which are necessary to survive the stresses provoked by chromosome aberrations. The CIN tolerance mechanisms are currently not completely understood [18], but one important recurring event is a loss of TP53 function, which otherwise prevents the propagation of CIN cells [19].

The CIN 70 signature [1] of genes whose expression is correlated with CIN is enriched by genes involved in mitosis and cell cycle regulation. It has later been critisized [20] for being rather a marker for cell proliferation than for CIN, because it reflects evolved aneuploid cancer cell populations which have adapted their genome instead of a primary response to CIN [20]. These studies highlighted, that we have to distinguish between acute responses to aneuploidy and CIN [11], mechanisms for tolerating CIN [18] and the cellular programme [21, 22] and genetic alterations [23] acquired by evolved CIN cells. These cellular programmes might differ between cancer cell lines and tumours, partially as a result of treatment effects and partially as a result of the tumour environment. Recently, it has been discovered that chromosome segregation errors as well as replication stress activate the anti-viral immune cGAS-STING pathway, which responds to genomic double-stranded DNA in the cytosol [2, 24]. This interesting research links cell intrinsic processes with cell to cell communication and immune response in the tumour microenvironment

The phenotypic plasticity in combination with tumour heterogeneity enables CIN tumours to rapidly adapt to diverse stress conditions. It has been shown that CIN permits and accelerates the acquisition of resistance against

anti-cancer therapies by acquiring recurrent copy number changes [25, 26]. This acquired drug resistance could potentially exacerbate the intrinsic drug resistance [27] of many CIN cells, which highlights the need to better understand somatic point mutations and specific copy number changes of CIN tumours in the context of cancer treatment.

Computational studies of cancer genomic data [28] have provided valuable insights into CIN [1, 20–23] and aneuplody [28] and guided experimental and clinical testing. However, most of these studies did not differentiate between W-CIN and S-CIN. Here we analyse cancer genomic data to better understand commonalities and differences between both types of CIN. In particular, we analyse, across multiple cancer types, the genomic landscape of S-CIN and W-CIN, their relationship to prognosis and drug sensitivity, the relationship between CIN, somatic point mutations and specific copy number variations and report a new link between CIN and the PI3K oncogenic pathway.

2 Results

2.1 Landscape of W-CIN and S-CIN across human cancers

CIN is a dynamic feature of abnormal chromosomes, rendering its assessment in routine experimental settings difficult [29, 30]. However, there are established surrogate scores based on copy number data which have previously been evaluated as good proxy measures for CIN in bulk tumour data [7]. To quantify the degree of W-CIN of a given tumour sample we used the WGII score, which is the average percentage of changed genome relative to the sample ploidy [7]. The degree of S-CIN was assessed by the SCIN score, which is the number of structurally aberrant regions in the genome of a sample. In total, we calculated WGII and SCIN scores for 21633 samples including 10308 primary tumours, 391 metastatic tumours and 10934 normal tissues derived from 32 cancer types.

The distribution of WGII varies drastically across cancer types (**Fig. 1(A)**), but shows a characteristic bimodal pattern, see also the pan-cancer histogram on the right hand side. The colour coding of the whole genome doubling (WGD) status indicates, that tumour samples with high levels of WGII are often also characterised by a whole genome doubling event. This is not an artefact of the WGII score, which is measured relative to the sample ploidy. This suggests that whole genome doubling is an important mechanism for W-CIN in many cancer types. However, the exception is Kidney chromophobe (KICH), where whole genome doubling seems to be rare, but high levels of the WGII score can still be observed. In this cancer type, there is also no clear bimodal pattern, suggesting that other mechanisms than whole genome doubling drive CIN in KICH. Even in cancers where the bimodal pattern suggests a clear distinction between numerically unstable and numerically stable tumours, it is difficult to define a universal threshold for tumours distinguishing W-CIN from stable tumours across cancer types. For example,

in Ovarian serous cystadenocarcinoma (OV) one can distinguish two groups of low and high WGII scores with whole genome doubling, but the overall level of the WGII score is much higher than in other cancer types. Similarly, for Breast invasive carcinoma (BRCA) there are many many patients with high levels of WGII even in the group of samples which did not undergo whole genome doubling. This suggests, that other processes than whole genome doubling can drive a certain degree of W-CIN in these tumours.

In contrast to the WGII distribution, the pan-cancer distribution of SCIN scores peaks at low values and is right skewed (**Fig. 1(B)**). This indicates that many tumours are structurally stable, but some can exhibit extreme levels of SCIN. These extreme levels of SCIN are typically seen in tumours which did not undergo whole genome doubling.

The distribution of SCIN-scores indicates a high degree of inter-tumour heterogeneity within the same cancer type and across cancer types. Ovarian serous cystadenocarcinoma (OV), Uterine carcinosarcoma (UCS) and Sarcoma (SARC) are the cancers exhibiting the highest levels of SCIN scores (**Fig. 1(B)**) and these tumours also exhibit many samples with W-CIN, compare **Fig. 1(A)**. Both types of CIN occur also in many OV, Esophageal carcinoma (ESCA) and BRCA samples, whereas Thyroid carcinoma (THCA), Thymoma (THYM) and Acute myeloid leukemia (LAML) samples are typically both structurally and numerically stable. Cancer types previously appreciated as those dominated by the CIN phenotype [31], including Stomach adenocarcinoma (STAD), Colon adenocarcinoma (COAD), Uterine corpus endometrial carcinoma (UCEC), OV, UCS and Prostate adenocarcinoma (PRAD) have extremely heterogenous SCIN scores.

We also tested for associations of both types of CIN with other types of genetic instability by correlating the WGII and SCIN scores with different features: aneuploidy score (AS), homologous recombination deficiency (HRD), silent mutation rate (SMR), non-silent mutation rate (NSMR)and intra-tumour heterogeneity (ITH). The WGII score is uniformly positively associated with the aneuploidy score see **Fig. 1(C)**, which is expected, because aneuploidy can be both a cause and a consequence of CIN. Intriguingly, HRD is consistently positively associated with the SCIN score (**Fig. 1(D)**), suggesting that impaired repair of DNA double strand breaks (DSBs) might be a key driver of S-CIN. There is no consistent relationship for the other genetic instability scores across different cancer or tissue types.

To check for a potential link between CIN and proliferation, we used a proliferation index [32] and the expression of the MKI67 marker for proliferation. In many cancers including Kidney chromophobe (KICH), Kidney renal clear cell carcinoma (KIRC), SARC, Pancreatic adenocarcinoma (PAAD) and BRCA, increasing levels of WGII and SCIN scores go along with increasing levels of these proliferation markers (**Fig. 1(C)**). However, this is not the case for other cancers, indicating again the complex relationship between CIN and proliferation [33–35]. The balance between the proliferation promoting effect of CIN as a template for Darwinian selection and the cellular burden of

chromosomal aberrations accompanied by CIN might be highly cancer type dependent.

Both S-CIN scores and WGII tend to be higher in metastatic tumours than in primary tumours and in normal samples (**Fig. S1(A)**), consistent with previous findings linking CIN and metastasis [24]. The small sample size for metastatic tumours prevents a cancer type specific analysis of the relationship between CIN and metastatic disease.

These results highlight that W-CIN and S-CIN are two related but distinct phenotypes with different distributions across cancer types. Whole genome doubling is often accompanied by W-CIN, but this does not completely explain the elevated levels of WGII in some cancer types or individual tumours. The bimodal distribution of the WGII score in most cancer types suggests that a separation of samples into a high and a low W-CIN group is possible. However, in some cancers like OV, even the low W-CIN group can exhibit substantial levels of whole genome aberrations. In contrast, S-CIN is a continuous trait which is strongly associated with homologous recombination, but not with whole genome doubling.

2.2 Clinical significance of CIN

To analyse the relationship between W-CIN and prognosis, we divided the tumour samples in each individual cancer type into two distinct groups of high (WGII^{high}) and low (WGII^{high}) using the median as a threshold, compare **Fig. 1(A)**. For eight of the 32 cancer types we found that (WGII^{high}) patients had a significantly shorter overall survival (Log rank test, $p < 0.05$) than patients in the WGII^{low} group (**Fig. 2(A)**). This includes BRCA, Head and neck squamous cell carcinoma (HNSC), Brain lower grade glioma (LGG), Liver hepatocellular carcinoma (LIHC), OV, SARC, TGCT and UCEC. Disease free survival is lower in the WGII^{high} group for LIHC, OV, PRAD and UCEC patients, see (**Fig. S2(A)**) and progression free survival is negatively associated with high WGII in LGG, LIHC, OV, PRAD, SKCM, THYM, UCEC and UVM (**Fig. S4(A)**).

For SCIN, there is no clear cut separation into two different groups, see **Fig. 1(B)**. When separating the tumour samples into a SCIN^{low} and SCIN^{high} group using the median of the SCIN score in each tumour type we found that the overall survival of patients in 10 out of 32 cancers (Log rank test, $p < 0.05$) is negatively associated with SCIN (**Fig. 2(B)**). High SCIN scores are linked to impaired disease free survival in Adrenocortical carcinoma (ACC), KIRC, Kidney chromophobe (KIRP), Lung squamous cell carcinoma (LUSC), PRAD, THCA and UCEC (**Fig. S2(B)**). For OV, the survival analysis for S-CIN provides very variable results. Patients with high SCIN scores tend to have slightly better overall survival (**Fig. 2(B)**), but the effect is very small and at the edge of statistical significance. Disease free and progression free survival of OV patients is, however, worse for the SCIN^{high} group (**Fig. S2(B)** and **Fig. S4(B)**).

