Chromosome 7 gain compensates for chromosome 10 loss in glioma

Nishanth Ulhas Nair[1,#], Alejandro A. Schäffer[1,#], E. Michael Gertz[1,#], Kuoyuan Cheng[1,2], Johanna Zerbib[3], Avinash Das Sahu[4], Gil Leor[3], Eldad D. Shulman[1], Kenneth D. Aldape[5], Uri Ben-David[3], Eytan Ruppin[1,6,*]

1. Computational Precision Oncology Section, Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.
2. MSD, Beijing, China.
3. Department of Human Molecular Genetics and Biochemistry, Faculty of Medicine, Tel Aviv University, Tel Aviv-Yafo, Israel.
4. The University of New Mexico, Comprehensive Cancer Center, Albuquerque, NM, USA.
5. Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.
6. Lead contact

# -- equally contributing co-first authors
* -- corresponding author
Contact Information for the Corresponding Author:
Dr. Eytan Ruppin
NCI/NIH
Bldg. 15-C1
Bethesda, MD 20892 USA
E-mail: eytan.ruppin@nih.gov
Phone: 240-858-3169

Running Title: Compensatory aneuploidy in glioma

Statement of Significance: Increased expression of multiple rescuer genes on gained chromosome 7 could compensate for downregulation of several vulnerable genes on lost chromosome 10, resolving the long-standing mystery of this frequent co-occurrence in gliomas.

Declaration of interests

E.R. is a co-founder of MedAware Ltd and a co-founder (divested) and non-paid scientific consultant of Pangea Therapeutics. K.C. is currently an employee at MSD. All other authors have no competing interests.

ABSTRACT

The co-occurrence of chromosome 10 loss and chromosome 7 gain in gliomas is the most frequent loss-gain co-aneuploidy pair in human cancers. This phenomenon has been investigated since the late 1980s without resolution. Expanding beyond previous gene-centric studies, we investigated the co-occurrence in a genome-wide manner taking an evolutionary perspective. Mining of large-scale tumor aneuploidy data confirmed the previous finding of a small-scale longitudinal study that the most likely order is chromosome 10 loss followed by chromosome 7 gain. Extensive analysis of genomic and transcriptomic data from both patients and cell lines revealed that this co-occurrence can be explained by functional rescue interactions that are highly enriched on chromosome 7, which could potentially compensate for any detrimental consequences arising from the loss of chromosome 10. Transcriptomic data from various normal, non-cancerous human brain tissues were analyzed to assess which tissues may be most predisposed to tolerate compensation of chromosome 10 loss by chromosome 7 gain. The analysis indicated that the pre-existing transcriptomic states in the cortex and frontal cortex, where gliomas arise, are more favorable than other brain regions for compensation by rescuer genes that are active on chromosome 7. Collectively, these findings suggest that the phenomenon of chromosome 10 loss and chromosome 7 gain in gliomas is orchestrated by a complex interaction of many genes residing within these two chromosomes and provide a plausible reason why this co-occurrence happens preferentially in cancers originating in certain regions of the brain.

INTRODUCTION

Gliomas, including glioblastoma (GBM) and lower-grade gliomas (LGG), have poor prognosis (1). Since the 1980's it has been known that loss of chromosome 10 (10 loss) and gain of chromosome 7 (7 gain) often co-occur in GBM (2-4). These two copy number alterations (CNAs) occur predominantly in IDH-wildtype GBM (5, 6), has worse prognosis than the *IDH1/IDH2*-mutant form (7). Within the IDH-wildtype GBM category, the presence of 10 loss/7 gain was not significantly associated with prognosis in a multivariate clustering approach (8), but segmental loss of intervals within chromosome 10 was associated with a better prognosis than the loss of the entire chromosome 10 (9). Previous studies aiming to order the two CNAs suggested that both 10 loss and 7 gain are early events in tumorigenesis that can occur in either order, but with 10 loss more commonly preceding 7 gain (3, 5, 6, 10). Via longitudinal data analysis, Körber et al. (10) showed that either 10 loss first or 7 gain first is possible. Our analysis focuses on trying to explain why 7 gain and 10 loss co-occur more often than would be expected by chance, which was not considered in their analysis.

Genomic studies in the 1990's and 2000's used microsatellite marker genotyping and comparative genomic hybridization (CGH) to characterize segmental aneuploidies to pinpoint "critical regions" on the two chromosomes, especially on chromosome 10 (11-14). More recent studies used genotyping arrays and high-throughput sequencing (15-17). Position-based marker techniques were coupled with gene cloning techniques to identify possible culprit genes on chromosomes 10 and 7. One idea was to search for genes whose expression change correlates with the copy number change, and sometimes considering additional functional evidence (6, 15-17). A related method was to seek genes that have frequent somatic mutations correlated with changes in expression (18).

Analyses of chromosome 10 identified three regions that are most frequently lost, one on 10p and two on 10q (11, 13, 19-22). In the region closest to the 10q telomere, the tumor suppressor gene *PTEN* was cloned (23-25) and identified as a likely suspect in GBM pathogenesis (12, 23, 24). The fact that *PTEN* is often the target of somatic mutations in the retained copy of chromosome 10 bolstered the evidence (26), but a meta-analysis of over 10,000 patients suggested that *PTEN* loss is unlikely sufficient to explain all chromosome 10 losses (27).

Even though 10q is lost in approximately 80% of GBMs, the biallelic inactivation of *PTEN* occurs only in about 40% of cases, and GBMs without *PTEN* biallelic inactivation showed similar *PTEN* expression levels as samples in which 10q was not lost (28). The gene *ANXA7* has been suggested as another possible culprit in a frequently deleted region of 10q closer to the centromere (29). *ADARB2* and *KLF6* have been suggested as possible culprit genes on 10p (20).

Gain of chromosome 7 usually covers the whole chromosome (6), but some segmental CNAs of various intervals on chromosome 7 were detected (14). The most commonly suggested culprit genes are *EGFR* on 7p (4, 27, 30, 31) and *MET* on 7q (31). Several studies have claimed that specific gene pairs such as *PDGFA* on chromosome 7 and *PTEN* on chromosome 10 (6) or small sets of genes fully explain the co-occurrence of 7 gain and 10 loss (6, 32-34). EGFR is one of the most studied human receptor tyrosine kinases, but the mysteries of EGFR and its relationship to chromosome 7 gains are still being elucidated (35). *EGFR* amplifications often co-occurs with aberrant transcripts, especially a transcript denoted EGFRvIII, which lead to constitutive EGFR signaling (36). *EGFR* has kinase-dependent and kinase-independent pro-tumorigenic functions (35). *EGFR* aberrations in GBM may manifest in at least six ways that are not mutually exclusive: i) activating point mutations, ii) genomic amplification, iii) chromosomal rearrangements, iv) autocrine signaling, especially between mutant forms of EGFR, such as EGFRvIII, v) intragenic duplication of the kinase domain, vi) *EGFR* fusions with pieces of other genes (35). High-level *EGFR* amplifications may occur either on double minute chromosomes or via insertions of extra copies of *EGFR* on chromosome 7 (37). *EGFR* aberrations in GBM may be clonal or subclonal (37, 38). A study of 86 GBMs found that *EGFR* amplification occurs with higher probability in samples that have a gain of chromosome 7 (82.1%) compared to samples that do not (66.7%), but all four combinations for *EGFR* amplification or not and chromosome 7 gain or not were observed (39), which qualitatively confirmed an earlier study (40). Chromosome 7 copy number gains to trisomy or moderate polysomy show no clear association with *EGFR* expression, although *EGFR* amplifications to double-digit numbers of gene copies do show a significant association with *EGFR* expression (41). The mixed association results yield doubt that *EGFR* is the sole culprit gene among GBMs harboring a gain of 7p encompassing *EGFR.*

In summary, no small set of genes has yet been identified that can fully explain the frequent 10 loss/7 gain co-occurrence. Recent literature supports the notion that aneuploidy is shaped by the accumulation of multiple haploinsufficient and triplosensitive genes (42, 43). Therefore, in this work, instead of aiming to identify specific genes whose losses or gains on chromosome 10 and chromosome 7 can provide an explanation to the 10 loss/7 gain double event, we take a non-reductionist approach by analyzing chromosomes 10 and 7 as two large collections of genes that may interact, driving the 10 loss/7 gain combination. First, we develop a mathematical model that shows under plausible probabilistic assumptions that 7 gain after 10 loss is significantly more likely than the opposite case and that the less frequent order of 10 loss after 7 gain can be treated as random. These analytical findings based on a large-scale database confirm two longitudinal, wet-lab studies (5, 10) and remarkably reach an estimate comparable to that of Körber et al. (10) that 10 loss occurs first in ~60% of tumors with both aneuploidies while 7 gain occurs first in 40% of such tumors. Next, by analyzing patient tumor and cell line datasets, we provide an evolutionary explanation of why the combination of 10 loss and 7 gain aneuploidies is so prevalent in gliomas. Finally, by analyzing transcriptomic data from normal, non-cancerous human brain tissues, we provide a plausible reason why 10 loss and 7 gain co-occurrence happens preferentially in cancers originating in certain regions of the brain.

MATERIALS AND METHODS

Analysis of Progenetix data

We downloaded the metadata from tumors from 118,238 patients from the Progenetix (44) database on January 5, 2023 (https://progenetix.org/). The metadata included histological diagnosis identifiers, which are specified using a hierarchical system of National Cancer Institute (NCI) Thesaurus Terms. We searched this ontology of terms to identify patients that could be unambiguously assigned to 24 TCGA cancer types. Specifically, we used the “rols” package in GNU R to download and search the ontology. NCI Terms and TCGA cancer types are not exactly one-to-one matches, but GBM was an unambiguous diagnosis in Progenetix.

We then downloaded segmental copy-number data from Progenetix, also on January 5, 2023. A total of 50,392 patients could be unambiguously identified with a unique TCGA type and had CNV data. We further eliminated those samples in the TCGA projects and randomly

selected one tumor sample from each person having more than one sample, which left 39,085 persons and 39,085 samples. We used only tumor samples, not matched normal, but some persons had more than one tumor sample. We chose one primary tumor sample per person randomly, yielding 39,085 tumors. Samples in Progenetix sometimes have more than one set of CNV calls, possibly due to multiple sequencing runs, but this only affects individuals in TCGA. The segment data in the Progenetix database do not distinguish ambiguous regions from calls of neutral copy number. Nor does the technology used for much of the Progenetix database, comparative genomic hybridization (CGH), identify the ploidy of the sample. It records deviations from neutral copy-number for that sample.

Calling arm changes from Progenetix data

We identified arm changes using the following rules. We ordered the segments between the telomere of the arm and the centromere by position. Segments do not overlap, so there is no ambiguity in this order. We treated any segments not called by Progenetix for which

$$|\log_2(\text{fold change})| < \log_2\frac{5}{4} - 10^{-4}$$

as having neither a gain nor loss call. The constant 5/4 was selected because it represents a gain on a ploidy background of 4 (the actual ploidy is unknown). The tolerance 10^{-4} compensates for the limited precision with which Progenetix saves its fold change numbers.

The most telomeric number change in this filtered ordered list determined the putative direction of arm change and must have been called within the telomeric 5 Mbp region of the chromosome arm. We then scanned the list from telomeric end to centromere, counting any contiguous sequence of segments either called in the opposite direction or having no call as a single gap. If the copy number change could be extended toward the centromere allowing at most 3 gaps and the number of bases called in the extension (not the length of the extension) covers 60% of the chromosome arm, then that arm was considered to be lost or gained. Progenetix does not distinguish between calls of copy number neutrality and regions that could not be called, which makes it impossible to use a threshold derived from the percentage of bases in the chromosome. We found a threshold that 60% of an arm's bases be gained or lost to be a reasonable value because it captured large losses on the telomeric end of an arm while allowing

for large uncalled regions. We experiment with a stricter 80% threshold, which was the threshold used by Taylor et al. (45) for the TCGA dataset, but it did not result in a subjective difference in GBM. Using a rule that gain or loss of a chromosome arm was recorded a gain or loss of a chromosome, we performed a similar series of Fisher exact tests and once again found that loss of 10 with gain of 7 stood out as highly significant and the most strongly significant event.

Calling chromosome loss/gain co-occurrences

To test for all co-occurring chromosome arm changes, for each cancer type, we used a Fisher exact test to determine whether the pair of arm changes occur more often than by chance given the rate of occurrence of the individual arm changes. Tests were one-sided because we only kept cooccurrences that happen more often than expected. We only included samples for which there was evidence of aneuploidy, which we define here as the evidence of an arm change on any (non-acrocentric) autosome. We only considered loss/gain pairs because the CGH technology is relative to overall chromosome content, and we can be most confident that a loss/gain pair, having a change in each direction, reflects true events and not changes in the content of unrelated chromosomes. We use the FDR correction to compensate for multiple hypothesis testing (46).

Comparison of Progenetix analysis with previously published data

We compared our results to loss/gain pairs in Supplementary Table 4 in Prasad et al. (47). We considered only loss/gain pairs, and therefore took the cancer specific pairs and p-values from Prasad et al, restrict them to the 22 TCGA cancer types we were able to identify, and recompute the FDR correction based on the pairs retained. From our own list, and the list of Prasad et al., we keep only those items with FDR-adjusted $P \leq 0.05$. To test whether the list in Prasad et al. is enriched within our own, we use a single two-sided Fisher exact test for which we consider the universe to be all possible autosome arm gain/loss pairs among the 22 cancer types, ignoring the short ends of the five acrocentric chromosomes (13, 14, 15, 21, 22).

DU-SR analysis

INCISOR is a computational method(48) that identifies clinically relevant SR interactions that are supported by all these four steps to predict DU SR gene pairs.

a) Essentiality Screens: INCISOR mines essentiality data from hundreds of pan-cancer cell lines such as (49) to identify genes that, when knocked down, can be rescued by the upregulation of another gene. In each gene pair, the first gene is called the vulnerable (V) gene and the second gene is called the rescuer (R) gene.
b) Survival of the Fittest: INCISOR uses hundreds or thousands TCGA patient tumor (genomic and transcriptomic) data to find SR pairs that appear more frequently in their rescued state (i.e., gene V is inactive and gene R is specifically upregulated), indicating positive selection.
c) Patient Survival: Using a stratified Cox proportional hazard model, INCISOR selects SR pairs (when gene V is inactive and gene R is upregulated) linked to worse patient survival, accounting for various confounding factors.
d) Phylogenetic Screening: Finally, INCISOR picks SR pairs with high phylogenetic similarity as the most likely candidates.

More details regarding the INCISOR pipeline can be found in Sahu et al. (48). We ran INCISOR on 664 TCGA GBM and LGG patient samples and on hundreds of pan-cancer cell lines at FDR < 0.1 to identify DU-SR interactions in brain tumors. We also ran INCISOR on 8,085 TCGA pan-cancer patient tumor samples. For the pan-cancer analysis, we removed brain cancers from both cell line and patient datasets to avoid identifying brain-specific synthetic rescues which should be enriched in the previous TCGA test. Given the larger sample sizes a more stringent threshold of FDR-adjusted $P < 0.05$ was used in the pan-cancer analysis to identify DU-SR interactions.

