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 W-CIN and S-CIN can co-occur, they are initiated by different molecular events. By analysing tumour genomic data from 33 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 with homologous recombination deficiency. Both CIN phenotypes are prognostic in several cancer types. Most drugs are less efficient in high CIN cell lines, we also report compounds and drugs which could specifically target W-CIN or S-CIN. By analysing associations between CIN and bio-molecular entities at pathway and gene expression levels, we complement gene signatures of CIN and report that the drug resistance gene *CKS1B* is strongly associated with both W-CIN and S-CIN. Finally, we propose a potential copy number dependent mechanism to activate the *PI3K* pathway in high CIN tumours.

Keywords: whole chromosomal instability; structural chromosomal instability; whole genome doubling; integrative analysis; *PI3K* oncogenic activation

1. Simple summary

Many cancer cells are chromosomally unstable, a phenotype describing a tendency for accumulating chromosomal aberrations. Entire chromosomes tend to be gained or lost, which is called whole chromosome instability (W-CIN). Structural chromosomal instability (S-CIN) describes an increased rate of gaining, losing or translocating smaller parts of chromosomes. Here we analyse data from 33 cancer types to find differences and commonalities between W-CIN and S-CIN. We find, that W-CIN is strongly linked to whole genome doubling (WGD), whereas S-CIN is associated with a specific DNA damage repair pathway. Both W-CIN and S-CIN are difficult to target using currently available compounds and have distinct prognostic values. The activity of the drug resistance gene *CKS1B* is associated with both CIN types, which merits further investigation as potential CIN target. In addition, we identify a potential copy number based mechanism promoting signalling of the important *PI3K* cancer pathway in high CIN tumours.

2. 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 chromosomal changes. Accordingly, two major forms of CIN can be distinguished: Numerical or whole chromosome instability (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 S-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 not only induces CIN [9, 10], but also occurs as an immediate short term response to aneuploidy and CIN [11].

Aneuploidy and CIN have typically detrimental effects on cell fitness and proliferation [5, 11, 12]. Therefore, it was unclear why CIN is often associated with poor patient survival and more aggressive disease progression [1, 13, 14]. Stratification of breast cancer patient samples into low, intermediate and high CIN groups revealed that patients with intermediate levels of CIN had the worst survival, whereas the low and high CIN groups had a better prognosis [15, 16]. These results hinted at mechanisms for tolerating CIN in order to survive the stresses provoked by chromosomal aberrations. The CIN tolerance

mechanisms are currently not completely understood [17], but one important recurring event is a loss of *TP53* function, which otherwise prevents the propagation of CIN cells [18].

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 criticised 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 [19]. These studies highlighted, that we have to distinguish between acute responses to aneuploidy and CIN [11], mechanisms for tolerating CIN [17] and the cellular programme [20, 21] and genetic alterations [22] acquired by evolved CIN cells. These cellular programmes might differ between cancer cell lines and tumours, partially as a result of treatment effects or as a result of interactions with the tumour microenvironment. Recently, it was 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, 23]. This interesting research links cancer 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 [24, 25]. This acquired drug resistance could potentially exacerbate the intrinsic drug resistance [26] of many CIN cells, which highlights the need to better understand genomic changes of CIN tumours in the context of anti-cancer treatment.

Computational studies of cancer genomic data have provided valuable insights into CIN [1, 19–22] and aneuploidy [27] 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 propose a new link between CIN and the PI3K oncogenic pathway.

3. Materials and Methods

3.1. TCGA Pan-cancer clinical and molecular data

We analysed chromosome instability of 33 primary tumour types: adrenocortical carcinoma (ACC, $n = 89$); bladder urothelial carcinoma (BLCA, $n = 399$); breast invasive carcinoma (BRCA, $n = 1039$); cervical and endocervical cancers (CESC, $n = 294$); cholangiocarcinoma (CHOL, $n = 36$); colon adenocarcinoma (COAD, $n = 420$); lymphoid neoplasm diffuse large b-cell lymphoma (DLBC, $n = 47$); esophageal carcinoma (ESCA, $n = 162$);

glioblastoma multiforme (GBM, $n = 556$); head and neck squamous cell carcinoma (HNSC, $n = 510$); kidney chromophobe (KICH, $n = 65$); kidney renal clear cell carcinoma (KIRC, $n = 480$); kidney renal papillary cell carcinoma (KIRP, $n = 280$); acute myeloid leukemia (LAML, $n = 124$); brain lower grade glioma (LGG, $n = 506$); liver hepatocellular carcinoma (LIHC, $n = 361$); lung adenocarcinoma (LUAD, $n = 490$); lung squamous cell carcinoma (LUSC, $n = 482$); mesothelioma (MESO, $n = 81$); ovarian serous cystadenocarcinoma (OV, $n = 550$); pancreatic adenocarcinoma (PAAD, $n = 165$); pheochromocytoma and paraganglioma (PCPG, $n = 160$); prostate adenocarcinoma (PRAD, $n = 471$); rectum adenocarcinoma (READ, $n = 154$); sarcoma (SARC, $n = 244$); skin cutaneous melanoma (SKCM, $n = 104$); stomach adenocarcinoma (STAD, $n = 427$); testicular germ cell tumours (TGCT, $n = 133$); thyroid carcinoma (THCA, $n = 463$); thymoma (THYM, $n = 106$); uterine corpus endometrial carcinoma (UCEC, $n = 512$); uterine carcinosarcoma (UCS, $n = 56$); uveal melanoma (UVM, $n = 80$).

We also calculated CIN scores for 391 metastatic tumour tissues, 8719 blood derived normal tissues and 2207 solid normal tissues.

The TCGA pan-cancer molecular and clinical data were downloaded from the Pan-Cancer Atlas [28]. The file names for different data modalities are: copy number segment data from `broad.mit.edu_PANCAN_Genome_Wide_SNP_6_whitelisted.seg`; ABSOLUTE [29] inferred ploidy data from `TCGA_mastercalls.abs_tables_JSedit.fixed.txt`; normalised and batch-effect corrected gene expression profile from `EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv`; clinical data from `TCGA-CDR-SupplementalTableS1.xlsx`; PARADIGM [30] inferred pathway activity data from `merge_merged_reals.tar.gz`.

3.2. CCLE molecular and sample annotation data

Cell line multi-omics data were downloaded from the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE) [31]. In particular, the copy number segment data are located in `CCLE_copynumber_2013-12-03.seg.txt`. Gene expression profiles and sample annotations are located in `CCLE_RNAseq_genes_rpkm_20180929.gct.gz` and `Cell_lines_annotations_20181226.txt`. The binary alteration matrix is located in `CCLE_MUT_CNA_AMP_DEL_binary_Revealer.gct`. Sample ploidy data estimated using the ABSOLUTE algorithm [29] are located in `CCLE_ABSOLUTE_combined_20181227.xlsx`.

3.3. CTRP drug screening data

We collected cell line pharmacological profiling data from the Cancer Therapeutics Response Portal (CTRP [32], `CTRPv2.0_2015_ctd2_ExpandedDataset.zip`). Drug sensitivity measured by the area under the dose response curve (AUC) was further min-max normalised. From this, we computed the drug sensitivity index as $1 -$ normalised AUC with values in the range between 0 (highest resistance) and 1 (most sensitive).

3.4. Quantifying W-CIN and S-CIN

We implemented three different karyotypic complexity scores [7] to quantify CIN in both TCGA bulk tumours and CCLE cell lines: the numerical chromosome instability score (NCIN score), the structural chromosome instability score (SCIN score) and the weighted genome instability Index (WGII). The NCIN scores 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. We found that the WGII is less sensitive to outliers and leads to less variable estimates. Therefore, we used it as a more robust proxy measure for W-CIN instead of the NCIN score.

3.5. Association analysis between CIN and genome instability

Aneuploidy scores (AS) of samples are taken from [27], Supplementary Table S2, 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 [33], Supplementary Table S1. Microsatellite instability (MIN) scores are collected from [34], Supplementary Table S5. The correlations of these genome instability scores and WGII or SCIN scores were quantified by Spearman correlation coefficients.

3.6. Survival analysis

We performed survival analysis using the survival R package [35]. 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 log rank 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 survminer R package [36].

3.7. Treatment response analysis

We labelled patients with complete/partial response to chemotherapy or radiation-therapy as responders and the other patients as non-responders. A Wilcoxon rank sum test was used to evaluate the differences of the WGII and SCIN scores in the responder and non-responder groups.

3.8. Identification of candidate compounds selectively targeting CIN

Spearman correlation coefficients between drug sensitivity (defined in 3.3) and CIN were computed for 545 CTRP compounds. Compounds with multiple testing adjusted $p \leq 0.05$ and median drug sensitivity > 0.5 were considered as candidate compounds selectively targeting low CIN cancer cells (compounds

with negative correlation coefficients) or high CIN cancer cells (compounds with positive correlation coefficients).

3.9. Association analysis between CIN and PARADIGM pathway activities

We collected the sample-wise PARADIGM pathway activity matrix from the Pan-Cancer Atlas [28] with the file name merge_merged_reals.tar.gz. For each cancer type we computed the 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 whose PARADIGM pathway activities are strongly positively correlated with WGII or SCIN score (correlation coefficient greater than 0.3) in more than eight cancer types.

3.10. Association analysis between somatic alterations and CIN

We used the limma R package [37] for multiple linear regression analysis on CIN scores, using alteration status (mutation, copy number amplification or copy number deletion versus wild type) and cohort as predictor variables. To achieve sufficient statistical power, only alterations which occurred in more than 20 samples were included as predictors.

4. Results

4.1. 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 [38, 39]. However, there are established surrogate scores based on copy number segment 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, 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 33 cancer types.

The distribution of WGII varies drastically across cancer types (**Figure 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 characterised by a WGD event. Please note that this is not an artefact of the WGII, which is measured relative to the sample ploidy. This suggests that WGD is an important mechanism inducing W-CIN in many cancer types. However, the exception is kidney chromophobe (KICH), where WGD events seem to be rare, but high levels of the WGII can still be observed. In this cancer type, there is also no clear bimodal pattern, suggesting that other mechanisms than WGD drive W-CIN in KICH. Even in cancers where

the bimodal pattern suggests a clear separation between numerically unstable and numerically stable tumours, it is difficult to define an universal WGII threshold distinguishing numerically stable from W-CIN tumours across cancer types. For example, in ovarian serous cystadenocarcinoma (OV) one can distinguish low and high WGII groups with WGD, but the overall level of the WGII is much higher than that in other cancer types. Similarly, for breast invasive carcinoma (BRCA) there are many patients with high levels of WGII even in the group of samples which did not undergo WGD. This suggests, that other processes than WGD 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 (**Figure 1(B)**). This indicates that most tumours are structurally chromosomal stable, but some can exhibit extreme levels of S-CIN. These extreme levels of S-CIN are typically seen in tumours which did not undergo WGD.

The distribution of SCIN scores indicates a high degree of tumour heterogeneity within the same cancer type and across cancer types. Ovarian serous cystadenocarcinoma (OV), uterine carcinosarcoma (UCS) and sarcoma (SARC) show the highest SCIN scores (**Figure 1(B)**) and many samples within these tumours also exhibit high WGII (compare **Figure 1(A)**). Both types of CIN occur 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 recognised as those dominated by the CIN phenotype [40], including stomach adenocarcinoma (STAD), colon adenocarcinoma (COAD), uterine corpus endometrial carcinoma (UCEC), OV, UCS and prostate adenocarcinoma (PRAD) have extremely heterogeneous SCIN scores.

We also checked for associations 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 is uniformly positively associated with the aneuploidy score (**Figure 1(C)**), which is expected, because aneuploidy can be both a cause and a consequence of CIN [41]. Intriguingly, HRD is consistently positively associated with the SCIN score (**Figure 1(D)**), suggesting that impaired repair of DNA double strand breaks might be a key driver of S-CIN.

CIN and microsatellite instability (MIN) are usually considered mutually exclusive [38]. Indeed, most MIN tumours have low WGII and SCIN scores, but some MIN samples which underwent WGD can also exhibit signs of W-CIN and S-CIN (**Figure A1(A)**).

To check for a potential link between CIN and proliferation, we used a proliferation index [33] 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 (**Figure 1(C)**). However, this is not the case for other cancers, reflecting again the complex relationship between CIN and proliferation [42–44]. 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 WGII and SCIN scores tend to be higher in metastatic tumours than in primary tumours and in normal samples (**Figure A1(B)**), consistent with previous findings linking CIN and metastasis [23]. The small sample size for metastatic tumours prevents a cancer type specific analysis of the relationship between CIN and metastatic disease.

