Title: Cell-specific expression buffering promotes cell survival and cancer robustness

Hao-Kuen Lin[a,b,1], Jen-Hao Cheng[a,c,1], Chia-Chou Wu[a,1], Feng-Shu Hsieh[a], Carolyn Dunlap[a], and Sheng-hong Chen[a,c*]

[a] Lab for Cell Dynamics, Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan

[b] College of Medicine, National Taiwan University, Taipei 106, Taiwan

[c] Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

[1] equal contribution

[*] Correspondence: shengchen@gate.sinica.edu.tw

Summary blurb

This study explores a genome-wide functional buffering mechanism, Cell-specific Expression Buffering (CEBU), where gene expression contributes to functional buffering in specific cell types and tissues. CEBU is prevalent in humans, linking to genetic interactions, tissue homeostasis and cancer robustness.

Abstract

Functional buffering ensures biological robustness critical for cell survival and physiological homeostasis in response to environmental challenges. However, in multicellular organisms, the mechanism underlying cell- and tissue-specific buffering and its implications for cancer development remain elusive. Here, we propose a Cell-specific Expression-BUffering (CEBU) mechanism, whereby a gene's function is buffered by cell-specific expression of a buffering gene, to describe functional buffering in humans. The likelihood of CEBU between gene pairs is quantified using a C-score index. By computing C-scores using genome-wide CRISPR screens and transcriptomic RNA-seq of 684 human cell lines, we report that C-score-identified putative buffering gene pairs are enriched for members of the same pathway, protein complex and duplicated gene family. Furthermore, these buffering gene pairs contribute to cell-specific genetic interactions and are indicative of tissue-specific robustness. C-score derived buffering capacities can help predict patient survival in multiple cancers. Our results reveal CEBU as a critical mechanism of functional buffering contributing to cell survival and cancer robustness in humans.

Running title: Expression buffering for cancer robustness

Keywords: functional buffering, expression buffering, buffering capacity, genetic interaction, tissue homeostasis, cancer robustness

Introduction

Robustness in biological systems is critical for organisms to carry out vital functions in the face of environmental challenges [1, 2]. A fundamental requirement for achieving biological robustness is functional buffering, whereby the biological functions performed by one gene can also be attained via other buffering genes. These buffering genes thereby enhance the robustness of the biological system. Multiple mechanisms underlie this functional buffering, either via the responsive or intrinsic activities of buffering genes. Responsive buffering mechanisms activate buffering genes only if the function of a buffered gene is compromised. Therefore, the mechanism incorporates a control system that senses-a function has been compromised and then activates the buffering genes, usually by up-regulating their gene expression [3]. One classical responsive buffering mechanism is genetic compensation among duplicated genes, whereby functional compensation is achieved by activating expression of paralogs when the functions of the original duplicated genes is compromised [4]. Genetic analyses of duplicated genes in *Saccharomyces cerevisiae* have revealed up-regulated gene expression in ~10% of paralogs when cell growth is compromised due to deletions of their duplicated genes [5-7]. Apart from duplicated genes, non-orthologous/analogous genes can also be activated for functional buffering [8]. For example, multiple growth signaling pathways coordinate cellular growth in parallel, and inactivation of one pathway can lead to activation of others for better survival [3, 4]. In cases of both duplicated and analogous genes, expression of their buffering genes is not activated unless their buffered functions are compromised [3, 7], so responsiveness follows a “need-based” principal. This responsive buffering mechanism has been observed for stress responses of both unicellular and multicellular organisms [3].

Although numerous studies over recent decades have focused on investigating such responsive buffering mechanisms, a few recent studies have revealed another less-characterized type of buffering mechanism in multicellular organisms, i.e. intrinsic buffering through cell-specific gene expression. In *Caenorhabditis elegans,* essentiality of a gene can depend on the natural variation in expression of other genes in the same pathway, suggesting that functionally analogous genes in the same pathway can buffer each other [9]. The essentiality of duplicated genes across different human cells can be associated with the expression levels of their paralogs [10-12]. Specifically, higher expression of the buffering paralogs is associated with diminished essentiality of their corresponding buffered duplicated genes, indicating that cell-specific expression of paralogs can contribute to different levels of functional buffering. Thus, expression of buffering genes, either non-duplicated or duplicated, is constitutively active even if the cell does not experience compromised function. Accordingly, this scenario represents an intrinsic buffering mechanism whereby expression of the buffering genes is constitutive in specific cells and tissue types. It is likely that this intrinsic buffering mechanism is controlled by cell- and tissue-specific transcriptional or epigenetic regulators, leading to variable essentiality across different cells and tissues.

In this study, we directly investigated if cell-specific gene expression can act as an intrinsic buffering mechanism (which we have termed “Cell-specific Expression-BUffering” or CEBU, **Fig. 1A**) to buffer functionally related genes in the genome, thereby strengthening cellular plasticity for cell- and tissue-specific tasks. To estimate buffering capability, we quantified the proposed CEBU mechanism using a quantitative index, the C-score. This index calculates the adjusted correlation between expression of a buffering gene and the essentiality of the buffered gene (**Fig. 1B**), utilizing transcriptomics data [13] and gene dependency data from the DepMap project [13, 14] across 684 human cell lines. Our results suggest that intrinsic

buffering by CEBU plays a critical role in cell-specific survival, tissue homeostasis, and cancer robustness.

Results

Development of the C-score to infer cell-specific expression buffering

In seeking an index to infer cell-specific expression buffering, we postulated a buffering relationship where the essentiality of a buffered gene (G1) increases when expression of its buffering gene (G2) decreases. Given that G2 expression differs among cells, the strength of functional buffering varies across cells, thereby conferring on G1 cell-specific essentiality (**Fig. 1A**). This cell-specific expression buffering mechanism, here named CEBU, is the basis for our development of C-score. The C-score of a gene pair is derived from the correlation between the essentiality of a buffered gene (G1) and expression of its buffering gene (G2) (see C-score plot, **Fig. 1B**), and is formulated as

$$\text{C-score} = \rho_{G1,G2}\left(1 + b\frac{slope_{min}}{slope_{G1,G2}}\right)$$

where ρ denotes the Pearson correlation coefficient between essentiality of G1 and expression of G2. Their regression slope ($slope_{G1,\,G2}$) is normalized to $slope_{min}$, which denotes the minimum slope of all considered gene pairs in the human genome (see **Methods**). The normalized slope can be weighted by cell- or tissue-type specific b. In our current analysis, b is set as 1 for a pan-cell- and pan-cancer-type analysis. Gene essentiality is represented by gene dependency data (dependency scores, D.S.) from the DepMap project [14], where the effect of each gene on cell proliferation was quantified after its knockout using the CRISPR/Cas-9 approach. Specifically, a more negative dependency reflects slower cell proliferation when the gene is knocked out, thus reflecting stronger essentiality. Expression data was obtained through RNA-seq [13]. We anticipated that the higher the C-score of a gene pair, the more likely G2 would buffer G1 based on the proposed CEBU mechanism.

We conducted a genome-wide analysis to calculate C-scores for all gene pairs across 684 human cell lines. The calculated C-score was compared to a bootstrapped null distribution generated by random shuffling of G2 expression among cell lines (**Fig. S1A**). The bootstrapped null distribution can be modeled as a normal distribution (**Fig. S1B**). For our analysis, we considered gene pairs to have a high C-score with a strong likelihood of intrinsic buffering when their C-scores were > 0.25 (0.058% of gene pairs in the human genome, significant with a *q*-value < 2.2e-16 after multiple testing correction, **Fig. S1A**). Based on our C-score definition, a high C-score should be indicative of marked variability in cell-specific essentiality and expression. Indeed, we observed higher variability in both G1 dependency and G2 expression for high C-score gene pairs (**Fig. S2A**). Nevertheless, high variation alone is insufficient to grant a high C-score. A high C-score requires consistent pairing between G1 and G2 across cell lines and, as anticipated, disrupting the pairing between G1 dependency and G2 expression by shuffling G2 expression amongst cell lines (without changing variability) abolished the C-score relationship (**Fig. S2B**). Moreover, both mean G1 dependency and mean G2 expression in high C-score gene pairs were lower than those parameters in all gene pairs (**Fig. S2C**), implying that G1s and G2s in high C-score gene pairs tend to be more essential and less expressed, respectively.