We also did the DU-SR analysis on TCGA pan-cancer patient tumor samples (after excluding brain tumors) by separately running INCISOR on TCGA samples with (n = 792) and without (n = 6130) whole genome duplications with FDR threshold of 0.05. The analysis was also repeated by randomly sampling 792 out of 6130 samples without whole genome duplications. The annotations for presence/absence of WGD were reused from Taylor et al. (45).

Enrichment analysis of rescuers on a specific chromosome

For all the genes on specific chromosome arms (e.g., chromosome arm 10q), we predicted DU-SR interactions in a genome-wide manner. Then, we looked for enrichment of rescuer genes among the genes on some chromosome (e.g., chromosome 7) and checked if they occurred more often on that chromosome than expected by random chance (Fisher exact test). We did a similar enrichment analysis for each chromosome arm. Statistical tests were corrected for multiple hypothesis testing using the FDR (false discovery rate) correction (46).

Pairwise pathway enrichment analysis

We identified the pathway pairs enriched by using DU-SR gene pairs (generated from TCGA GBM+LGG data) and carried out a pairwise pathway enrichment analysis among the vulnerable genes located on chromosome 10 and their rescuer genes on chromosome 7, using 50 Cancer Hallmark pathways (MSigDB's hallmark gene sets (50)). We employed a one-sided Fisher exact test to examine whether the gene pairs (comprising vulnerable and rescuer genes) are statistically enriched in specific pairs of pathways. The aim of this analysis was to identify pathways whose dysfunction can be mitigated by the activity of rescuer genes in other pathways. FDR correction was applied to adjust for multiple hypothesis testing.

Cell line essentiality analysis

We used CRISPR gene-knockdown essentiality screen data from DepMap (49) for our cell line essentiality analysis. Arm change event information for cell lines was taken from previous literature (51) (bioRxiv 2023.01.27.525822). We then divided CNS cancer cell lines into two groups based on arm-level copy-number events. All CNS cell lines in the essentiality screen were wildtype for *IDH1/2* mutations. Genes on either chromosome 10 or chromosome 7 with differential essentiality between the two groups are identified using a one-sided Wilcoxon rank-sum test ($P < 0.05$). A similar analysis is also done by considering all genes. The number of more/less essential genes obtained for a specific chromosome or arm is then compared with the corresponding numbers obtained via the all-gene analysis using a two-sided Fisher exact test.

For the analysis of the isogenic RPE1 clones, CERES dependency scores of 3 clones – RPE1-SS48 (diploid), RPE1-SS6 (trisomy 7) and RPE1-SS119 (trisomy 8) – were retrieved from Zerbib et al. (bioRxiv 2023.01.27.525822). Genes with CERES score < -0.2 in the WT clone

SS48 were included in the analysis. A non-parametric paired ANOVA was performed to determine the difference between the clones.

Activity scores

Arm level copy-number data for TCGA patient tumors were sourced from Taylor et al. (45). Taylor et al. considered chromosome arms with ≥ 80% change as aneuploid (value of +1 for gain or -1 for loss) and <20% as non-aneuploid (value 0) (details in (45)). Gene expression levels were categorized as high or low based on percentiles across all TCGA GBM+LGG patients: > 67th percentile for high expression and ≤ 33rd percentile for low expression. An 'Activity Score' was computed as the fraction of genes on chromosome 7 with high expression within specific gene subsets: (i) rescuer genes, (ii) top hub rescuer genes (those that mapped to ≥10 or ≥70 vulnerable genes on chromosome 10 using TCGA GBM+LGG or pan-cancer data, respectively), (iii) non-rescuer genes on chromosome 7, and (iv) all genes on chromosome 7.

Normal non-cancerous brain tissue analysis

We used RNA-seq TPM data (log-scale) from 13 GTEx brain tissues containing 2642 normal non-cancerous samples. For each tissue, we further rank normalized (from 0 to 1) the expression data first across genes for every sample, and then across samples for each gene. A gene was considered to have low expression in a sample if its expression was less than or equal to the 33rd percentile of its expression across all samples in that tissue. The DU-SR network was derived using either TCGA brain cancer (GBM, LGG) data or using pan-cancer data without brain tumors. cSR load in a normal sample is defined as the fraction of DU-SR pairs (out of all DU SR pairs on chromosomes 10 and 7) where the vulnerable gene on chromosome 10 has low expression (inactive) and the rescuer gene on chromosome 7 does not have low expression (active). Tissue cSR load was computed as the median value of all sample specific cSR loads in a specific normal tissue. These tissue cSR loads were computed for all 13 brain tissues and rank normalized (from 0 to 1).

Control experiments were carried out by first randomly generating 1000 ‘cSR’ (sub)networks of the same size (i.e., same number of gene pairs) as our original network from a large network of all possible distinct gene pairs from across the genome. As before, rank

normalized random-cSR loads were computed for all 13 brain tissues for each random network. An empirical p-value of both these two tissues having the two high values they have in the observed data based on their cSR load in comparison to random controls was computed as follows: the number of random networks in which the rank normalized random-cSR loads of both these two tissues (frontal cortex and cortex) are greater than or equal to the minimum observed rank normalized cSR loads of either one of these two tissues (using the actual cSR network) divided by 1000.

Data/code availability

The TCGA data was obtained from (https://portal.gdc.cancer.gov/). The GTEx data was obtained from (https://xenabrowser.net/datapages/). The DepMap data was obtained from (https://depmap.org/portal/data_page/?tab=overview). The RPE1 data was obtained from this publication (https://www.biorxiv.org/content/10.1101/2023.01.27.525822v1). All the biological patient and *in vitro* data used for this study was taken from previous literature (no new biological data was generated). Tumor metadata and segmental copy number data were obtained from the Progenetix database at (https://progenetix.org/).

The R code used to generate the main results in this paper has been made available for reproducibility: https://hpc.nih.gov/~Lab_ruppin/code_for_paper_braincancerproject.zip

RESULTS

Analysis Overview

We use the phrasing "10 loss" or "7 gain" to denote loss of either arm (or the whole chromosome) of chromosome 10 or gain of either arm of chromosome 7, respectively (precise rules for calling an arm as lost or gained are provided in **Methods**). The terms *10 cnn* or *7 cnn* indicate a copy number neutral state where there is no arm loss and no arm gain of chromosome 10 or chromosome 7, respectively, compared to the modal ploidy (**Methods**).

Our analysis proceeds in four main steps:

1. *Mathematical Modeling:* We developed mathematical models to study the order of co-occurring aneuploidies. Applying these models, we showed that the probability of 7 gain occurring after 10 loss is significantly greater than 10 loss after 7 gain, finding that the less frequent sequence 10 loss after 7 gain can be treated as occurring by random chance.
2. *Data Mining for identifying Genetic Interactions:* We analyzed hundreds of genomic and transcriptomic samples from brain cancer patients and cell lines. Our aim was to identify a category of clinically significant genetic interactions that functionally compensate for the genes located on the lost arm (synthetic rescues). This step revealed that genes on chromosome 10, when lost, can be best functionally compensated by the upregulation of genes on chromosome 7, compared to genes on other arms. We also showed that glioma patients with 10 loss and 7 gain events tend to have worse overall survival, testifying to enhanced underlying tumor fitness.
3. *CRISPR Screening for Fitness Effects:* Analyzing large-scale *in vitro* CRISPR essentiality screens in central nervous system (CNS) cancer cell lines, we showed that the fitness effects of different possible sequences of chromosomal arm alterations further support the capacity of 7 gain to compensate for 10 loss.
4. *Normal tissue analysis:* Although some of our analyses in parts 1-3 use brain cancer data, those analyses do not explain why the co-occurrence of 10 loss and 7 gain is so high in specific brain cancers (such as GBM) which preferentially arise from certain regions of the brain, but not in cancers from other body locations. Therefore, we analyzed expression data from thousands of normal non-cancerous samples from human brain tissues. From these analyses, we provide evidence that the normal transcriptome state in the cortex or frontal cortex (which are common tissues of origin for GBM), is permissive of a 10 loss and 7 gain co-occurrence in cancers arising from these tissues.

Chromosome 10 loss and chromosome 7 gain co-occurrence frequency and their associations with patient survival in gliomas

We analyzed over 39,000 patient tumors in the Progenetix database (44, 52), omitting tumors in the database that were also in The Cancer Genome Atlas (TCGA) cohort (53) to avoid replicating analysis performed on the TCGA. We quantified the relative prevalences of the 10 loss and 7 gain chromosomal events in glioblastoma in comparison to other loss-gain co-

aneuploidy events occurring in other types of cancer. In this analysis, we considered p and q arms separately and focused on autosomes because Progenetix does not curate the sex of subjects. We did not consider specific gene alterations, because aside from TCGA data, no data on gene-specific alterations are available in Progenetix.

We identified 1,280 significantly co-occurring chromosome arm loss-gain event pairs out of 31,122 possibilities (Fisher exact test, false discovery rate or FDR < 0.05; **Table S1**; **Methods**). We find a significant overlap (Fisher exact test, Odds ratio (OR) = 15.0, P = 1.65e-39, **Methods**) between our identified co-occurring loss-gain events and events described by Prasad et al., which examined the TCGA cohort (53). Notably, we find that the four pairs (7q gain, 10p loss), (7p gain, 10q loss), (7p gain, 10p loss), (7q gain, 10q loss) in GBM are the four most significantly enriched co-occurring chromosome loss-gain events across all such co-aneuploidy events occurring in any cancer type (**Fig. 1A**; robustness studies for different parameters shown in **Fig. S1**). Since the four two-arm co-aneuploides are at the top of the list, we infer all four are important, even though specific tumors may have an aneuploidy affecting only one arm on either chromosome 7 or 10. The numbers of GBM patient samples in the Progenetix data where loss or gain occurs on chromosomes 10 or 7 are summarized in **Fig. 1B**. In follow-up analyses we considered chromosome-arm and whole-chromosome events together, that is, an event is a 10 loss if either 10p or 10q arm (or both) is lost and an event is a 7 gain if either 7p or 7q (or both) is gained.

Probabilistic model to estimate the probability of 10 loss occurring before or after 7 gain

Given the distribution of 10 losses and 7 gains that we found in the Progenetix data, we asked whether we could determine which of the two changes, 10 loss or 7 gain, occurs preferentially, but not exclusively, earlier. We hypothesize in our subsequent analysis that the aneuploidy that occurs second may compensate for the aneuploidy that occurs first. Intuitively, if there are many more tumors that have 10 loss without 7 gain than tumors that have 7 gain without 10 loss, this suggests that 10 loss is more likely to come first.

To study if 10 loss is likely to occur before 7 gain, one may compute the fraction of tumors with only a loss of 10 among all tumors for which one of these events occurs (either 10

loss or 7 gain but not both, **Fig. 1C,** see mathematical details in **Supplementary note 1**). Considering the counts of 7 gain and 10 loss in Progenetix, we find that fraction is $555/(555 + 339) \approx 0.625$ (based on the counts from **Fig. 1B**), which is different from the 1/2 expected (if the two aneuploidies occur in either order with equal probability) by a one-sided binomial test with a p-value of 2.50e-13. This indicates that 10 loss occurs significantly more frequently before 7 gain than the opposite order. Our model confirms at scale the empirical findings of Körber et al. (10) that 10 loss first and 7 gain second is the more likely order.

Next, we study why the co-occurrence of 10 loss and 7 gain happens more frequently than expected if the two aneuploidy events were happening independently without any functional compensatory forces. To this end, we develop a 'mixture model' in which we posit that there are two types of tumors that have both 7 gain and 10 loss: one type has the events occurring at random in either order and the other has the events occurring predictably in the order 10 loss first and 7 gain second. Intuitively, if these 'predictably ordered' tumors are frequent enough then they are sufficient to explain the excess co-occurrence of 10 loss and 7 gain. The idea that a formal mixture of two models for tumor progression can explain the data better than a single model been used repeatedly in tumor phylogenetics (54-57). Our intent is to prove that a mixture of two models explains the data better than one model, but two models suffice to allow us to proceed to the next step of analysis, concerning synthetic rescue interactions.

Using the counts given in **Fig. 1B**, we compute that the excess probability is entirely attributable to 10 loss after 7 gain if the probability of 10 loss coming first is 0.625 (details of the computation are provided in **Supplementary note 2**; **Fig. S2**), which exactly matches the estimate given above (and detailed in **Supplementary note 1**) from the Progenetix observed data. Reassuringly, the data provided in Körber et al. (10) suggest an empirical estimate of $x \geq 0.595$, which is not significantly different from 0.625 (binomial test). The analysis of Körber et al. uses time-series data on tumors not in Progenetix, so there is no circularity or self-reference when we compare our data analysis based on Progenetix to the previous study. We conclude that all the excess probability of co-occurrence can be explained by tumors in which deterministically 10 loss is first and 7 gain is second. This suggests that the opposite event of 7 gain first and 10 loss second can be treated as occurring by chance.

Co-occurrence of 10 loss and 7 gain is associated with worse survival in glioma patients

We analyzed the TCGA GBM and LGG clinical genomic data (28, 58, 59) to ask whether 10 loss, 7 gain, and their co-occurrence, are associated with the patients' clinical outcome. This analysis reveals that patients with 10 loss and 7 gain events have significantly shorter overall survival than patients whose tumors with 10 loss and 7 cnn ($P = 0.03$), 10 cnn and 7 gain ($P < 0.0001$), 10 cnn and 7 cnn ($P < 0.0001$) (**Fig. 2A**, with more detailed pairwise survival analysis provided in **Fig. S3A-E**). Our survival analysis differs substantially from two previous publications (8, 9), which grouped the patients using other genomic features. Survival keeps worsening with each event in the following order: 10 cnn and 7 cnn (maximum survival), 10 cnn and 7 gain, 10 loss and 7 cnn, 10 loss and 7 gain (worst survival). These trends remain if we consider only patients without an *IDH1/2* mutation or GBM patients only (**Fig. S3F-G**). This testifies that 10 loss solely leads to a more aggressive form (lower survival) than 7 gain only, and 10 loss and 7 gain is the most aggressive state. This suggests that the co-occurrence of 10 loss and 7 gain increases the fitness of tumors, in comparison to each of the individual aneuploidies.

Genes on chromosome 10 are enriched for functional rescuer genes on chromosome 7, testifying for 7 gain capacity to compensate for 10 loss

Loss of many genes on a chromosomal arm is likely to be disadvantageous to cancer cells (60-63). We investigated whether at least some negative effects for gliomas due to the loss of many genes on chromosome 10 could be rescued by gains of genes on chromosome 7.