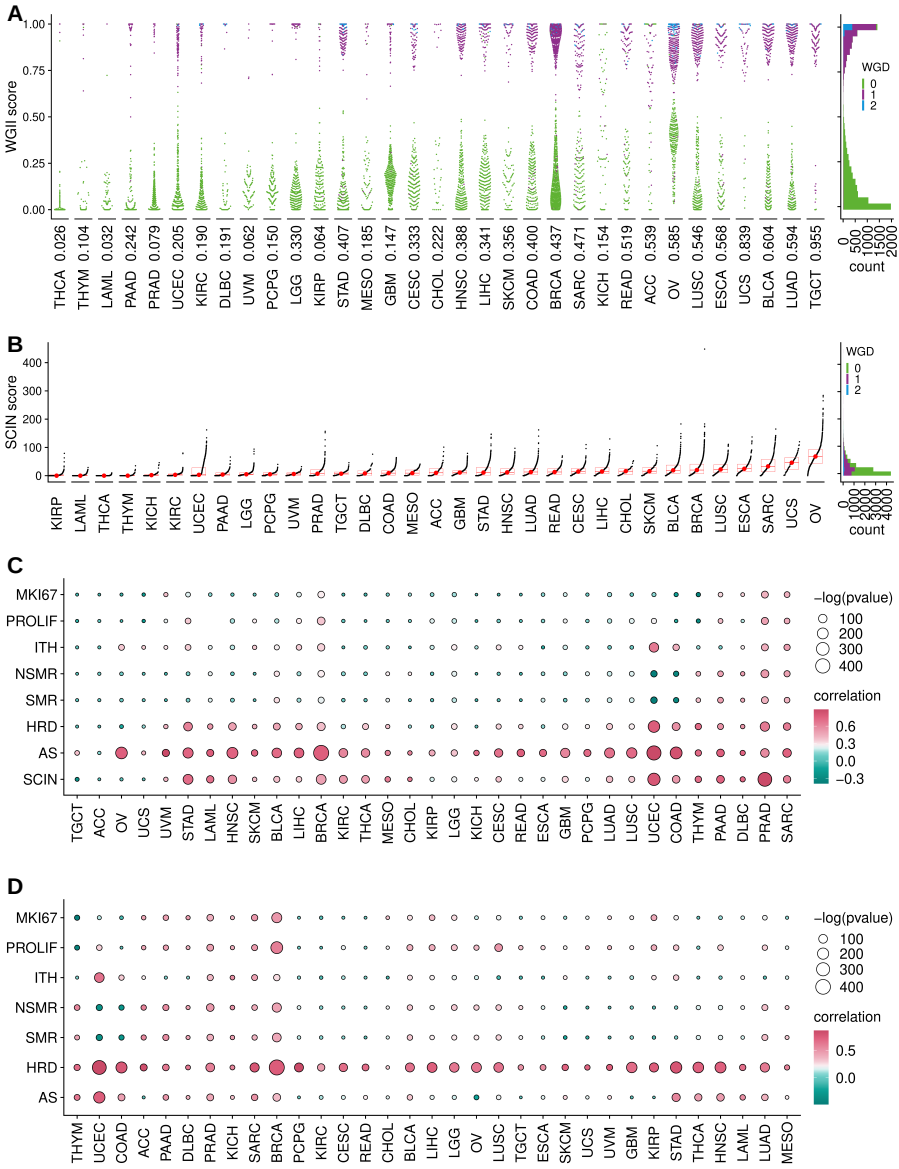


Figure 1. Distribution of CIN scores and their association with genetic instability. (A) Left: WGII for TCGA tumour samples (dots) from different cancer types, sorted according to median WGII. The colour coding indicates the WGD status and the number below each beeswarm plot is the proportion of samples which underwent WGD. Right: Pan-cancer histogram of WGII. **(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 **(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 these indices were collected from [27, 33].

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 in most cancer types separates high W-CIN from low W-CIN samples, but does not provide a universal threshold valid across cancer types. However, in some cancers like OV, even the non-WGD samples can exhibit substantial levels of W-CIN. In contrast, S-CIN is a continuous trait which is strongly associated with HRD, but not with WGD. Please note that these patterns are also observed in cell lines (**Figure A1(C, D)**).

4.2. Clinical significance of CIN in different cancer types

To analyse the relationship between W-CIN and prognosis, we divided the tumour samples in each individual cancer type into disjoint WGII^{high} and WGII^{low} groups using the median as a threshold. For eight of the 33 cancer types we found that WGII^{high} patients had a significantly shorter overall survival than patients in the WGII^{low} group (**Figure 2(A)**, **Table A1**, Log rank test, $p < 0.05$). 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 (**Figure A2(A)**, **Table A3**, Log rank test, $p < 0.05$) and progression free survival is negatively associated with high WGII in LGG, LIHC, OV, PRAD, SKCM, THYM, UCEC and UVM (**Figure A3(A)**, **Table A5**, Log rank test, $p < 0.05$).

Using an analogous separation of the tumour samples into SCIN^{low} and SCIN^{high} groups using the median SCIN score in each tumour type we found that the overall survival of patients in 10 out of 33 cancers is negatively associated with S-CIN (**Figure 2(B)**, **Table A2**, Log rank test, $p < 0.05$). High SCIN scores are linked to impaired disease free survival in adrenocortical carcinoma (ACC), KIRC, kidney renal papillary cell carcinoma (KIRP), lung squamous cell carcinoma (LUSC), PRAD, THCA and UCEC (**Figure A2(B)**, **Table A4**, Log rank test, $p < 0.05$). For OV, patients with high SCIN scores tend to have slightly better overall survival (**Figure 2(B)**, **Table A2**). However, the effect is very small and at the edge of statistical significance. In addition, the analysis of disease free survival (**Figure A2**, **Table A3**, **Table A4**) and progression free survival (**Figure A3**, **Table A5**, **Table A6**) does not provide any evidence for an effect of S-CIN on the prognosis of OV patients.

To further explore the clinical relevance of both types of CIN in therapy, we studied the association between CIN and response to radiotherapy or chemotherapy. Radiotherapy responders tend to have lower WGIIs than radiotherapy non-responders (Wilcoxon rank test, $p = 10e - 5$), whereas SCIN is not significantly associated with radiotherapy response (**Figure 2(C)**). On a pan-cancer level, we did not find a significant difference between WGIIs in the

group of chemotherapy responders versus non-responders (**Figure 2(D)**). The median SCIN score of chemotherapy responders is slightly higher. One possible explanation is that high S-CIN samples tend to have defective homologous recombination repair (see **Figure 1(B)**), which renders them slightly more sensitive to chemotherapy [45, 46].

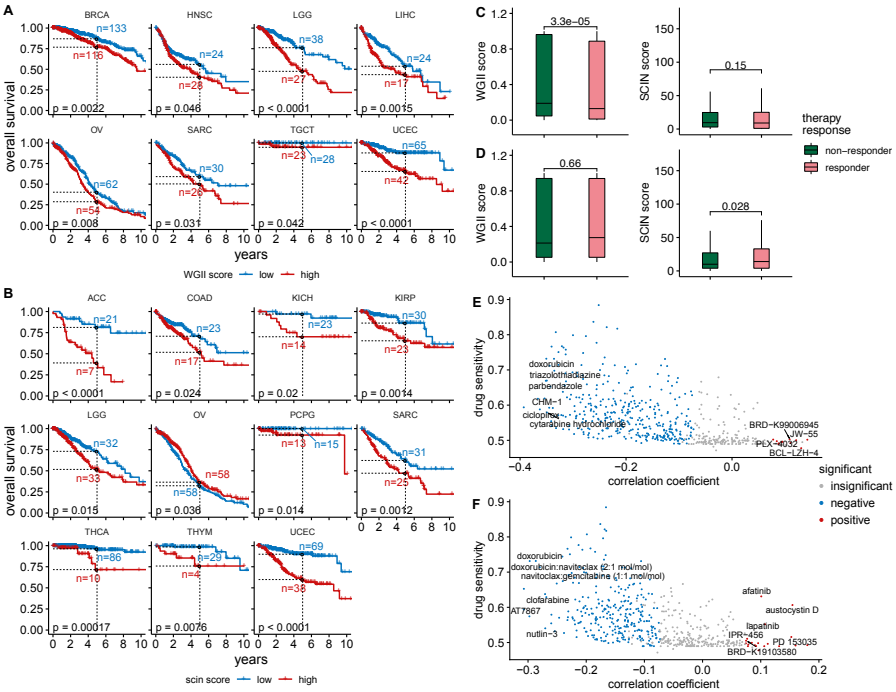


Figure 2. Clinical significance of CIN in different cancer types. (A) For eight cancer types there are significant differences in overall survival between patient samples with low WGII (blue) and high WGII (red). Dashed lines indicate the five year overall survival probability of the two groups. **(B)** 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 and SCIN score between radiotherapy responders and non-responders using a Wilcoxon rank sum test. **(D)** Comparison of the WGII and SCIN score between chemotherapy responders and non-responders using a Wilcoxon rank sum test. **(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. Compounds whose sensitivity is significantly negatively or positively correlated with SCIN score are highlighted in blue and red, respectively.

Next, we asked whether there are drugs suitable for targeting CIN [47]. 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 dose response 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 the highest resistance level, whereas values of one indicate the highest possible sensitivity. We then computed Spearman rank correlation coefficients between the drug sensitivity of each compound 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 (**Figure 2(E)**) or SCIN score (**Figure 2(F)**).

For the majority of compounds we found negative correlations between their sensitivity and both types of CIN (**Figure 2(E, F)**), highlighting that for many compounds CIN confers an intrinsic drug resistance [26]. Only a few compounds are more potent in high CIN cell lines than in low CIN cell lines. However, their overall levels of sensitivity are typically low in comparison to drugs more efficient in low CIN cell lines.

The strongest positive correlations between sensitivity and WGII were found for the compounds BRD-K99006945, JW-55 (targeting *TNK*), PLX-40932 (targeting *BRAF*) and BCL-LZH-4 (Spearman correlation coefficient larger than 0.3 and FDR adjusted $p < 5\%$, **Figure 2(E)**). None of these compounds has been approved for clinical use by the FDA.

Drugs showing increasing sensitivity with the SCIN-score include Afatinib and Lapatinib (Spearman correlation coefficient larger than 0.3 and FDR adjusted $p < 5\%$, **Figure 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 [48]. Austocystin D is a natural cytotoxic agent and also more efficient in high S-CIN tumours. Further details about the correlations between both CIN scores and drug sensitivity can be found in the **Table S1** and **Table S2**.

Overall, the analysis shows that the prognostic value of CIN depends on cancer types and that S-CIN and W-CIN provide distinct prognostic information. The prognosis for many cancer types worsens with increased levels of CIN. Only for OV we found a slightly better overall survival for patients with high levels of S-CIN. It is possible, that a stratification of patients according to cancer subtypes might reveal more fine grained insights regarding the prognostic value of CIN [15, 16]. Our drug sensitivity analysis reveals that most compounds are less efficient in high CIN tumours than in low CIN tumours. There are a few drugs to which high CIN cells are more sensitive than low CIN cells. In particular, we suggest that Afatinib, Lapatinib and Austocystin D merit further investigation for targeting S-CIN tumours. However, current drug sensitivity screens do not include many highly potent drugs specifically targeting CIN.

4.3. PARADIGM pathway activity and CIN

To identify pathways with altered activity in W-CIN or S-CIN tumours, we used the PARADIGM framework [30]. PARADIGM is a computational model which represents interactions between biological entities as a factor graph. PARADIGM integrates copy number and gene expression data and computes activities for each PARADIGM pathway feature in an individual tumour sample. These features refer to protein coding genes, protein complexes, abstract processes and gene families. We focused on the PARADIGM features for protein coding genes, because these are easier to interpret and can be used to generate experimentally testable predictions. We correlated the PARADIGM pathway features with the WGII or SCIN scores and filtered features with a significant (FDR adjusted $p < 5\%$) Spearman correlation coefficient larger than 0.4 in at least eight of the 32 cancer types (WGII: **Figure 3(A)**, SCIN score: **Figure 3(C)**).

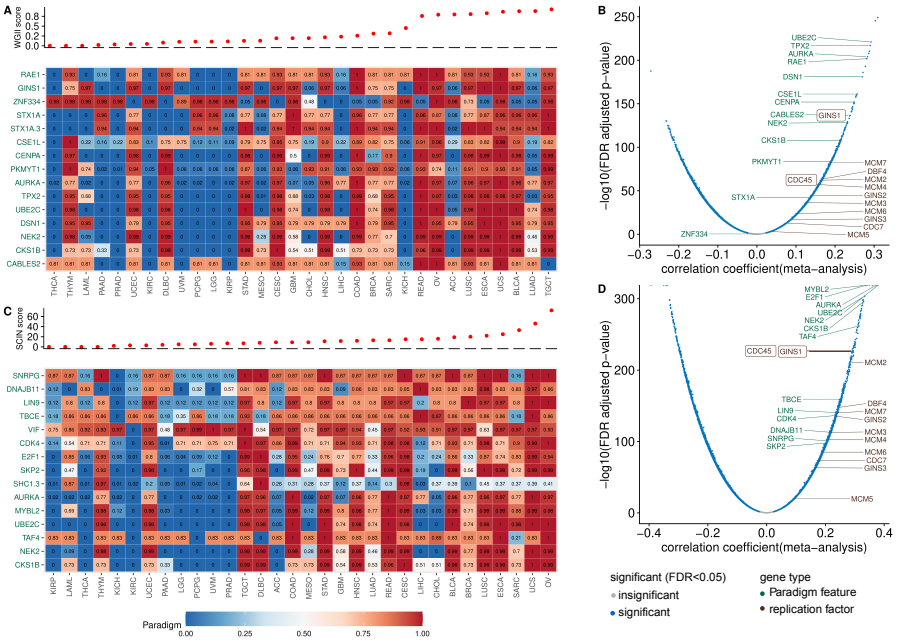


Figure 3. PARADIGM pathway activity and gene expression associated with CIN. (A) The PARADIGM pathway-level activities corresponding to protein-coding genes (rows) were correlated with the WGII. Only pathways with a significant correlation (FDR adjusted $p < 5\%$) larger than 0.4 in at least eight cancer 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, see top panel. **(B)** Volcano plot for the correlation between gene expression and WGII. **(C)** Analogous to **(A)**, but for the SCIN score instead of the WGII. **(D)** Correlation of SCIN score and gene expression, analogous to **(B)**.