To further explore the clinical relevance of both types of CIN, we compared the distribution of CIN scores between patients responding to radiotherapy and to chemotherapy. Indeed, we find that tumour samples from radiotherapy responders tend to have lower WGII scores (Wilcoxon $p = 10e - 5$) than samples from radiotherapy non-responders, whereas SCIN is not significantly associated with radiotherapy response (**Fig. 2(C)**). On a pan-cancer level, we did not find a significant difference between WGII scores in the group of chemotherapy responders versus non-responders (**Fig. 2(D)**). The median SCIN score of chemotherapy responders is slightly higher. This might be a confounding effect of the treatment itself. An alternative explanation is that high SCIN samples tend to have defective homologous recombination (HR) DNA repair (compare **Fig. 1(B)**), which renders them more sensitive to chemotherapy. On the basis of the available data, we can not exclude either of these explanations.

Next, we asked whether there are drugs suitable for specifically targeting CIN. To this end, we combined data from the Cancer Therapeutics Response Portal (CTRP) and the Cancer Cell Line Encyclopedia (CCLE). We normalised the area under the curve (AUC) values of 545 compounds and small molecules in all cell lines to values between zero and one and defined drug sensitivity as one minus the normalised AUC. Values of zero indicate highest resistance, whereas values of one indicate the highest possible sensitivity. Then, we computed Spearman rank correlation coefficients between the drug sensitivity of each compound or small molecule with the WGII or SCIN scores. To analyse the typical drug sensitivity as a function of CIN, we plotted the median drug sensitivity of each compound or small molecule across cell lines against their correlation coefficients with WGII (**Fig. 2(E)**) or SCIN score (**Fig. 2(F)**).

For the majority of compounds and small molecules we found negative correlations between their AUC and both types of CIN (**Fig. 2(E,F)**), highlighting that for many compounds CIN confers an intrinisic drug resistance [27]. Only a few compounds are more sensitive in high CIN cell lines than in stable cell lines, but their overall efficiency is also typically lower. The sensitivity of the compounds BRD-K99006945, JW-55 (targeting TNK), PLX-40932 (targeting BRAF) and BCL-LZH-4 have the strongest positive correlation with WGII across cancer types (Spearman correlation coefficient larger than 0.3 and FDR threshold of 5%, **Fig. 2(E)**). None of these compounds have been approved for clinical use by the FDA.

Drugs showing increasing sensitivity with the SCIN-score included Afatinib and Lapatinib (Spearman correlation coefficient larger than 0.3 and FDR threshold of 5%, **Fig. 2(F)**). Lapatinib targets HER2/neu and is used in combination treatment of HER2-positive breast cancer. Afatinib is used to treat non small lung cancers with EGFR mutations. Austocystin D is a natural cytotoxic agent found to be more sensitive in cell lines with higher SCIN scores. IPR-456, PD 153035 and BRD-K19103580 are further compounds with drug sensitivity positively associated with SCIN scores. Further details about

the correlations between both CIN scores and sensitivity can be found in the **Supplemental Table 1**.

Overall, the analysis shows that the prognostic value of CIN depends on cancer type and that S-CIN and W-CIN provide distinct prognostic information. The only cancer were we found a slightly better prognosis for patients with high levels of S-CIN is OV. It has been reported, that women with BRCA-associated ovarian carcinoma have a better prognosis [36]. This is consistent with our observation that SCIN scores are strongly associated with HRD (**Fig. 1(B)**). Our analysis also reveals specific compounds and drugs whose *in vitro* sensitivity is higher in S-CIN or W-CIN tumours.

3 PARADIGM pathway activity and CIN

To identify gene related pathways with altered activity in S-CIN or W-CIN tumours, we used the PARADIGM framework [37], which represents a gene as a factor graph encoding the interaction of the expression and activity of a gene and its products. We used gene expression data to assess the activity of each PARADIGM pathway corresponding to a protein coding gene in each individual tumour sample. These pathway activity scores were then correlated with W-CIN or S-CIN scores (see **Fig. 3**). We filtered pathways corresponding to protein coding genes which showed a significant ($FDR < 5\%$) correlation larger than 0.4 in at least eight of the 32 cancer types analysed.

Pathways corresponding to the mitotic genes *TPX2, RAE1, UBE2C, AURKA, NEK2* (see **Fig. 3(A)**) show increased activity in tumours with high WGII-scores, consistent with the known role of chromosome segregation errors in W-CIN [1, 21]. Accordingly, the PARADIGM pathways corresponding to the genes *CSE1L, DSN1* and *CENPA* have higher activity in tumours with high levels of WGII across many cancers. *CSE1L* is a gene associated with the spindle checkpoint and the separation of centrosomes during cell division as well as a direct effector of p53. *DSN1* and *CENPA* are required for kinetochore assembly.

The STX1 (SYNTAXIN 1A) pathway is overexpressed in W-CIN tumours. This finding is surprising, because the STX1 gene is normally expressed in brain cells and STX1 is key molecule in synaptic exocytosis and ion channel regulation. The reason why STX1 expression is increased in W-CIN tumours needs further investigation.

It is interesting to note the positive association of the PARADIGM pathway for the GINS1 complex with WGII-score [10]. The GINS1 protein is essential for the formation of the replication origin firing complex. The CMG (Cdc45-MCM-GINS) complex is a component of the replicative helicase, which functions to unwind DNA ahead of the replication fork [38]. As detailed in [10], overexpression of GINS *in vitro* increases replication origin firing and triggers whole chromosome missegregation and W-CIN . Indeed, when we complement our PARADIGM pathway analysis with simple gene wise correlation of the

WGII and gene expression, we find many genes involved in DNA replication and replication origin firing (see **Fig. 3(B)**).

The analysis of the SCIN score associated PARADIGM pathways (**Fig. 3(C)**) revealed again proteins involved in kinetochore function, mitotic progression and spindle assembly and chromosome segregation (AURKA, UBE2C NEK2, TBCE) or cell cycle progression (CDK4, E2F1). The activity of the CKS1B pathway is positively associated to both the SCIN score and WGII . CKS1B is an essential protein for the positive regulation of the cyclin dependent kinase activity and has recently been linked to cancer drug resistance and discussed as a new therapeutic target [39]. Our results suggest, that the CIN status should be included, when CKS1B is analysed as a new target gene.

To check for the robustness of these findings, we also performed a gene wise correlation of the SCIN score and gene expression (**Fig. 3(C)**). There, we also highlighted genes involved in DNA replication. Gene set enrichment analysis indicates that the set of genes annotated as replication genes is enriched at the top of a list ranked by their correlation with S-CIN (**Fig. S5**).

Taken together, our analysis of PARADIGM pathway activity and gene expression in the context of CIN not only recovered known CIN genes involved in mitotic processes and spindle assembly, but highlighted, amongst others, the replication factor GINS1 to be associated with W-CIN [10] and the CDK regulator and drug resistance protein CKS1B as strongly associated with both types of CIN. In addition, we observed that the expression of genes involved in DNA replication is positively associated with CIN.

4 Somatic point mutation frequencies in high CIN tumours

To investigate the relationship between somatic point mutations and CIN, we checked for genes that are more frequently or less frequently mutated in high CIN tumours. We included 19171 gene mutations, for which there were at least 20 samples in the wild type or mutant group across different cancer types. We then used a bivariate linear model to regress WGII or S-CIN score against somatic point mutation status (present or absent) and cancer type. The regression coefficient for mutation status was used as measure for association with CIN, adjusted for tumour type.

As expected, mutation of TP53 shows the strongest association with CIN. The WGII score of tumours harbouring a TP53 mutation is approximately 23% (ANOVA p-value $< 10e - 100$, **Fig. 4(A)**) higher than for the wild type. The mean difference of the SCIN-score in a tumour sample with a TP53 mutation compared to wild type samples is approximately 11 structural aberrations **Fig. 4(B)**. This corroborates the well known role of TP53 as a gatekeeper of genome stability.

Contrary to the enrichment of TP53 mutations in both types of CIN, we find that the presence of mutations in 10749 different genes is negatively

associated with both WGII and SCIN score (**Fig. 4(A)**). A similar negative correlation between the frequencies of recurrent copy number alterations and somatic mutations has previously been reported [40]. Later it was realised that this negative relationship can be reversed, when the confounding effect of samples with high levels of microsatellite instability (MSI) or POLE mutation [28, 41] is removed. When we exclude these hypermutated samples we observe a more even distribution between genes more or less frequently mutated in high CIN as compared to low CIN tumours (**Fig. 4(C,D)**). This is also consistent with **Fig. 1(C,D)**, where we found that neither the silent nor the non-silent mutation rate (SMR and NSMR) are uniformly positively associated with WGII and SCIN scores.

Intriguingly, even after excluding hypermutated samples, we find somatic point mutations of important cancer genes including PI3KCA, PTEN and ARID1A to be underrepresented in both high S-CIN and W-CIN tumours. HRAS and JAK1 mutations are less frequent in tumours with high WGII score and KRAS mutations are underrepresented in samples with high S-CIN score. The underrepresentation of somatic mutations in these key cancer genes in high CIN tumours can not be explained by differences in the overall mutation rates of these samples.

5 Copy number gains and losses associated with CIN

Given that somatic mutations of many genes are underrepresented in high CIN tumours, we asked for copy number alterations which are specifically linked to CIN (**Fig. 5(A)**). One of the strongest associations between a copy number gain and both WGII and SCIN score was found for the MYC proto-oncogene. The candidate oncogene PVT1 is also specifically gained in tumours with high levels of S-CIN . PVT1 is involved in the regulation of MYC [42] and carries a p53-binding site. In addition, we found several genes from the WFDC-EPPIN family.