Characterization of C-score-identified gene pairs

Since several duplicated gene pairs have been implicated to have functional buffering via gene expression [10-12], we started out by characterizing the duplicated genes among C-score-identified gene pairs. We found that duplicated gene pairs are enriched among gene pairs with C-scores > 0.085 (*p*-value = 0.05 using a hypergeometric test), suggesting that CEBU is a prevalent buffering mechanism among duplicated genes. Interestingly, the

138 majority of high C-score gene pairs are non-duplicated (>90%, **Fig. 2A**). In these cases, G2s
139 can be G1s' functional analogs as surrogate genes. Accordingly, we examined if the high C-
140 score gene pairs are more likely to participate in the same function or biological pathway or
141 physically interact. We calculated the enrichment of curated gene sets in terms of Gene
142 Ontology (GO) [15] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [16] from the
143 Molecular Signatures Database [17] (**Fig. 2B**). Gene pairs with higher C-scores consistently
144 exhibited greater functional enrichment. Likewise, we observed a monotonic increase in the
145 enrichment of protein-protein interactions (PPI, STRING [18] and CORUM [19] databases)
146 between G1 and G2 in accordance with increasing C-score cutoff (**Fig. 2C**). The enriched
147 functions mostly include house-keeping functions such as nitrogen metabolism and
148 biosynthetic processes (**Fig. 2D**). Moreover, the observed pathway and PPI enrichments are
149 mostly due to the non-duplicated gene pairs since duplicated gene pairs appear to be
150 independent of C-scores (**Fig. S3A** and **S3B)**. In summary, duplicated gene pairs are enriched
151 among high C-score gene pairs and, for the non-duplicated gene pairs having high C-scores,
152 they are likely members of the same biological pathways and/or are encoding proteins with
153 physical interactions, supporting the possibility that functional buffering between these gene
154 pairs via CEBU.
155
156 **Experimental validation of C-score-inferred cell-specific expression buffering**
157 To validate putative C-score-inferred buffering gene pairs, we conducted experiments on the
158 highest C-score gene pair, i.e., *FAM50A-FAM50B*, which is a duplicated gene pair. Based on
159 the C-score plot of *FAM50A-FAM50B* (**Fig. 3A**), we expected that *FAM50B* would have a
160 stronger buffering effect for *FAM50A* in cell lines located at the top-right of the plot (e.g.
161 A549 and MCF7) relative to those at the bottom-left (e.g. U2OS). Consequently, growth of
162 the cell lines at top-right would be more sensitive to dual suppression of *FAM50A* and

FAM50B. Using gene-specific small hairpin RNAs (shRNAs), we suppressed expression of *FAM50A* and *FAM50B*, individually and in combination. We observed stronger suppression in the A549 and MCF7 cell lines relative to the U2OS cell line, as indicated by relative cell growth (**Fig. 3B**). Using these results, we quantified genetic interactions between *FAM50A* and *FAM50B* using Bliss score [20], whereby lower scores indicate stronger synergistic interactions (see **Methods**). Consistent with our C-score prediction, the *FAM50A-FAM50B* gene pair in the A549 and MCF7 cell lines exhibited stronger synergy than in the U2OS cell line (Bliss score: 1.07 in A549, 1.06 in MCF7 and 1.44 in U2OS). Note that a recently study showed synthetic lethality between *FAM50A* and *FAM50B* in the A375 cell line [21], which is also located at the top-right of the *FAM50A-FAM50B* C-score plot (**Fig. 3A**).

Although duplicated genes are better recognized for their buffering relationship, there is no direct evidence of intrinsic buffering among non-duplicated genes. Thus, we also examined a non-duplicated gene pair, *POP7-RPP25*, representing one of the high C-score gene pairs with physical interaction. These two genes encode protein subunits of the ribonuclease P/MRP complex. In the C-score plot of *POP7-RPP25* (**Fig. 3C**), the HT29 cell line lies in the top-right region and the U2OS and LN18 cell lines are in the bottom-left region, indicating the likelihood of a stronger buffering effect in the HT29 cell line. Indeed, when we suppressed expression of *POP7* and *RPP25* using gene-specific shRNAs in these three cell lines, we observed that dual suppression of *POP7* and *RPP25* resulted in stronger synergistic effects for the HT29 cell line but not for the U2OS or LN18 cell lines (Bliss scores for *POP7-RPP25* genetic interactions are 1.02 in HT29, 1.13 in U2OS, and 1.19 in LN18, **Fig. 3D**).

Harnessing C-score to infer cell-specific genetic interactions

C-score describes the likelihood of functional buffering between two genes across hundreds of cell lines. Thus, for a particular cell line, it is possible to use C-scores to infer genetic interactions. As revealed by our experimental results in **Figure 3**, cell lines located at the upper right of a C-score plot are more sensitive to dual gene suppression. We reasoned that these cells can have greater buffering capacity relative to other cell lines, so they are more likely to exhibit synergistic genetic interactions. To infer cell-specific genetic interactions using C-score, we calculated buffering capacities as the relative G2 expression (compared to that of all other cell lines) of the cell line of interest adjusted by the C-score of the gene pair (**Fig. S4A**). We validated these C-score-derived buffering capacities for predicting genetic interactions using experimental results from four independent studies in human cells (**Table S1)** [22-25]. Using the receiver operating characteristic (ROC) curve to assess the performance of buffering capacity predictions, the area under curve (AUC) is significantly larger than random (Mann-Whitney U test with p-value < 0.05). Furthermore, the prediction performance increased with higher C-score cutoffs, as indicated by the increase in AUC (**Fig. 4**). Moreover, in addition to the qualitative predictions described above, C-score-derived buffering capacity appears to be quantitatively related to the strength of genetic interaction. We observed a negative correlation between C-score-derived buffering capacities and experimentally validated genetic interactions (C-score cutoff = 0.25, correlation = -0.231, p-value = 0.034, **Fig. S4B**), and this correlation is stronger for higher C-score cutoffs (**Fig. S4C**). Note that this inference on genetic interaction is robust since changing threshold for calculating buffering capacity did not qualitatively alter the results (**Methods** and **Fig. S4D**). Accordingly, based on these results, the CEBU mechanism can be used to infer cell-specific genetic interactions in human cells.

<u>Tissue-specificity of C-score-derived buffering capacity</u>

Typically, the same cell and tissue types share similar gene expression profiles. Thus, next we explored if expression of G2s in high C-score gene pairs exhibit tissue specificity. For each G1 and G2 gene pair, we grouped cell lines based on their tissue types and applied a tissue specificity index, τ [26], which quantifies tissue specificity (low τ, broadly expressed across tissues; high τ, only expressed in one or a few specific tissues). As shown in **Figure 5A**, G2s generally presented higher tissue specificity compared to G1s (significant with *t*-test, $p < 2.2e\text{-}16$) and compared to the control of randomly shuffled cell lines in G2 (**Fig. S5**). Since G2 expression is a critical determinant of our C-score-derived buffering capacity (**Fig. S4A**), we wondered if buffering capacity may also be tissue-specific. To investigate this possibility, we calculated the average of buffering capacity for each tissue type based on high C-score gene pairs (**Fig. 5B**). Our results showed distinct C-score-derived buffering capacities across different tissues, with the top three buffered tissues being the central nervous system, bone and the peripheral nervous system. In contrast, blood cells—including multiple myeloma, lymphoma, and leukemia cell lines—exhibited the lowest buffering capacities.

C-score-derived buffering capacity is indicative of cancer aggressiveness

If tissue-specific buffering is indeed a natural phenomenon, we wondered if cancers derived from different tissues utilize the buffering capacities endowed by the CEBU mechanism for more robust proliferation. In other words, would higher buffering capacity render cancers more robust and aggressive, thereby resulting in a poorer prognosis? To test this hypothesis, we aimed to gain a ground truth of expression-based cancer patient prognosis by analyzing patient gene expression and survival data for all available cancer types from The Cancer Genome Atlas (TCGA) [27] and then used this ground truth to examine the performance of buffering capacity predictions. As an example, **Figure 6A** presents the C-score plot of

NAMPT dependency and *CALD1* expression, indicating stronger buffering capacity in cell lines of the central nervous system (yellow) and weaker buffering capacity in leukemia cell lines (blue). Importantly, when we stratified patients based on *CALD1* expression, we observed a considerable difference in survival for patients with lower grade glioma (LGG, and see **Table S2** for cross-referencing between cell lines and TCGA cancers and for the full names of cancer abbreviations), but not for patients with acute myeloid leukemia (LAML, **Fig. 6B** left panel for LGG and right panel for LAML), suggesting that buffering capacity derived from the expression of *CALD1* and its C-score with *NAMPT* in human cell lines could be used to infer differences in cancer patient survival for LGG but not LAML.