Synthetic rescue (DU-SR) interactions are a form of functional interplay in which the detrimental impact on cell fitness caused by the inactivation of a specific gene (referred to as the *'vulnerable gene'*) is compensated by upregulation of another gene known as the *'rescuer gene*' (48) (**Fig. 2B**). We previously published a computational pipeline called 'IdeNtification of ClinIcal Synthetic Rescues in cancer' (INCISOR), which mines genomic, transcriptomic, and phenotypic data from hundreds of cancer patients and cell lines to identify DU-SR interactions that may be clinically relevant (48, 64) (**Methods**). INCISOR identifies DU-SR interactions using hundreds of *in vitro* essentiality screens from cancer cell lines, and then filters those interactions for positive selection and worsening survival association in patients' tumors, and for

interactions between gene pairs with high phylogenetic similarity across divergent eukaryotic species (48). Many of the predicted clinically relevant SR interactions were validated through experimental *in vitro* screens and by their ability to predict cancer drug response in patients (48, 65). While the terms 'vulnerable' and 'rescuer' are used for historical reasons (rescuer genes are identified for 'vulnerable' drug targets), INCISOR identifies any gene pairs where the downregulation of a gene and the upregulation of the partner gene is beneficial for cancer, over and above their individual gene effect (48).

We applied INCISOR to analyze 664 TCGA GBM and LGG patients to predict genome-wide clinically relevant DU-SR interactions for the (vulnerable) genes on chromosome 10 (**Methods**, **Table S2A-D**). Remarkably, we found that among all chromosome arms, 7p and 7q have the highest enrichment of rescuer genes for the vulnerable genes in 10p and 10q, in comparison to what is expected by chance (overlap enrichment test using one-sided Fisher exact test, FDR < 10^{-19}, **Fig. 2C, Table S3A,** with more results in **Table S3B-H, Methods**). All four candidate arm pairs are the most significantly enriched co-occurring arm pairs in the order (7q, 10p), (7q, 10q), (7p, 10p), (7p, 10q). The order of the four pairs is not identical to the order in the Progenetix analysis (**Fig. 1A**), but the pair (7q, 10p) is most significant in both the Progenetix and the synthetic rescue analysis. We observe that the rescuers on chromosome 7 are positionally spread throughout the chromosome (**Fig. 2D;** plot generated using karyoploteR (66)), which may explain why gains of chromosome 7 in brain cancer often involve whole chromosome arms rather than smaller regions (**Discussion**). This analysis testifies that the loss of genes on chromosome 10 in gliomas could be compensated by the amplification of many rescuer genes on chromosome 7.

To avoid concerns that the DU-SR results were only due to the high prevalence of brain cancers with 7 gain and 10 loss, we repeated the DU-SR analysis on TCGA GBM and LGG patients while removing an INCISOR step that searches patient data that looks for gene pairs with positive selection based on their expression or copy-number status. Reassuringly, this analysis yielded similar results (**Fig. S4**; there are no circularity concerns for the other steps of INCISOR). Next, we performed the DU-SR prediction analysis on the pan-cancer TCGA data *after removing GBM and LGG tumor samples* (**Methods**, **Table S4A-E**). The strong enrichment

of rescuer genes in chromosome 7 for the vulnerable genes in chromosome 10 persisted (**Table S3B**).

Previous attempts to identify specific culprit genes (see **Introduction**) focused on the paradigm of tumor suppressor genes, such as *PTEN*, on chromosome 10 and on oncogenes, such as *EGFR*, on chromosome 7. In contrast, the paradigm of synthetic rescue is not limited to rescuing tumor suppressor genes that have been inactivated. Among the genes on chromosome 7, the COSMIC Cancer Gene Census (67) lists 23 oncogenes and among the genes on chromosome 10, COSMIC lists 16 tumor suppressors, most of which are not implicated in brain cancer of any subtype (**Table S3F,G**). Indeed, consistent with our hypothesis that tumor suppressors on 10 need not be rescued by oncogenes on 7, we find that among the 903 DU-SR pairs) listed in **Table S2B**, there are only four (oncogene on 7, tumor suppressor on 10) pairs, *(TRRAP, CCDC6), (TRIM24, KAT6B), (EZH2, SUFU) (TRRAP, SUFU)*, none of which involves a gene implicated (by COSMIC) in brain cancer. One might also wonder whether rescuer genes are important if oncogenes on chromosome 10 are lost but none of the 12 COSMIC oncogenes on chromosome 10 has been implicated (according to COSMIC) in any form of adult brain cancer (**Table S3H**). Accordingly, these 12 chromosome 10 oncogenes have at most 3 rescuer genes on chromosome 7 (**Tables S2C, S3H**).

Next, we compared gene expression patterns between tumor samples from TCGA GBM/LGG patients who have lost chromosome 10 with those with a 10 copy-number neutral (cnn) state (chromosome 7 is in a copy number neutral state in both groups). We formulated an "activity score" based on the proportion of a defined set of genes on chromosome 7 that have high expression levels when chromosome 10 is lost (**Methods**). There was overall increased expression of genes on 7 (i.e., higher activity scores) when chromosome 10 is lost. Heightened expression is pronounced amongst the predicted rescuer genes — especially those capable of rescuing many vulnerable genes on chromosome 10 — compared to non-rescuer genes (**Fig. 2E**). In other words, when chromosome 10 is lost, even when chromosome 7 is not gained, *the rescuer genes that lie on 7 are highly expressed.* Repeating this analysis while excluding brain tumors yielded congruent results (**Fig. 2E**).

The DU-SR analysis (by analyzing GBM and LGG data) predicts 237 vulnerable genes and 272 rescuer genes on chromosomes 10 and 7, respectively. Some previously suggested candidate genes *PTEN* and *ADARB2* were among the predicted vulnerable genes on chromosome 10 (see **Introduction**), whereas some genes previously reported to drive 7 gain, such as *EGFR*, *MET*, *BRAF*, where among the identified rescuers (**Table S2B**). **Table S2D** provides a list of rescuers, ranked by the number of vulnerable genes they interact with.

We carried out a pairwise pathway enrichment analysis among the vulnerable genes located on chromosome 10 and their rescuer genes on chromosome 7, using 50 Cancer Hallmark pathways (MSigDB's hallmark gene sets (50); **Methods**). This aims to identify pathways whose dysfunction can be mitigated by the activity of rescuer genes in other pathways. Vulnerable genes on chromosome 10 that are involved in glycolysis, heme metabolism, and hypoxia are rescued by chromosome 7 genes involved in pathways including Notch signaling, Wnt beta-catenin signaling, and epithelial-mesenchymal transition. The full list of such enriched pathway pairs is summarized in **Fig. 2F, Fig. S5, Table S5A-B**.

To test whether the loss or gain of these chromosomes could be driven by a few key genes, we removed previously reported well-known genes among those present in DU-SR network (*PTEN*, *ADARB2* on chromosome 10 and *MET*, *BRAF*, *EGFR* on chromosome 7). Rescuer interactions for the vulnerable genes on chromosome 10 were still highly enriched with genes in chromosome 7 (FDR < 10^{-19}, **Table S3C**). This suggests that the chromosomal events of 10 loss and 7 gain are orchestrated by a broad network of genes.

Our study analyzes of genes on chromosome 7 and chromosome 10 as broadly interacting sets of genes. Yet, it is also worthy to consider the genes that participate in many synthetic rescue pairs (**Table S4C-E**) individually and to explore previous evidence for why that may be the case. We present this analysis in **Supplementary note 3**.

Though not a focus of this paper, our analyses have some possibility of identifying chromosome arms that (like 7p and 7q) have unexpected co-occurrence of gains in conjunction with 10 loss and a concentration of rescuer genes. We identified co-occurring gains of either

arm of chromosome 19 as significant events (**Supplementary note 4,** based partly on **Table S3E**).

Synthetic rescuer interactions are detectable and functionally important in patients without whole genome duplications

The effects of 10 loss and 7 gain may depend on whether the tumor has undergone whole genome duplication (WGD): if a cell gains a second copy of the genome before either a chromosome 10 loss or a chromosome 7 gain, then the copy number proportion is at most 3 copies of chromosome 10 to at least 5 copies of chromosome 7, whereas if no WGD happened, the resulting copy number proportion is 1 copy of chromosome 10 to 3 or more copies of chromosome 7. Using the WGD presence/absence predictions of Taylor et al. (45) for TCGA samples, we tested whether the co-aneuploidy occurs more often in the absence of WGD since the effect would be greater. Among TCGA GBM and LGG samples, the combination of 7 gain and 10 loss occurs in 35/183 (19%) of samples predicted by Taylor et al. (45) to have WGD and a higher 139/449 (31%) of samples predicted not have WGD. The difference is moderately significant ($P < 0.003$) by a two-sided Fisher's exact test. However, the difference in proportions is not significant for GBM and LGG separately possibly due to smaller sample sizes. For GBM samples only, the combination of 7 gain and 10 loss occurs in 20/24 (83%) of samples with WGD and 91/119 (76%) of samples without WGD; for LGG samples only, the combination of 7 gain and 10 loss occurs in 15/159 (9%) of samples with WGD and 48/330 (15%) samples without WGD.

Furthermore, we aimed to find DU-SR interactions for the genes on chromosome 10 by using pan-cancer TCGA data (excluding brain tumors) while separately analyzing patients predicted to have or to lack whole genome duplication (WGD). For samples without WGD (n = 6,130), there was an enrichment of rescuer genes on chromosome 7 for the genes on chromosome 10 (one-sided Fisher exact test, Odds ratio = 1.38, P = 0.0018). Rescuers on 7q and 7p for genes on 10p and 10q were specifically highly enriched. (**Table S3D**). The number of tumors with WGD is much smaller (n=792), and within the WGD set, we were unable to find DU-SR interactions to carry out enrichment analysis. To check whether DU-SR interactions

were not detectable due to smaller sample size, we repeated the analysis on a random sample of 792 non-WGD tumors (out of 6,130) and found that we were unable to identify DU-SR interactions. This analysis suggests that DU-SR interactions are functionally relevant (and hence detectable) in cancers without WGD, however for cancers with WGD it is difficult to detect DU-SR interactions possibly because of low sample size.

Essentiality screens in CNS cell lines further testify to the fitness benefits of 7 gain in the presence of 10 loss

Next, we turned to study the potential selective advantage of 10 loss and 7 gain by analyzing the DepMap dataset of gene essentiality screens (via CRISPR-based knockouts) in IDH-wildtype CNS cancer cell lines (49, 68). First, we partitioned CNS cell lines with 10 loss into 7 gain and 7 cnn groups, and compared the relative essentiality between the two groups for the genes on chromosomes 10 and 7. We find that chromosome 10 genes tend to be less essential in the 7 gain group than in the 7 cnn group which is in line with our hypothesis that chromosome 7 gain is protective against chromosome 10 loss (two-sided Fisher exact test, P = 0.0034 in comparison to the all-gene analysis; **Methods**; **Fig. 3A, Table S6A**); however, chromosome 7 genes tend to be more essential in the 7 gain group (P = 0.0019, **Fig. 3A**). Besides, chromosome 7 genes also tend to be more essential in 10 loss and 7 gain cell lines than in the 10 cnn and 7 gain cell lines (P = 1.11e-23, **Fig. 3B, Table S6B**). In contrast, when comparing 7 gain vs 7 cnn groups in 10 cnn cell lines, chromosome 7 genes tend to be less essential with 7 gain (two-sided Fisher exact test, P = 9.75e-18 in comparison to the all-gene analysis; **Fig. 3C, Table S6C**). Since 7 gain in the presence of 10 loss is associated with decreased essentiality to chromosome 10 genes and increased essentiality to chromosome 7 genes, this suggests an evolutionary pressure to maintain 7 gain in the presence of 10 loss, consistent with an associated selective advantage. Furthermore, we analyzed a CRISPR screen of recently developed isogenic system of RPE1 cells molecularly engineered to have distinct single whole-chromosome trisomies (bioRxiv 2023.01.27.525822). The chromosome 10 genes that were essential in the near-diploid clone (RPE1-SS48) were significantly less essential in the clone with trisomy 7 (RPE1-SS6), but not in the clone with trisomy 8 (RPE1-SS119) (**Fig. 3D; Methods**). The RPE1 cell line analysis further supports the likely evolutionary benefit of gaining chromosome 7 after the loss of 10. Taken together, our

analysis of *in vitro* CRISPR essentiality screens show that the fitness effects of different possible sequences of chromosomal arm alterations further support the capacity of 7 gain to compensate for 10 loss, thus reinforcing the findings of our synthetic rescue work.

Why do 10 loss and 7 gain happen in certain regions of the brain rather than other regions?

Among the different kinds of brain cancers, 10 loss and 7 gain mainly co-occur in GBMs, which are more likely to arise in the supratentorial regions of the cerebral hemisphere (e.g., cortex or frontal cortex) (69, 70) than in other regions of the brain. While previous sections of this paper focused on studying cancer samples, here, to understand why this specific co-occurrence is likely to happen in cancers arising from certain regions of the brain, we analyzed gene expression data from 2,642 samples from 13 types of normal non-cancerous brain tissues from the Genotype-Tissue (GTEx) dataset (71). The GTEx analysis is motivated by a previous study that showed that the inferred synthetic lethality activity in normal tissues can explain tumor suppressors' role in cancers arising more frequently in specific tissues vs other tissues (72).

We had previously applied INCISOR to identify genetic interactions (DU-SR network) in chromosomes 10 and 7 where the inactivation of gene on chromosome 10 and its activation of its partner gene on chromosome 7 is beneficial for cancer. To study the potential effects of the INCISOR-derived DU-SR network (derived from brain tumor data) in normal non-cancerous tissues in analogous fashion to the analysis conducted in Cheng et al. (72), we defined a measure called *"cancer synthetic rescue (cSR) load*", which is a quantitative measurement of naturally occurring 'potent' DU-SR interactions in normal samples between gene pairs on chromosomes 10 and 7 (**Methods;** an overview of the approach is shown in **Fig. 4A**). cSR load in a single sample is defined as the fraction of DU-SR pairs on chromosomes 10 and 7 in which the gene on chromosome 10 has low expression (i.e., inactive) and the partner gene on chromosome 7 is not lowly expressed and hence likely to be active. Tissue cSR load is the median value of all single-sample cSR loads across all samples in a specific normal tissue.

Since 10 loss and 7 gain are early events in cancer development, we hypothesized that normal tissues with higher tissue cSR loads are more likely to develop 10 loss and 7 gain co-

occurrence during the process of carcinogenesis, since the latter testifies to the compensatory potential of cell survival after the loss of 10. As in (72), the DU-SR interactions on chromosomes 10 and 7 are derived from cancer and not normal samples, but we hypothesize that they could become functionally relevant as cancer develops.

We computed the Tissue cSR loads of 13 normal non-cancerous brain regions in GTEx samples. Our results show that among all these 13 brain tissues, the frontal cortex and cortex tissues have the highest tissue cSR loads, and notably, these are the two most relevant tissues to GBM in the GTEx cohort (**Fig. 4B**). Control experiments using randomly generated cSR networks did not show a similar pattern (empirical p-value of both these two tissues being ranked in top two based on their cSR load in comparison to random controls is 0.007; **Fig. 4B; Methods**). We repeated this analysis by computing DU-SR interactions (on chromosomes 10 and 7) in a pan-cancer manner by explicitly removing any brain tumor data. Reassuringly, we again see the cortex and frontal cortex as the two tissues with the highest cSR loads (empirical $P < 0.001$; **Fig. 4C; Methods**). These results suggest that the pre-existing transcriptomic state in the cortex and frontal cortex might tolerate and compensate for the loss of 10 by rescuer genes that are active on 7, whose active upregulated transcriptional state is then further fixed by the gain of 7.