Genes which were specifically lost in tumour samples with high WGII score include STAMBPL1, which is linked to the TGF-beta pathway, MIER2 and FAS (**Fig. 5**(B)). Annotations for the FAS receptor indicate that it contains a death domain and plays a central role in the regulation of apoptosis.

The top genes whose copy number loss is most strongly associated with SCIN score are PDE40, RB1 and PTEN. The tumour suppressor RB1 is a key regulator of the G1/S transition of the cell cycle and required for the stabilisation of heterochromatin. PTEN loss is also frequent in tumours with high levels of WGII.

6 PI3KCA copy number gains in high CIN tumours suggest a gene dosage dependent mechanism for PI3K pathway activation

In section 4 we observed that somatic point mutations of PTEN and PIK3CA were reproducibly scarce in high CIN tumours. In addition, copy number loss of PTEN is very frequent in tumour samples with high levels of both types of CIN. This led us to ask whether there is a link between CIN and specific gene copy number alterations in these two genes to activate the PI3K oncogenic pathway. The *PIK3CA* gene encodes the catalytic subunit of phosphatidylinositol 3-kinase (PI3K) and the PI3K oncogenic pathway is frequently deregulated in many cancers. PTEN is a tumour suppressor gene and negatively regulates the growth-promoting PI3K/AKT/mTOR signal transduction pathway.

The oncoprint in **Fig. 6** displays tumour samples from all 32 TCGA cancer types in our investigation, which harbour at least one of the following genetic alterations: somatic mutation of PIK3CA or PTEN, copy number amplification of PIK3CA, deletion of PTEN. It is apparent, that there is only a small number of cancers with an amplification of PKI3CA or an deletion of PTEN, which simultaneously harbour somatic mutations in any of these genes. The copy number of both genes is also strongly associated with their gene expression. In particular, amplification and simultaneous overexpression of PIK3CA is associated with higher levels of both WGII and CIN score.

To check, whether this effect could be driven by cancer therapeutic treatment, we used cell line data from CCLE and found a very similar pattern. Copy number gains of PIK3CA are linked to high levels of gene expression, do rarely co-occur with somatic mutations, but are associated with high levels of S-CIN and W-CIN.

Taken together, we suggest a gene dosage effect on PI3K pathway activity, which is facilited in high CIN tumours. This effect seems to be cell intrinsic, because it can also be observed in cell line data.

7 Discussion

Here, we used statistical data analysis to better understand the roles of S-CIN and W-CIN across 32 different cancer types. We used the WGII score as a proxy measure for W-CIN and the SCIN score to quantify the degree of S-CIN and associated these karyotypic complexity scores with various other molecular and clinical features.

Our analysis reveals that the majority of tumours with high levels of W-CIN underwent whole genome doubling. Whole genome doubling is an early event in tumourigenesis and has been discussed as a way to rapidly accumulate numerical and structural chromosomal abnormalities and to buffer against negative effects of mutations and aneuploidy [13, 43, 44]. Our analysis results suggest, that whole genome doubling is typically accompanied by W-CIN, but not S-CIN. Instead, we find that high levels of S-CIN are linked to homologous

recombination repair deficiency, highlighting the different processes involved in these two different CIN phenotypes [6].

Although whole genome doubling is observed in many tumour samples with high levels of W-CIN, it is not sufficient to explain the elevated levels of W-CIN in many tumour samples which did not undergo whole genome doubling, as most prominently observed in BRCA, KICH, OV, SARC or UCEC. We speculate that replication stress is an alternative mechanism for these elevated levels of CIN. This is based on ample evidence that replication stress can induce CIN [7, 45] and our observation that replication factors are overexpressed in tumours with high levels of W-CIN. Additional support for this hypothesis comes from our observed prevalence of copy number gains for MYC in W-CIN and S-CIN tumours and its regulator PVT1 in samples with high levels of S-CIN. MYC has been implicated as one possible driver for replication stress [46].

We find that WGII and S-CIN score are associated with poor prognosis in different cancer types. Only for ovarian cancer we find that S-CIN might provide a slightly better prognosis. In addition, we observe slightly higher levels of the WGII score in patients resistant to radiotherapy. However, the relationship between CIN and prognosis is multifaceted and depends on details of the cellular physiology [3]. For instance, extreme levels of CIN in breast cancer subtypes [16, 17] were associated with better prognosis. This indicates, that a subtype specfic analysis of W-CIN and S-CIN and prognosis might potentially be an interesting future project. This might also apply for the response against radiotherapy, where improved sensitivity against radiotherapy in transplanted human glioblastoma tumours has been reported [47].

From the association of WGII and SCIN score with *in vitro* drug sensitivity it is apparent, that both types of CIN are linked to intrinsic drug resistance, corroborating earlier results in colon cancer [3, 27]. However, as a new contribution we filtered small molecules and compounds for which drug sensitivity is positively associated with S-CIN or W-CIN. For S-CIN this includes the approved drugs Afatinib and Lapatinib and the natural cytotoxic agent Austocystin D. It remains to be tested, whether these drugs or compounds are indeed efficient against high CIN tumours *in vivo*.

In addition to well known CIN genes including TPX2, UBE2C and AURKA we identified a number of new candidate CIN genes or corresponding PARADIGM pathways [37]. One interesting new finding is the chemotherapeutic drug resistance inducing gene CKS1B [39], which we found to be strongly associated with both S-CIN and W-CIN. CKS1B is cell cycle progression gene, which is discussed as a new drug target. Here we show, that CKS1B is overexpressed in S-CIN and W-CIN tumours, which might be important for the stratification of patients. We also note that the activity of the origin firing factor GINS1 is linked to CIN, which was mechanistically verified in a recent collaboration [10]. In this context, we also found many genes involved in DNA replication to be overexpressed in tumours with high levels of W-CIN and S-CIN.

Both S-CIN and W-CIN are strongly anticorrelated with somatic point mutations of TP53. We find many copy number gains of important onogenes and loss of tumour suppressor genes [48], which are strongly associated with S-CIN and W-CIN. Most strikingly, copy number gains of the oncogene PIK3CA and deletion of the tumours suppressor PTEN do rarely occur in combination with somatic mutations in these genes. In addition, copy number gain of PIK3CA is linked to increased gene expression and strongly associated with S-CIN and W-CIN. Intriguingly, it has recently been reported that mutations in PIK3CA increased in vitro cellular tolerance to spontaneous genome doubling [49]. Our results, however, suggest a gene dosage effect for the activation of the PI3K pathway in the context of high CIN. This copy number dependent activation of PIK3CA was observed in both tumour derived patient samples and in cancer cell lines, indicating that it is an intrinsic property of CIN cells. We suggest that copy number gains of PIK3CA should be further investigated for both their mechanistic role in CIN and for their clinical implications regarding treatment strategies and patient stratification.

Methods

7.1 Clinical and molecular data

We analysed chromosome instability of 33 cancers: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical and endocervical cancers; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large b-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIPAN, kidney cancer; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumours; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

TCGA pan-cancer molecular and clinical data were downloaded from the Pan-Cancer Atlas (PanCanAtlas) https://gdc.cancer.gov/node/905/. Copy number data are based on Affymetrix SNP 6.0 for 10,308 primary tumours. For 9,973 of these samples, gene expression profiles by RNA sequencing is available. To analyse metastasis-CIN and tumourigenesis-CIN correlation, we also calculated CIN scores for 391 metastatic tumour tissues, 8719 blood derived normal tissues and 2207 solid normal tissues.

7.2 Quantifying structural and numerical CIN

We implemented three different scores to quantify CIN in bulk tumour samples [7]: the numerical chromosome instability score (NCIN score), the structural chromosome instability score (SCIN score) and the weighted genome instability Index (WGII). The NCIN and SCIN scores are defined as normalised counts of focal and gross chromosomal changes over 22 autosomal chromosomes, respectively. The normalisation accounts for chromosome length. The WGII is defined as the average percentage of changed genome relative to the sample ploidy over 22 autosomal chromosomes. WGII is highly correlated to the NCIN score and only long chromosome segments contribute to the total WGII score. We found that the WGII score has better statistical properties and is less sensitive to errors outliers. Therefore, we used it as a proxy measure for W-CIN.

7.3 Survival analysis

We performed survival analysis using the r-package survival. Patients were stratified according to their median CIN score of all patients from the same cohort. An univariate cox-proportional hazards model was fitted to evaluate the association between patient survival and CIN and the Likelihood ratio test was applied to calculate the p-value for the survival difference between high CIN and low CIN groups. Survival curves were visualised using ggsurvplot implemented in the R-package survminer.

7.4 Drug response analysis

We labeled patients with complete/partial response to therapy as responders and the other patients as non-responders. A Wilcoxon rank sum test was used to asses the differences of the WGII and SCIN score in the responder and non-responder group.

7.5 Association analysis between CIN and PARADIGM pathway activities

We collected the sample-wise PARADIGM pathway activity matrix from the pan-cancer atlas [50] https://gdc.cancer.gov/about-data/publications/pancanatlas with the file name merge_merged_reals.tar.gz. Then we calculated the cancer type wise Spearmann correlation coefficient between CIN score (WGII or SCIN score) and PARADIGM pathway activity and selected the top 15 pathways corresponding to significant protein coding genes. We filtered genes/proteins with PARADIGM pathways strongly positively correlated with WGII or and SCIN score (coefficient greater than 0.3) in more than seven cancer types.