To gain a ground truth of expression-based cancer patient prognosis for all cancer types, we assessed differential patient survival against gene expression using Cox regression and controlling for clinical characteristics including age, sex, pathological stage, clinical stage, and tumor grade, followed by multiple testing correction (false discovery rate < 0.2). We then systematically assessed how buffering capacity from C-score-identified gene pairs could help predict cancer patient survival for all cancer types. In **Figure 6C**, we show examples of ROC curves for two cancer types: pancreatic adenocarcinoma (PAAD) and mesothelioma (MESO). Examining all 30 cancer types, we found that for 20 of them, expression of at least one gene predicted patient survival with statistical significance and, for 10 out of those 20 cancer types, performance of buffering capacity for at least one C-score cutoff was significantly better than random (AUC > 0.5, false discovery rate < 0.2) (**Fig. S6A**). In addition, buffering capacity is also predictive of pathological stage (**Fig. 6D** and **S6B**), clinical stage (**Fig. S6C**), and tumor grade (**Fig. S6D**) in multiple cancer types. In general, the buffering capacity-based predictions performed better for greater C-score cutoffs. Taken together, our results show that

buffering capacity derived from our C-score index can be indicative of cancer aggressiveness, as illustrated by patient survival, cancer pathological stage, clinical stage and tumor grade.

Discussion

Unlike the responsive buffering mechanism whereby the buffering gene is only activated when its buffered function is compromised, the intrinsic buffering or CEBU mechanism proposed herein maintains a constitutively active state with both cell- and tissue-specificity and, thus, displays variability in buffering capacity, as illustrated in **Figure 1**. Since the buffering gene (G2) is continuously expressed in certain cell types, there is no need for a control system. Thus, the response time of intrinsic buffering is much quicker than responsive buffering. By means of the CEBU mechanism, intrinsic buffering capacity can be adjusted by altering the expression level of buffering genes, likely via epigenetic regulation, thereby meeting cell- or tissue-specific functional needs. Therefore, CEBU describes a simple, efficient and versatile mechanism for functional buffering in humans and other multicellular organisms. It is likely that as an intrinsic buffering mechanism, CEBU serves distinct kinds of functions compared to those of responsive buffering. Due to its integral and “hard-wired” nature of cell- and tissue-specific buffering, we suspect that CEBU could play an important role in maintaining tissue- and physiological homeostasis.

Based on our analyses, the G1s of high C-score gene pairs exhibit stronger essentiality and are broadly expressed across tissues, whereas the G2s of such gene pairs are less essential and expressed in specific tissues (**Fig. 5A** and **Fig. S2A**). Previously, the essentiality of genes has been correlated with their expression level and tissue specificity [28-30]. Housekeeping genes that are broadly expressed in most cells exhibit stronger essentiality, whereas genes expressed in specific cell types are considered to have weaker essentiality. However, it

remains elusive as to how these two types of genes are functionally associated with each other. Here, CEBU represents a putative mechanistic link between these two types of genes, enabling their cooperation to regulate cellular functions through intrinsic buffering. In particular, house-keeping functions like metabolic and cell-cycle-related processes are highly enriched among high C-score gene pairs (**Fig. 2E**), indicating that house-keeping functions can be more robustly maintained via CEBU-mediated regulation. Moreover, we found that the tissues predicted to have the strongest buffering capacities are the nervous and bone systems (**Fig. 5B**), both of which exhibit relatively low regenerative capacities. In contrast, blood cells, which are frequently regenerated in humans, are predicted to have the weakest buffering capacities (**Fig. 5B**). It is possible that cell types with weaker regenerative capacities can be sustained robustly through the stronger functional buffering afforded by CEBU, thereby maintaining their tissue homeostasis.

CEBU describes an intrinsic buffering mechanism that is expected to function under normal physiological conditions. Indeed, when we examined if our C-score could be biased due to our usage of cancer cell lines, we found that only a low percentage (2.3% per gene pair, **Fig. S7A**) of mutant cell lines contributed to our C-score measurements. Moreover, excluding mutant cell lines did not qualitatively affect our C-score measurements, especially for high C-score gene pairs (**Fig. S7B**). The same trend holds for cancer-related genes (**Fig. S7B**). These results suggest mutant cell lines are not the major determinants of C-scores. Similarly, since copy number variation (CNV) is a major mechanism for oncogenic expression, we checked if CNV contributes to G2 expression. As shown in **Figure S7C**, the correlation between G2 expression and copy number decreases with increasing C-score, indicating that CNV is not a primary mechanism regulating G2 expression. Thus, the C-score measurement is likely not biased by the utilization of cancer cell lines.

The high C-score gene pairs are enriched with duplicated genes, supporting the notion that duplicated genes contribute to the context-dependent essentiality of their paralogous genes [10-12]. Although redundant duplicated genes are thought to be evolutionarily unstable [31], divergence in expression provides a mechanism to retain them [5-7, 12]. It is possible that this divergence in duplicated gene expression facilitates the cell- and tissue-specific expression of buffering genes, thus representing a critical evolutionary mechanism for CEBU in multicellular organisms. On the other hand, despite numerous observations of functional buffering among duplicated genes [32-34], how widely a gene can buffer another non-orthologous gene is less well known. Our C-score index identified a high percentage of non-duplicated gene pairs with high buffering capacities (**Fig. 2A**), and these non-duplicated gene pairs tend to belong to the same pathways and/or protein complexes (**Fig. 2B** and **2C**). Therefore, it is possible that many of these G1s and G2s represent non-orthologous functional analogs. One simple scenario could be that G1 and G2 physically interact with each other to form a protein complex, wherein G1's function can be structurally substituted by G2. Indeed, we identified the *POP7* and *RPP25* gene pair as an example of this scenario (**Fig. 3C** and **3D**). More sophisticated and indirect functional buffering can also occur between G1s and G2s given the complex interactions among biological functions [35]. We expect that additional types of molecular mechanisms can account for the buffering effects of CEBU, which remain to be tested experimentally.

C-score-derived cell-specific buffering capacities comply well with experimentally validated genetic interactions in human cells (**Fig. 4** and **Fig. S4**), indicating that CEBU may represent a critical mechanism for synthetic lethality in human cells. In practice, it remains a daunting challenge to systematically characterize genetic interactions in organisms with complex

genomes due to large numbers of possible gene pairs, i.e. ~200 million gene pairs in humans. Previous efforts have used computational approaches on conserved synthetically lethal gene pairs in the budding yeast *Saccharomyces cerevisiae* or employ data mining on multiple large datasets to infer human synthetic lethalities [36, 37]. Nevertheless, predictions emanating from different studies exhibit little overlap [38], evidencing the marked complexity of synthetic lethality in humans. The CEBU mechanism proposed here can contribute both experimentally and computationally to a better characterization of human genetic interactions.

Using G2 expression of a single high C-score gene pair to stratify cancer patients, we observed a significant difference in cancer patient survival, indicating that stronger buffering capacity (as derived from C-scores and relative G2 expression in human cell lines) could be predictive of cancer aggressiveness in patients (see **Fig. 6A** and **6B** for an example). Indeed, buffering capacity helps predict cancer patient survival in ten cancer types, and generally the performance of prediction is better with a higher C-score cutoff (**Fig. 6C** and **S6A**). Apart from patient survival, buffering capacity is also indicative of pathological stage, clinical stage, and tumor grade of cancers (**Fig. 6D** and **S6B-6D**). These results support our hypothesis that stronger buffering capacity via higher G2 expression contributes to cancer robustness in terms of proliferation and drug resistance. Given the complexity of cancers, it is surprising to see such general predictivity of cancer prognosis by individual high C-score gene pairs. Accordingly, we suspect that some cancer cells may adopt this cell- and tissue-specific buffering mechanism to enhance their robustness in proliferation and stress responses by maintaining or even upregulating the gene expression of buffering genes. Clinically, the expression of such buffering genes could represent a unique feature for evaluating cancer progression when applied alongside other currently used clinical characteristics. Finally, experimental validation of C-score-predicted genetic interactions in cancers will help identify

potential drug targets for tailored combination therapy based on cancer types and their expression profiles.

Materials and Methods

Dependencies and gene expression data retrieval and processing

Data on dependencies and CCLE (Cancer Cell Line Encyclopedia) gene expression were downloaded from the DepMap database (DepMap Public 19Q4) [13, 14]. Dependencies modeled from the CERES computational pipeline based on a genome-wide CRISPR loss-of-function screening were selected. CCLE expression data was quantified as $\log_2$ TPM (Transcripts Per Million) using RSEM (RNA-seq by Expectation Maximization) with a pseudo-count of 1 in the GTEx pipeline (https://gtexportal.org/home/documentationPage). Only uniquely mapped reads in the RNA-seq data were used in the GTEx pipeline. Integrating and cross-referencing the dependency and expression datasets yielded 18239 genes and 684 cell lines. Genes lacking dependencies for any one of the 684 cell lines were discarded from our analyses.