PARADIGM pathway features corresponding to the mitotic genes *TPX2, RAE1, UBE2C, AURKA, NEK2* (see **Figure 3(A)**) show increased activity in tumours with high WGII, consistent with the known role of chromosome segregation errors in W-CIN [1, 20]. Additionally, the PARADIGM features 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 *TP53*. *DSN1* and *CENPA* are required for kinetochore assembly.

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

It is interesting to note the positive association of the PARADIGM feature for *GINS1* with WGII [10]. The *GINS1* protein is essential for the formation of the *Cdc45-MCM-GINS* (CMG) complex which functions to unwind DNA ahead of the replication fork [49]. 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 **Figure 3(A, B)**).

The analysis of the SCIN score associated PARADIGM features (**Figure 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* (cyclin-dependent kinase regulatory subunit 1B) pathway is positively associated with both the WGII and SCIN score. *CKS1B* has recently been linked to cancer drug resistance and was discussed as a new therapeutic target [50]. Our results suggest that the *CKS1B* activity is closely linked to CIN, which needs to be considered when studying *CKS1B* as a new target gene or as a marker of drug resistance.

To check for the robustness of these findings, we also performed a gene wise correlation of the SCIN score and gene expression (**Figure 3(C)**, **Figure A4(C)**). There, we also highlighted genes involved in DNA replication. Gene set enrichment analysis indicates that the top high S-CIN associated genes are enriched with replication origin factors (**Figure A4(A, B)**).

Please note, that the analysis of genes and PARADIGM pathways negatively associated with CIN did not reveal a similarly consistent pattern across cancer types (see **Figure A5**).

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 over-expression of genes involved in DNA replication is positively associated with high CIN.

4.4. Somatic point mutation frequencies in high CIN tumours

To investigate the relationship between somatic point mutations and CIN, we identified genes that are more frequently or less frequently mutated in high CIN tumours. From the 19171 gene mutations, we included only those occuring in more than 19 samples in the wild type or mutant group across different cancer types. We fitted a linear regression model using WGII or SCIN score as response and somatic point mutation status (present or absent) and cancer type as predictors. The estimated regression coefficient for mutation status was used to measure its association with CIN, adjusted for tumour type.

As expected, at the pan-cancer level, *TP53* mutation shows the strongest association with CIN. The mean WGII of tumours harbouring a *TP53* mutation is approximately 23% higher than that of the wild type (ANOVA p-value $< 2.2e-16$, **Figure 4(A)**). 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 (**Figure 4(B)**). In line with this, *TP53* mutation is positively associated with high CIN in many individual cancer types (**Figure A6(A)**). In fact, even after removing MIN samples, this correlation still holds (**Figure A6(B)**), corroborating the well known role of *TP53* as a gatekeeper of genome stability (see e.g. [51]).

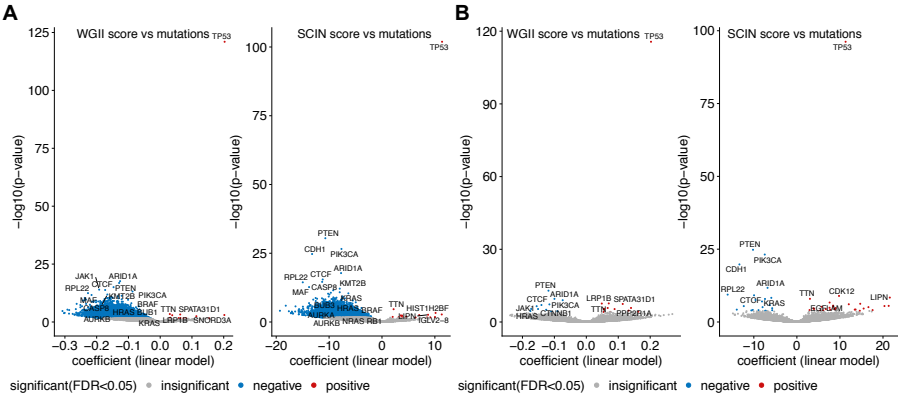


Figure 4. Pan-cancer somatic mutations and CIN. (A) The volcano plots show the association between somatic mutations and the WGII (left) or the SCIN score (right). The linear model coefficient indicates the mean difference of the respective CIN score when the mutation is present in a tumour sample relative to the wild type. Genes with lowest p-values, well-known CIN genes and cancer driver genes are highlighted. The analysis was performed on genes for which samples sizes for both wild type group and mutated group are larger than 19. Mutations significantly associated (FDR $< 5\%$) with higher or lower CIN score are highlighted in blue and red, respectively. **(B)** The same as **(A)**, but hypermutated MIN samples are excluded.

Contrary to the enrichment of *TP53* mutation 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 (**Figure 4(A)**). A similar negative correlation between the frequencies of recurrent copy number alterations and somatic mutations has previously been reported [52]. Later it was realised that this negative relationship can be reversed, when the confounding effect of MIN [21, 27] is removed. When we exclude these hypermutated samples, we observe a more even distribution between genes more or less frequently mutated in high CIN compared to low CIN tumours (**Figure 4(B)**). This is also consistent with **Figure 1(C, D)**, where we found that neither the silent mutation rate nor the non-silent mutation rate is 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 under-represented in high CIN bulk tumours (**Figure 4(B)**) and high CIN cancer cell lines (**Figure A6(D)**). *HRAS* and *JAK1* mutations are less frequent in tumours with high WGII and *KRAS* mutations are under-represented in samples with high S-CIN score. More remarkably, when only considering validated cancer driver somatic mutations [53], the above observed relationship between *PI3KCA* mutation, *PTEN* mutation and CIN still holds (**Figure A6(C)**). The under-representation 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.

4.5. Copy number gains and losses associated with CIN

Given that somatic mutations of many genes are underrepresented in high CIN tumours, we next investigated copy number alterations which are specifically linked to CIN (**Figure 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 CIN. *PVT1* is involved in the regulation of *MYC* [54] and carries a *TP53*-binding site. In addition, we found copy number gains for genes encoding members of the *WFDC-EPPIN* family, which have been linked to proliferation, metastasis, apoptosis and invasion in ovarian cancer (reviewed in [55]).

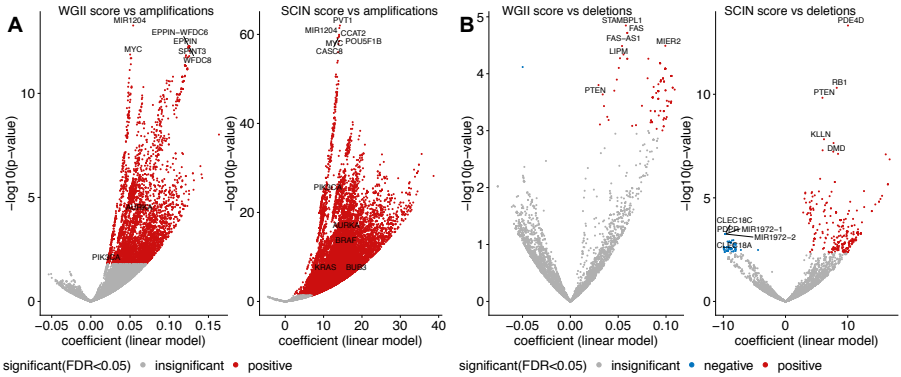


Figure 5. Copy number amplifications and deletions enriched in high CIN samples. (A) The volcano plots show the gene wise associations between copy number amplification status and WGII (left) and SCIN score (right), obtained from a regression model adjusted by cancer type. The linear model coefficient indicates the mean difference of the respective CIN score when the alteration is present in a tumour sample relative to the wild type. Genes with the lowest p-values and well-known CIN genes are highlighted. Blue and red colours encode genes with a significantly higher alteration frequency ($FDR < 0.05$) in samples with low and high CIN scores, respectively. The analysis was performed on 122222 genes with sample sizes greater than 19 for both wild type and amplified groups. **(B)** Pan-cancer copy number deletions associated with SCIN score displayed in an analogous way as in **(A)**.

Genes specifically lost in tumour samples with high WGII include *STAMBPL1*, which is linked to the *TGF-β* pathway, *MIER2* and *FAS* (**Figure 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 strongly associated with SCIN score are *PDE4D*, *RB1* and *PTEN*. The tumour suppressor *RB1* is a key regulator of the G1/S transition of the cell cycle and is required for the stabilisation of heterochromatin. *PTEN* loss is also frequent in tumours with high levels of WGII.

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

In section 4.4 we observed that somatic point mutations of *PTEN* and *PIK3CA* were scarce in high CIN tumours. In addition, copy number amplification of *PIK3CA* and copy number loss of *PTEN* are 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 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 **Figure 6(A)** displays tumour samples from all 33 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 SCIN score.

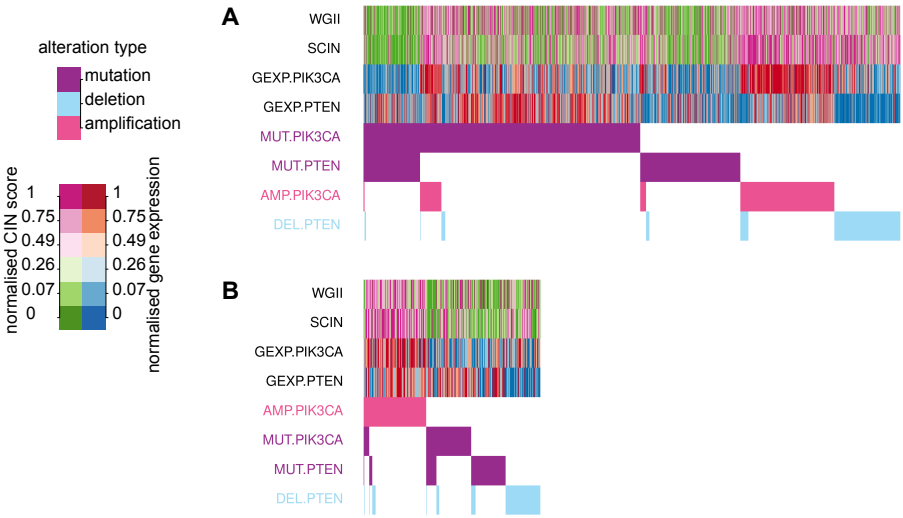


Figure 6. Oncoprint for *PIK3CA* and *PTEN* in relation to CIN. (A) The bottom panel depicts the presence or absence of somatic mutations, copy number amplifications of *PIK3CA* and deletions of *PTEN* in TCGA tumour samples (columns). Alterations are sorted by their frequency. The upper panel shows the WGII, 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.

To check, whether this effect is preserved in pure cancer cells, we used cell line data from CCLE and found a very similar pattern. Copy number gains of *PIK3CA* are linked to high levels of its gene expression, do rarely co-occur with somatic mutations, but are associated with high levels of W-CIN and S-CIN.

Taken together, we suggest a gene dosage effect on *PI3K* pathway activity, which is facilitated in high CIN tumours. This effect is cancer cell intrinsic, because it can also be observed in cancer cell lines.

5. Discussion

W-CIN and S-CIN are two distinct but related phenotypes triggered by different biological mechanisms and leading to diverse consequences. A large majority of pan-cancer association studies have focused on CIN in general or exclusively on W-CIN. Here we presented an integrative statistical analysis for 33 cancer types distinguishing between W-CIN and S-CIN. We used the WGII 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 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 [12, 56, 57]. The results of our analysis 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 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, 58] 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 of *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 [59].

We find that WGII and S-CIN score are associated with poor prognosis in different cancer types. Only in the case of ovarian cancer we find that high S-CIN patients have a slightly longer overall survival, but the difference is very small and at the edge of statistical significance. In addition, we observe slightly higher levels of the WGII 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 [15, 16] 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 to radiotherapy, where improved sensitivity against radiotherapy in transplanted human glioblastoma tumours has been reported [60].

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, 26]. However, as a new contribution we filtered small molecules and compounds of 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 and corresponding PARADIGM pathway features [30]. One interesting new finding is the chemotherapeutic drug resistance inducing gene *CKS1B* [50], which is strongly associated with both W-CIN and S-CIN. *CKS1B* is cell cycle progression gene, which is discussed as a new drug target. Here we show, that *CKS1B* is overexpressed in W-CIN and S-CIN tumours, which might be important for the

stratification of patients. We also note that the activity of the replication 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 W-CIN and S-CIN are strongly correlated with somatic point mutation of *TP53*. We find many copy number gains of important onogenes and loss of tumour suppressor genes [61] are strongly associated with W-CIN and S-CIN. Most strikingly, copy number gains of the oncogene *PIK3CA* and deletion of the tumour suppressor gene *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 W-CIN and S-CIN. Intriguingly, it has recently been reported that mutations in *PIK3CA* increased *in vitro* cellular tolerance to spontaneous genome doubling [62]. 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 *PI3K* signalling was observed in both bulk tumours and 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.

6. Conclusions

In summary, our pan-cancer analysis provides insights into the distinct and common molecular, prognostic and therapeutic characteristics of W-CIN and S-CIN. Our results suggest that whole genome doubling and homologous recombination deficiency might be the most important drivers for W-CIN and S-CIN, respectively. The predictive value of W-CIN and S-CIN depends on the cancer type. We report that most of the existing compounds preferrably kill low CIN cells, but we also suggest a few compounds with increased efficiency in high CIN cells. High activity of *CKS1B* might be a promising CIN target, because its expression is linked to both high W-CIN and high S-CIN. We propose a new copy number dependent mechanism for an increased activity of the oncogenic *PI3K* pathway in high CIN cancer cells, which merits experimental investigation.