7.6 Association analysis between CIN and genome instability

Aneuploidy scores (AS) of samples are taken from [28] https://www.sciencedirect.com/science/article/pii/S1535610818301119#app2, tumour characteristics including homologous recombination deficiency (HRD), silent mutation rate (SMR), non-silent mutation rate (NSMR), proliferation and intra-tumour heterogeneity (ITH) were collected from [32] https://www.sciencedirect.com/science/article/pii/S1074761318301213?via%3Dihub#app2. Microsatelite instability (MSI) scores are collected from [51] https://www.sciencedirect.com/science/article/pii/S009286741830237X#app2. Sample-wise Spearman correlation coefficients between CIN score and the above mentioned characteristics are calculated.

Author contributions statement

Xiaoxiao Zhang performed data analysis and wrote the manuscript. Maik Kschischo designed and supervised the research and wrote the manuscript.

Data availability

All data used in this study are publicly available. The sources are detailed in the section on methods.

Supplementary information. See the file Supplemental Information for Supplemental figures and tables. Spreadsheet tables of contain further analysis results.

Acknowledgments. This work was supported by the FOR2800 funded by the Deutsche Forschungsgemeinschaft (sub-project 3).

References

[1] Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043–1048 (2006). URL http://dx.doi.org/10.1038/ng1861 .

[2] Bakhoum, S. F. & Cantley, L. C. The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. *Cell* **174** (6), 1347–1360 (2018). URL https://linkinghub.elsevier.com/retrieve/pii/S0092867418310432. https://doi.org/10.1016/j.cell.2018.08.027 .

[3] Sansregret, L., Vanhaesebroeck, B. & Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol* **15** (3), 139–150 (2018). URL http://www.nature.com/articles/nrclinonc.2017.198. https://doi.org/10.1038/nrclinonc.2017.198 .

[4] Roschke, A. V. & Rozenblum, E. Multi-Layered Cancer Chromosomal Instability Phenotype. *Frontiers in Oncology* **3** (2013). https://doi.org/10.3389/fonc.2013.00302 .

[5] Thompson, S. L., Bakhoum, S. F. & Compton, D. A. Mechanisms of Chromosomal Instability. *Current Biology* **20** (6), R285–R295 (2010). URL https://linkinghub.elsevier.com/retrieve/pii/S096098221000076X. https://doi.org/10.1016/j.cub.2010.01.034 .

[6] Siri, S. O., Martino, J. & Gottifredi, V. Structural Chromosome Instability: Types, Origins, Consequences, and Therapeutic Opportunities. *Cancers* **13** (12), 3056 (2021). URL https://www.mdpi.com/2072-6694/13/12/3056. https://doi.org/10.3390/cancers13123056 .

[7] Burrell, R. A. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492 EP – (02 2013/02/27/online) .

[8] Wilhelm, T., Said, M. & Naim, V. DNA Replication Stress and Chromosomal Instability: Dangerous Liaisons. *Genes* **11** (6), 642 (2020). URL https://www.mdpi.com/2073-4425/11/6/642. https://doi.org/10.3390/genes11060642 .

[9] Böhly, N., Kistner, M. & Bastians, H. Mild replication stress causes aneuploidy by deregulating microtubule dynamics in mitosis. *Cell Cycle* **18** (20), 2770–2783 (2019). URL https://www.tandfonline.com/doi/full/10.1080/15384101.2019.1658477. https://doi.org/10.1080/15384101.2019.1658477 .

[10] Schmidt, A.-K. *et al.* Dormant replication origin firing links replication stress to whole chromosomal instability in human cancer. preprint, Cell Biology (2021). URL http://biorxiv.org/lookup/doi/10.1101/2021.10.11.463929.

[11] Passerini, V. *et al.* The presence of extra chromosomes leads to genomic instability. *Nature Communications* **7** (1) (2016). https://doi.org/10.1038/ncomms10754 .

[12] Thompson, S. L., Bakhoum, S. F. & Compton, D. A. Mechanisms of chromosomal instability. *Current Biology* **20**, R285–R295 (2010). URL http://www.sciencedirect.com/science/article/pii/S096098221000076X. https://doi.org/http://dx.doi.org/10.1016/j.cub.2010.01.034 .

[13] Dewhurst, S. M. *et al.* Tolerance of Whole-Genome Doubling Propagates Chromosomal Instability and Accelerates Cancer Genome Evolution. *Cancer Discovery* **4** (2), 175–185 (2014). URL http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-13-0285. https://

doi.org/10.1158/2159-8290.CD-13-0285 .

[14] Bakhoum, S. F., Danilova, O. V., Kaur, P., Levy, N. B. & Compton, D. A. Chromosomal Instability Substantiates Poor Prognosis in Patients with Diffuse Large B-cell Lymphoma. *Clinical Cancer Research* **17** (24), 7704–7711 (2011). URL http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-11-2049. https://doi.org/10.1158/1078-0432.CCR-11-2049 .

[15] Tijhuis, A. E., Johnson, S. C. & McClelland, S. E. The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. *Molecular Cytogenetics* **12** (1), 17 (2019). URL https://doi.org/10.1186/s13039-019-0429-1. https://doi.org/10.1186/s13039-019-0429-1 .

[16] Roylance, R. *et al.* Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **20** (10), 2183–2194 (2011) .

[17] Birkbak, N. J. *et al.* Paradoxical Relationship between Chromosomal Instability and Survival Outcome in Cancer. *Cancer Res* **71** (10), 3447–3452 (2011). URL http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-10-3667. https://doi.org/10.1158/0008-5472.CAN-10-3667 .

[18] Gronroos, E. & López-García, C. Tolerance of Chromosomal Instability in Cancer: Mechanisms and Therapeutic Opportunities. *Cancer Res* **78** (23), 6529–6535 (2018). URL http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-18-1958. https://doi.org/10.1158/0008-5472.CAN-18-1958 .

[19] Thompson, S. L. & Compton, D. A. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *Journal of Cell Biology* **188** (3), 369–381 (2010). URL https://rupress.org/jcb/article/188/3/369/35591/Proliferation-of-aneuploid-human-cells-is-limited. https://doi.org/10.1083/jcb.200905057 .

[20] Sheltzer, J. M. A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis. *Cancer Research* **73** (21), 6401–6412 (2013). URL http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-13-0749. https://doi.org/10.1158/0008-5472.CAN-13-0749 .

[21] Endesfelder, D. *et al.* Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in er+ breast cancer. *Cancer Research* **74** (17), 4853–4863 (2014). URL https:

//cancerres.aacrjournals.org/content/74/17/4853. https://doi.org/10.1158/0008-5472.CAN-13-2664, https://arxiv.org/abs/https://cancerres.aacrjournals.org/content/74/17/4853.full.pdf .

[22] Buccitelli, C. *et al.* Pan-cancer analysis distinguishes transcriptional changes of aneuploidy from proliferation. *Genome Res.* **27** (4), 501–511 (2017). URL http://genome.cshlp.org/lookup/doi/10.1101/gr.212225.116. https://doi.org/10.1101/gr.212225.116 .

[23] Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355** (6322), eaaf8399 (01 2017/01/20). https://doi.org/10.1126/science.aaf8399 .

[24] Bakhoum, S. F. *et al.* Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467 (2018). URL https://doi.org/10.1038/nature25432 .

[25] Salgueiro, L. *et al.* Acquisition of chromosome instability is a mechanism to evade oncogene addiction. *EMBO Mol Med* **12** (3) (2020). URL https://onlinelibrary.wiley.com/doi/10.15252/emmm.201910941. https://doi.org/10.15252/emmm.201910941 .

[26] Lukow, D. A. *et al.* Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Developmental Cell* **56** (17), 2427–2439.e4 (2021). URL https://linkinghub.elsevier.com/retrieve/pii/S153458072100592X. https://doi.org/10.1016/j.devcel.2021.07.009 .

[27] Lee, A. J. *et al.* Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research* **71**, 1858–1870 (2011). URL http://cancerres.aacrjournals.org/content/71/5/1858. https://doi.org/10.1158/0008-5472.CAN-10-3604, https://arxiv.org/abs/http://cancerres.aacrjournals.org/content/71/5/1858.full.pdf .

[28] Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33** (4), 676–689.e3 (2018). https://doi.org/https://doi.org/10.1016/j.ccell.2018.03.007 .

[29] Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396** (6712), 643–649 (1998). URL http://www.nature.com/articles/25292. https://doi.org/10.1038/25292 .

[30] McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: Therapeutic and diagnostic challenges. *EMBO reports* **13** (6), 528–538 (2012). https://doi.org/10.1038/embor.2012.61 .

[31] Delaney, J. R. *et al.* Haploinsufficiency networks identify targetable patterns of allelic deficiency in low mutation ovarian cancer. *Nature Communications* **8**, 14423 (2017) .

[32] Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48** (4), 812–830.e14 (2018). URL http://www.sciencedirect.com/science/article/pii/S1074761318301213. https://doi.org/https://doi.org/10.1016/j.immuni.2018.03.023 .

[33] Sheltzer, J. M. & Amon, A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends in Genetics* **27** (11), 446–453 (2011). URL https://linkinghub.elsevier.com/retrieve/pii/S0168952511001181. https://doi.org/10.1016/j.tig.2011.07.003 .

[34] Salmina, K. *et al.* The Cancer Aneuploidy Paradox: In the Light of Evolution. *Genes* **10** (2), 83 (2019). https://doi.org/10.3390/genes10020083 .

[35] Chunduri, N. K. & Storchová, Z. The diverse consequences of aneuploidy. *Nature Cell Biology* **21** (1), 54–62 (2019). URL http://www.nature.com/articles/s41556-018-0243-8. https://doi.org/10.1038/s41556-018-0243-8 .