C-score calculation

Our C-score index integrates the dependencies of buffered genes (G1) and gene expression of buffering genes (G2) to determine the buffering relationship between gene pairs. Genes with mean dependencies > 0 or mean gene expression < 0.5 $\log_2$ TPM were discarded, yielding 9196 G1s and 13577 G2s. The C-score integrates the correlation (ρ) and slope between the dependency of gene $G1$ and the gene expression of gene $G2$, defined as:

$$\text{C-score} = \rho_{G1,G2}\left(1 + b\frac{slope_{min}}{slope_{G1,G2}}\right)$$

where ρ denotes the Pearson correlation coefficient and $slope_{min}$ denotes the minimum slope of all considered gene pairs that present a statistically significant positive correlation. The normalized slope can be weighted by cell- and tissue-type specific b. In our analysis, b is set as 1 for a pan-cell or pan-cancer analysis.

Duplicated gene assignment

Information on gene identity was obtained from ENSEMBL (release 98, reference genome GRCh38.p13) [39]. Two genes are considered duplicated genes if they have diverged from the same duplication event.

Enrichment analysis for buffering gene pairs

For the enrichment analysis of gene pairs, we used a previously described methodology [40]. Briefly, gene sets of GO and KEGG were downloaded from the Molecular Signatures Database. The number of total possible gene pairs is 9196 (G1) x 13577 (G2). The condition of G1 and G2 being the same gene was not considered as a buffering gene pair under all C-score cutoffs. Enrichment was calculated as:

$$\log \frac{\frac{e_{ac}}{e_a}}{\frac{e_c}{e_t}}$$

where e_{ac} represents the number of gene pairs that are both annotated and with buffering capability, e_a is the number of annotated gene pairs, e_c is the number of buffering gene pairs, and e_t is the total number of gene pairs.

Protein-protein interaction (PPI) data was downloaded from the STRING database (version 11) [18]. Only high-confidence interactions (confidence > 0.7) in human were considered. The STRING database determines confidence by approximating the probability that a link exists between two enzymes in the KEGG database. Data on protein core complexes were

409 downloaded from CORUM (http://mips.helmholtz-muenchen.de/corum). The enrichment
410 calculation is the same as for GO and KEGG, except that e_{ac} represents the number of gene
411 pairs that have PPI or are in the same complex and have buffering capability, and e_a is the
412 number of gene pairs that have PPI or are in the same complex.
413
414 **Experimental validation**
415 A549, H4, HT29, LN18, MCF7, and U2OS cell lines were selected based on their locations
416 in the C-score plots (**Fig. 2A and 2C**), indicating different buffering capacity. All cell lines
417 were purchased from ATCC and they were cultured in Dulbecco's Modified Eagle Media
418 (H4 and LN18), Ham's F-12K Medium (A549), or RPMI 1640 media (HT29, MCF7, and
419 U2OS) supplemented with 5% fetal bovine, serum, 100 U/mL penicillin, 100 µg/mL
420 streptomycin, and 250 ng/mL fungizone (Gemini Bio-Products). Cell growth was monitored
421 by time-lapse imaging using Incucyte Zoom, taking images every 2 hours for 2-4 days. To
422 suppress *FAM50A, FAM50B*, *POP7* and *RPP25* expression, lentivirus-based shRNAs were
423 delivered individually or in combination. The gene-specific shRNA sequences are: *FAM50A*
424 - CCAACATTGACAAGAAGTTCT and GAGCTGGTACGAGAAGAACAA; *FAM50B* –
425 CACCTTCTACGACTTCATCAT; *POP7* – CTTCAGGGTCACACCCAAGTA and
426 CGGAGACCCAATGACATTTAT; and *RPP25* – CCAGCGTCCAAGAGGAGCCTA. To
427 ensure better knockdown of gene expression, shRNAs were delivered twice (7 days and 4
428 days before seeding). Equal numbers of cells were seeded for cell growth measurements by
429 time-lapse imaging using Incucyte Zoom. The lentivirus-based shRNAs were purchased from
430 the RNAi core of Academia Sinica. The growth rate of each condition was measured by
431 fitting cell confluence to an exponential growth curve using the Curve Fitting Toolbox in
432 MATLAB.
433

434 **Bliss independence model**

435 The synergy of cytotoxicity was measured using the Bliss independent model [20]. The Bliss
436 model is presented as a ratio of the expected additive effect to the observed combination
437 effect:

$$E_{bliss} = \frac{E_A + E_B - E_A \times E_B}{E_{AB}}$$

438 where *E* is the effect of drug *A*, *B*, or a combination of *A* and *B*. Effect was measured by the
439 relative cell growth, based on the fold-change of confluency between 0 and 72 hours upon
440 suppression of *FAM50A* and *FAM50B* or suppression of *POP7* and *RPP25* in all cell lines
441 except MCF7, and between 0 and 96 hours upon suppression of *FAM50A* and *FAM50B* in
442 MCF7.

443

444 **Cell-specific buffering capacity and comparison with experimental genetic interactions**

445 Cell-specific buffering capacity was derived from the C-score of a given gene pair and gene
446 expression of the buffering gene (G2) in the cell line of interest following the equation:

447

$$\text{buffering capacity} = \frac{\text{cell line expression} - \text{ 25th percentile of all expression (G2)}}{slope_{mod}}$$

$$\text{where } slope_{mod} = \text{C-score} \times \frac{sd(G2\ expression)}{sd(G1\ dependency)}$$

448

449 where *sd* = standard deviation. The 25th percentile cutoff for expression is determined
450 empirically, although different percentile cutoffs do not qualitatively affect the measurements
451 of buffering capacities (**Fig. S4A**).

452 Combinatorial CRISPR screen-derived genetic interaction scores were pooled from four
453 literature sources [22-25] (**Table S1**). We only considered cell lines that appear in DepMap
454 CERES 19Q4. There were two C-scores for each gene-pair of the experimental dataset (either

gene could be a G1), and we assigned the higher C-score for that gene-pair. Overall, we curated 10,222 genetic interaction scores in various cell lines from the literature, and 1986 out of 10,222 genetic interaction scores had a C-score > 0.1. To evaluate the validity of buffering capacity, we generated a ground truth table by assigning gene-pairs with a positive genetic interaction as false for buffering and a negative genetic interaction as true for buffering. The qualitative performance of buffering capacity to the ground truth data was assessed by ROC curve. Additionally, we correlated the buffering capacity directly via a ground truth genetic interaction score for quantitative evaluation.

Tissue specificity

To calculate tissue-specificity, cell lines were grouped by their respective tissues, and expression of genes in cell lines of the same tissue were averaged. Tissue specificity was calculated as tau (τ) [26], where τ is defined as:

$$\tau = \frac{\sum_{i-1}^{N}(1 - \frac{x_i}{x_{max}})}{N - 1}$$

with N denoting the number of tissues, x_i denoting the expression of a gene, and x_{max} denoting the highest gene expression across all tissues. Note, expression values were log-transformed, so $\log_2$ TPM < 1 was considered as 0 in tissue specificity calculations [41].

Cancer-specific survival prediction according to C-score gene pairs

Gene expression and survival data from The Cancer Genome Atlas (TCGA) [27] was retrieved from Xena [42]. The DepMap cancer cell lines were mapped to TCGA cancers based on the annotation in **Table S2** (cancers that do not have a matched cancer type in

CERES 19Q4 were not analyzed). To systematically analyze cancer prognosis, we first performed a multiple test correction on the *p*-values from Cox regression controlling for age, sex, pathological stage, clinical stage and tumor grade. We calculated the false discovery rate (FDR) using the Benjamini–Hochberg procedure with a threshold < 0.2. The ground truth table for each cancer was constructed using the adjusted *p*-value. AUC of ROC curves were used to assess the performance of survival based on buffering capacity. AUCs and ROCs were generated using python. The statistical significance of AUC was assessed by using the Mann-Whitney U test [43] to evaluate whether gene expression with a positive Cox coefficient (poorer prognosis) reflected significantly higher buffering capacities in each cancer with different C-score cut-offs. The *p*-values of the Mann-Whitney U test were adjusted using the Benjamini-Hochberg procedure with a threshold < 0.2. We conducted a similar approach to the prognosis analysis for buffering capacities and clinical features. We calculated the p-values of correlation of gene expression and clinical features, staging and grade, followed by correcting the p-values using the Benjamini-Hochberg procedure with a threshold < 0.2. We then used Mann-Whitney U tests to evaluate if gene pairs with a significant positive correlation between gene expression and tumor aggressiveness presented a significantly higher buffering capacity in each cancer for different C-score cut-offs.

Data Availability

All data and code used for this article are available in the FigShare repository https://figshare.com/s/6f8929c6543687a6062f, and https://figshare.com/s/b778489bb2f6fc3b0069, respectively.