DISCUSSION

The mystery behind the frequent co-occurring loss of chromosome 10 and gain of chromosome 7 in many gliomas has been investigated since the late 1980's. 43.2% and 77.6% of the GBM samples in the Progenetix (without TCGA) and TCGA databases respectively have a 10 loss and 7 gain co-occurrence. Prior studies have tried with limited success to explain these frequently co-occurring events focusing on a few driver genes as culprits. In contrast, our genome-wide analysis suggested a different perspective. Although a few drivers likely play important roles in 10 loss and 7 gain co-occurring gliomas, our analysis suggests that this phenomenon is further orchestrated by a complex interaction of many genes residing within these two chromosomes,

where increased expression of multiple rescuer genes on the gained chromosome 7 compensates for the down-regulation of multiple vulnerable genes on the lost chromosome 10.

Analyzing data from Progenetix, we developed a model that elucidated the probability of these chromosome co-occurring events, by estimating the probability of 7 gain after 10 loss and vice-versa. We found that 7 gain after 10 loss is the more likely event, and that the opposite event can be treated as occurring by random chance. This analysis uses snapshot data on aneuploidies in thousands of samples to confirm the findings of Körber et al. (10), which analyzed longitudinal data on 21 samples. In our study, we assumed that 10 loss and 7 gain occur as separate events because we could not find any evidence among the dozens of previous relevant papers that the two aneuploidies occur simultaneously via a mechanism such as breakage-fusion-bridge cycles. Next, we found that while it remains possible that the loss of chromosome 10 in many GBM/LGG tumors is driven by a few drivers, this loss may be enabled by the tumor's ability to compensate and mitigate the adverse effects it triggers. Even when a few driver genes underlie the recurrence of a lost chromosome, it is widely expected that the loss of many other genes on the same chromosomal-arm would be disadvantageous to cell survival, thereby reducing the positive selection towards that chromosome-arm loss (60-63). We show that gaining chromosome 7 after the loss of 10 increases the fitness advantage conferred to GBM/LGG tumors by 10 loss primarily through intricate genetic rescue interplays between genes on these chromosomes. Even during instances where a specific gene loss on chromosome 10 is neutral or beneficial for cancer (instead of being detrimental), INCISOR can identify partner genes on chromosome 7 whose activation can provide a synergistic benefit to cancer cells. Further independent investigation in IDH wild-type CNS cancer cell line essentiality screens reinforced the conclusions of our synthetic rescue analysis that the gain of 7 in the presence of 10 loss is associated with a selective advantage. Lastly, we analyzed the transcriptomic data from normal non-cancerous human brain tissues and found that the preexisting transcriptomic state of the cortex and frontal cortex may make them predisposed to cancers with 10 loss and 7 gain co-occurrence possibly due to the high activity of rescuer genes on chromosome 7.

Of course, there are also gliomas where 10 loss alone happens without the gain of chromosome 7. Interestingly, we do see that in such cases, the rescuer genes we identified on chromosome 7 tend to have higher expression than tumors where 10 is not lost, testifying to the compensatory role of the genes on chromosome 7.

Our findings address another puzzle about the co-occurrence of 10 loss and 7 gain. Previous studies noted that the losses on chromosome 10 are often segmental, but the gains on chromosome 7 more often encompass the entire chromosome, albeit with exceptions (6, 9, 13, 31). One explanation suggested previously for this distinction is that an epistatic oncogenic benefit is achieved when both the oncogenes *EGFR* and *MET* and other genes on chromosome 7 (e.g., *HGF* and *PDGFA*) are gained (73). Since these two gene pairs located far apart, the evolving cancer is likely to gain the entire chromosome 7. Extending upon this earlier suggestion, we find that the predicted rescuer genes on chromosome 7 are distributed throughout the entire chromosome, pointing to a potential compensatory role of the gain of the whole chromosome. Our analysis testifies that DU-SR interactions are also functionally relevant in cancer samples that have not undergone a whole genome duplication.

Limitations of the study

Our study has several limitations. The Progenetix analysis is inherently limited to analysis of copy number alterations, but it may be better to consider copy number alterations together with gain-of-function gene-specific alterations that might have a similar functional effect on one gene as a copy number gain. In our synthetic rescue identification analysis, we used gene expression and copy-number information of individual genes. While gene expression or copy-number variations are reasonably correlated with protein expression (74, 75), these correlations are not always high and hence these copy-number changes may not always translate to actual changes at the protein levels. Wet lab experiments using human cell lines and animal models are obviously required to further test and validate individual DU-SR gene pairs or rescuer/vulnerable genes. While the analysis on normal non-cancerous GTEx brain tissues was done on 2,642 samples from 13 regions of the brain, the sample size is limited for each specific region (median number of samples is 205).

Conclusions

Beyond providing a comprehensive evolutionary account of the frequent 10 loss/7 gain co-occurrence, our analysis provides specific predictions of genes involved. Of special interest are major rescuer genes that reside on chromosome 7, whose targeting may have a therapeutic potential, if further corroborated in future gene-specific experimental studies. The experimental testing of such predictions is out of the scope of the current investigation, which is focused on presenting a first of its kind holistic evolutionary explanation of this fundamental co-occurring event. To facilitate such a future exploration, we cataloged a prioritized list of vulnerable/rescuer genes that reside on chromosome 10/7 (**Tables S2, S4**).

In conclusion, this analysis presents a new multi-pronged approach to analyze co-occurring aneuploidy events in cancer, shedding new light on a long-contemplated chromosomal co-occurrence event in gliomas. It could be applied in future studies of other common co-occurring aneuploidies across many cancer types.

AUTHOR CONTRIBUTIONS

Conceptualization: N.U.N., E.R.

Resources: N.U.N., A.A.S., E.M.G., J.Z., G.L., U.B-D, E.R.

Data curation: N.U.N., A.A.S., E.M.G., J.Z., G.L.

Software: N.U.N., A.A.S., E.M.G.

Formal analysis: N.U.N., A.A.S., E.M.G., J.Z., G.L.

Supervision: E.R.

Funding acquisition: E.R.

Validation: N.U.N., A.A.S., E.M.G., U.B-D, E.R.

Investigation: N.U.N., A.A.S., E.M.G., K.C., K.D.A., U.B-D, E.R.

Visualization: N.U.N., A.A.S., E.M.G, U.B-D, E.R.

Methodology: N.U.N., A.A.S., E.M.G., K.C., A.D.S., E.D.S., U.B-D., E.R.

Writing-original draft: N.U.N., A.A.S., E.M.G., E.R.

Project administration: E.R.

Writing-review and editing: N.U.N., A.A.S., E.M.G., K.C., U.B-D., E.R.

ACKNOWLEDGEMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute. Work in the Ben-David lab was supported by the was supported by the European Research Council Starting Grant (grant #945674 to U.B-D.). This work used the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We thank Dr. Mark R. Gilbert, Dr. Orieta Celiku, Dr. Kun Wang, Mr. Neel Sanghvi for their help on this work. The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The authors used ChatGPT 4 to improve the quality of writing for some sentences.

REFERENCES

1. Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. JAMA. 2013;310(17):1842-50.
2. Bigner SH, Mark J, Burger PC, Mahaley MS, Jr., Bullard DE, Muhlbaier LH, et al. Specific chromosomal abnormalities in malignant human gliomas. Cancer Res. 1988;48(2):405-11.
3. Höglund M, Gisselsson D, Mandahl N, Johansson B, Mertens F, Mitelman F, et al. Multivariate analyses of genomic imbalances in solid tumors reveal distinct and converging pathways of karyotypic evolution. Gene Chromosome Canc. 2001;31(2):156-71.
4. Venter DJ, Thomas DG. Multiple sequential molecular abnormalities in the evolution of human gliomas. Br J Cancer. 1991;63(5):753-7.
5. Barthel FP, Johnson KC, Varn FS, Moskalik AD, Tanner G, Kocakavuk E, et al. Longitudinal molecular trajectories of diffuse glioma in adults. Nature. 2019;576(7785):112-20.
6. Ozawa T, Riester M, Cheng YK, Huse JT, Squatrito M, Helmy K, et al. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell. 2014;26(2):288-300.
7. Sanson M, Marie Y, Paris S, Idbaih A, Laffaire J, Ducray F, et al. Isocitrate dehydrogenase 1 codon 132 mutation Is an important prognostic biomarker in gliomas. J Clin Oncol. 2009;27(25):4150-4.
8. Galbraith K, Kumar A, Abdullah KG, Walker JM, Adams SH, Prior T, et al. Molecular correlates of long survival in IDH-wildtype glioblastoma cohorts. J Neuropath Exp Neur. 2020;79(8):843-54.

9. Stichel D, Ebrahimi A, Reuss D, Schrimpf D, Ono T, Shirahata M, et al. Distribution of *EGFR* amplification, combined chromosome 7 gain and chromosome 10 loss, and T*ERT* promoter mutation in brain tumors and their potential for the reclassification of IDHwt astrocytoma to glioblastoma. Acta Neuropathol. 2018;136(5):793-803.
10. Körber V, Yang J, Barah P, Wu Y, Stichel D, Gu Z, et al. Evolutionary trajectories of IDH^{wt} glioblastomas reveal a common path of early tumorigenesis instigated years ahead of initial diagnosis. Cancer Cell. 2019;35(4):692-704 e12.
11. Fults D, Pedone C. Deletion mapping of the long arm of chromosome-10 in glioblastoma-multiforme. Gene Chromosome Canc. 1993;7(3):173-7.
12. del Mar Inda M, Fan X, Munoz J, Perot C, Fauvet D, Danglot G, et al. Chromosomal abnormalities in human glioblastomas: gain in chromosome 7p correlating with loss in chromosome 10q. Mol Carcinog. 2003;36(1):6-14.
13. Karlböm AE, James CD, Boethius J, Cavenee WK, Collins VP, Nordenskjold M, et al. Loss of heterozygosity in malignant gliomas involves at least 3 distinct regions on chromosome-10. Hum Genet. 1993;92(2):169-74.
14. Mohapatra G, Bollen AW, Kim DH, Lamborn K, Moore DH, Prados MD, et al. Genetic analysis of glioblastoma multiforme provides evidence for subgroups within the grade. Genes Chromosomes Cancer. 1998;21(3):195-206.
15. Bidinotto LT, Torrieri R, Mackay A, Almeida GC, Viana-Pereira M, Cruvinel-Carloni A, et al. Copy number profiling of Brazilian astrocytomas. G3 (Bethesda). 2016;6(7):1867-78.
16. Crespo I, Tão H, Nieto AB, Rebelo O, Domingues P, Vital AL, et al. Amplified and homozygously deleted genes in glioblastoma: impact on gene expression levels. PLoS One. 2012;7(9):e46088.
17. Crespo I, Vital AL, Nieto AB, Rebelo O, Tão H, Lopes MC, et al. Detailed characterization of alterations of chromosomes 7, 9, and 10 in glioblastomas as assessed by single-nucleotide polymorphism arrays. J Mol Diagn. 2011;13(6):634-47.
18. Masica DL, Karchin R. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. Cancer Res. 2011;71(13):4550-61.
19. Durand K, Guillaudeau L, Pommepuy I, Mesturoux L, Chaunavel A, Gadeaud E, et al. Alpha-internexin expression in gliomas: relationship with histological type and 1p, 19q, 10p and 10q status. J Clin Pathol. 2011;64(9):793-801.
20. Hata N, Yoshimoto K, Yokoyama N, Mizoguchi M, Shono T, Guan YL, et al. Allelic losses of chromosome 10 in glioma tissues detected by quantitative single-strand conformation polymorphism analysis. Clin Chem. 2006;52(3):370-8.
21. Rasheed BK, McLendon RE, Friedman HS, Friedman AH, Fuchs HE, Bigner DD, et al. Chromosome 10 deletion mapping in human gliomas: a common deletion region in 10q25. Oncogene. 1995;10(11):2243-6.
22. Roversi G, Pfundt R, Moroni RF, Magnani I, van Reijmersdal S, Pollo B, et al. Identification of novel genomic markers related to progression to glioblastoma through genomic profiling of 25 primary glioma cell lines. Oncogene. 2006;25(10):1571-83.
23. Li DM, Sun H. *TEP1,* encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. Cancer Res. 1997;57(11):2124-9.

24. Liu W, James CD, Frederick L, Alderete BE, Jenkins RB. *PTEN/MMAC1* mutations and *EGFR* amplification in glioblastomas. Cancer Res. 1997;57(23):5254-7.
25. Steck PA, Pershouse MA, Jasser SA, Yung WK, Lin H, Ligon AH, et al. Identification of a candidate tumour suppressor gene, *MMAC1*, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet. 1997;15(4):356-62.
26. Tohma Y, Gratas C, Biernat W, Peraud A, Fukuda M, Yonekawa Y, et al. *PTEN (MMAC1)* mutations are frequent in primary glioblastomas (de novo) but not in secondary glioblastomas. J Neuropathol Exp Neurol. 1998;57(7):684-9.
27. Thuy MN, Kam JK, Lee GC, Tao PL, Ling DQ, Cheng M, et al. A novel literature-based approach to identify genetic and molecular predictors of survival in glioblastoma multiforme: Analysis of 14,678 patients using systematic review and meta-analytical tools. J Clin Neurosci. 2015;22(5):785-99.
28. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. Cell. 2013;155(2):462-77.
29. Yadav AK, Renfrow JJ, Scholtens DM, Xie H, Duran GE, Bredel C, et al. Monosomy of chromosome 10 associated with dysregulation of epidermal growth factor signaling in glioblastomas. JAMA. 2009;302(3):276-89.
30. Libermann TA, Nusbaum HR, Razon N, Kris R, Lax I, Soreq H, et al. Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. Nature. 1985;313(5998):144-7.
31. Wullich B, Sattler HP, Fischer U, Meese E. Two independent amplification events on chromosome 7 in glioma: amplification of the epidermal growth factor receptor gene and amplification of the oncogene MET. Anticancer Res. 1994;14(2A):577-9.
32. Baysan M, Woolard K, Cam MC, Zhang W, Song H, Kotliarova S, et al. Detailed longitudinal sampling of glioma stem cells in situ reveals Chr7 gain and Chr10 loss as repeated events in primary tumor formation and recurrence. Int J Cancer. 2017;141(10):2002-13.
33. Bredel M, Scholtens DM, Harsh GR, Bredel C, Chandler JP, Renfrow JJ, et al. A network model of a cooperative genetic landscape in brain tumors. JAMA. 2009;302(3):261-75.
34. Gao F, Zhang P, Zhou CX, Li JF, Wang Q, Zhu FL, et al. Frequent loss of *PDCD4* expression in human glioma: Possible role in the tumorigenesis of glioma. Oncol Rep. 2007;17(1):123-8.
35. Rodriguez SMB, Kamel A, Ciubotaru GV, Onose G, Sevastre AS, Sfredel V, et al. An overview of EGFR mechanisms and their implications in targeted therapies for glioblastoma. Int J Mol Sci. 2023;24(13).
36. Collins VP. Amplified genes in human gliomas. Semin Cancer Biol. 1993;4(1):27-32.
37. Lopez-Gines C, Gil-Benso R, Ferrer-Luna R, Benito R, Serna E, Gonzalez-Darder J, et al. New pattern of EGFR amplification in glioblastoma and the relationship of gene copy number with gene expression profile. Mod Pathol. 2010;23(6):856-65.
38. Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc Natl Acad Sci U S A. 2012;109(8):3041-6.
39. McNulty SN, Cottrell CE, Vigh-Conrad KA, Carter JH, Heusel JW, Ansstas G, et al. Beyond sequence variation: assessment of copy number variation in adult glioblastoma through targeted tumor somatic profiling. Hum Pathol. 2019;86:170-81.