Supplementary Materials: The following are available online. Table S1: the relationship between drug sensitivity and W-CIN, Table S2: the relationship between drug sensitivity and S-CIN.

Author Contributions: X.Z.: methodology, formal analysis, investigation, software, data curation, writing, review and editing, visualization; M.K.: conceptualization, methodology, formal analysis, investigation, writing, review and editing, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

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

Data Availability Statement: The source code and data used to reproduce this work are available at https://github.com/mcmzxx/pancan_cin.

Acknowledgments: This study used data generated by TCGA Research Network: https://www.cancer.gov/tcga and Broad-Novartis CCLE: https://sites.broadinstitute.org/ccle

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CIN	Chromosomal instability
W-CIN	Whole chromosome instability
S-CIN	Structural chromosomal instability
WGD	Whole genome doubling
WGII	Whole genome integrity index
SCIN score	Structural chromosomal instability score
AS	Aneuploidy score
HRD	Homologous recombination deficiency
SMR	Silent mutation rate
NSMR	Non-silent mutation rate
ITH	Intra-tumour heterogeneity
MIN	Microsatellite instability
AUC	Area under the dose response curve

Appendix A

Table A1. Association between W-CIN and overall survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
LGG	509	0.00	0.76	0.48	38.00	27.00
UCEC	518	0.00	0.88	0.65	65.00	42.00
LIHC	366	0.00	0.54	0.44	24.00	17.00
BRCA	1066	0.00	0.87	0.77	133.00	116.00
OV	558	0.01	0.40	0.29	62.00	54.00
SARC	252	0.03	0.59	0.50	30.00	26.00
TGCT	133	0.04	1.00	0.95	28.00	23.00
HNSC	516	0.05	0.56	0.40	24.00	28.00
UVM*	80	0.07	0.94	0.92	33.00	31.00
THCA	497	0.09	0.97	0.89	51.00	45.00
THYM	122	0.09	0.98	0.88	18.00	15.00
KIRP	282	0.10	0.84	0.68	34.00	19.00
PCPG	161	0.14	0.98	0.95	16.00	12.00
GBM	571	0.17	0.07	0.06	10.00	9.00
ACC	89	0.18	0.69	0.56	14.00	14.00
DLBC	48	0.19	0.72	0.90	6.00	3.00
STAD	433	0.25	0.41	0.35	10.00	8.00
CESC	294	0.27	0.73	0.60	25.00	16.00
LAML	179	0.34	0.24	0.17	6.00	3.00
PRAD	489	0.41	0.99	0.97	35.00	49.00
KIRC	506	0.44	0.62	0.63	66.00	81.00
CHOL*	36	0.45	0.71	0.88	12.00	15.00
READ*	154	0.48	0.94	0.96	51.00	64.00
ESCA*	182	0.49	0.82	0.70	65.00	50.00
MESO*	86	0.59	0.65	0.71	28.00	28.00
LUAD	491	0.63	0.42	0.40	28.00	25.00
COAD	425	0.75	0.66	0.54	25.00	15.00
KICH	65	0.77	0.86	0.85	17.00	20.00
SKCM*	104	0.77	0.84	0.91	33.00	37.00
PAAD	183	0.85	0.16	0.33	3.00	5.00
BLCA	405	0.87	0.42	0.42	24.00	23.00
UCS*	56	0.88	0.79	0.81	22.00	20.00
LUSC	481	0.96	0.50	0.45	38.00	42.00

[a] 5-year overall survival probability in low W-CIN group.
[b] 5-year overall survival probability in high W-CIN group.
[c] number of samples at risk in low W-CIN group at 5th year.
[d] number of samples at risk in high W-CIN group at 5th year.
* 1-year overall survival statistics was reported in these cancer types due to short survival.

Table A2. Association between S-CIN and overall survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
UCEC	518	0.00	0.90	0.60	69.00	38.00
ACC	89	0.00	0.81	0.39	21.00	7.00
THCA	497	0.00	0.96	0.72	86.00	10.00
SARC	252	0.00	0.62	0.47	31.00	25.00
KIRP	282	0.00	0.86	0.65	30.00	23.00
THYM	122	0.01	0.98	0.76	29.00	4.00
PCPG	161	0.01	1.00	0.92	15.00	13.00
LGG	509	0.01	0.73	0.52	32.00	33.00
KICH	65	0.02	0.97	0.70	23.00	14.00
COAD	425	0.02	0.71	0.52	23.00	17.00
OV	558	0.04	0.32	0.37	58.00	58.00
ESCA*	182	0.06	0.81	0.72	60.00	55.00
LUAD	491	0.08	0.48	0.35	28.00	25.00
TGCT	133	0.09	0.95	1.00	25.00	26.00
UCS*	56	0.09	0.75	0.85	20.00	22.00
READ*	154	0.10	0.96	0.93	64.00	51.00
LAML*	179	0.10	0.58	0.45	66.00	22.00
PAAD	183	0.11	0.25	0.26	5.00	3.00
KIRC	506	0.17	0.65	0.59	83.00	64.00
BRCA	1066	0.22	0.83	0.81	129.00	120.00
GBM	571	0.23	0.05	0.08	7.00	12.00
CHOL*	36	0.39	0.76	0.83	13.00	14.00
PRAD	489	0.45	0.99	0.97	40.00	44.00
BLCA	405	0.46	0.40	0.44	21.00	26.00
UVM*	80	0.50	0.93	0.93	38.00	26.00
HNSC	516	0.56	0.51	0.44	27.00	25.00
LUSC	481	0.57	0.46	0.49	36.00	44.00
SKCM*	104	0.63	0.84	0.93	36.00	34.00
MESO*	86	0.64	0.62	0.75	28.00	28.00
LIHC	366	0.73	0.48	0.50	23.00	18.00
CESC	294	0.77	0.68	0.66	20.00	21.00
STAD	433	0.90	0.35	0.43	11.00	7.00
DLBC*	48	0.93	0.96	0.89	23.00	15.00

[a] 5-year overall survival probability in low S-CIN group.
[b] 5-year overall survival probability in high S-CIN group.
[c] number of samples at risk in low S-CIN group at 5th year.
[d] number of samples at risk in high S-CIN group at 5th year.
* 1-year overall survival statistics was reported in these cancer types due to short survival.

Table A3. Association between W-CIN and disease free survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
UCEC	406	0.00	0.90	0.72	55.00	28.00
OV	279	0.00	0.25	0.10	17.00	8.00
PRAD	332	0.02	0.87	0.78	24.00	27.00
LIHC	315	0.03	0.41	0.22	14.00	4.00
COAD	175	0.05	0.83	0.62	11.00	3.00
LGG*	130	0.06	0.97	0.98	61.00	41.00
LUSC	295	0.08	0.73	0.63	24.00	24.00
MESO*	15	0.08	0.67	1.00	6.00	2.00
CESC	170	0.11	0.85	0.76	17.00	9.00
KICH	29	0.16	0.91	1.00	5.00	12.00
BRCA	927	0.25	0.86	0.82	105.00	85.00
CHOL*	24	0.26	0.71	0.50	10.00	4.00
DLBC	28	0.28	1.00	0.92	5.00	3.00
UCS*	26	0.28	1.00	0.77	11.00	9.00
PAAD*	68	0.35	0.86	0.80	21.00	21.00
KIRC	107	0.38	0.92	0.76	19.00	19.00
TGCT	104	0.39	0.68	0.82	9.00	15.00
SARC	148	0.44	0.56	0.45	16.00	11.00
GBM*	3	0.48	1.00	1.00	1.00	2.00
KIRP	180	0.52	0.74	0.89	17.00	14.00
READ*	42	0.53	0.90	1.00	15.00	20.00
LUAD	291	0.59	0.63	0.54	19.00	18.00
THCA	352	0.76	0.91	0.90	36.00	33.00
ESCA*	87	0.80	0.75	0.82	28.00	22.00
ACC	52	0.83	0.68	0.72	11.00	10.00
BLCA	187	0.88	0.72	0.71	11.00	15.00
HNSC	130	0.94	0.69	0.54	7.00	4.00
STAD	255	0.96	0.59	0.70	7.00	6.00
PCPG	144	0.97	0.95	0.97	12.00	10.00

[a] 5-year disease free survival probability in low W-CIN group.
[b] 5-year disease free survival probability in high W-CIN group.
[c] number of samples at risk in low W-CIN group at 5th year.
[d] number of samples at risk in high W-CIN group at 5th year.
* 1-year disease-free survival statistics was reported in these cancer types due to short survival.

Table A4. Association between S-CIN and disease free survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
UCEC	406	0.00	0.92	0.65	61.00	22.00
ACC	52	0.00	0.81	0.48	16.00	5.00
PRAD	332	0.00	0.88	0.75	27.00	24.00
KIRP	180	0.02	0.88	0.72	18.00	13.00
LUSC	295	0.03	0.71	0.64	25.00	23.00
THCA	352	0.03	0.92	0.81	64.00	5.00
KIRC	107	0.04	0.94	0.70	24.00	14.00
READ*	42	0.09	1.00	0.89	20.00	15.00
BRCA	927	0.10	0.88	0.80	100.00	90.00
PAAD*	68	0.14	0.91	0.74	23.00	19.00
LIHC	315	0.15	0.33	0.31	12.00	6.00
GBM*	3	0.16	1.00	1.00	1.00	2.00
KICH	29	0.18	0.94	1.00	11.00	6.00
COAD	175	0.27	0.79	0.71	9.00	5.00
CESC	170	0.30	0.86	0.75	15.00	11.00
HNSC	130	0.31	0.66	0.59	8.00	3.00
LUAD	291	0.37	0.63	0.54	19.00	18.00
DLBC*	28	0.45	1.00	1.00	14.00	11.00
MESO*	15	0.46	0.69	1.00	6.00	2.00
ESCA*	87	0.49	0.80	0.76	26.00	24.00
PCPG	144	0.49	0.93	1.00	12.00	10.00
LGG	130	0.61	0.76	0.65	3.00	3.00
OV	279	0.64	0.22	0.14	17.00	8.00
CHOL*	24	0.65	0.67	0.57	10.00	4.00
BLCA	187	0.69	0.73	0.69	14.00	12.00
TGCT	104	0.70	0.75	0.75	14.00	10.00
STAD	255	0.76	0.67	0.64	8.00	5.00
UCS*	26	0.76	0.92	0.83	10.00	10.00
SARC	148	0.79	0.53	0.48	14.00	13.00

[a] 5-year disease free survival probability in low S-CIN group.
[b] 5-year disease free survival probability in high S-CIN group.
[c] number of samples at risk in low S-CIN group at 5th year.
[d] number of samples at risk in high S-CIN group at 5th year.
[*] 1-year disease-free survival statistics was reported in these cancer types due to short survival.

Table A5. Association between W-CIN and progression free survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
UCEC	518	0.00	0.85	0.56	60.00	34.00
LGG	509	0.00	0.55	0.28	26.00	10.00
PRAD	489	0.00	0.75	0.66	27.00	36.00
OV	558	0.00	0.18	0.08	24.00	11.00
LIHC	366	0.00	0.35	0.19	13.00	5.00
UVM*	79	0.01	0.94	0.69	31.00	24.00
THYM	122	0.01	0.91	0.65	17.00	12.00
SKCM*	104	0.03	0.78	0.64	28.00	24.00
CHOL*	36	0.06	0.63	0.34	10.00	5.00
KIRC	504	0.06	0.68	0.59	55.00	63.00
CESC	294	0.06	0.71	0.62	22.00	14.00
ESCA*	182	0.07	0.66	0.58	47.00	37.00
ACC	89	0.08	0.50	0.39	11.00	10.00
SARC	252	0.14	0.44	0.34	20.00	15.00
HNSC	516	0.20	0.51	0.46	20.00	24.00
BRCA	1066	0.22	0.80	0.76	119.00	102.00
BLCA	406	0.29	0.38	0.44	19.00	17.00
READ*	154	0.32	0.90	0.88	47.00	57.00
GBM	571	0.33	0.04	0.02	5.00	3.00
TGCT	133	0.43	0.69	0.80	17.00	20.00
COAD	425	0.43	0.63	0.55	21.00	9.00
DLBC	48	0.43	0.70	0.77	6.00	3.00
KIRP	281	0.48	0.72	0.79	28.00	17.00
LUSC	482	0.53	0.55	0.54	31.00	34.00
LUAD	491	0.68	0.39	0.39	20.00	18.00
THCA	497	0.74	0.85	0.83	42.00	40.00
KICH	65	0.75	0.87	0.87	17.00	20.00
STAD	435	0.82	0.40	0.48	10.00	8.00
PCPG	161	0.83	0.80	0.87	11.00	11.00
PAAD*	183	0.85	0.62	0.64	42.00	50.00
UCS*	56	0.92	0.50	0.58	14.00	14.00
MESO*	84	0.97	0.59	0.52	22.00	19.00

[a] 5-year progression free survival probability in low W-CIN group.
[b] 5-year progression free survival probability in high W-CIN group.
[c] number of samples at risk in low W-CIN group at 5th year.
[d] number of samples at risk in high W-CIN group at 5th year.
[*] 1-year progression-free survival statistics was reported in these cancer types due to short survival.