[36] Zhong, Q., Peng, H.-L., Zhao, X., Zhang, L. & Hwang, W.-T. Effects of *BRCA1* - and *BRCA2* -Related Mutations on Ovarian and Breast Cancer Survival: A Meta-analysis. *Clin Cancer Res* **21** (1), 211–220 (2015). URL http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-14-1816. https://doi.org/10.1158/1078-0432.CCR-14-1816 .

[37] Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26** (12), i237–i245 (2010). URL https://doi.org/10.1093/bioinformatics/btq182. https://doi.org/10.1093/bioinformatics/btq182, https://arxiv.org/abs/https://academic.oup.com/bioinformatics/article-pdf/26/12/i237/16893608/btq182.pdf .

[38] Kamada, K. in *The GINS Complex: Structure and Function* (ed.MacNeill, S.) *The Eukaryotic Replisome: a Guide to Protein Structure and Function*, Vol. 62 135–156 (Springer Netherlands, Dordrecht, 2012). URL http://link.springer.com/10.1007/978-94-007-4572-8_8. Series Title: Subcellular Biochemistry.

[39] Shi, W. *et al.* CKS1B as Drug Resistance-Inducing Gene—A Potential Target to Improve Cancer Therapy. *Front. Oncol.* **10**, 582451 (2020). URL https://www.frontiersin.org/article/10.3389/fonc.2020.582451/full. https://doi.org/10.3389/fonc.2020.582451 .

[40] Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics* **45** (10), 1127–1133 (2013). URL http://www.nature.com/articles/ng.2762. https://doi.org/10.1038/ng.2762 .

[41] Buccitelli, C. *et al.* Pan-cancer analysis distinguishes transcriptional changes of aneuploidy from proliferation. *Genome Research* **27** (4), 501–511 (2017). URL http://genome.cshlp.org/content/27/4/501.abstract .

[42] Carramusa, L. *et al.* The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* **213** (2), 511–518 (2007). URL https://onlinelibrary.wiley.com/doi/10.1002/jcp.21133. https://doi.org/10.1002/jcp.21133 .

[43] López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nature Genetics* **52** (3), 283–293 (2020). URL https://doi.org/10.1038/s41588-020-0584-7. https://doi.org/10.1038/s41588-020-0584-7 .

[44] Quinton, R. J. *et al.* Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature* **590** (7846), 492–497 (2021). URL https://www.nature.com/articles/s41586-020-03133-3. https://doi.org/10.1038/s41586-020-03133-3 .

[45] Bakhoum, S. F., Kabeche, L., Murnane, J. P., Zaki, B. I. & Compton, D. A. DNA-Damage Response during Mitosis Induces Whole-Chromosome Missegregation. *Cancer Discovery* **4** (11), 1281–1289 (2014). https://doi.org/10.1158/2159-8290.CD-14-0403 .

[46] Kotsantis, P., Petermann, E. & Boulton, S. J. Mechanisms of Oncogene-Induced Replication Stress: Jigsaw Falling into Place. *Cancer Discov* **8** (5), 537–555 (2018). URL http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-17-1461. https://doi.org/10.1158/2159-8290.CD-17-1461 .

[47] Bakhoum, S. F. *et al.* Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat Commun* **6** (1), 5990 (2015). URL http://www.nature.com/articles/ncomms6990. https://doi.org/10.1038/ncomms6990 .

[48] Davoli, T. *et al.* Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. *Cell* **155** (4), 948–962 (2013). URL https://linkinghub.elsevier.com/retrieve/pii/S0092867413012877. https://doi.org/10.1016/j.cell.2013.10.011 .

[49] Berenjeno, I. M. *et al.* Oncogenic PIK3CA induces centrosome amplification and tolerance to genome doubling. *Nat Commun* **8** (1), 1773 (2017).

URL http://www.nature.com/articles/s41467-017-02002-4. https://doi.org/10.1038/s41467-017-02002-4 .

[50] Chang, K. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* **45** (10), 1113–1120 (2013). URL https://doi.org/10.1038/ng.2764. https://doi.org/10.1038/ng.2764 .

[51] Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173** (2), 371–385.e18 (2018). https://doi.org/10.1016/j.cell.2018.02.060 .

Figures

Fig. 1 Distribution of CIN scores and their association with genetic instability. **(A)** Left: WGII scores for tumour samples from different cancer types, sorted according to median WGII. The colour coding indicates the WGD status and the numbers below the proportion of samples which underwent WGD. Right: Pan-cancer histogram of WGII scores. **(B)** SCIN scores for tumour samples from different cancer types, ordered by their median SCIN-score. Right: Pan-cancer histogram of SCIN scores with colours indicating WGD status. **(C, D)** Correlation between WGII score (C) or SCIN score (D) with different indices for genetic instability, intra-tumour heterogeneity and proliferation: MKI67 expression, proliferation rates (PROLIF), intra-tumour heterogeneity (ITH), non-silent mutation rate (NSMR), silent mutation rate (SMR), homologous recombination deficiency (HRD) and aneuploidy score (AS). Data for PROLIF, ITH, NSMR, SMR, HRD, and AS were collected from [32].

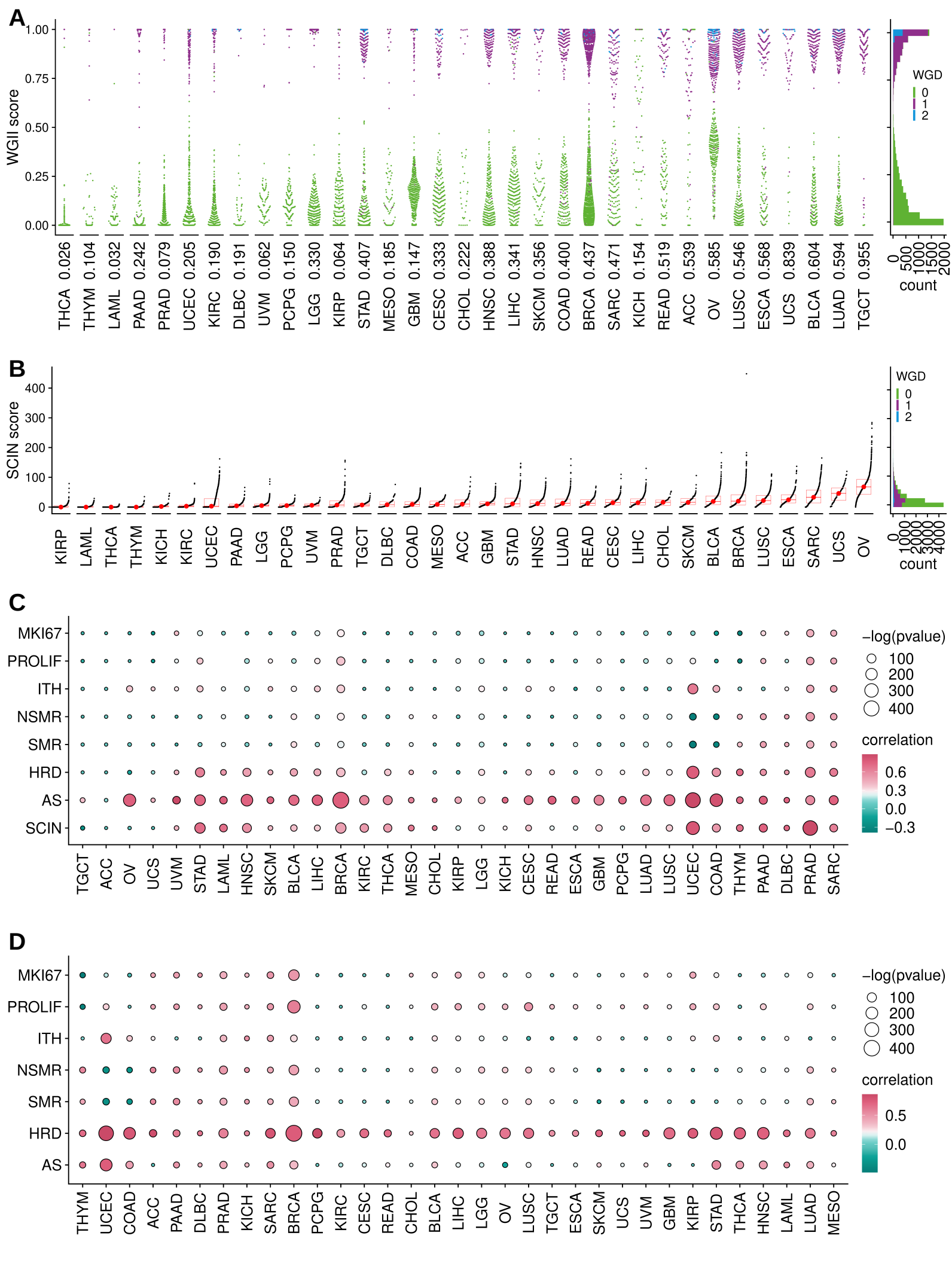
Fig. 2 Clinical significance of CIN. (A) For eight cancer types there are significant differences (log-rank p-value$< 5\%$) in overall survival between patient samples with low WGII score (blue) and high WGII score (red). Dashed lines indicate the five year overall survival probability of the two groups. **(B)** The SCIN score is associated with overall survival in 11 cancer types (low SCIN score group in blue and high SCIN group in red). **(C)** Comparison of the WGII score and SCIN score between radiotherapy responders and non-responders. **(D)** Comparison of the WGII score and SCIN score between chemotherapy responders and non-responders. **(E)** The median drug sensitivity of a compound plotted against the correlation coefficient between drug sensitivity and WGII. Drugs with significant positive and negative correlations between their sensitivity and WGII are highlighted in red and blue, respectively. **(F)** The median drug sensitivity of a compound plotted against the correlation coefficient between drug sensitivity and SCIN score. Drugs that have a sensitivity which is significantly correlated with SCIN score are highlighted in blue (negative) and red (positive).