40. Crespo I, Vital AL, Nieto AB, Rebelo O, Tao H, Lopes MC, et al. Detailed characterization of alterations of chromosomes 7, 9, and 10 in glioblastomas as assessed by single-nucleotide polymorphism arrays. J Mol Diagn. 2011;13(6):634-47.
41. Gonzalez-Tablas M, Arandia D, Jara-Acevedo M, Otero A, Vital AL, Prieto C, et al. Heterogeneous *EGFR*, *CDK4*, *MDM4,* and *PDGFRA* gene expression profiles in primary GBM: No association with patient survival. Cancers (Basel). 2020;12(1).
42. Jubran J, Slutsky R, Rozenblum N, Rokach L, Ben-David U, Yeger-Lotem E. Machine-learning analysis reveals an important role for negative selection in shaping cancer aneuploidy landscapes. Genome Biol. 2024;25(1):95.
43. Sack LM, Davoli T, Li MZ, Li Y, Xu Q, Naxerova K, et al. Profound tissue specificity in proliferation control underlies cancer drivers and aneuploidy patterns. Cell. 2018;173(2):499-514 e23.
44. Gao B, Baudis M. Signatures of discriminative copy number aberrations in 31 cancer subtypes. Front Genet. 2021;12:654887.
45. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell. 2018;33(4):676-89 e3.
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57(1):289-300.
47. Prasad K, Bloomfield M, Levi H, Keuper K, Bernhard SV, Baudoin NC, et al. Whole-genome duplication shapes the aneuploidy landscape of human cancers. Cancer Res. 2022;82(9):1736-52.
48. Sahu AD, J SL, Wang Z, Zhang G, Iglesias-Bartolome R, Tian T, et al. Genome-wide prediction of synthetic rescue mediators of resistance to targeted and immunotherapy. Mol Syst Biol. 2019;15(3):e8323.
49. 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. Nat Genet. 2017;49(12):1779-84.
50. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.
51. Cohen-Sharir Y, McFarland JM, Abdusamad M, Marquis C, Bernhard SV, Kazachkova M, et al. Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. Nature. 2021;590(7846):486-91.
52. Cai H, Kumar N, Ai N, Gupta S, Rath P, Baudis M. Progenetix: 12 years of oncogenomic data curation. Nucleic Acids Res. 2014;42(Database issue):D1055-62.
53. Network CGAR, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113-20.
54. Beerenwinkel N, Rahnenfuhrer J, Kaiser R, Hoffmann D, Selbig J, Lengauer T. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. Bioinformatics. 2005;21(9):2106-7.
55. Beerenwinkel N, Schwarz RF, Gerstung M, Markowetz F. Cancer evolution: mathematical models and computational inference. Syst Biol. 2015;64(1):e1-25.
56. Pathare S, Schaffer AA, Beerenwinkel N, Mahimkar M. Construction of oncogenetic tree models reveals multiple pathways of oral cancer progression. Int J Cancer. 2009;124(12):2864-71.

57. Roman T, Nayyeri A, Fasy BT, Schwartz R. A simplicial complex-based approach to unmixing tumor progression data. BMC Bioinformatics. 2015;16:254.
58. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(7216):1061-8.
59. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N Engl J Med. 2015;372(26):2481-98.
60. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. Nature Reviews Genetics 2020;21:44–62.
61. Chunduri NK, Storchova Z. The diverse consequences of aneuploidy. Nat Cell Biol. 2019;21(1):54-62.
62. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet. 2012;13(3):189-203.
63. Jubran J, Slutsky R, Rozenblum N, Ben-David U, Yeger-Lotem E. Jubran, J., Slutsky, R., Rozenblum, N., Rokach, L., Ben-David, U., & Yeger-Lotem, E. (2023). Machine-learning analysis of factors that shape cancer aneuploidy landscapes reveals an important role for negative selection. bioRxiv. 2023.
64. Wang X, Vizeacoumar FS, Sahu AD. INCISOR: An algorithm to identify synthetic rescue mediators of resistance to targeted and immunotherapy. Methods Mol Biol. 2021;2381:203-15.
65. Lee JS, Nair NU, Dinstag G, Chapman L, Chung Y, Wang K, et al. Synthetic lethality-mediated precision oncology via the tumor transcriptome. Cell. 2021;184(9):2487-502 e13.
66. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 2017;33(19):3088-90.
67. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18(11):696-705.
68. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell. 2017;170(3):564-76 e16.
69. Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko IV, et al. Aberrant signaling pathways in glioma. Cancers (Basel). 2011;3(3):3242-78.
70. Syafruddin SE, Nazarie W, Moidu NA, Soon BH, Mohtar MA. Integration of RNA-Seq and proteomics data identifies glioblastoma multiforme surfaceome signature. BMC Cancer. 2021;21(1):850.
71. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.
72. Cheng K, Nair NU, Lee JS, Ruppin E. Synthetic lethality across normal tissues is strongly associated with cancer risk, onset, and tumor suppressor specificity. Sci Adv. 2021;7(1).
73. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A. 2007;104(50):20007-12.
74. Gry M, Rimini R, Strömberg S, Asplund A, Pontén F, Uhlén M, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. Bmc Genomics. 2009;10:365.
75. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. Bmc Med Genet. 2019;20:175.

FIGURE LEGENDS

Figure 1: Chromosome arms co-occurring loss-gain event statistics and probabilistic analysis. ***(A)*** *Plot charting the landscape of chromosomal arm loss-gain enrichments across cancer types from patient tumors in Progenetix. Each point represents a chromosome arm loss/gain event for one cancer type (n =* 5,886 *data points in total); the Odds Ratio (OR) and FDR values are obtained using Fishers exact test. The dashed vertical line is at OR = 1; the dashed horizontal line is at FDR = 0.01. Events with FDR < 0.01 and OR > 1 are colored red while the rest are colored gray. In this figure, only events with at least 50 co-occurring cases in a specific cancer type are considered, and only a few top events are described (full results described in* ***Table S1****). There were no LGG cancers in the Progenetix data.* ***(B)*** *Bar plot showing the number of GBM patients in Progenetix with or without a 10 loss or 7 gain. 10 loss or 7 gain implies either the p/q arm is lost or gained, respectively. The odds ratio and p-value for 10 loss and 7 gain co-occurrence using a two-sided Fisher exact test is shown. For the 705 cancers without 7 gain or 10 loss, an arm gain or loss was reported on another chromosome indicating the tumor was aneuploid. Keywords – GBM: Glioblastoma multiforme, COAD: Colon adenocarcinoma.* ***(C)*** *Illustration of approximating the fraction of cooccurrences having 10 loss first by the number of tumors that have only 10 loss. The square to the left of the dotted line represents tumors with only one of 7 gain or 10 loss; the square to the right represents tumors with both. The solid blue area shows the proportion of each region for which 10 loss occurs first under the assumption that the events are independent. The blue striped area represents the greater number of tumors with 10 loss first followed by 7 gain that empirical evidence suggests would occur.*

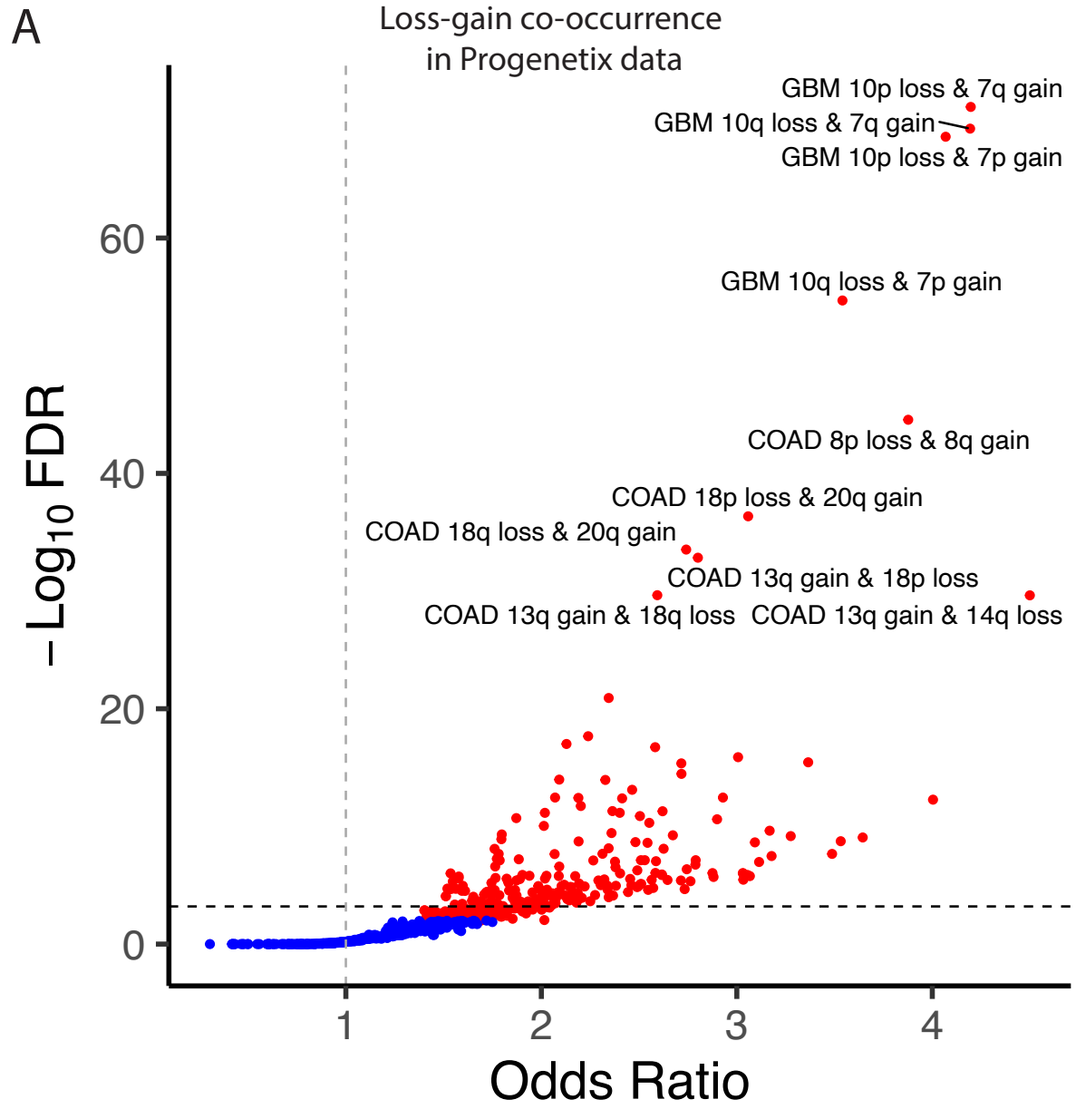
Figure 2: Evolutionary insights into the loss of chromosome 10 and gain of chromosome 7 in gliomas by analyzing patient tumor data. ***(A)*** *Survival analysis (Kaplan-Meier curves) for TCGA GBM and LGG patients with the following occurrences: 10 cnn and 7 cnn (n=259), 10 loss and 7 cnn (n=38), 10 cnn and 7 gain (n=40), 10 loss and 7 gain (n=174). Log-rank test p-values (P) are shown. Pairwise survival analysis between patients with 10 loss and 7 gain with the three other groups are shown in* ***Fig. S3A-C****.* ***(B)*** *An illustration showcasing a DU-SR interaction between two genes A and B (adapted from Fig. 1 in (48)).* ***(C)*** *Volcano plot showing the overlap enrichment of rescuer genes (from GBM-LGG DU-SR analysis) with the genes in*

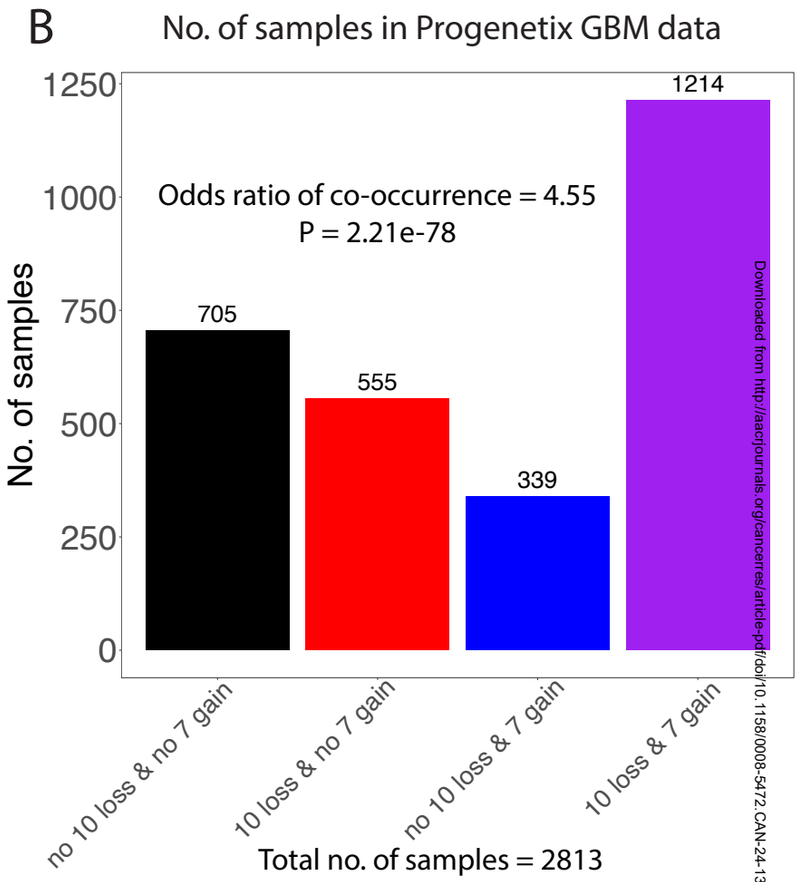
each chromosome arm (one-sided Fisher exact test). The dashed horizontal line is FDR = 0.1, and the dashed vertical line is at odds ratio = 1. The rescuer genes are identified for all the (vulnerable) genes on a specific arm and then tested for overlap enrichment across all chromosome arms. Rescuers of vulnerable genes on 10p or 10q are more enriched on chromosomes 7p or 7q. ***(D)*** *A plot of rescuer genes on 7 against ideograms of chromosome 7 generated using karyoploteR (66). The full extent of each gene, from 5' most exon in any transcript to 3' most end of any transcript, is shown in blue; because of the scale of genes versus the scale of chromosomes, these may appear as lines. On rare occasions, genes overlap, and one gene is plotted in a second row above the ideogram.* ***(E)*** *Box plots showing activity scores on chromosome 7 between two groups of GBM and LGG patients: 10 loss and 7 cnn (n=38) versus 10 cnn and 7 cnn (n=259). Activity score is the fraction of a defined set of genes on chromosome 7 that have high expression (**Methods**). The defined sets of genes are: (i) rescuer genes on chromosome 7 (chr7); (ii) top hub rescuers on chromosome 7 (i.e., those that map to ≥ 10 vulnerable genes on chromosome 10); (iii) genes on chromosome 7 which are not rescuers; (iv) all chromosome 7 genes. DU-SR network is computed using TCGA GBM+LGG data or TCGA pan-cancer data without using brain tumors.* ***(F)*** *Pairwise pathway enrichment analysis for the DU-SR network derived from GBM+LGG data for the vulnerable genes in chromosome 10 and rescuer genes in chromosome 7 (using all the genes on these chromosomes as the universal background set for enrichment). Statistical test used is one-sided Fisher exact test (P < 0.05, adjusted-P/FDR < 0.2). The established set of 50 Cancer Hallmark pathways from MSigDB was used for this analysis (50)). Keywords: 'padj' implies 'adjusted p-value'. GBM: Glioblastoma multiforme, LGG: brain lower grade glioma; chr7: chromosome 7; 10 cnn or 7 cnn: copy number neutral state for chromosomes 10 and 7, respectively; '&' implies 'and'.*