Table A6. Association between S-CIN and progression free survival across cancer types

cohort	sample_number	pvalue	low_surv5[a]	high_surv5[b]	low_surv5_n[c]	high_surv5_n[d]
UCEC	518	0.00	0.86	0.50	65.00	29.00
ACC	89	0.00	0.66	0.22	16.00	5.00
KIRP	281	0.00	0.86	0.61	27.00	18.00
THCA	497	0.00	0.87	0.64	75.00	7.00
PRAD	489	0.00	0.76	0.65	32.00	31.00
LGG	509	0.00	0.50	0.33	18.00	18.00
CHOL*	36	0.02	0.69	0.27	11.00	4.00
THYM	122	0.03	0.85	0.60	26.00	3.00
GBM*	571	0.06	0.29	0.32	73.00	70.00
COAD	425	0.07	0.65	0.53	19.00	11.00
SARC	252	0.08	0.44	0.34	19.00	16.00
PAAD*	183	0.09	0.67	0.59	46.00	46.00
KICH	65	0.09	0.97	0.72	23.00	14.00
KIRC	504	0.11	0.66	0.60	67.00	51.00
ESCA*	182	0.12	0.61	0.63	41.00	43.00
LIHC	366	0.12	0.26	0.27	11.00	7.00
BRCA	1066	0.12	0.81	0.75	118.00	103.00
READ*	154	0.12	0.93	0.84	59.00	45.00
UVM*	79	0.21	0.82	0.80	32.00	23.00
OV	558	0.25	0.14	0.13	22.00	13.00
BLCA	406	0.31	0.44	0.39	16.00	20.00
SKCM*	104	0.31	0.74	0.67	28.00	24.00
LUAD	491	0.35	0.40	0.38	19.00	19.00
UCS*	56	0.41	0.54	0.54	14.00	14.00
CESC	294	0.41	0.70	0.63	19.00	17.00
STAD	435	0.41	0.45	0.44	11.00	7.00
PCPG	161	0.45	0.84	0.84	13.00	9.00
DLBC*	48	0.51	0.83	0.84	19.00	15.00
MESO*	84	0.73	0.63	0.47	24.00	17.00
TGCT	133	0.77	0.72	0.77	19.00	18.00
LUSC	482	0.79	0.51	0.57	29.00	36.00
HNSC	516	0.93	0.49	0.49	23.00	21.00

[a] 5-year progression free survival probability in low S-CIN group.
[b] 5-year progression free survival probability in high S-CIN group.
[c] number of samples at risk in low S-CIN group at 5th year.
[d] number of samples at risk in high S-CIN group at 5th year.
[*] 1-year progression-free survival statistics was reported in these cancer types due to short survival.

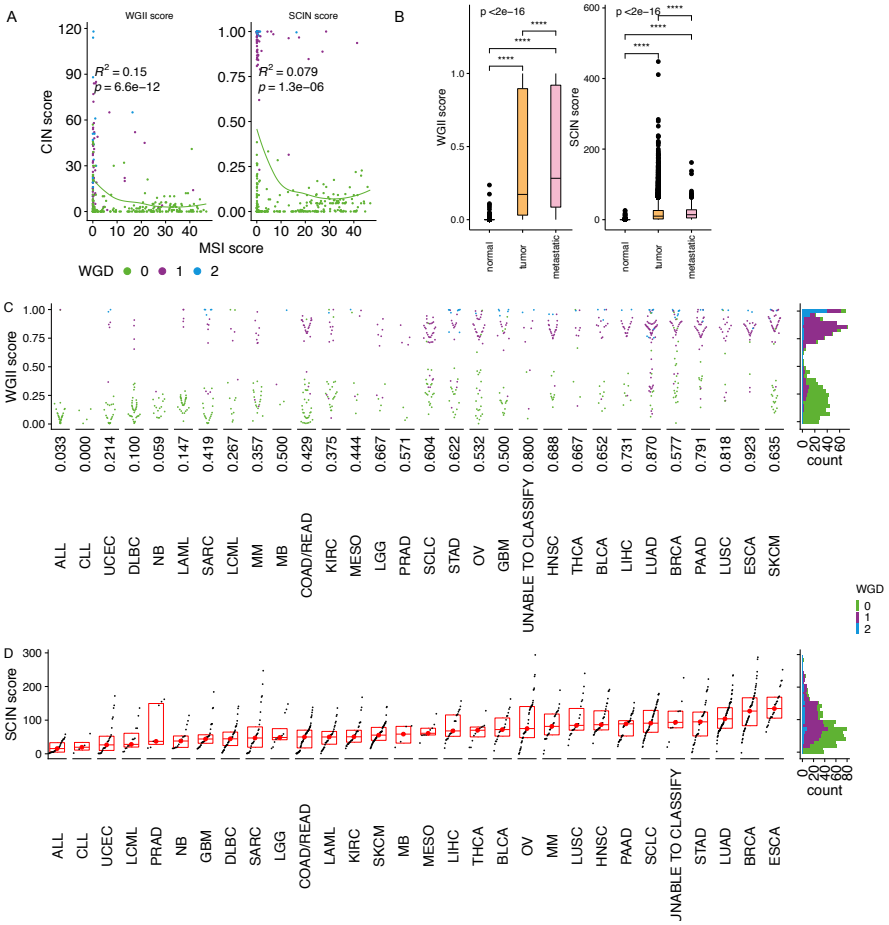


Figure A1. (A) The relationship between microsatellite instability (MIN) scores and the WGII and the SCIN score. **(B)** Comparison of the WGII and the SCIN score between normal samples with primary and metastatic tumour samples. **(C)** Cancer-type-wise WGII distribution in CCLE cell lines, cancer types are ordered by the median WGII; whole genome doubling (WGD) status is encoded by colours. The number reported on x axis is the proportion of samples that underwent WGD. **(D)** SCIN score distribution in CCLE cell lines, cancer types are ordered by their median SCIN score.

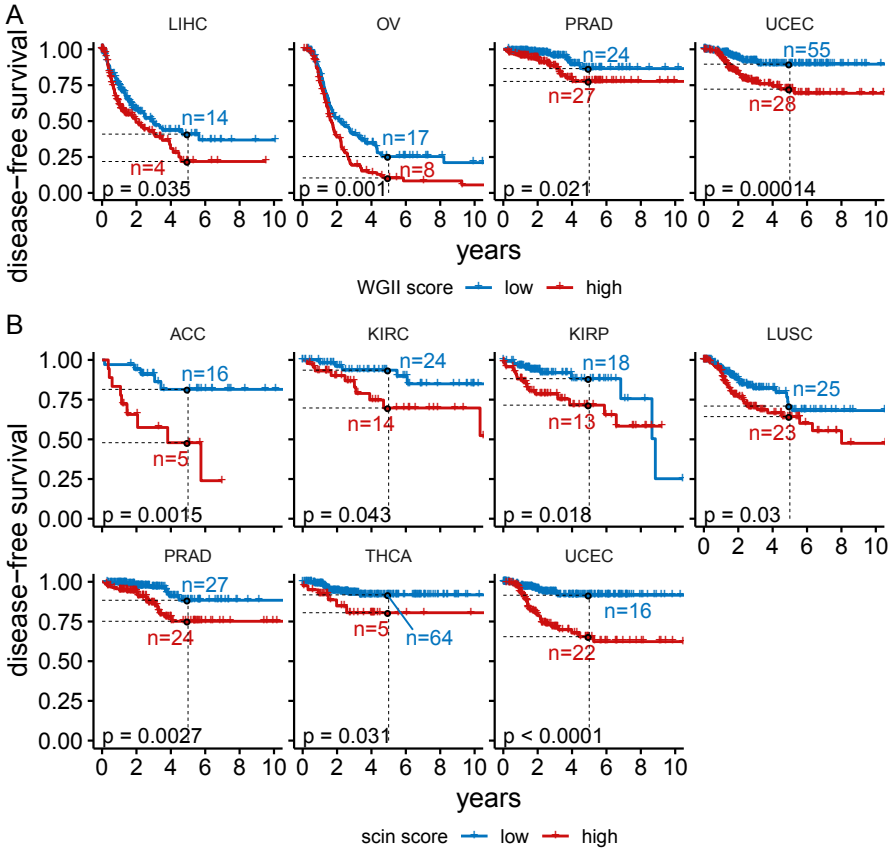


Figure A2. (A) Disease free survival in four cancer types where significant differences between high and low WGII groups were observed. **(B)** Disease free survival in seven cancer types where significant differences between high and low SCIN score group were observed.

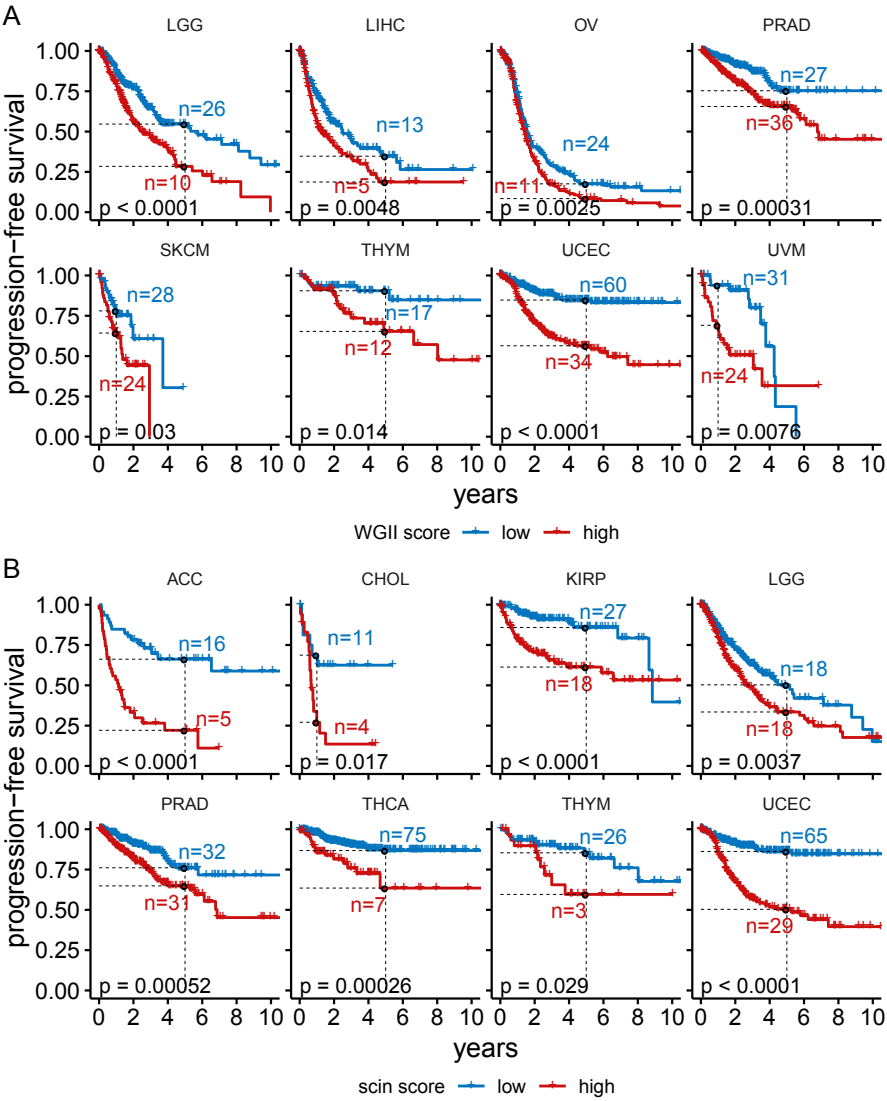


Figure A3. **(A)** Progression free survival in eight cancer types where significant differences between high and low WGII group were observed. **(B)** Progression free survival in eight cancer types where significant differences between high and low SCIN score group were observed.

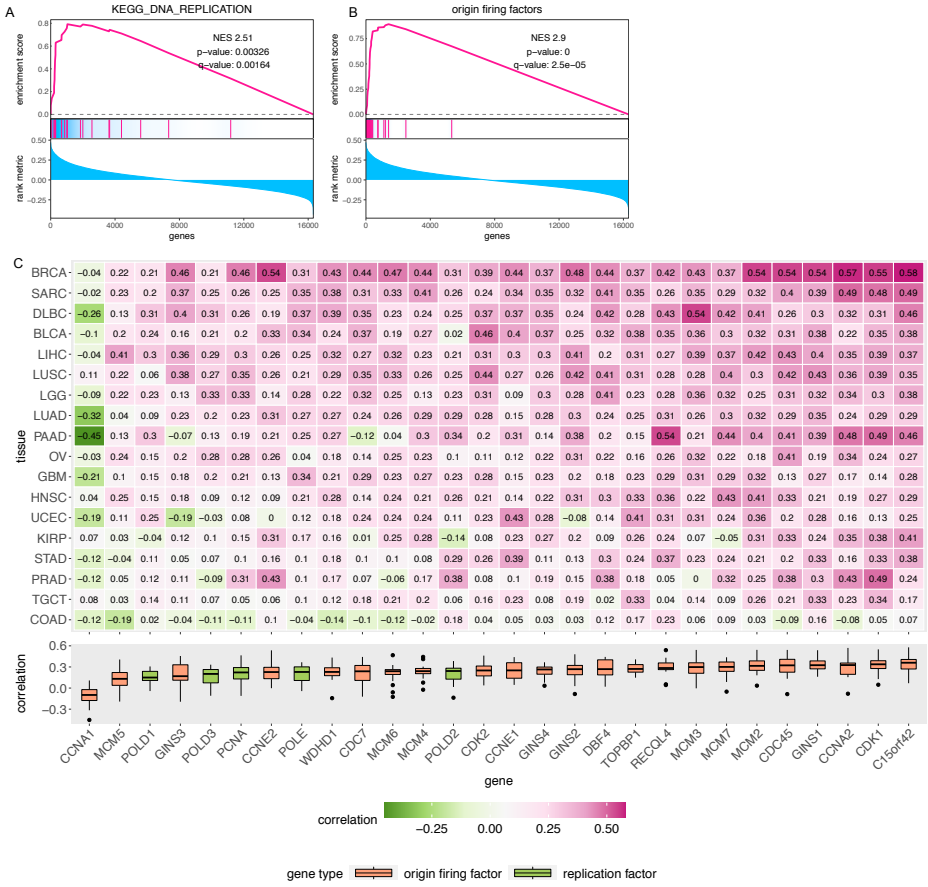


Figure A4. (A) Gene set enrichment analysis (GSEA) [63] for KEGG [64] DNA replication gene set. All genes were ordered according to the correlation of their expression with the SCIN score and enrichment significance is evaluated using permutation test. **(B)** GSEA analysis for manually curated origin firing factor gene set(curated in [10]). **(C)** Gene expression of many origin firing factors is positively correlated with SCIN scores in many cancer types. Rows and columns of the heatmap represent cancer types and origin firing factor genes, respectively. Cancer types are clustered based on their correlation coefficients with origin firing factors. Genes are ordered based on the median correlation coefficient. Colour and values encoded in the heatmap represent the Spearmann correlation coefficient.