Acknowledgement

We thank members of the Lab for Cell Dynamics for helpful discussions. We are grateful to Jose Reyes, Hannah Katrina Co, and Jun-Yi Leu for their comments and suggestions on the manuscript, and Ann Mikaela Lynne Ong Co, Jo-Hsi Huang, Su-Ping Lee, the Imaging Core at the Institute of Molecular Biology, and the RNAi Core at Academia Sinica for their technical support.

Author Contributions

H.-K.L., J.-H.C., and S.-h.C. conceived the project. H.-K.L., J.-H.C., C.-C.W., and S.-h.C. designed and conducted the computational and statistical analyses. H.-K.L., J.-H.C., C.-C.W., and S.-h.C. wrote the manuscript. F.-S.H., C.D. and S.-h.C. designed the experiments, and F.-S.H., and C.D. conducted the experiments.

Conflict of Interests

J.-H.C. is an employee of ACT Genomics

References and Citations

1. Kitano H. Biological robustness. Nat Rev Genet. 2004;5(11):826-37. Epub 2004/11/03. doi: 10.1038/nrg1471. PubMed PMID: 15520792.
2. Masel J, Siegal ML. Robustness: mechanisms and consequences. Trends Genet. 2009;25(9):395-403. doi: 10.1016/j.tig.2009.07.005. PubMed PMID: WOS:000270185900004.
3. El-Brolosy MA, Stainier DYR. Genetic compensation: A phenomenon in search of mechanisms. PLoS Genet. 2017;13(7):e1006780. Epub 2017/07/14. doi: 10.1371/journal.pgen.1006780. PubMed PMID: 28704371; PubMed Central PMCID: PMCPMC5509088.
4. Diss G, Ascencio D, DeLuna A, Landry CR. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. J Exp Zool B Mol Dev Evol. 2014;322(7):488-99. Epub 2014/01/01. doi: 10.1002/jez.b.22555. PubMed PMID: 24376223.
5. Kafri R, Bar-Even A, Pilpel Y. Transcription control reprogramming in genetic backup circuits. Nat Genet. 2005;37(3):295-9. Epub 2005/02/22. doi: 10.1038/ng1523. PubMed PMID: 15723064.
6. Kafri R, Levy M, Pilpel Y. The regulatory utilization of genetic redundancy through responsive backup circuits. P Natl Acad Sci USA. 2006;103(31):11653-8. doi:

10.1073/pnas.0604883103. PubMed PMID: WOS:000239616400042.
7. Kafri R, Springer M, Pilpel Y. Genetic Redundancy: New Tricks for Old Genes. Cell. 2009;136(3):389-92. doi: 10.1016/j.cell.2009.01.027. PubMed PMID: WOS:000263120600006.
8. Koonin EV, Mushegian AR, Bork P. Non-orthologous gene displacement. Trends Genet. 1996;12(9):334-6. Epub 1996/09/01. PubMed PMID: 8855656.
9. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, et al. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. Cell. 2015;162(2):391-402. Epub 2015/07/18. doi: 10.1016/j.cell.2015.06.037. PubMed PMID: 26186192.
10. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and characterization of essential genes in the human genome. Science. 2015;350(6264):1096-101. Epub 2015/10/17. doi: 10.1126/science.aac7041. PubMed PMID: 26472758; PubMed Central PMCID: PMCPMC4662922.
11. Dandage R, Landry CR. Paralog dependency indirectly affects the robustness of human cells. Mol Syst Biol. 2019;15(9):e8871. Epub 2019/09/27. doi: 10.15252/msb.20198871. PubMed PMID: 31556487; PubMed Central PMCID: PMCPMC6757259.
12. De Kegel B, Ryan CJ. Paralog buffering contributes to the variable essentiality of genes in cancer cell lines. Plos Genetics. 2019;15(10). doi: ARTN e1008466 10.1371/journal.pgen.1008466. PubMed PMID: WOS:000503176900046.
13. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503-+. doi: 10.1038/s41586-019-1186-3. PubMed PMID: WOS:000468844100040.
14. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature Genetics. 2017;49(12):1779-+. doi: 10.1038/ng.3984. PubMed PMID: WOS:000416480600018.
15. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9. Epub 2000/05/10. doi: 10.1038/75556. PubMed PMID: 10802651; PubMed Central PMCID: PMCPMC3037419.
16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30. Epub 1999/12/11. doi: 10.1093/nar/28.1.27. PubMed PMID: 10592173; PubMed Central PMCID: PMCPMC102409.
17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. Epub 2005/10/04. doi: 10.1073/pnas.0506580102. PubMed PMID: 16199517; PubMed Central PMCID: PMCPMC1239896.
18. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005;33(Database issue):D433-7. Epub 2004/12/21. doi: 10.1093/nar/gki005. PubMed PMID: 15608232; PubMed Central PMCID: PMCPMC539959.
19. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic Acids Res. 2010;38(Database issue):D497-501. Epub 2009/11/04. doi: 10.1093/nar/gkp914. PubMed PMID: 19884131; PubMed Central PMCID: PMCPMC2808912.
20. Bliss CI. The toxicity of poisons applied jointly. Annals of Applied Biology.

1939;26(3):585–615.
21. Thompson NA, Ranzani M, van der Weyden L, Iyer V, Offord V, Droop A, et al. Combinatorial CRISPR screen identifies fitness effects of gene paralogues. Nat Commun. 2021;12(1):1302. Epub 2021/02/28. doi: 10.1038/s41467-021-21478-9. PubMed PMID: 33637726; PubMed Central PMCID: PMCPMC7910459.
22. Rosenbluh J, Mercer J, Shrestha Y, Oliver R, Tamayo P, Doench JG, et al. Genetic and Proteomic Interrogation of Lower Confidence Candidate Genes Reveals Signaling Networks in beta-Catenin-Active Cancers. Cell Syst. 2016;3(3):302-+. doi: 10.1016/j.cels.2016.09.001. PubMed PMID: WOS:000395775300012.
23. Shen JP, Zhao DX, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat Methods. 2017;14(6):573-+. doi: 10.1038/nmeth.4225. PubMed PMID: WOS:000402291800016.
24. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol. 2018;36(2):179-89. Epub 2017/12/19. doi: 10.1038/nbt.4048. PubMed PMID: 29251726; PubMed Central PMCID: PMCPMC5800952.
25. Zhao D, Badur MG, Luebeck J, Magana JH, Birmingham A, Sasik R, et al. Combinatorial CRISPR-Cas9 Metabolic Screens Reveal Critical Redox Control Points Dependent on the KEAP1-NRF2 Regulatory Axis. Mol Cell. 2018;69(4):699-708 e7. Epub 2018/02/18. doi: 10.1016/j.molcel.2018.01.017. PubMed PMID: 29452643; PubMed Central PMCID: PMCPMC5819357.
26. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21(5):650-9. Epub 2004/09/25. doi: 10.1093/bioinformatics/bti042. PubMed PMID: 15388519.
27. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018;173(2):291-304 e6. Epub 2018/04/07. doi: 10.1016/j.cell.2018.03.022. PubMed PMID: 29625048; PubMed Central PMCID: PMCPMC5957518.
28. Hastings KEM. Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. J Mol Evol. 1996;42(6):631-40. doi: Doi 10.1007/Bf02338796. PubMed PMID: WOS:A1996UU51200002.
29. Subramanian S, Kumar S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics. 2004;168(1):373-81. doi: 10.1534/genetics.104.028944. PubMed PMID: WOS:000224393700029.
30. Zhang LQ, Li WH. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol. 2004;21(2):236-9. doi: 10.1093/molbev/msh010. PubMed PMID: WOS:000220083300005.
31. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151-5. Epub 2000/11/10. doi: 10.1126/science.290.5494.1151. PubMed PMID: 11073452.
32. Gu ZL, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH. Role of duplicate genes in genetic robustness against null mutations. Nature. 2003;421(6918):63-6. doi: 10.1038/nature01198. PubMed PMID: WOS:000180165500037.
33. DeLuna A, Springer M, Kirschner MW, Kishony R. Need-Based Up-Regulation of Protein Levels in Response to Deletion of Their Duplicate Genes. Plos Biol. 2010;8(3). doi: ARTN e1000347
10.1371/journal.pbio.1000347. PubMed PMID: WOS:000278125400021.