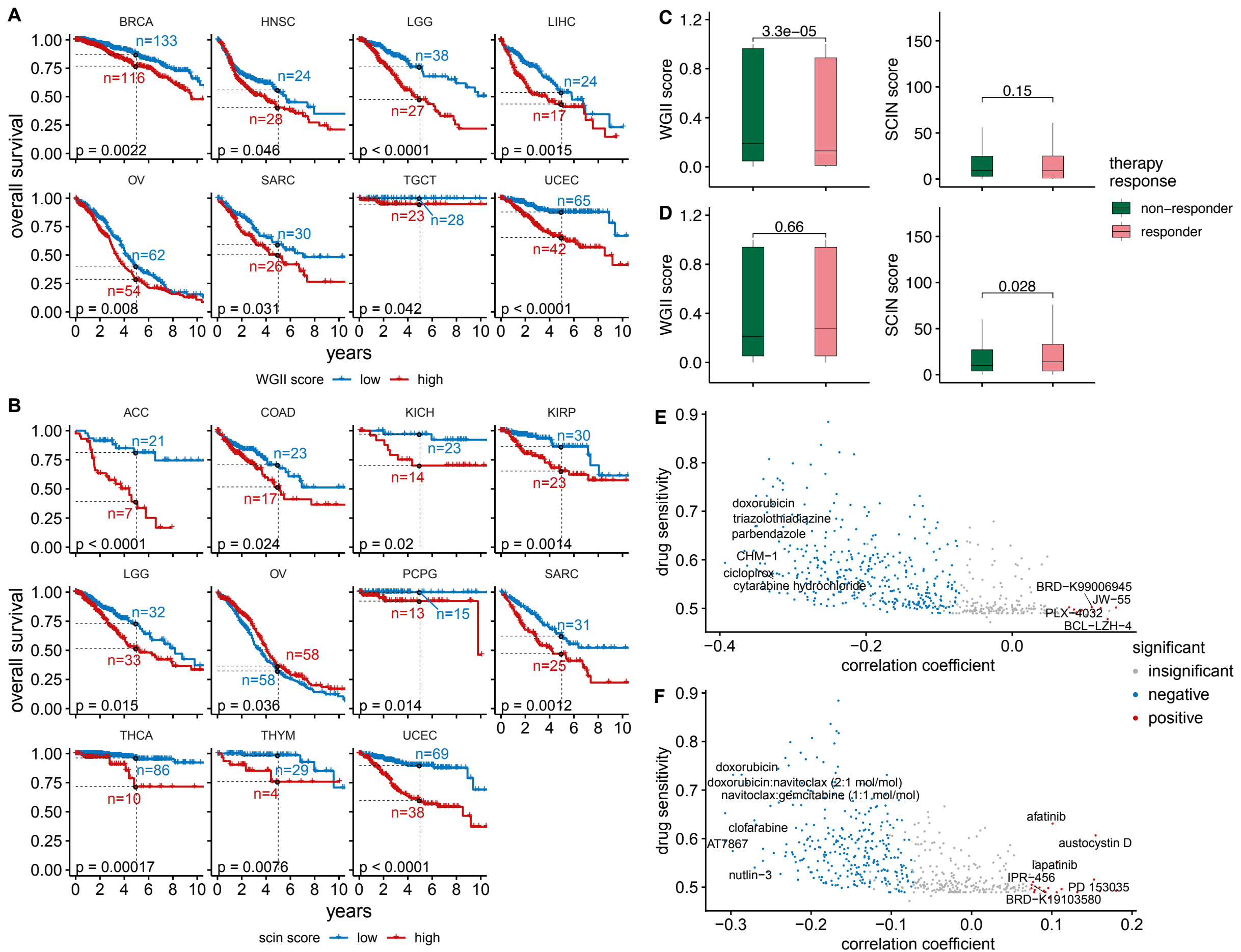
Fig. 3 Paradigm pathway activity and gene expression associated with CIN. (A) The PARADIGM pathway-level activities corresponding to protein-coding genes (rows) were correlated with the WGII score. Only pathways with a significant correlation ($p < 5\%$) larger than 0.4 in at least eight cancer types were included. The heatmap shows the normalised PARADIGM pathway activity (0-1 from low to high). Cancer types were ordered according to their median WGII score, see top panel. **(B)** Volcano plot for the correlation between gene expression and WGII. **(C)** Analogous to (A), but now for the SCIN score instead of the WGII score. **(D)** Correlation of SCIN scores and gene expression, analogous to (B).

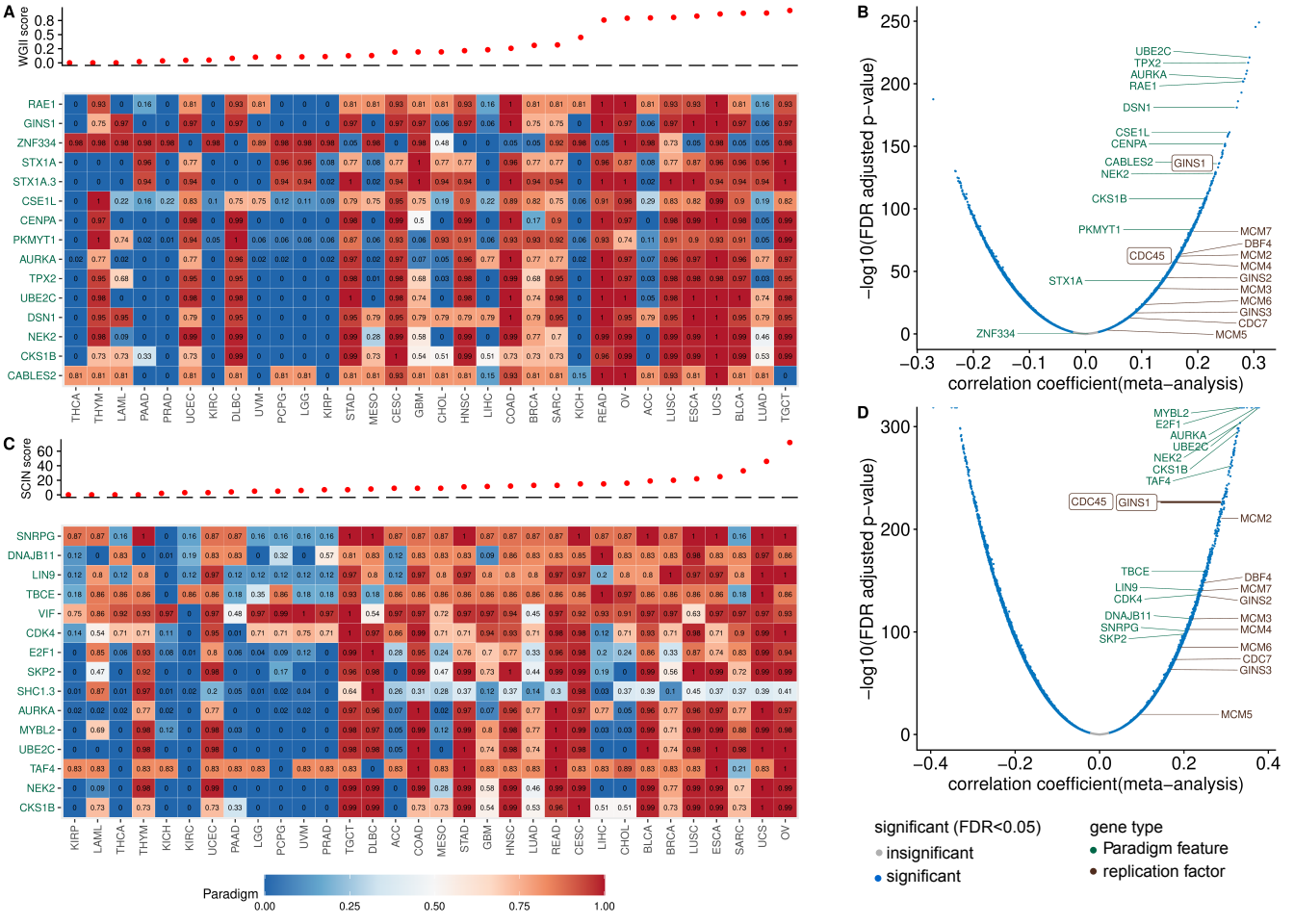
Fig. 4 Pan-cancer somatic mutations and CIN. (A) The volcano plots show the association of (small scale) somatic mutations and the WGII score (left) or the SCIN score (right). The linear model coefficient indicates the mean increase or decrease of the respective CIN score when the mutation is present in a tumour sample relative to the wild type. Genes with lowest p-values, well-known CIN genes and oncogenes are highlighted. The analysis was performed on genes for which samples size for both wild type group and mutated group are larger than 20. Genes were mutations are associated with a significantly (FDR $< 5\%$) higher or lower CIN score are highlighted in blue and red, respectively. **(B)** The same as (A), but hypermutated samples with high levels of MSI excluded.