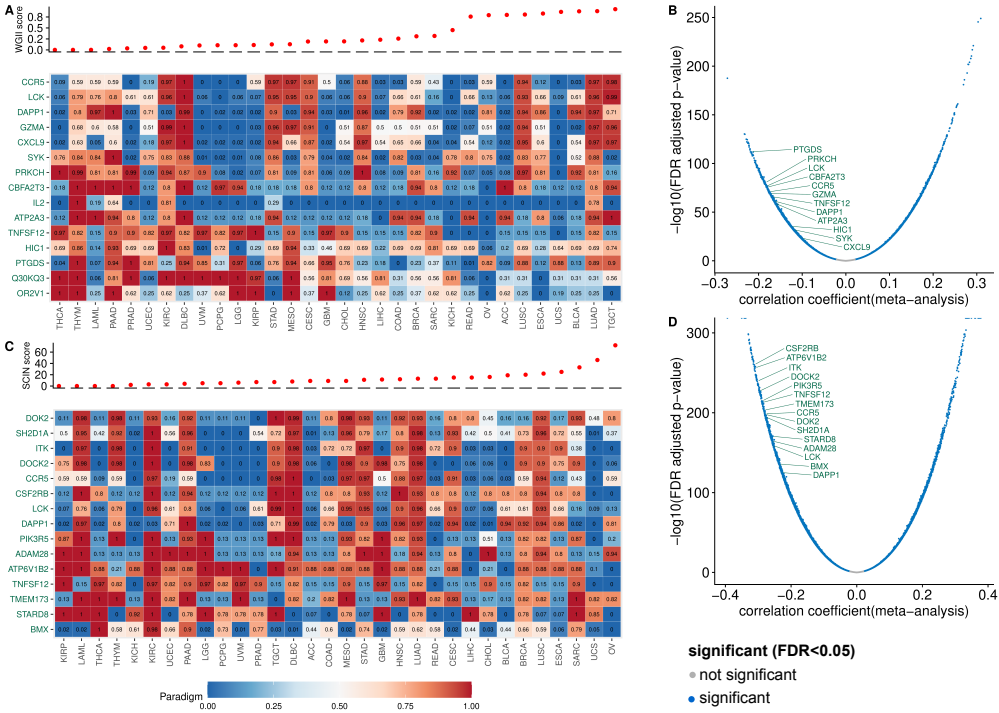


Figure A5. Paradigm pathway activity and gene expression negatively associated with CIN. (A) The PARADIGM pathway-level activities corresponding to protein-coding genes (rows) were correlated with the WGII. Only pathways with a significant negative correlation ($p < 5\%$) less than -0.4 in at least eight 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, see top panel. **(B)** Volcano plot for the correlation between gene expression and WGII, highlighting gene names corresponding to PARADIGM features that are significantly negatively associated with WGII. **(C)** Analogous to **(A)** , but now for the SCIN score instead of the WGII. **(D)** Correlation of SCIN scores and gene expression, analogous to **(B)**.

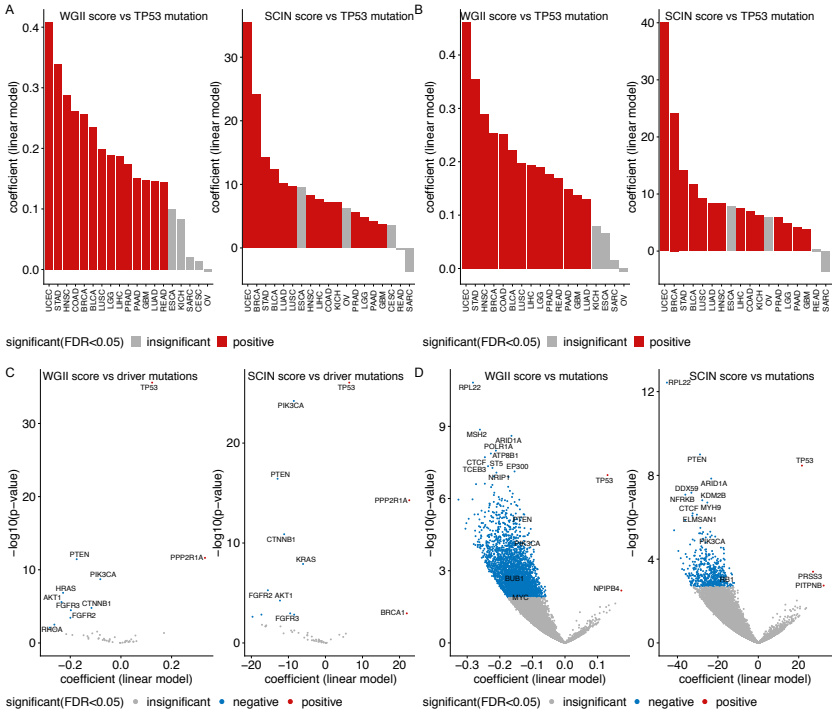


Figure A6. (A) *TP53* mutation is positively associated with high W-CIN in multiple cancer types. The bar shows the linear regression model coefficient using WGII as dependent variable and *TP53* mutation as explanatory variable. Only cancer types with both *TP53* mutant and *TP53* wild type in more than 20 samples are considered. **(B)** *TP53* mutation is positively associated with SCIN score, the association analysis is performed as in **(A)**, except excluding MIN samples. **(C)** The volcano plot shows the association between CIN and validated driver mutations, association analysis is performed using non-MIN samples only. **(D)** The volcano plot shows the correlation between CIN score and somatic mutations in CCLE cell line samples. For all association in **(A)-(D)**, red, grey and blue colors encode positive, insignificant and negative associations. $FDR > 0.05$ is considered as insignificant. Gene names of known important oncogenes and CIN driver genes are annotated in the volcano plot, if significantly associated with CIN.

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. https://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. & Rozenblum, E. Multi-layered cancer chromosomal instability phenotype. *Front. Oncol.* **3**, 302 (2013). URL https://www.frontiersin.org/article/10.3389/fonc.2013.00302. https://doi.org/10.3389/fonc.2013.00302 .

[5] Thompson, S. L., Bakhoum, S. F. & Compton, D. A. Mechanisms of chromosomal instability. *Curr. Biol.* **20**, R285–R295 (2010). URL http://www.sciencedirect.com/science/article/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**, 492–496 (2013). https://doi.org/10.1038/nature11935 .

[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. *bioRxiv* (2021). URL http://biorxiv.org/lookup/doi/10.1101/2021.10.11.463929. https://doi.org/10.1101/2021.10.11.463929 .

[11] Passerini, V. *et al.* The presence of extra chromosomes leads to genomic instability. *Nat. Commun.* **7** (1), 10754 (2016). https://doi.org/10.1038/ncomms10754 .

[12] Dewhurst, S. M. *et al.* Tolerance of Whole-Genome Doubling Propagates Chromosomal Instability and Accelerates Cancer Genome Evolution. *Cancer Discov* **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 .

[13] 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. *Clin. Cancer Res.* **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 .

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

[15] 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). https://doi.org/10.1158/1055-9965.EPI-11-0343 .

[16] 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 .

[17] 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 .

[18] Thompson, S. L. & Compton, D. A. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *J. Cell Biol.* **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 .

[19] Sheltzer, J. M. A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis. *Cancer Res.* **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 .

[20] Endesfelder, D. *et al.* Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in er+ breast cancer. *Cancer Res.* **74** (17), 4853–4863 (2014). URL https://cancerres.aacrjournals.org/content/74/17/4853. https://doi.org/10.1158/0008-5472.CAN-13-2664 .

[21] 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 .

[22] 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 (2017). https://doi.org/10.1126/science.aaf8399 .

[23] Bakhoum, S. F. *et al.* Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467–472 (2018). URL https://doi.org/10.1038/nature25432. https://doi.org/10.1038/nature25432 .

[24] Salgueiro, L. *et al.* Acquisition of chromosome instability is a mechanism to evade oncogene addiction. *EMBO Mol Med* **12** (3), e10941 (2020). URL https://onlinelibrary.wiley.com/doi/10.15252/emmm.201910941. https://doi.org/10.15252/emmm.201910941 .

[25] Lukow, D. A. *et al.* Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev. 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 .

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

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

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

[29] Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30** (5), 413–421 (2012). URL https://doi.org/10.1038/nbt.2203. https://doi.org/10.1038/nbt.2203 .

[30] 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 .

[31] Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483** (7391), 603–607 (2012). URL https://doi.org/10.1038/nature11003. https://doi.org/10.1038/nature11003 .

[32] Seashore-Ludlow, B. *et al.* Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* **5** (11), 1210–1223 (2015). URL https://cancerdiscovery.aacrjournals.org/content/5/11/1210. https://doi.org/10.1158/2159-8290.CD-15-0235 .

[33] 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/10.1016/j.immuni.2018.03.023 .

[34] 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 .

[35] Therneau, T. M. *A Package for Survival Analysis in R.* R package version 3.2-10 (2021). Available online: https://CRAN.R-project.org/package=survival (accessed on 4 October 2021).

[36] Kassambara, A., Kosinski, M. & Biecek, P. *survminer: Drawing Survival Curves using 'ggplot2'.* R package version 0.4.9 (2021). Available online: https://CRAN.R-project.org/package=survminer (accessed on 4 October 2021).

[37] Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** (7), e47 (2015). https://doi.org/10.1093/nar/gkv007 .

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

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

[40] Delaney, J. R. *et al.* Haploinsufficiency networks identify targetable patterns of allelic deficiency in low mutation ovarian cancer. *Nat. Commun.* **8** (1), 14423 (2017). URL https://doi.org/10.1038/ncomms14423. https://doi.org/10.1038/ncomms14423 .

[41] van Jaarsveld, R. H. & Kops, G. J. Difference makers: Chromosomal instability versus aneuploidy in cancer. *Trends Cancer* **2** (10), 561–571 (2016). URL https://www.sciencedirect.com/science/article/pii/S2405803316301200. https://doi.org/10.1016/j.trecan.2016.09.003 .

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

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

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

[45] Lord, C. J. & Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **16** (2), 110–120 (2016). URL https://doi.org/10.1038/nrc.2015.21. https://doi.org/10.1038/nrc.2015.21 .

[46] Turner, N., Tutt, A. & Ashworth, A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* **4** (10), 814–819 (2004). URL https://doi.org/10.1038/nrc1457. https://doi.org/10.1038/nrc1457 .

[47] Thompson, L., Jeusset, L., Lepage, C. & McManus, K. Evolving Therapeutic Strategies to Exploit Chromosome Instability in Cancer. *Cancers* **9** (12), 151 (2017). URL http://www.mdpi.com/2072-6694/9/11/151. https://doi.org/10.3390/cancers9110151 .

[48] Wu, Y.-L. *et al.* Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol.* **15** (2), 213–222 (2014). URL https://doi.org/10.1016/S1470-2045(13)70604-1. https://doi.org/10.1016/S1470-2045(13)70604-1 .

[49] 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, Dordrecht, The Netherlands, 2012). URL http://link.springer.com/10.1007/978-94-007-4572-8_8.

[50] 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 .

[51] Eischen, C. M. Genome stability requires p53. *Cold Spring Harb. Perspect. Med.* **6** (6), a026096 (2016). URL http://perspectivesinmedicine.cshlp.org/content/6/6/a026096.abstract. https://doi.org/10.1101/cshperspect.a026096 .

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

[53] Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10** (1), 25 (2018). URL https://doi.org/10.1186/s13073-018-0531-8. https://doi.org/10.1186/s13073-018-0531-8 .

[54] 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 .

[55] Zhang, C., Hu, H., Wang, X., Zhu, Y. & Jiang, M. Wfdc protein: A promising diagnosis biomarker of ovarian cancer. *J. Cancer* **12**, 5404–5412 (2021). URL https://www.jcancer.org/v12p5404.htm. https://doi.org/10.7150/jca.57880 .

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

[57] 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 .