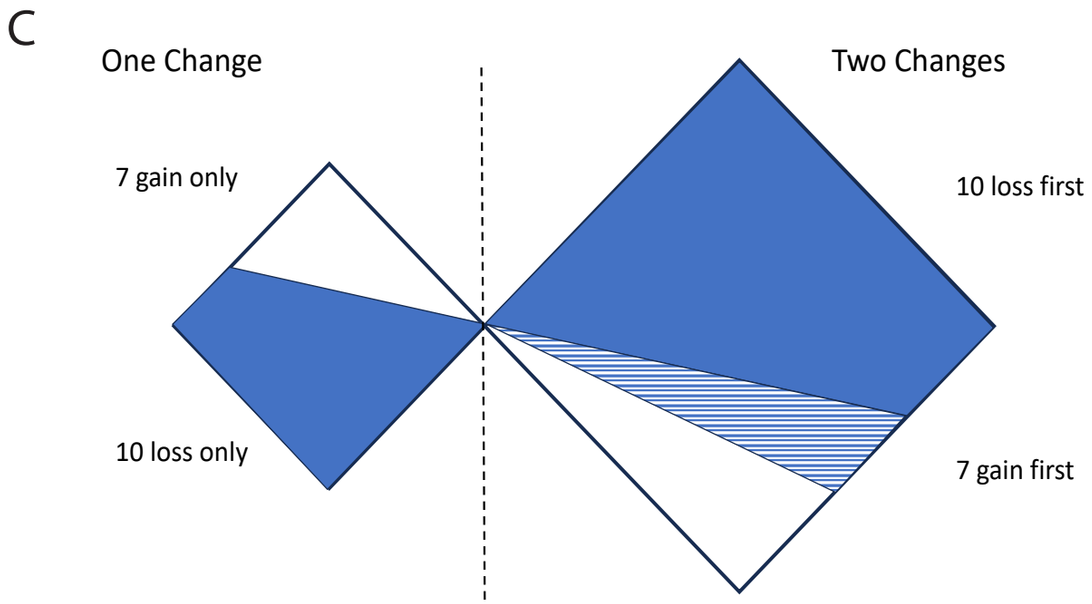
Figure 3: In vitro essentiality analysis associated with 10 loss and 7 gain on central nervous system (CNS) cancer cell lines. (A-C) *Bar plots showing the percentage of more (green) or less (yellow) differentially essential genes between two groups of CNS cell lines:* ***(A)*** *test group (10 loss and 7 gain , n=49) vs control group (10 loss and 7 cnn, n=8) of CNS cell lines;* ***(B)*** *test group (10 cnn and 7 gain, n=22) vs control group (10 cnn and 7 cnn, n=8);* ***(C)*** *test group (10 loss and 7 gain, n=49) vs control group (10 cnn and 7 gain, n=22).* ***(D)*** *Box plots comparing the dependency scores of chromosome 10 genes between a near-diploid RPE1 clone (SS48), a clone*

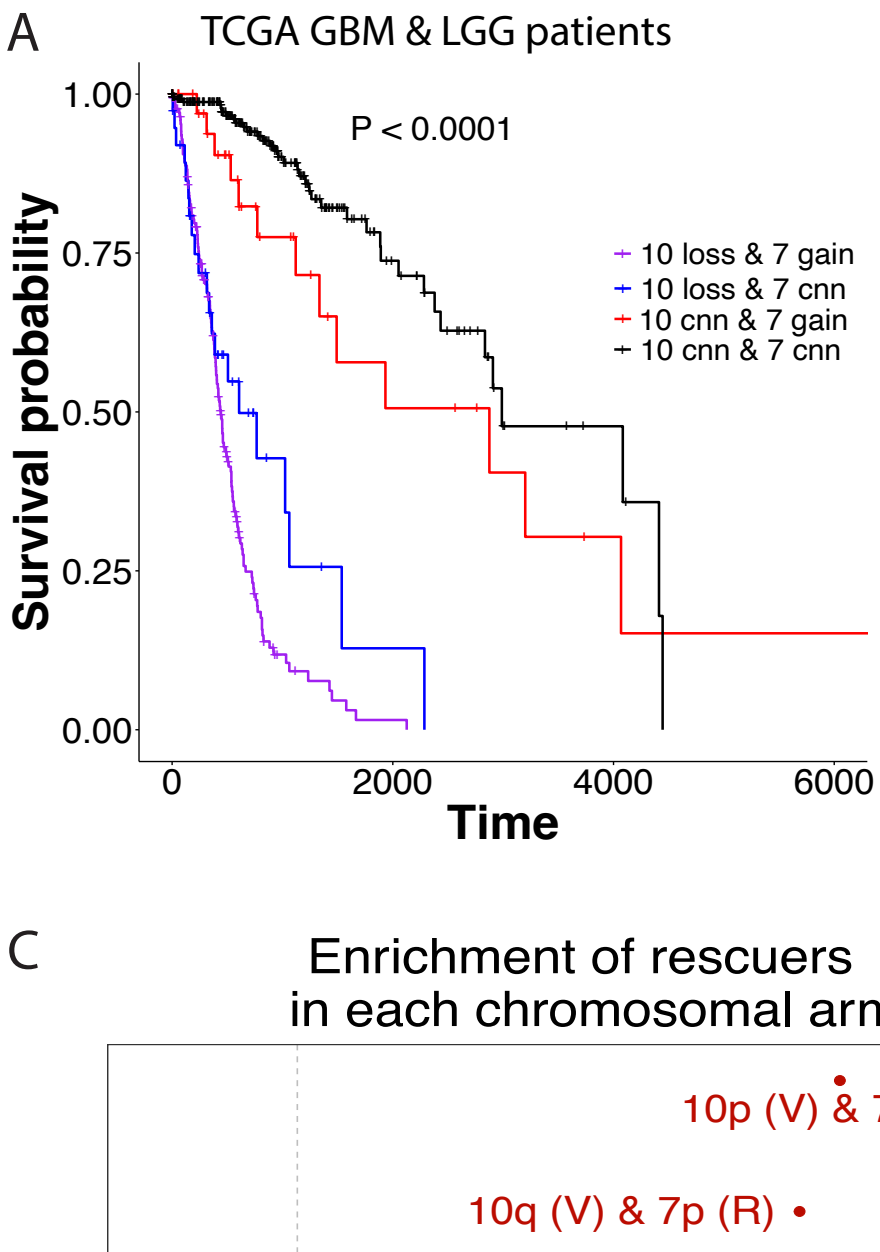
with trisomy of chromosome 7 (SS6) and a clone with trisomy of chromosome 8 (SS119). Only genes defined as essential in SS48 (CERES<-0.2) are considered for this analysis.

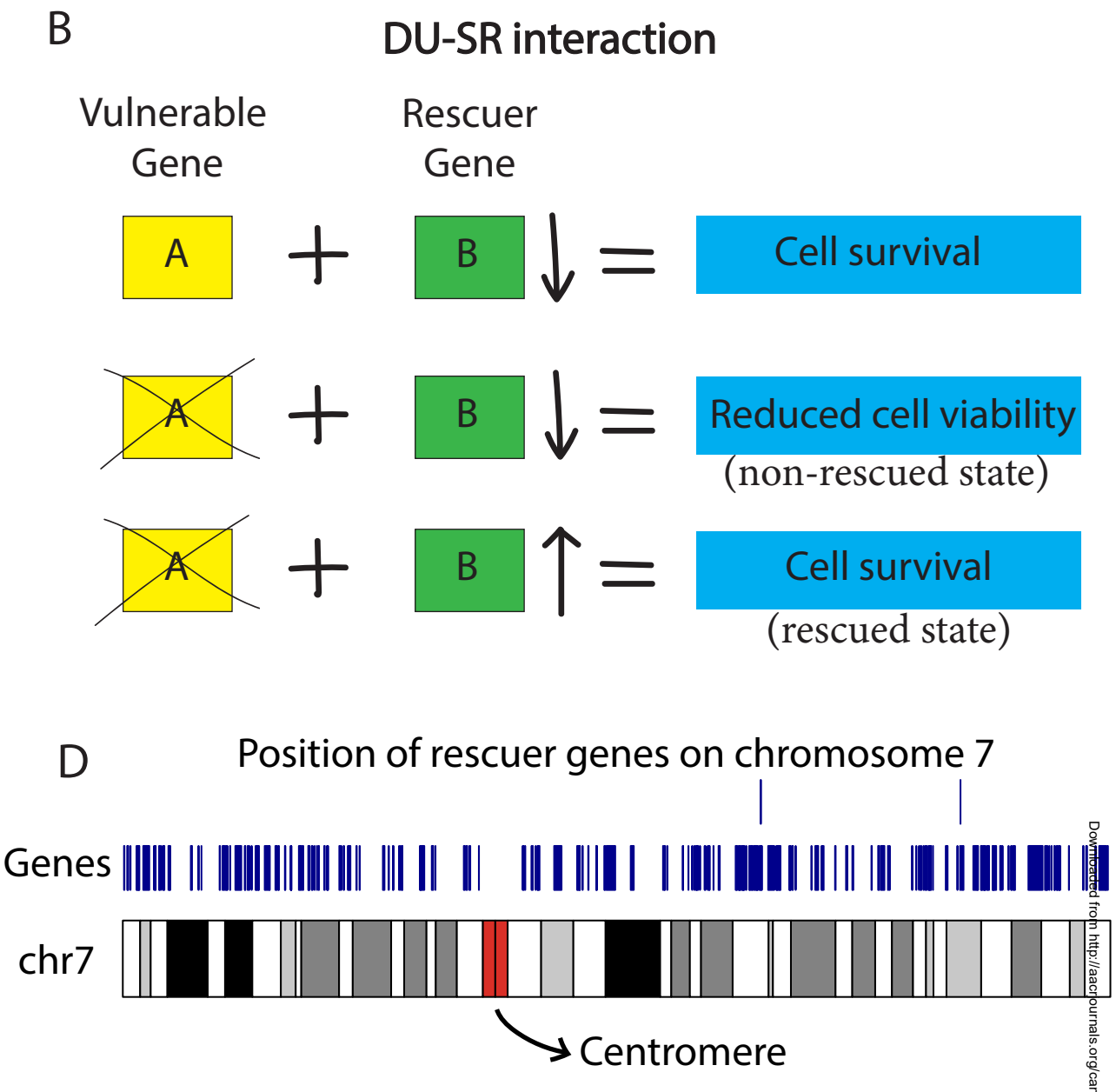
Figure 4: Normal non-cancerous brain tissue analysis. (A) *Toy example demonstrating the computation of cancer synthetic rescue (cSR) load. chr10 and chr7 stand for chromosomes 10 and 7, respectively.* ***(B, C)*** *Bar plots showing the variation of tissue cSR loads (rank normalized across tissues) across various brain tissues in the GTEx database, using DU-SR networks derived from:* ***(B)*** *brain cancer data;* ***(C)*** *pan-cancer data while excluding brain tumors. Frontal cortex and cortex tissues are shown in red as they are most relevant to GBMs. Empirical p-values (P) of both frontal cortex and cortex being ranked as the top two based on their tissue cSR loads in comparison to random controls are shown (****Methods****).*