34. Deutscher D, Meilijson I, Kupiec M, Ruppin E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat Genet. 2006;38(9):993-8. Epub 2006/08/31. doi: 10.1038/ng1856. PubMed PMID: 16941010.
35. Rancati G, Moffat J, Typas A, Pavelka N. Emerging and evolving concepts in gene essentiality. Nat Rev Genet. 2018;19(1):34-49. Epub 2017/10/17. doi: 10.1038/nrg.2017.74. PubMed PMID: 29033457.
36. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014;158(5):1199-209. Epub 2014/08/30. doi: 10.1016/j.cell.2014.07.027. PubMed PMID: 25171417.
37. Srivas R, Shen JP, Yang CC, Sun SM, Li JF, Gross AM, et al. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Molecular Cell. 2016;63(3):514-25. doi: 10.1016/j.molcel.2016.06.022. PubMed PMID: WOS:000381619900016.
38. Liu L, Chen X, Hu C, Zhang D, Shao Z, Jin Q, et al. Synthetic Lethality-based Identification of Targets for Anticancer Drugs in the Human Signaling Network. Sci Rep. 2018;8(1):8440. Epub 2018/06/02. doi: 10.1038/s41598-018-26783-w. PubMed PMID: 29855504; PubMed Central PMCID: PMCPMC5981615.
39. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. Nucleic Acids Res. 2019;47(D1):D745-D51. Epub 2018/11/09. doi: 10.1093/nar/gky1113. PubMed PMID: 30407521; PubMed Central PMCID: PMCPMC6323964.
40. Boyle EA, Pritchard JK, Greenleaf WJ. High-resolution mapping of cancer cell networks using co-functional interactions. Molecular Systems Biology. 2018;14(12). doi: ARTN e8594
10.15252/msb.20188594. PubMed PMID: WOS:000453861400001.
41. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2017;18(2):205-14. Epub 2016/02/20. doi: 10.1093/bib/bbw008. PubMed PMID: 26891983; PubMed Central PMCID: PMCPMC5444245.
42. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017;35(4):314-6. Epub 2017/04/12. doi: 10.1038/nbt.3772. PubMed PMID: 28398314; PubMed Central PMCID: PMCPMC5546205.
43. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of the Royal Meteorological Society. 2002;128(584):2145-66. doi: 10.1256/003590002320603584.
44. Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2020. Epub 2020/12/09. doi: 10.1093/nar/gkaa1106. PubMed PMID: 33290554.

Figure Legends

Figure 1. Genome-wide CEBU analysis using C-score

C-score plot: the x-axis is the dependency of the buffered gene (G1) and the y-axis is the expression of the buffering gene (G2). G1 is considered as being potentially buffered by G2, as quantified by C-score, which is an adjusted correlation for a given gene based on the C-score plot.

Figure 2. Properties of high C-score gene pairs

(A) Percentage of C-score-identified non-duplicated and duplicated gene pairs with different C-score cut-offs. (B-C) Functional enrichment of C-score gene pairs increases with C-score cutoff. (B) Enrichment of pairs of genes annotated with the same gene ontology biological process (GO:BP) term or KEGG pathway in C-score gene pairs, and enrichment increases with C-score cut-off. (C) Enrichment of pairs of genes with annotated protein-protein interactions from STRING and within the same protein complex from CORUM in C-score gene pairs, and enrichment increases with C-score cut-off. (D) Top 10 GO:BP enrichment of high C-score gene pairs with the same functional annotation. GO was annotated using PANTHER 15.0 [44] , and selected using GO Slim terms.

Figure 3. Experimental validation of cell-specific expression buffering between *FAM50A* and *FAM50B*, and *POP7* and *RPP25*

(A) C-score plot of the highest C-score gene pair, *FAM50A* (dependency score - D.S.) and *FAM50B* (G2) expression, labeled with the U2OS (predicted not synergistic), A549 (predicted synergistic), and MCF7 (predicted synergistic) cell lines. (B) Relative cell growth based on fold-change in confluency of the U2OS, A549, and MCF7 cell lines with or without *FAM50A* or *FAM50B* suppression. Bliss scores indicate strength of synergy between double suppression of *FAM50A* and *FAM50B* compared to either gene alone. Error bars indicate standard deviation of six technical repeats. (C) C-score plot of the non-duplicated gene pair,

POP7 D.S. and *RPP25* expression, labeled with the LN18 (predicted not synergistic), U2OS (predicted not synergistic), and HT29 (predicted synergistic) cell lines. (D) Relative cell growth based on fold-change in confluency of the HT29, U2OS and LN18 cell lines with or without shRNA-based *POP7* or *RPP25* suppression. Bliss scores indicate strength of synergy between double suppression of *POP7* and *RPP25* compared to either gene alone. Error bars indicate standard deviation of six technical repeats.

Figure 4. C-score-derived cell-specific buffering capacity predicts genetic interaction and cancer patient survival

Predictive performance shown as ROC curves for predicting genetic interactions using cell-specific buffering capacity. Prediction sets consist of 84 data-points (37 unique genetic interaction gene-pairs) across 8 cell lines with a C-score cut-off of 0.25.

Figure 5. C-score-derived tissue-specific buffering capacity

(A) Tissue specificity (τ) of G1 and G2 pairs. τ was calculated for G1 and G2 from high C-score gene pairs. Statistical significance was assessed by paired-*t* test. (B) Mean buffering capacity for each tissue type.

Figure 6. Harnessing cell-specific high C-score gene pairs for cancer patient prognosis

(A) C-score plot of *NAMPT* D.S. and *CALD1* expression (C-score = 0.447). Yellow circles represent central nervous system (CNS) cell lines and blue circles denote leukemia cell lines. (B) Kaplan-Meier overall survival plots for CNS (LGG, lower grade glioma, left panel) and leukemia (LAML, acute myeloid leukemia, right panel) cancer patients. Patients were stratified by high (>75%) or low (<25%) expression of *CALD1*, and *p*-values were calculated

using Cox regression controlling for age, sex, pathological staging, clinical staging, and tumor grade, and corrected for multiple testing (false discovery rate < 0.2). (C) ROC curves based on C-score gene pair-based buffering prediction of survival for PAAD (pancreatic adenocarcinoma, left panel) and MESO (mesothelioma, right panel) with different C-score cutoffs. (D) ROC curves based on C-score gene pair-based buffering prediction of pathological stage for PAAD with different C-score cutoffs.

Supporting Information Captions

Figure S1. Distribution of C-scores

Figure S2. Correlation properties of G1 dependency and G2 expression according to increasing C-score

Figure S3. Functional and pathway enrichments of C-score-identified duplicated genes

Figure S4. C-score-based prediction of cell-specific genetic interaction using buffering capacity

Figure S5. Shuffling the G2-tissue relationship disrupts G2 tissue-specificity

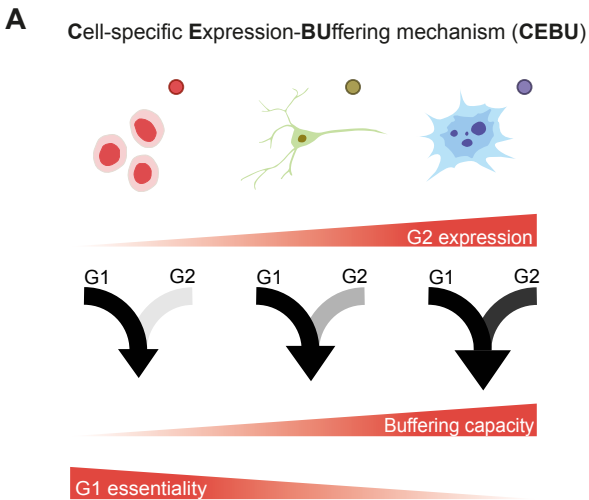
Figure S6. C-score-based prediction of cancer patient survival, pathological stage, clinical stage and tumor grade

Figure S7. Decreasing effects of mutational variation as C-score increases

Table S1. List of combinatorial CRISPR-screened gene pairs in published literature

Table S2. Cross-reference table for TCGA cancer type and DepMap cancer type

Figure 1


A
Cell-specific Expression-BUffering mechanism (CEBU)
G2 expression
G1
G2
G1
G2
G1
G2
Buffering capacity
G1 essentiality

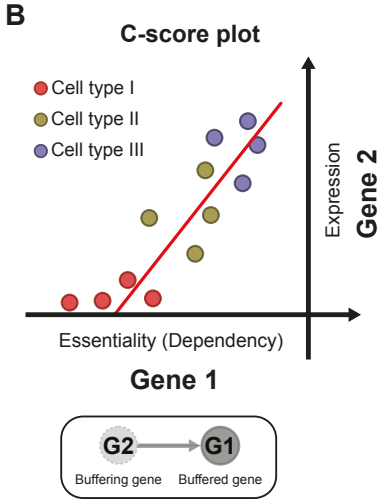

B
C-score plot
Cell type I
Cell type II
Cell type III
Expression
Gene 2
Essentiality (Dependency)
Gene 1
G2
G1
Buffering gene
Buffered gene