[58] Bakhoum, S. F., Kabeche, L., Murnane, J. P., Zaki, B. I. & Compton, D. A. DNA-Damage Response during Mitosis Induces Whole-Chromosome Missegregation. *Cancer Discov* **4** (11), 1281–1289 (2014).

https://doi.org/10.1158/2159-8290.CD-14-0403 .

[59] 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 .

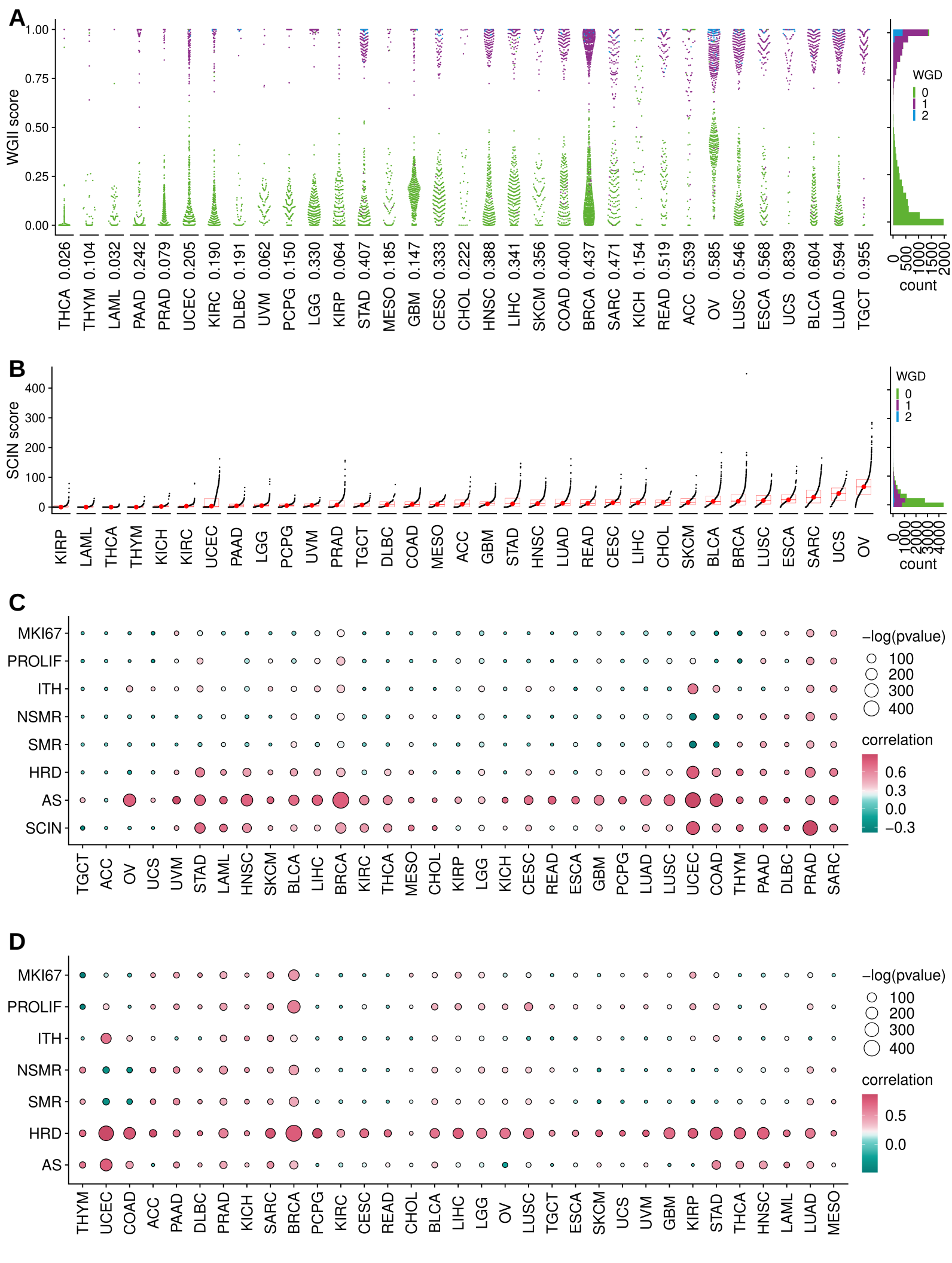
[60] Bakhoum, S. F. *et al.* Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat. Commun.* **6** (1), 5990 (2015). URL http://www.nature.com/articles/ncomms6990. https://doi.org/10.1038/ncomms6990 .

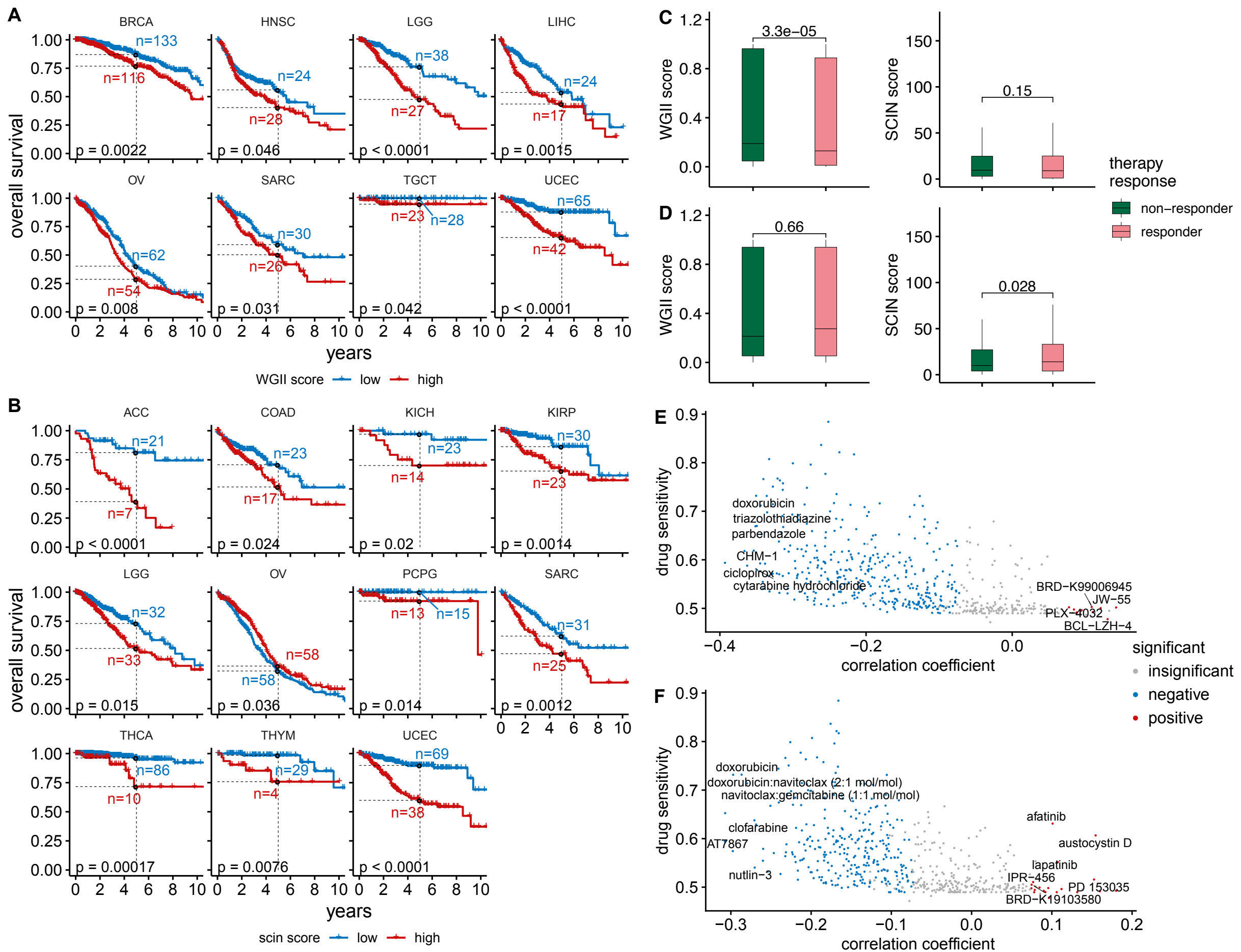
[61] Davoli, T. *et al.* Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. *Cell* **155** (4), 948–962 (2013). URL https://linkinghub.elsevier.com/retrieve/pii/S0092867413012877. https://doi.org/10.1016/j.cell.2013.10.011 .

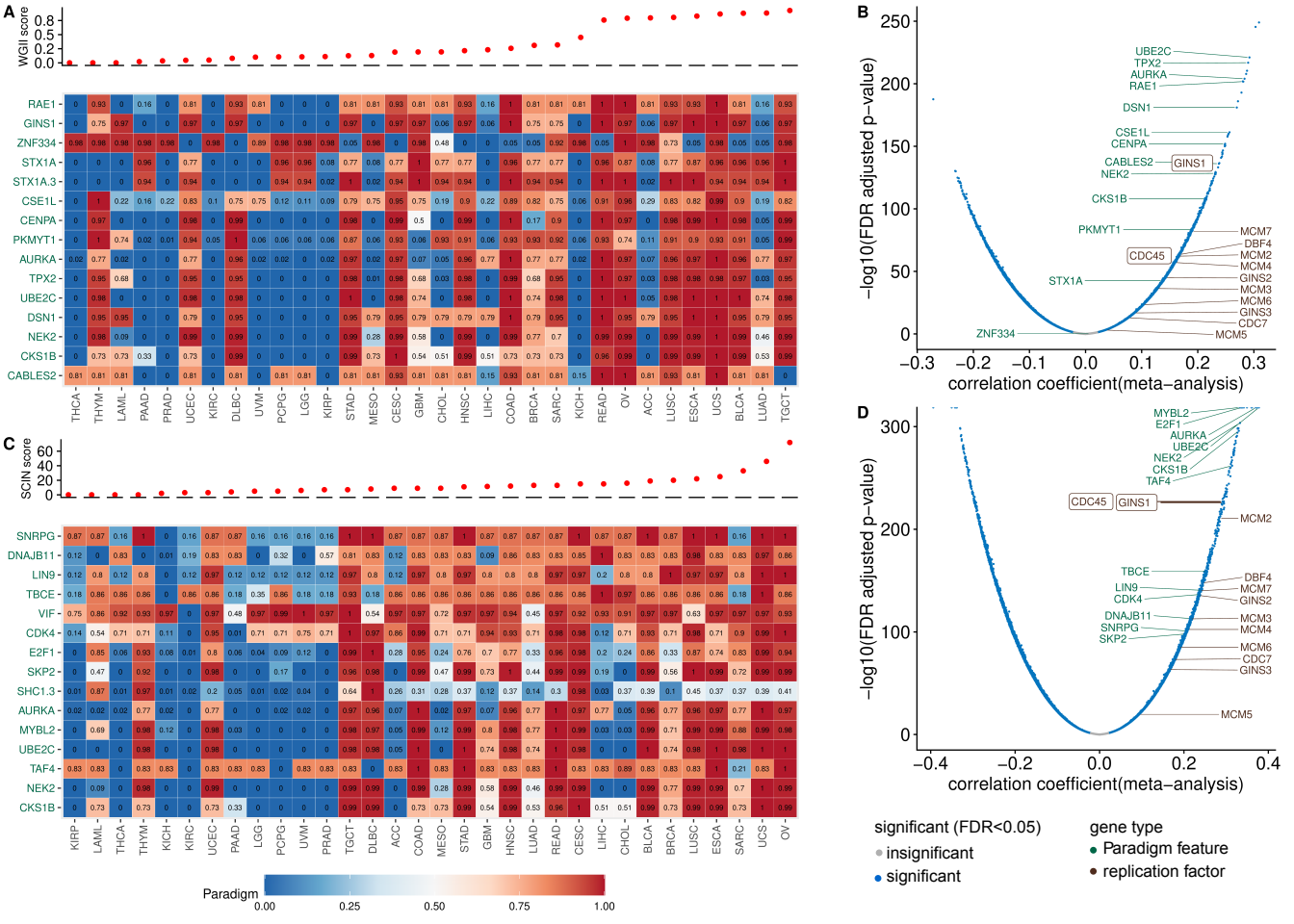
[62] Berenjeno, I. M. *et al.* Oncogenic PIK3CA induces centrosome amplification and tolerance to genome doubling. *Nat. Commun.* **8** (1), 1773 (2017). URL http://www.nature.com/articles/s41467-017-02002-4. https://doi.org/10.1038/s41467-017-02002-4 .