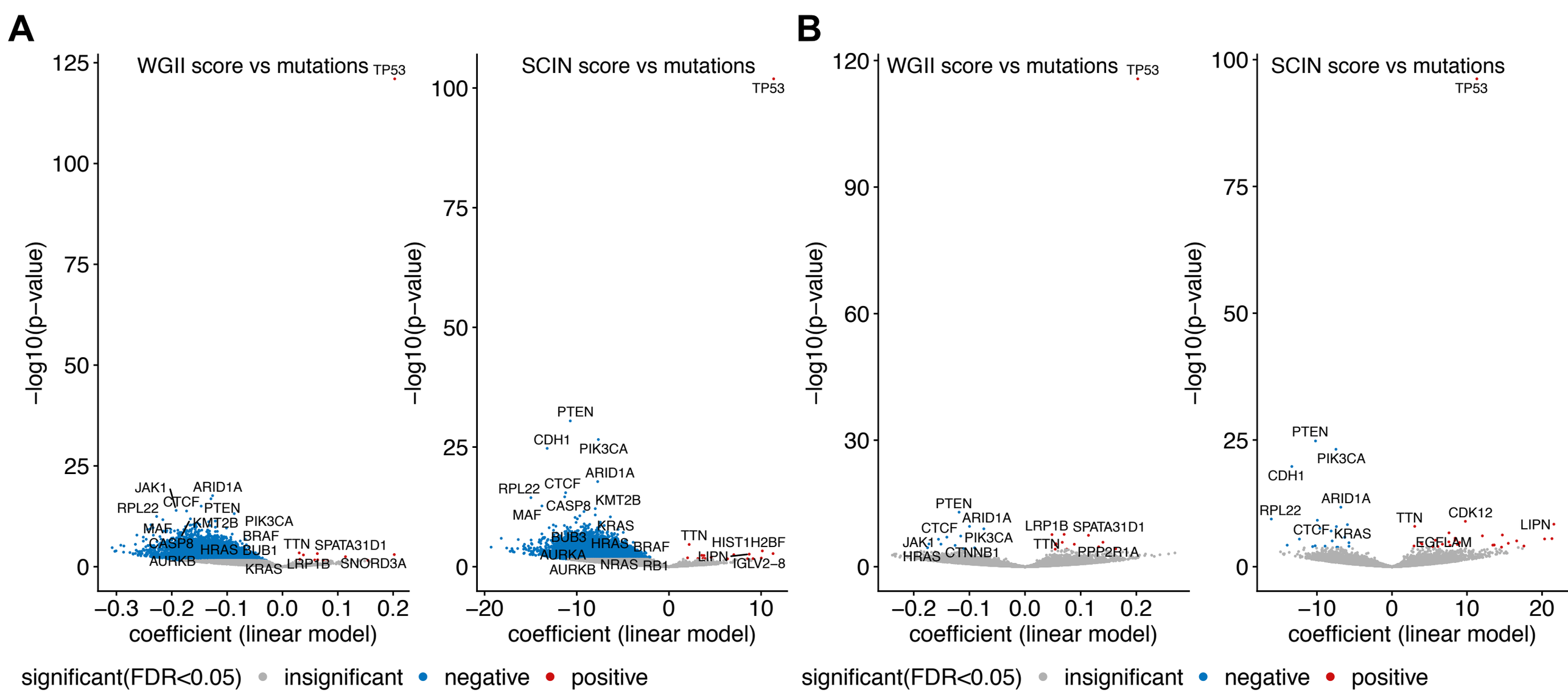
Fig. 5 Copy number amplifications and deletions enriched in high CIN samples. **(A)** The volcano plots show the gene wise associations between copy number amplification status and WGII score (left) and SCIN score (right), obtained from a regression model adjusted by cancer type. The linear model coefficient indicates the mean difference of the respective CIN score when the mutation is present in a tumour sample relative to the wild type. Genes with the lowest p-values and well-known CIN genes are highlighted and colours indicate genes above or below a 5% FDR threshold. Turquoise and fuchsia color encode genes with a significantly higher somatic mutation frequency ($FDR < 0.05$) in samples with low and high CIN scores, respectively. The analysis was performed on 122222 genes of which samples size for both wild type and amplified groups are greater than 20. **(B)** Pan-cancer copy number deletions associated with SCIN displayed in an analogous way as in (A).

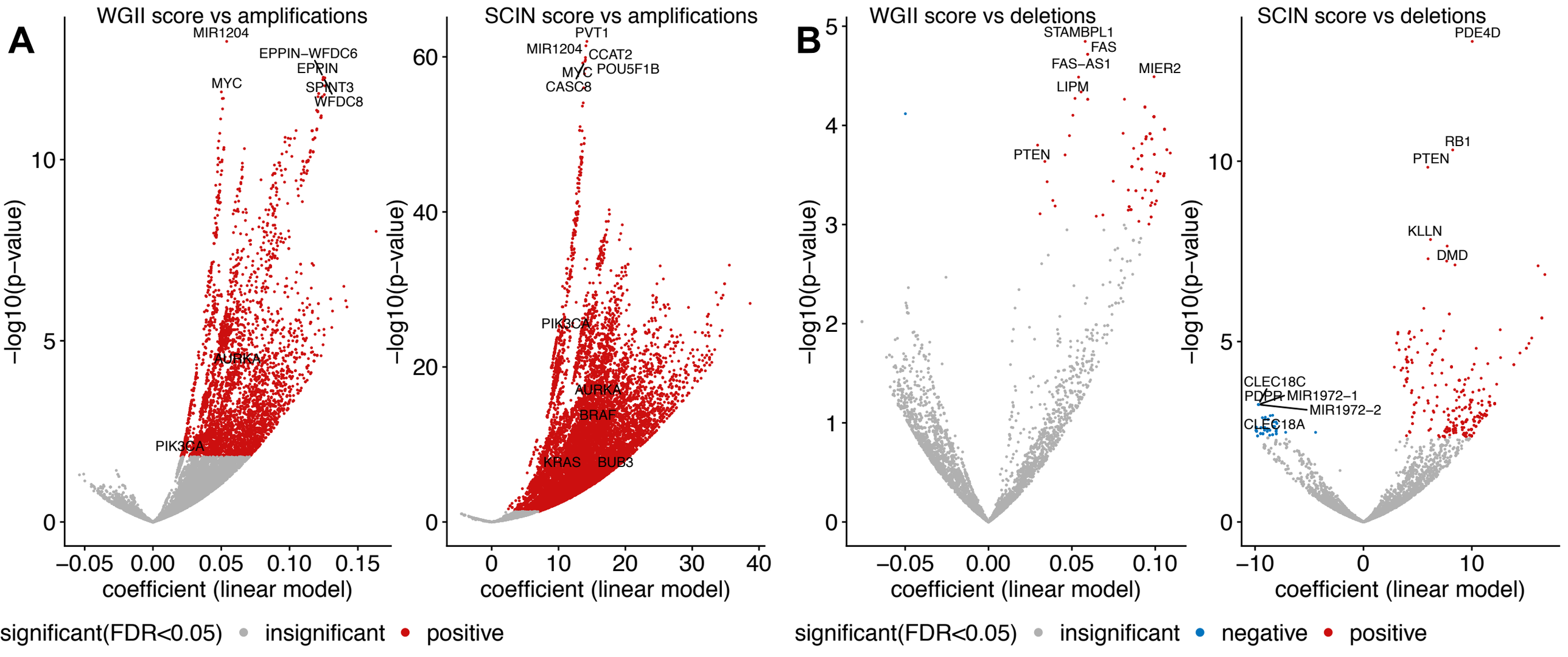
Fig. 6 Oncoprint for PIK3CA and PTEN in relation to CIN (A) The bottom panel depicts the presence or absence of somatic mutations, copy number amplifications of PIK3CA and deletions of PTEN in TCGA tumour samples (columns). Alterations are sorted by their frequency. The upper panel shows the WGII score, SCIN score, PI3KCA and PTEN gene expression. Different levels of CIN scores and gene expression are encoded by colours. **(B)** The corresponding oncoprint for cell line data from CCLE.