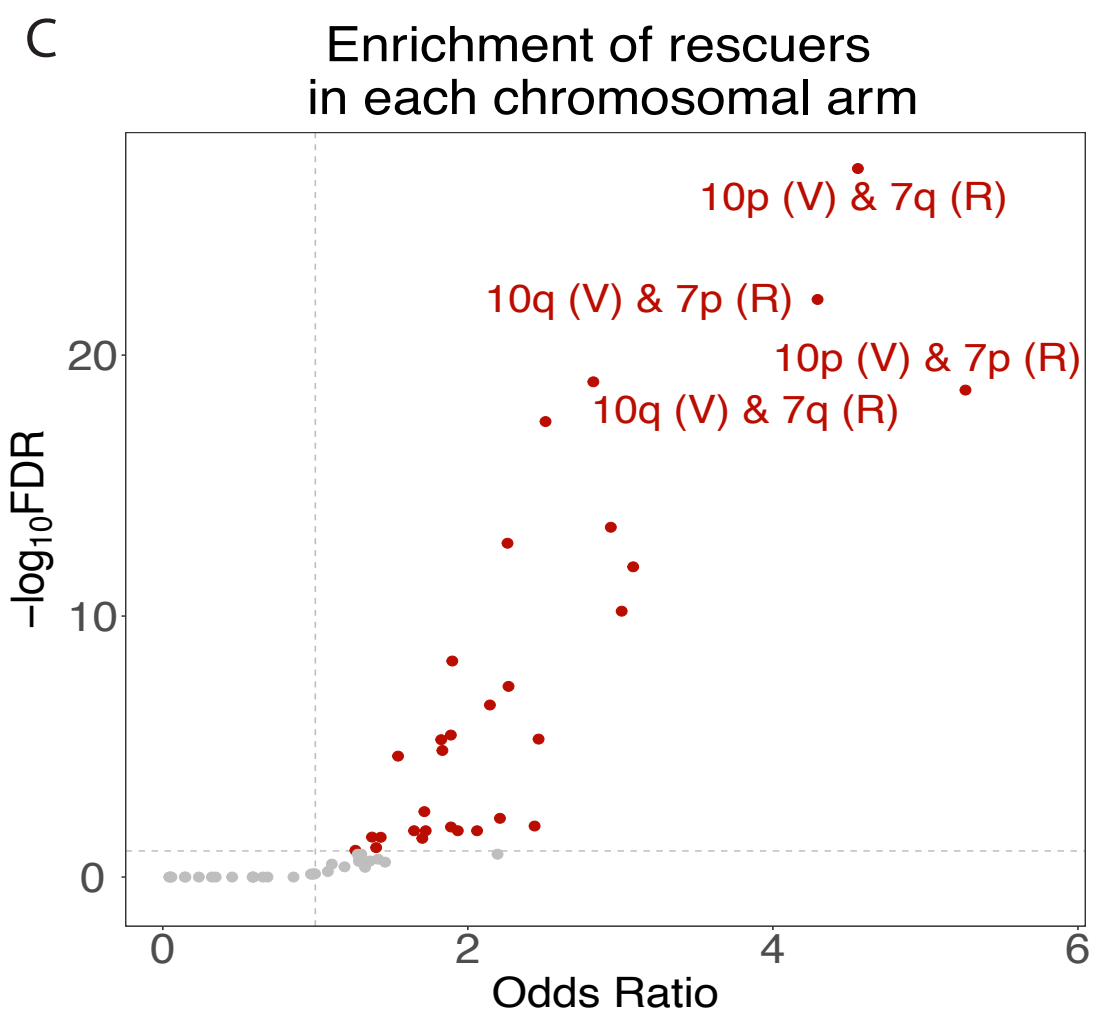


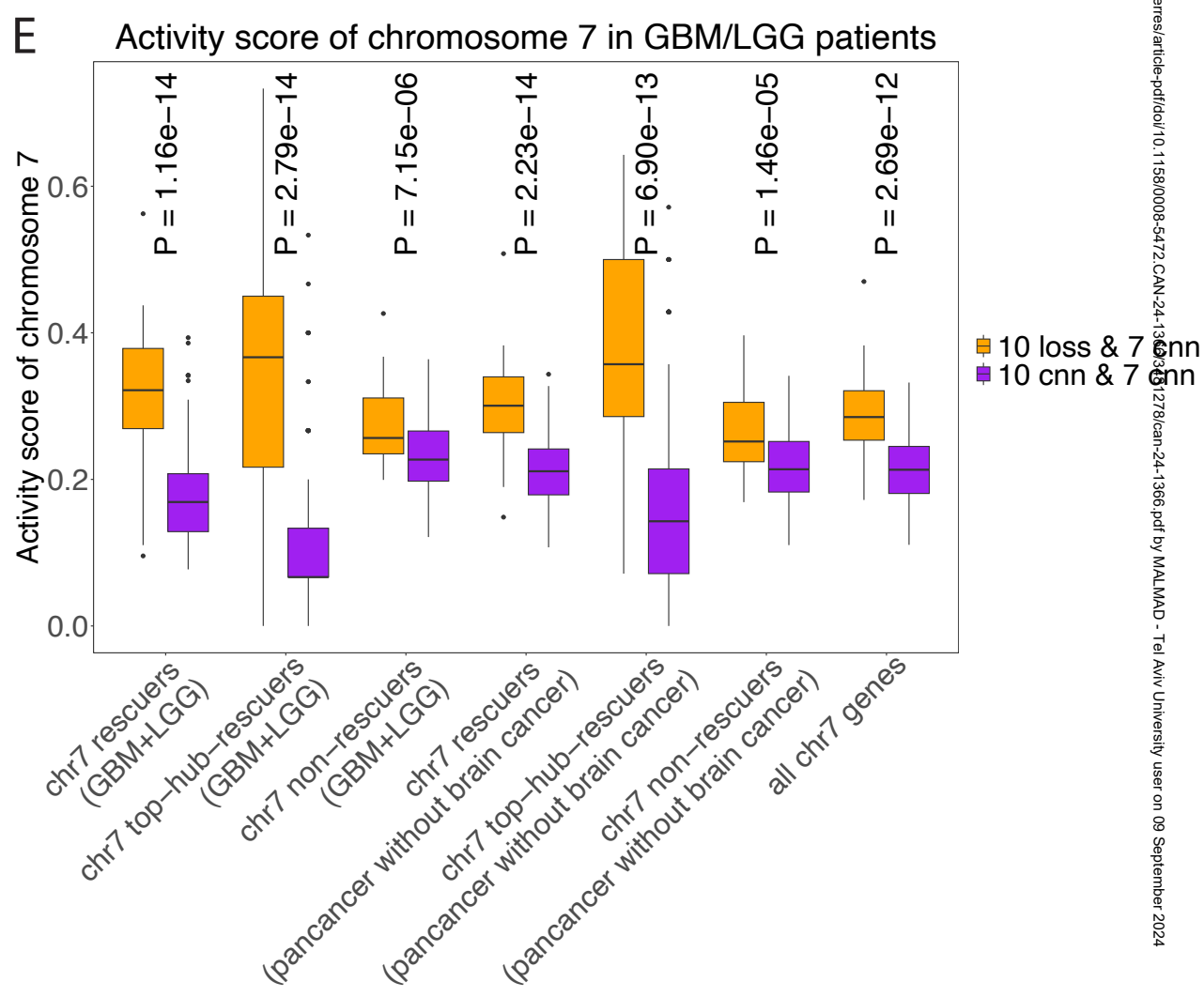


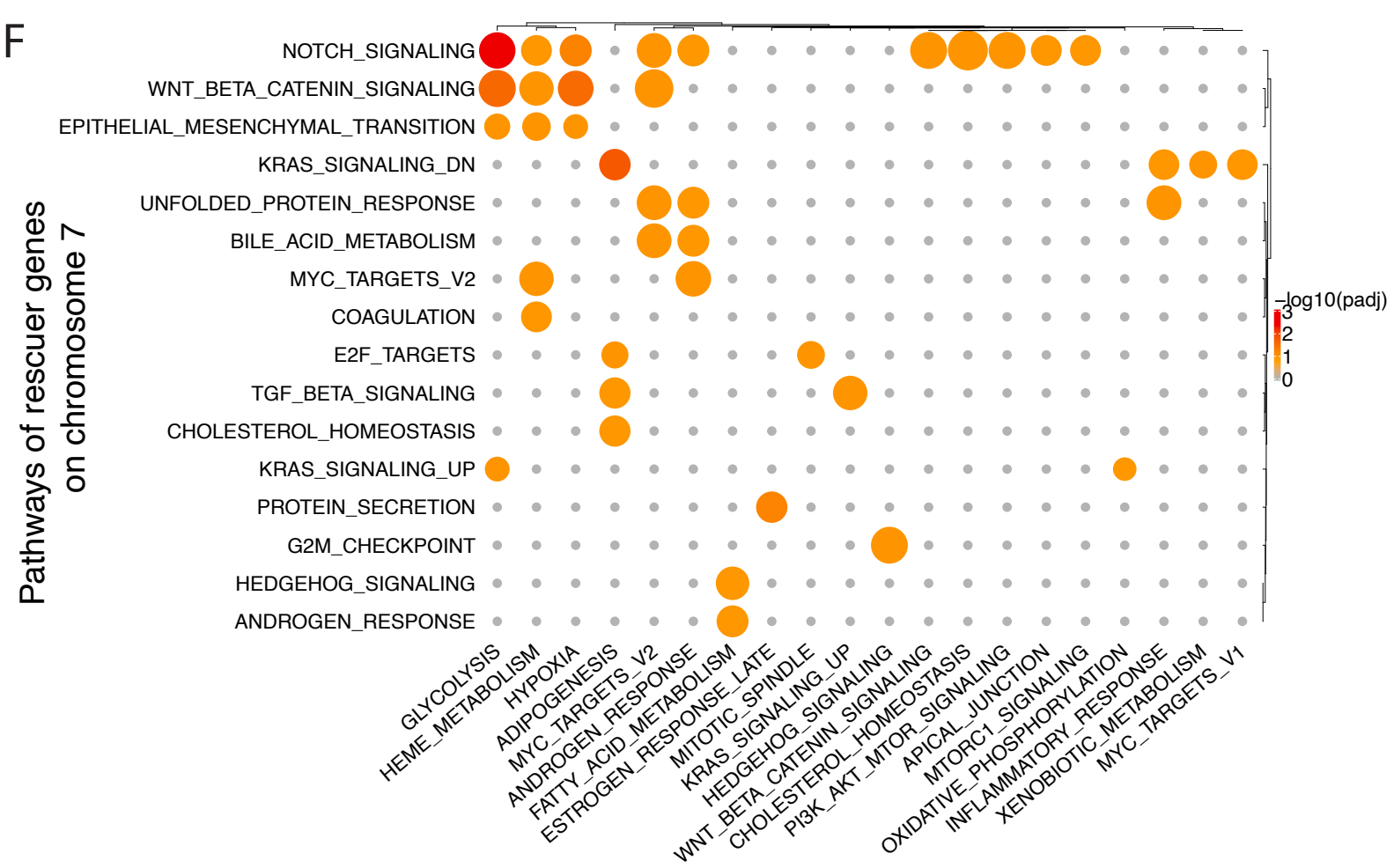




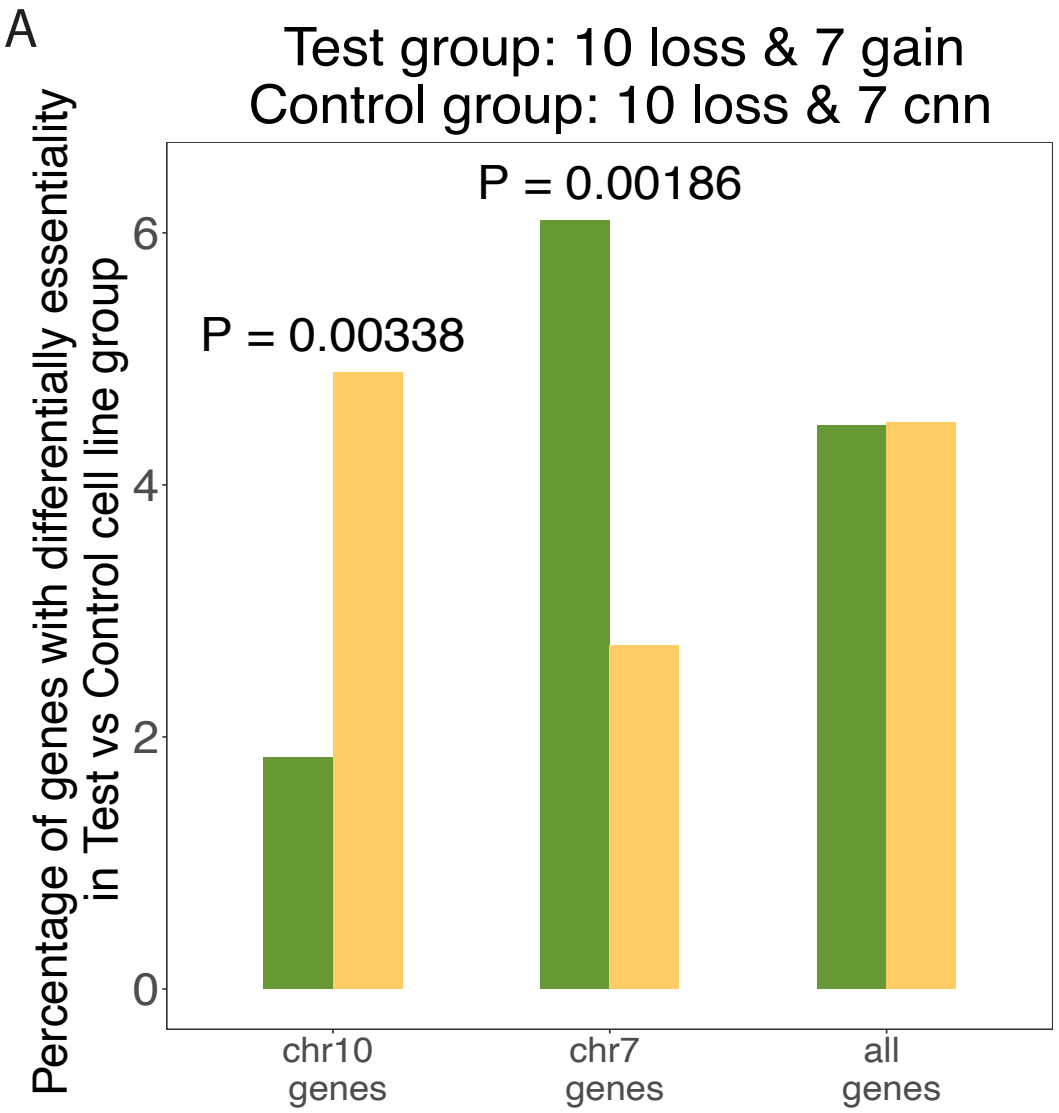




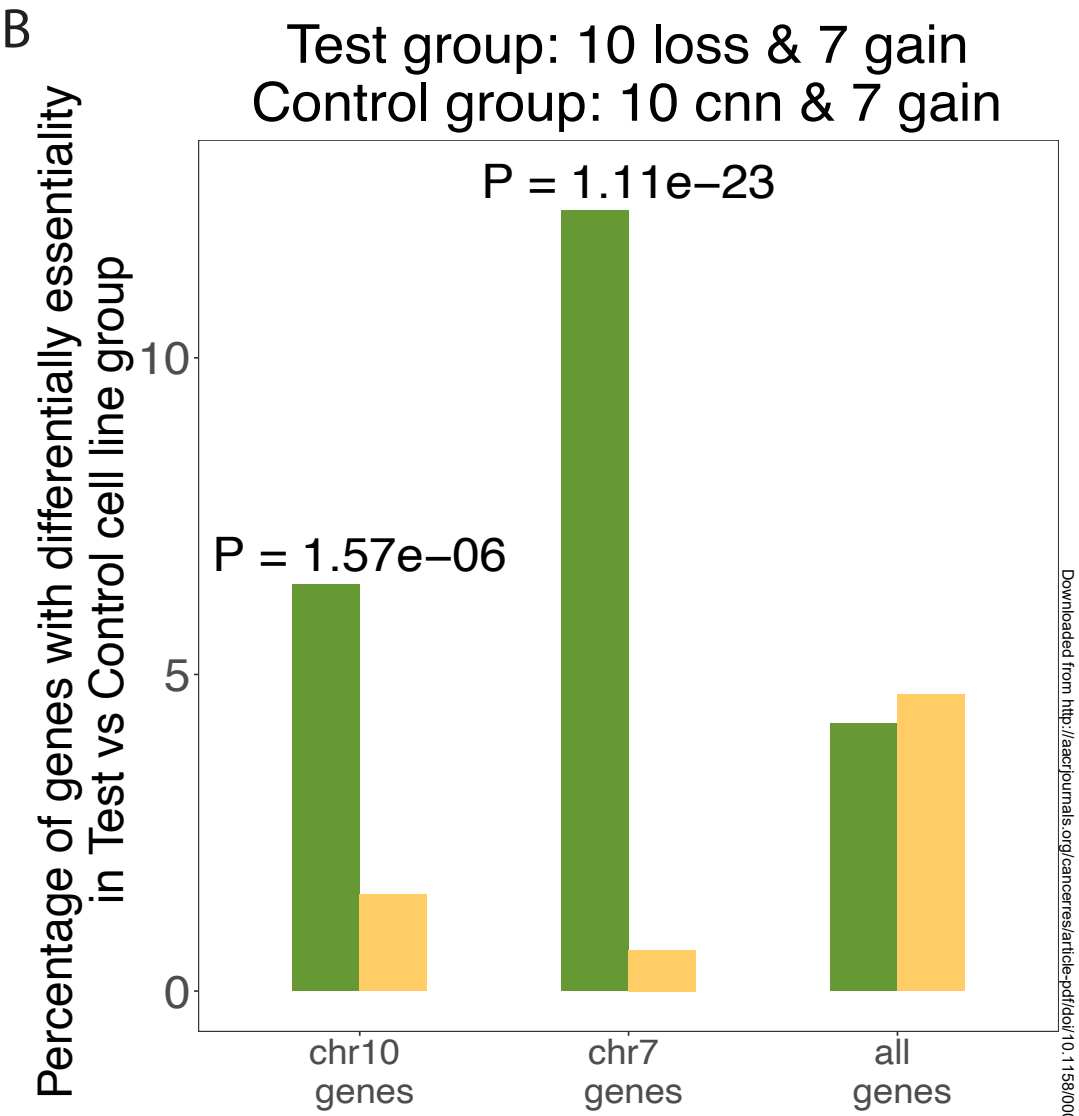


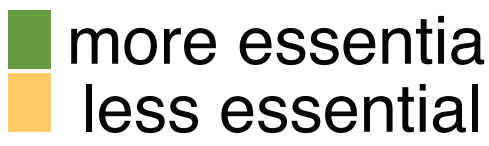


A

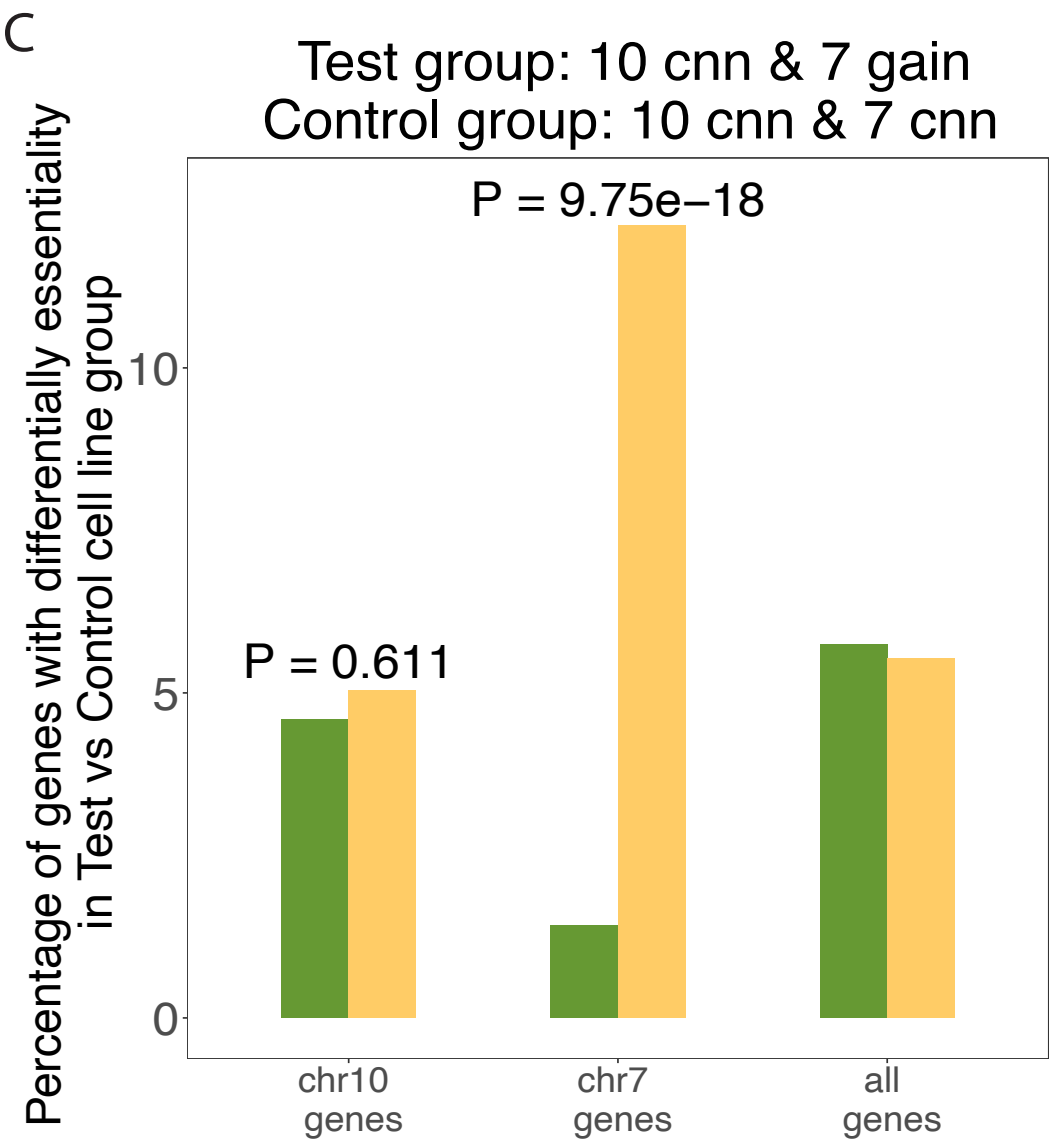


B

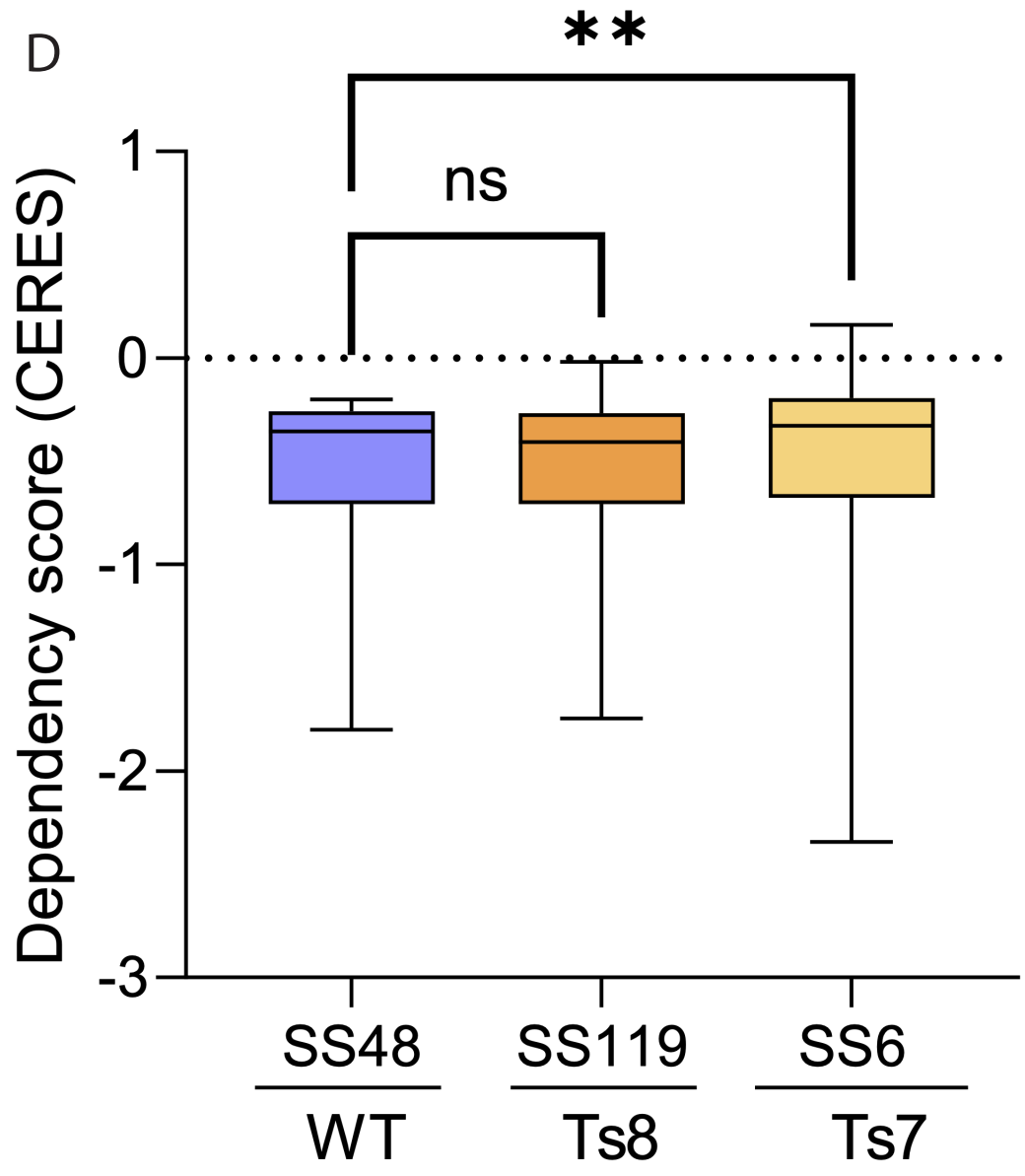




C



D

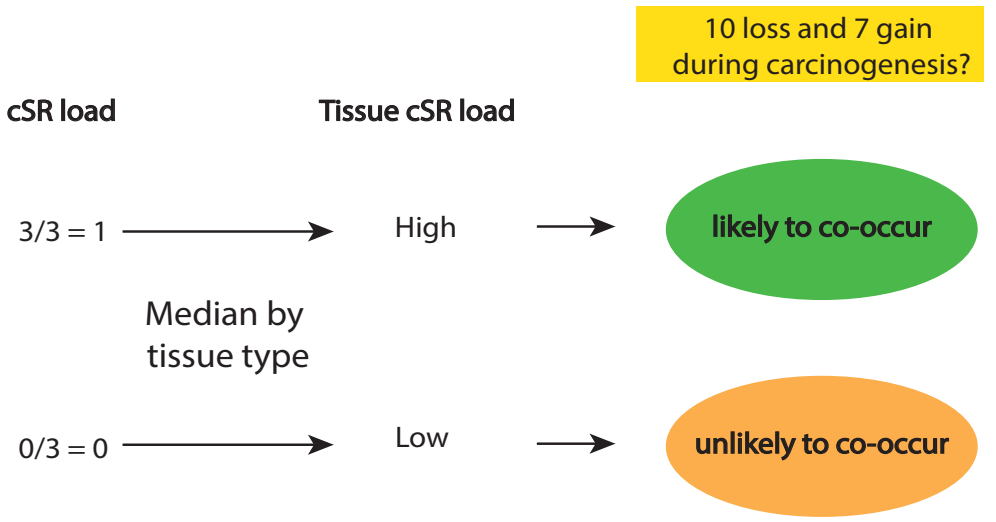


A