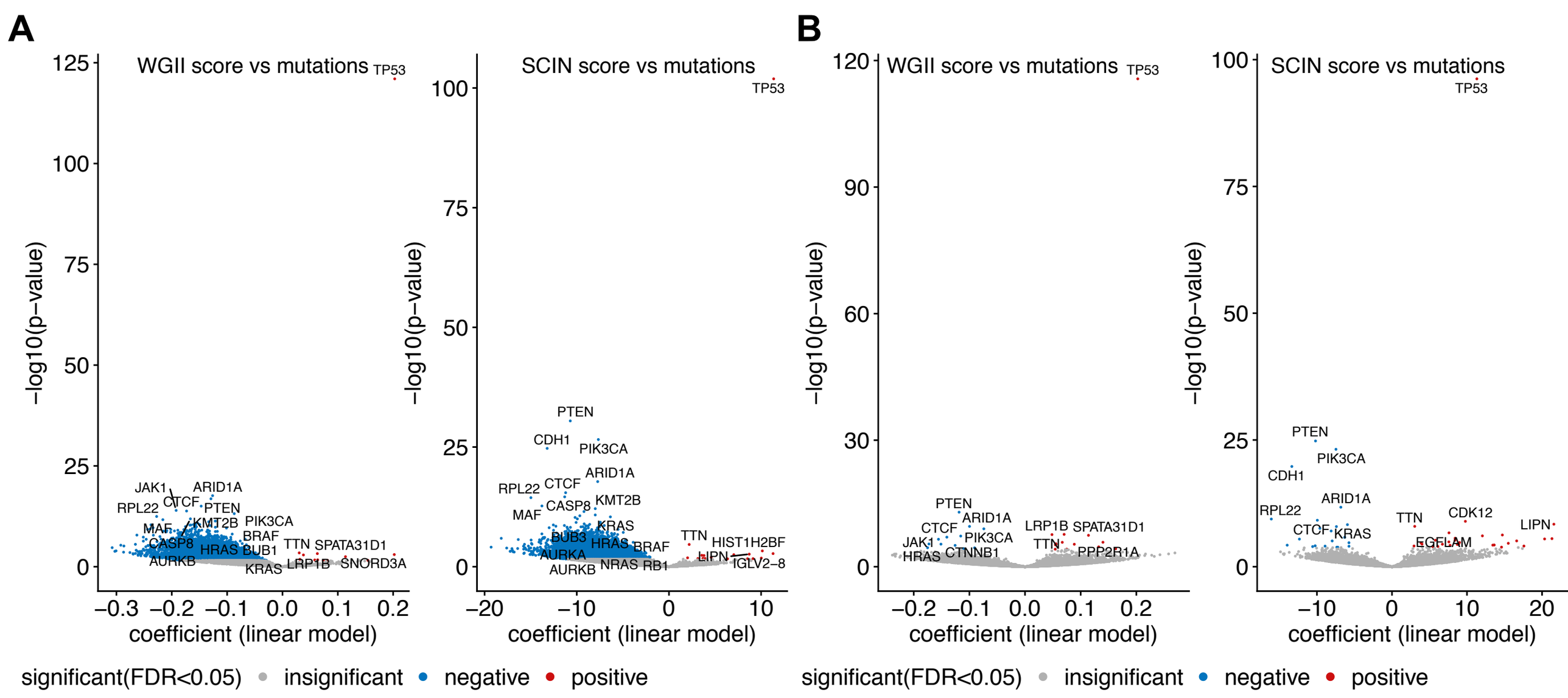
[63] Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102** (43), 15545–15550 (2005). URL https://www.pnas.org/content/102/43/15545. https://doi.org/10.1073/pnas.0506580102 .

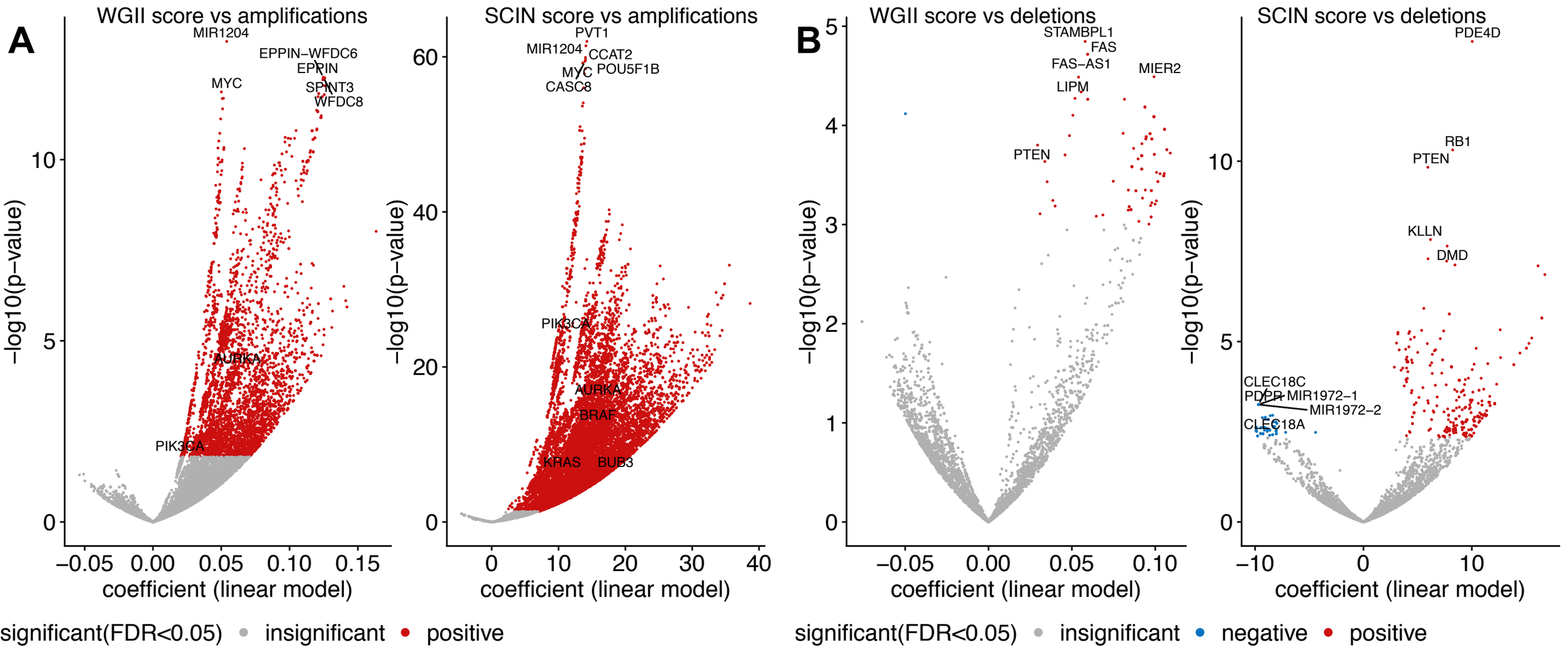
[64] Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000). URL https://doi.org/10.1093/nar/28.1.27. https://doi.org/10.1093/nar/28.1.27 .

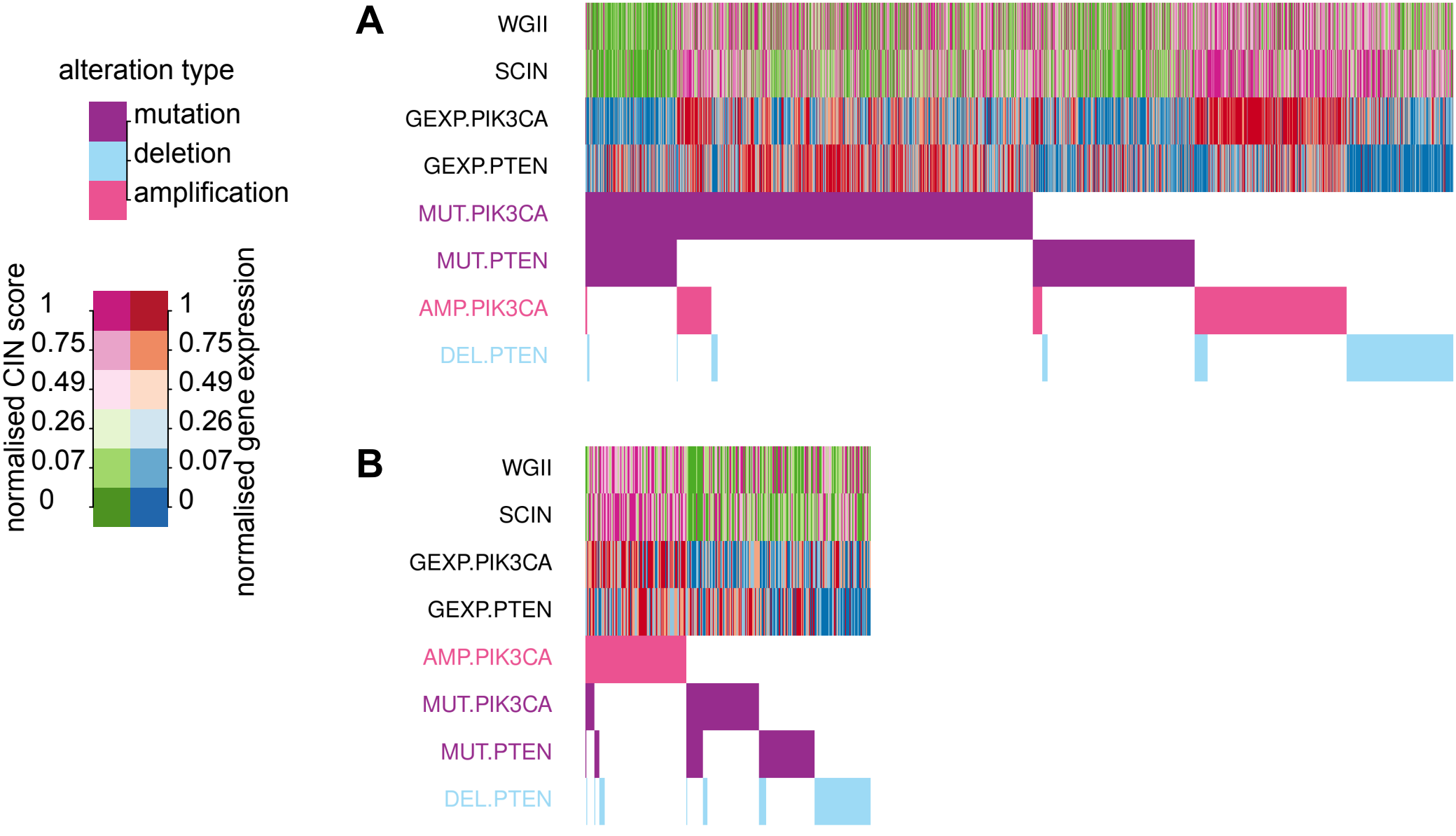

A
WGII score
1.00
0.75
0.50
0.25
0.00
THCA 0.026
THYM 0.104
LAML 0.032
PAAD 0.242
PRAD 0.079
UCEC 0.205
KIRC 0.190
DLBC 0.191
UVM 0.062
PCPG 0.150
LGG 0.330
KIRP 0.064
STAD 0.407
MESO 0.185
GBM 0.147
CESC 0.333
CHOL 0.222
HNSC 0.388
LIHC 0.341
SKCM 0.356
COAD 0.400
BRCA 0.437
SARC 0.471
KICH 0.154
READ 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

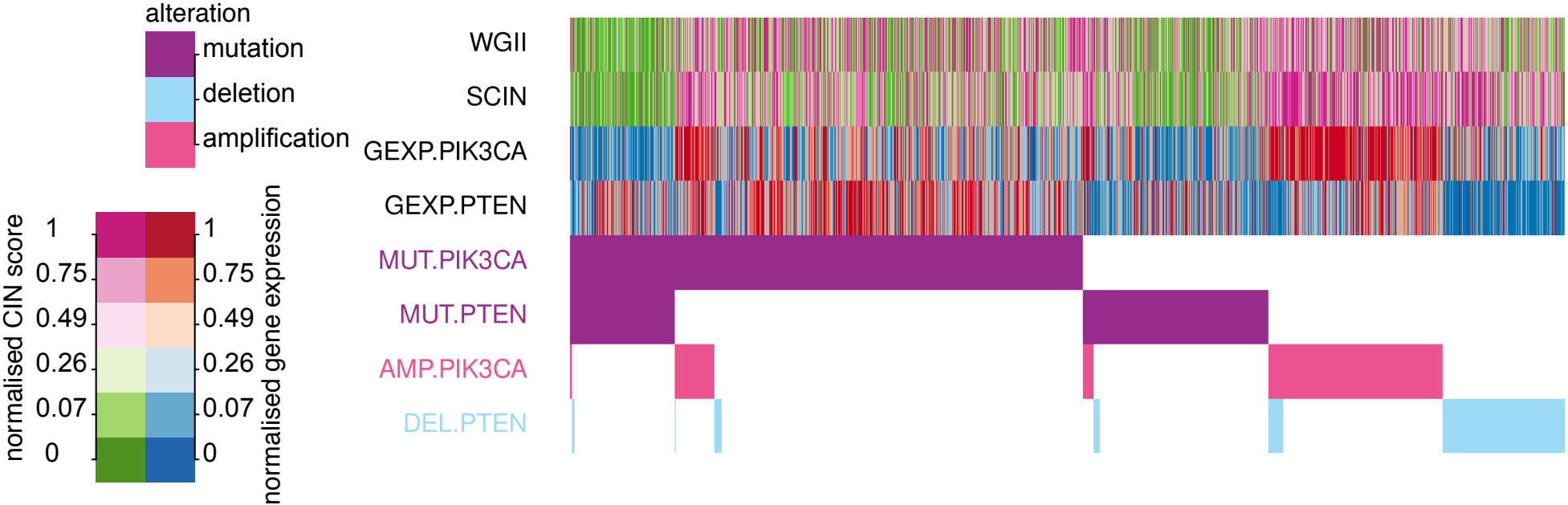

A
BRCA
HNSC
LGG
LIHC
OV
SARC
TGCT
UCEC
overall survival
years
WGII score low high
B
ACC
COAD
KICH
KIRP
LGG
OV
PCPG
SARC
THCA
THYM
UCEC
scin score low high
C
WGII score
SCIN score
3.3e−05
0.15
D
0.66
0.028
therapy response
non−responder
responder
E
drug sensitivity
correlation coefficient
doxorubicin
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


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