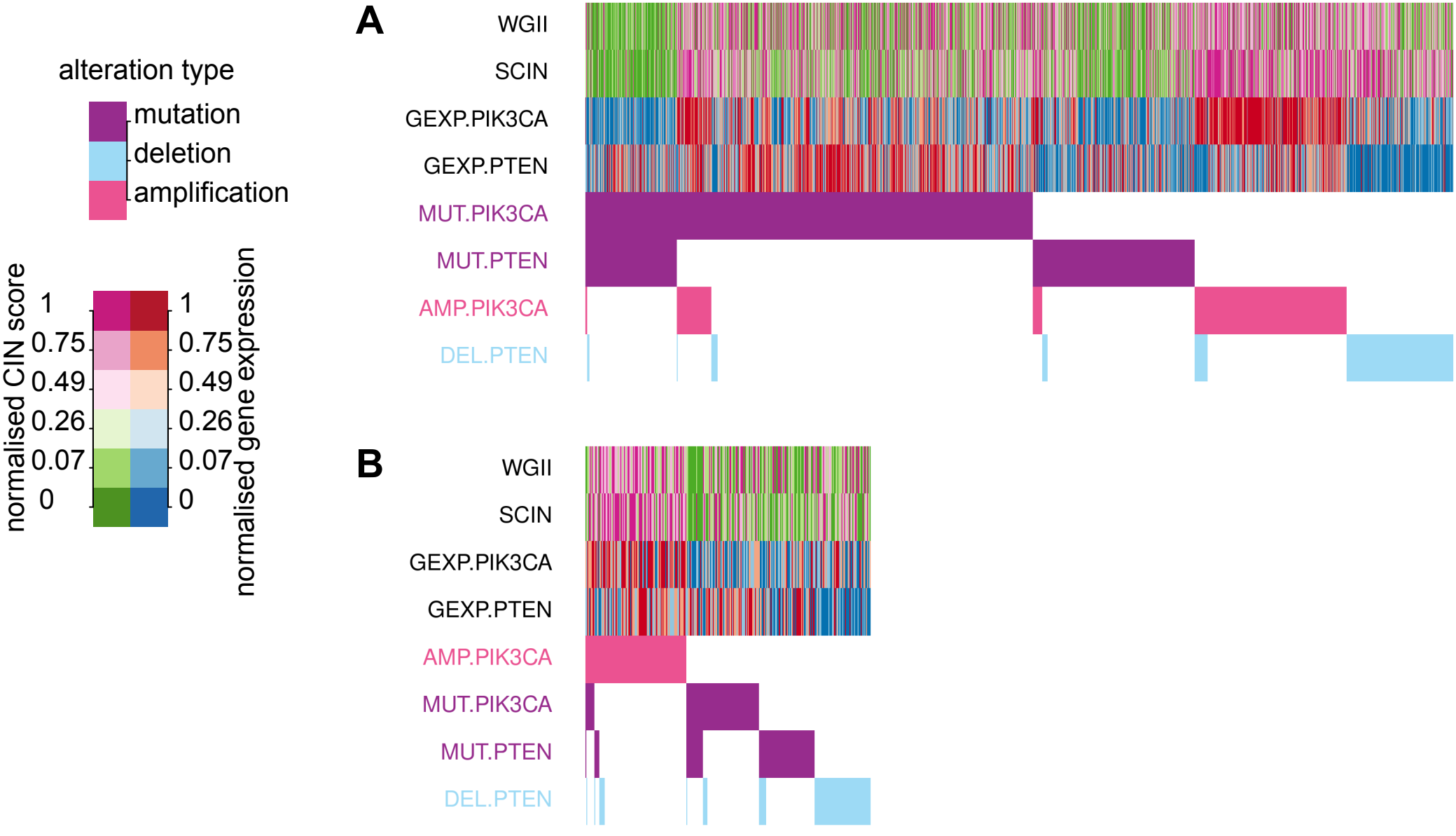

A
WGII score
1.00
0.75
0.50
0.25
0.00
THCA 0.026
THYM 0.104
LAML 0.032
PAAD 0.242
PRAD 0.079
UCEC 0.205
KIRC 0.190
DLBC 0.191
UVM 0.062
PCPG 0.150
LGG 0.330
KIRP 0.064
STAD 0.407
MESO 0.185
GBM 0.147
CESC 0.333
CHOL 0.222
HNSC 0.388
LIHC 0.341
SKCM 0.356
COAD 0.400
BRCA 0.437
SARC 0.471
KICH 0.154
READ 0.519
ACC 0.539
OV 0.585
LUSC 0.546
ESCA 0.568
UCS 0.839
BLCA 0.604
LUAD 0.594
TGCT 0.955
WGD
0
1
2
0
500
1000
1500
2000
count
B
SCIN score
400
300
200
100
0
KIRP
LAML
THCA
THYM
KICH
KIRC
UCEC
PAAD
LGG
PCPG
UVM
PRAD
TGCT
DLBC
COAD
MESO
ACC
GBM
STAD
HNSC
LUAD
READ
CESC
LIHC
CHOL
SKCM
BLCA
BRCA
LUSC
ESCA
SARC
UCS
OV
WGD
0
1
2
0
1000
2000
3000
4000
count
C
MKI67
PROLIF
ITH
NSMR
SMR
HRD
AS
SCIN
TGCT
ACC
OV
UCS
UVM
STAD
LAML
HNSC
SKCM
BLCA
LIHC
BRCA
KIRC
THCA
MESO
CHOL
KIRP
LGG
KICH
CESC
READ
ESCA
GBM
PCPG
LUAD
LUSC
UCEC
COAD
THYM
PAAD
DLBC
PRAD
SARC
−log(pvalue)
100
200
300
400
correlation
0.6
0.3
0.0
−0.3
D
MKI67
PROLIF
ITH
NSMR
SMR
HRD
AS
THYM
UCEC
COAD
ACC
PAAD
DLBC
PRAD
KICH
SARC
BRCA
PCPG
KIRC
CESC
READ
CHOL
BLCA
LIHC
LGG
OV
LUSC
TGCT
ESCA
SKCM
UCS
UVM
GBM
KIRP
STAD
THCA
HNSC
LAML
LUAD
MESO
−log(pvalue)
100
200
300
400
correlation
0.5
0.0


A
BRCA
HNSC
LGG
LIHC
OV
SARC
TGCT
UCEC
overall survival
years
WGII score low high
B
ACC
COAD
KICH
KIRP
LGG
OV
PCPG
SARC
THCA
THYM
UCEC
scin score low high
C
WGII score
SCIN score
3.3e−05
0.15
D
0.66
0.028
therapy response
non−responder
responder
E
drug sensitivity
correlation coefficient
doxorubicin
triazolothiadiazine
parbendazole
CHM−1
ciclopirox
cytarabine hydrochloride
BRD−K99006945
JW−55
PLX−4032
BCL−LZH−4
significant
insignificant
negative
positive
F
doxorubicin
doxorubicin:navitoclax (2:1 mol/mol)
navitoclax:gemcitabine (1:1 mol/mol)
clofarabine
AT7867
nutlin−3
afatinib
austocystin D
lapatinib
IPR−456
PD 153035
BRD−K19103580


A
WGII score
0.0
0.2
0.4
0.6
0.8
RAE1
GINS1
ZNF334
STX1A
STX1A.3
CSE1L
CENPA
PKMYT1
AURKA
TPX2
UBE2C
DSN1
NEK2
CKS1B
CABLES2
THCA
THYM
LAML
PAAD
PRAD
UCEC
KIRC
DLBC
UVM
PCPG
LGG
KIRP
STAD
MESO
CESC
GBM
CHOL
HNSC
LIHC
COAD
BRCA
SARC
KICH
READ
OV
ACC
LUSC
ESCA
UCS
BLCA
LUAD
TGCT
B
−log10(FDR adjusted p−value)
correlation coefficient(meta−analysis)
UBE2C
TPX2
AURKA
RAE1
DSN1
CSE1L
CENPA
CABLES2
GINS1
NEK2
CKS1B
PKMYT1
CDC45
STX1A
ZNF334
MCM7
DBF4
MCM2
MCM4
GINS2
MCM3
MCM6
GINS3
CDC7
MCM5
C
SCIN score
0
20
40
60
SNRPG
DNAJB11
LIN9
TBCE
VIF
CDK4
E2F1
SKP2
SHC1.3
AURKA
MYBL2
UBE2C
TAF4
NEK2
CKS1B
KIRP
LAML
THCA
THYM
KICH
KIRC
UCEC
PAAD
LGG
PCPG
UVM
PRAD
TGCT
DLBC
ACC
COAD
MESO
STAD
GBM
HNSC
LUAD
READ
CESC
LIHC
CHOL
BLCA
BRCA
LUSC
ESCA
SARC
UCS
OV
Paradigm
0.00
0.25
0.50
0.75
1.00
D
−log10(FDR adjusted p−value)
correlation coefficient(meta−analysis)
MYBL2
E2F1
AURKA
UBE2C
NEK2
CKS1B
TAF4
CDC45
GINS1
MCM2
TBCE
LIN9
CDK4
DNAJB11
SNRPG
SKP2
DBF4
MCM7
GINS2
MCM3
MCM4
MCM6
CDC7
GINS3
MCM5
significant (FDR<0.05)
insignificant
significant
gene type
Paradigm feature
replication factor


A
WGII score vs mutations
SCIN score vs mutations
−log10(p−value)
coefficient (linear model)
TP53
PTEN
CDH1
PIK3CA
ARID1A
RPL22
CTCF
KMT2B
CASP8
MAF
KRAS
BUB3
HRAS
BRAF
AURKA
AURKB
NRAS
RB1
TTN
HIST1H2BF
LIPN
IGLV2−8
JAK1
PTEN
BRAF
BUB1
SPATA31D1
LRP1B
SNORD3A
B
WGII score vs mutations
SCIN score vs mutations
CTNNB1
PPP2R1A
CDK12
EGFLAM
significant(FDR<0.05)
insignificant
negative
positive


A
WGII score vs amplifications
MIR1204
EPPIN−WFDC6
EPPIN
MYC
SPINT3
WFDC8
AURKA
PIK3CA
−log10(p−value)
coefficient (linear model)
SCIN score vs amplifications
PVT1
MIR1204
CCAT2
MYC
POU5F1B
CASC8
PIK3CA
AURKA
BRAF
KRAS
BUB3
B
WGII score vs deletions
STAMBPL1
FAS
FAS−AS1
MIER2
LIPM
PTEN
SCIN score vs deletions
PDE4D
RB1
PTEN
KLLN
DMD
CLEC18C
PDPR
MIR1972−1
MIR1972−2
CLEC18A
significant(FDR<0.05)
insignificant
negative
positive


alteration type
mutation
deletion
amplification
normalised CIN score
normalised gene expression
1
0.75
0.49
0.26
0.07
0
A
WGII
SCIN
GEXP.PIK3CA
GEXP.PTEN
MUT.PIK3CA
MUT.PTEN
AMP.PIK3CA
DEL.PTEN
B
WGII
SCIN
GEXP.PIK3CA
GEXP.PTEN
AMP.PIK3CA
MUT.PIK3CA
MUT.PTEN
DEL.PTEN