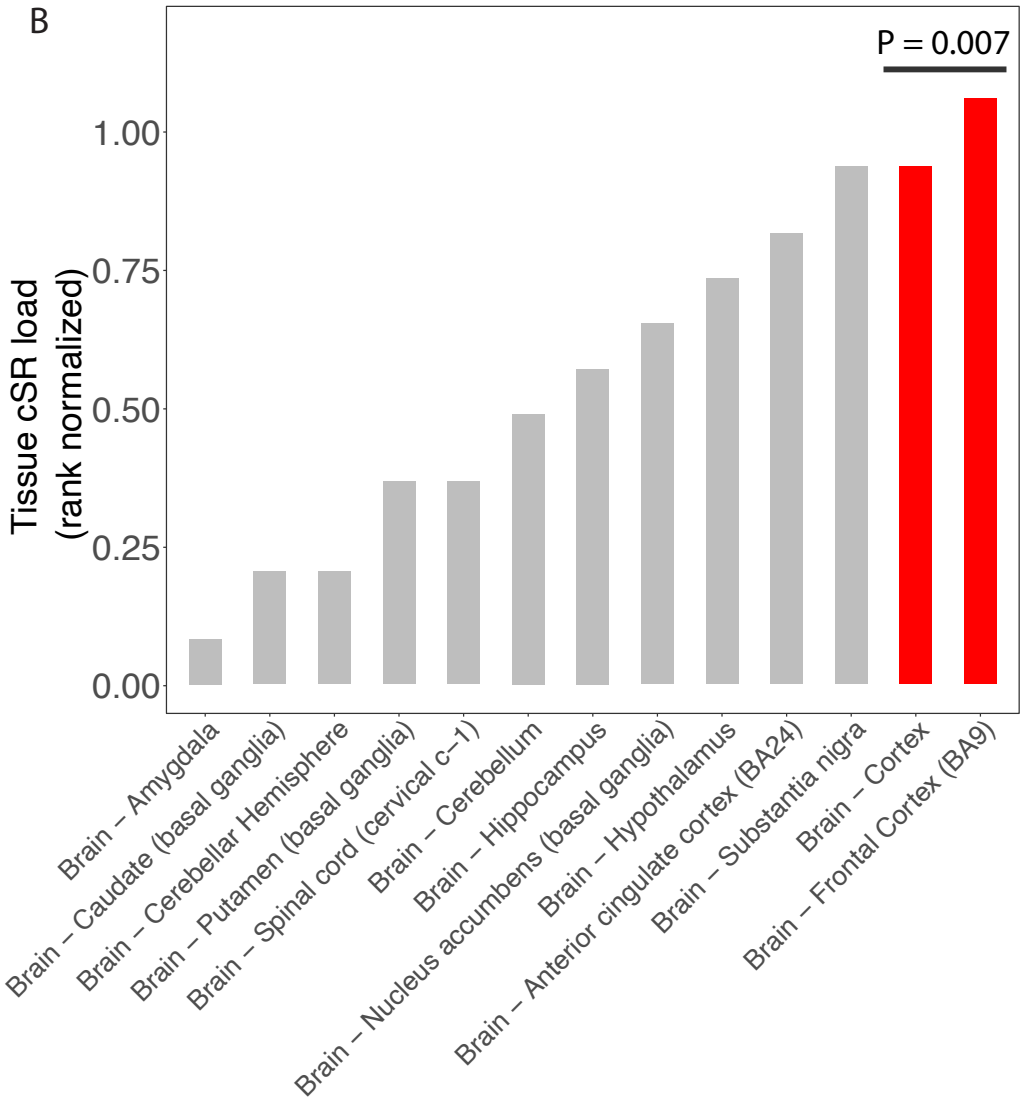
INCISOR applied on TCGA cancer data

GTEx data

DU-SR gene pairs		Low expression in normal human tissues?		
chr10 gene	chr7 gene	chr10 gene	chr7 gene	Gene pair potent?
Sample from Tissue 1				
gene A1	gene B1	yes	no	yes
gene A2	gene B2	yes	no	yes
gene A3	gene B3	yes	no	yes
Sample from Tissue 2				
gene A1	gene B1	no	no	no
gene A2	gene B2	yes	yes	no
gene A3	gene B3	no	yes	no



B



C