Figure 2

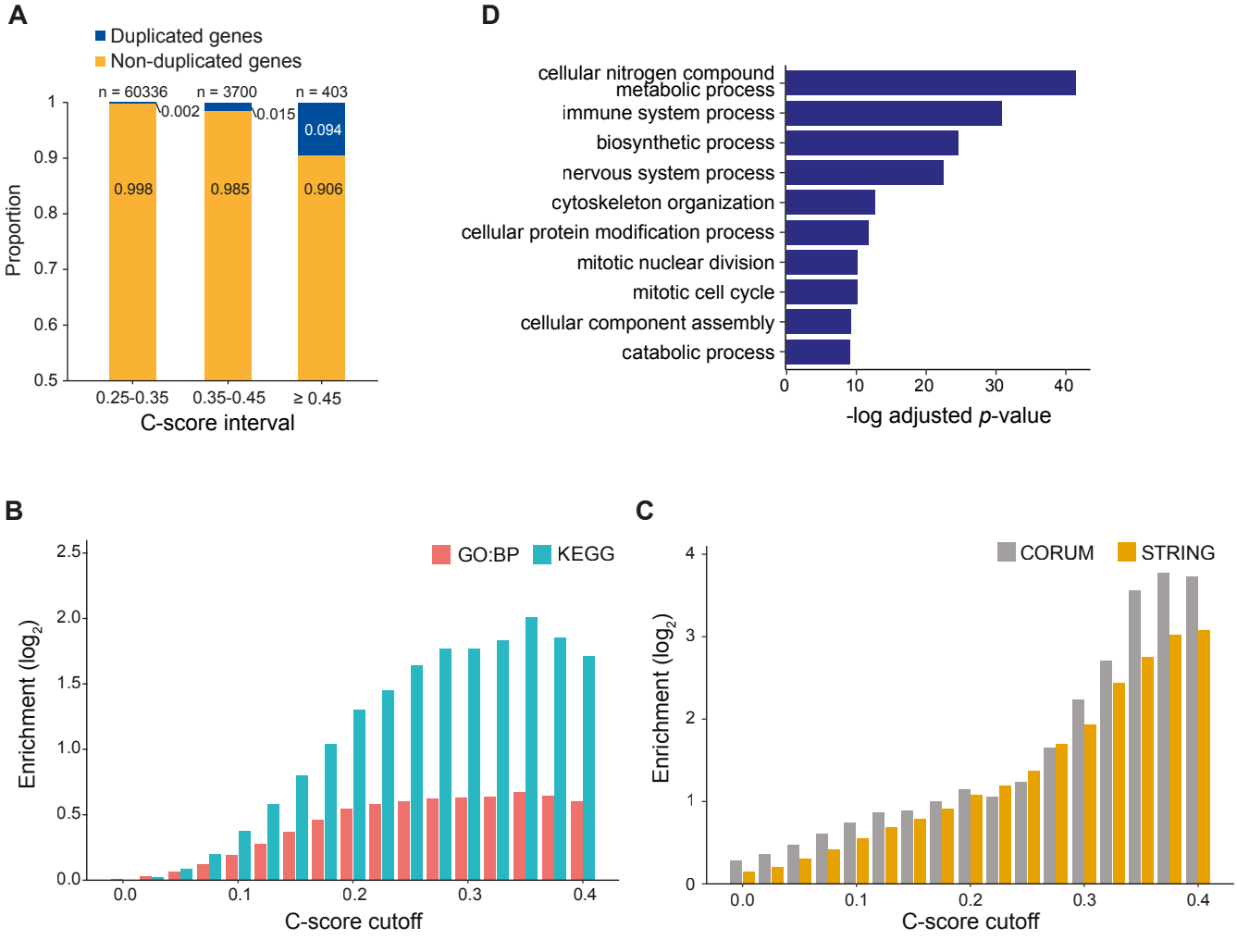

A
Duplicated genes
Non-duplicated genes
n = 60336
n = 3700
n = 403
0.002
0.015
0.094
0.998
0.985
0.906
Proportion
0.25-0.35
0.35-0.45
≥ 0.45
C-score interval
D
cellular nitrogen compound metabolic process
immune system process
biosynthetic process
nervous system process
cytoskeleton organization
cellular protein modification process
mitotic nuclear division
mitotic cell cycle
cellular component assembly
catabolic process
-log adjusted p-value
B
GO:BP
KEGG
Enrichment (log2)
C-score cutoff
C
CORUM
STRING
Enrichment (log2)
C-score cutoff

Figure 3

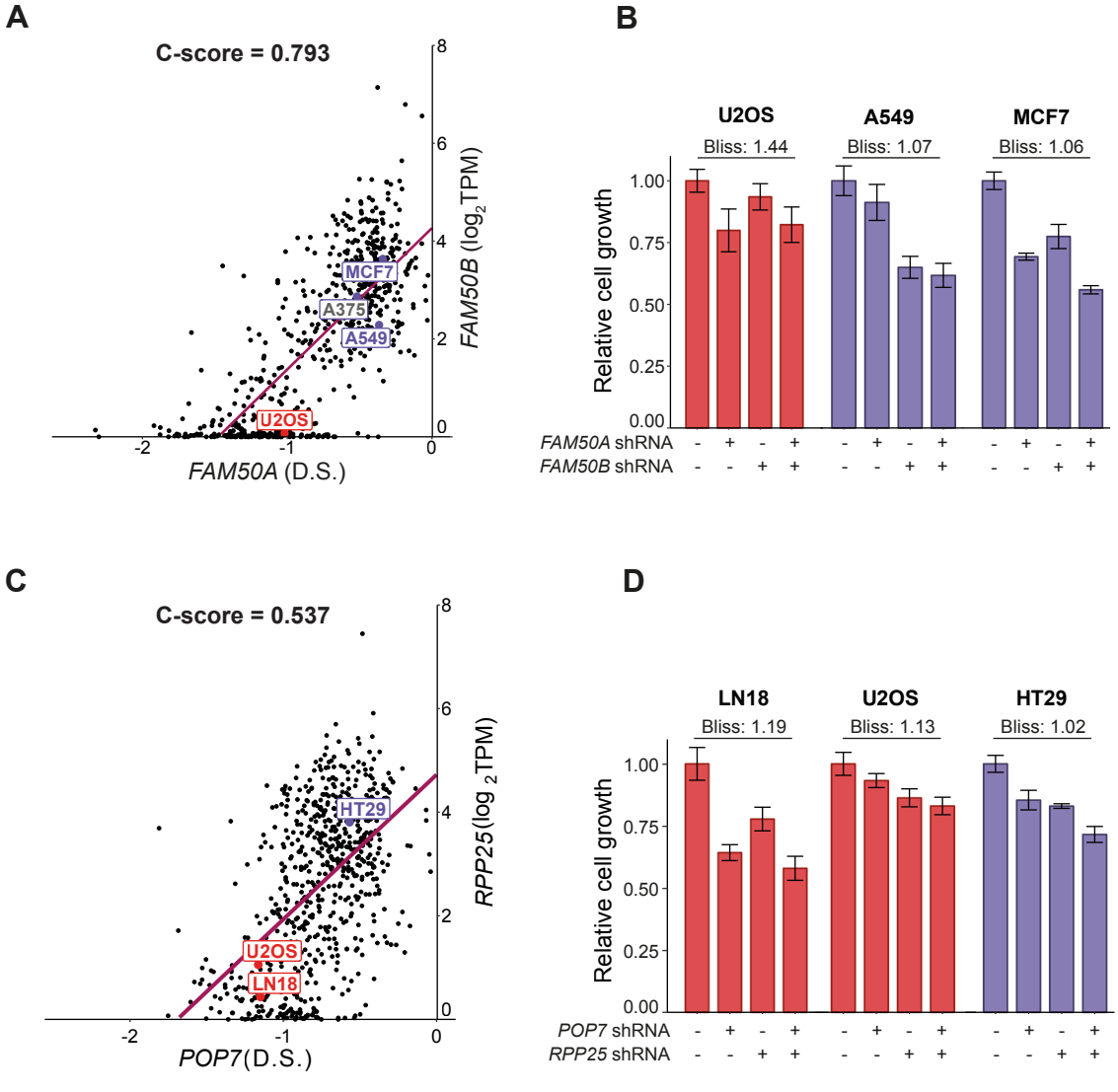

A
C-score = 0.793
MCF7
A375
A549
U2OS
FAM50A (D.S.)
FAM50B (log2TPM)
B
U2OS
A549
MCF7
Bliss: 1.44
Bliss: 1.07
Bliss: 1.06
Relative cell growth
FAM50A shRNA
FAM50B shRNA
C
C-score = 0.537
HT29
U2OS
LN18
POP7(D.S.)
RPP25(log 2TPM)
D
LN18
U2OS
HT29
Bliss: 1.19
Bliss: 1.13
Bliss: 1.02
Relative cell growth
POP7 shRNA
RPP25 shRNA

Figure 4

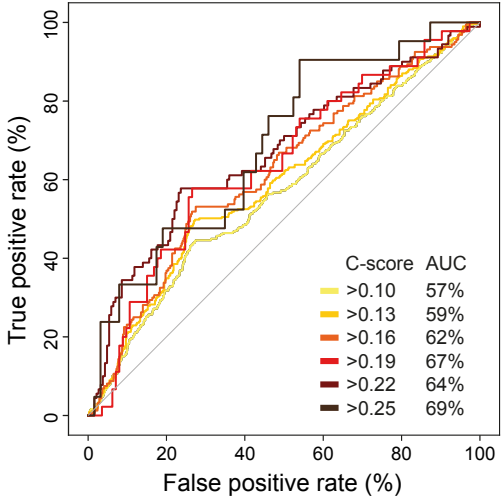

True positive rate (%)
False positive rate (%)
0
20
40
60
80
100
C-score AUC
>0.10 57%
>0.13 59%
>0.16 62%
>0.19 67%
>0.22 64%
>0.25 69%

Figure 5

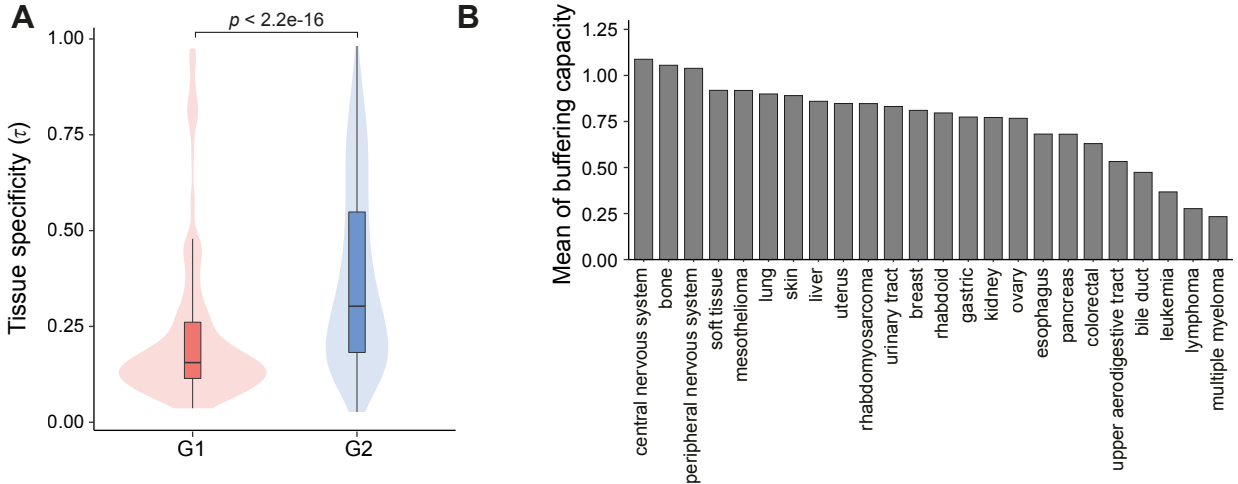
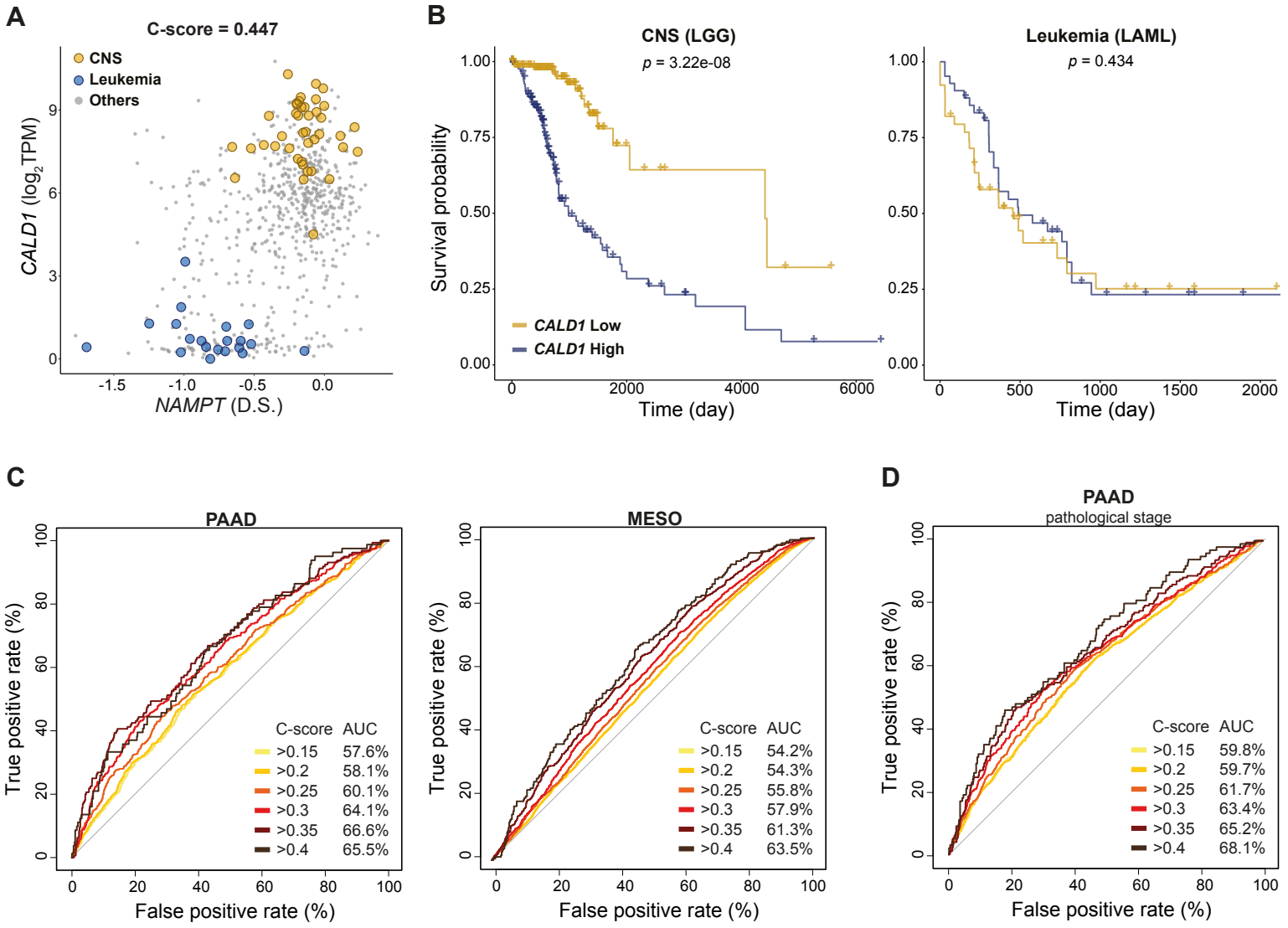

A
p < 2.2e-16
Tissue specificity (τ)
1.00
0.75
0.50
0.25
0.00
G1
G2
B
Mean of buffering capacity
1.25
1.00
0.75
0.50
0.25
0.00
central nervous system
bone
peripheral nervous system
soft tissue
mesothelioma
lung
skin
liver
uterus
rhabdomyosarcoma
urinary tract
breast
rhabdoid
gastric
kidney
ovary
esophagus
pancreas
colorectal
upper aerodigestive tract
bile duct
leukemia
lymphoma
multiple myeloma

Figure 6


A
C-score = 0.447
CNS
Leukemia
Others
CALD1 (log2TPM)
NAMPT (D.S.)
B
CNS (LGG)
p = 3.22e-08
Survival probability
CALD1 Low
CALD1 High
Time (day)
Leukemia (LAML)
p = 0.434
Time (day)
C
PAAD
True positive rate (%)
False positive rate (%)
C-score AUC
>0.15 57.6%
>0.2 58.1%
>0.25 60.1%
>0.3 64.1%
>0.35 66.6%
>0.4 65.5%
MESO
True positive rate (%)
False positive rate (%)
C-score AUC
>0.15 54.2%
>0.2 54.3%
>0.25 55.8%
>0.3 57.9%
>0.35 61.3%
>0.4 63.5%
D
PAAD
pathological stage
True positive rate (%)
False positive rate (%)
C-score AUC
>0.15 59.8%
>0.2 59.7%
>0.25 61.7%
>0.3 63.4%
>0.35 65.2%
>0.4 68.1%