Quantitative analysis of genetic interactions in human cells from genome-wide CRISPR-Cas9 screens

Maximilian Billmann[1,2]†*, Michael Costanzo[3]†, Mahfuzur Rahman[1], Katherine Chan[4], Amy Hin Yan Tong[3], Henry N. Ward[1,6], Arshia Z. Hassan[1], Xiang Zhang[1], Kevin R. Brown[4], Thomas Rohde[2], Angela H. Shaw[2], Catherine Ross[3], Jolanda van Leeuwen[3], Michael Aregger[3], Keith Lawson[3], Barbara Mair[3], Patricia Mero[4], Matej Usaj[3], Brenda Andrews[3,5]*, Charles Boone[3,5,7,8]*, Jason Moffat[3,4,5]*, Chad L. Myers[1,6]*

Affiliations

1. Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA.
2. Institute of Human Genetics, School of Medicine and University Hospital Bonn, University of Bonn, Bonn, Germany
3. Donnelly Centre, University of Toronto, Toronto ON, Canada
4. Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada
5. Department of Molecular Genetics, University of Toronto, Toronto ON, Canada
6. Graduate Program in Bioinformatics and Computational Biology (BICB), University of Minnesota, Minneapolis, MN, USA.
7. RIKEN Center for Sustainable Resource Science, Wako 351-0198, Saitama, Japan
8. Canadian Institute for Advanced Research (CIFAR)

† These authors contributed equally to this work

*Corresponding authors:
m.billmann@uni-bonn.de
brenda.andrews@utoronto.ca
charlie.boone@utoronto.ca
jason.moffat@sickkids.ca
chadm@umn.edu

Abstract

Genetic interaction (GI) networks in model organisms have revealed how combinations of genome variants can impact phenotypes and underscored the value of GI maps for functional genomics. To advance efforts toward a reference human GI network, we developed the quantitative Genetic Interaction (qGI) score, a method for precise GI measurement from genome-wide CRISPR-Cas9 screens in isogenic human cell lines. We found surprising systematic variation unrelated to genetic interactions in CRISPR screen data, including both genomically linked effects and functionally coherent covariation. Leveraging ~40 control screens and half a billion differential fitness effect measurements, we developed a CRISPR screen data processing and normalization pipeline to correct these artifacts and measure accurate, quantitative GIs. Finally, we comprehensively characterized GI reproducibility by recording 4 – 5 biological replicates for ~125,000 unique gene pairs. The qGI framework enables systematic identification of human GIs and provides broadly applicable strategies for analyzing context-specific CRISPR screen data.

Main

Genetic interaction (GI) analysis is a powerful approach for systematic interrogation of gene function. A GI occurs when the combined perturbation of two genes produces an unexpected phenotype not predictable from the individual effects. Such interactions reflect diverse functional relationships between genes[1]. For example, a strong negative GI may indicate synthetic lethality where the two genes buffer one another other, which can be therapeutically exploited, as has been done for PARP1 inhibitors in the context of BRCA1-mutant cancers [2,3]. Positive GIs can reflect phenotypic suppression, mechanisms underlying protective genetic variants in humans[1,4]. When GIs are systematically measured, the profile of a gene's GIs captures its functional relationships and can be used to cluster genes in an unbiased fashion on a genome-scale[1]. In the yeast model system, exhaustive double mutant perturbation screens have been applied to successfully mapped a global network of functional relations between genes [5–7]. Similar endeavors in higher eukaryotes like *C. elegans* and *Drosophila* have used combinatorial RNAi[8–10]. Such systematic GI studies have relied heavily on the synergistic evolution of experimental and computational methods[7,11–18].

In human cells, CRISPR-based functional screening technologies have now enabled systematic GI analysis[19–21]. Two primary strategies have emerged: combinatorial perturbation, where multiple gRNAs are co-expressed from a single vector[22–31] and single perturbation screens performed in cells carrying a stable pre-existing "query" mutation [32–35]. Combinatorial systems differ mainly in how gRNA expression is controlled and typically induce acute perturbation of the targeted genes. These studies are often limited in scale, targeting pairwise combinations of a few hundred genes within defined biological contexts with the currently largest study covering ~150,000 gene pairs[36,37]. In contrast, single perturbation screens combined with stable mutations often employ genome-wide gRNA libraries to screen for GIs between one or a few "anchor" mutations and the rest of the genome[32–34,38]. As an alternative to the strategies inserting both mutations experimentally, several efforts have exploited genome-wide CRISPR screens in panels

of cell lines that mimic an anchor mutation through naturally occurring mutations, DNA copy number or mRNA expression changes[39–43], an approach that limits the search space for GIs to the genetic makeup of cell lines.

The effectiveness of these platforms including the ability to scale up experiments depends on the development of tailored computational pipelines. Data from combinatorial gRNA expression systems pose unique challenges including positional gRNA expression effects, and often require customized pre-processing workflows. Initial comparisons of computational strategies for analyzing dual perturbation data have begun to merge[23,26,44–49]. Genome-scale query screens are often analyzed with adapted methods from standard CRISPR screens, which quantify how the abundance of cells with a certain perturbation changes over time[33,50–57]. These methods have proven highly effective for interpreting large-scale CRISPR screens across diverse cancer cell lines[53,54,58–60]. However, quantifying GIs requires measuring how CRISPR screen phenotypes change in the presence of a query mutation. Despite the availability of extensive genome-wide CRISPR screen datasets and refined analysis tools, comparable methods for systematic GI detection remain limited, likely due to the scarcity of large datasets involving multiple genome-wide query screens.

Here, we introduce the quantitative Genetic Interaction (qGI) score, a method for precise identification of GIs from genome-wide CRISPR-Cas9 screens in isogenic human cell lines. Developed using hundreds of screens with the TKOv3 gRNA library in HAP1 cells, each harboring a specific loss-of-function mutation, the qGI score robustly measures interaction effect sizes and significance between library genes and the query gene. The qGI scoring pipeline fits fitness effects measured in a given HAP1 clone against a panel of HAP1 reference screens in a batch-aware fashion. We identify and correct surprising chromosomal effects between isogenic HAP1 clones that otherwise introduce linkage-dependent false positives. Additionally, we developed a two-step singular value decomposition (SVD)-based correction to mitigate genetic interaction-independent, biologically coherent co-variation and experimental artifacts. Finally, we employed a novel Markov chain Monte Carlo (MCMC)-based method on seven query screens with 30 biological replicates to rigorously quantify false positive and false negative rates for qGI scores. Together, this computational pipeline enables the generation of a human reference genetic interaction network and provides broadly applicable tools for interpreting differential CRISPR screens.

Results

Genome-wide CRISPR-Cas9 screening across isogenic genetic backgrounds in HAP1 cells

A genetic interaction (GI) occurs when the combined perturbation of two genes produces a phenotype that cannot be predicted from the individual effects of each gene. To systematically map GIs in the human genome, we performed genome-wide CRISPR-Cas9 drop-out screens, measuring cell fitness as a composite phenotype following the single perturbation of every targetable protein-coding in a defined isogenic genetic background (Fig. 1a). In this setting, the isogenic background introduces the second (query) mutation, and interaction with each library gene are assessed. A GI is classified as positive when the observed cell fitness exceeds the expected

value and as negative when it falls below expectation, where the expectation is derived from an estimated null model. This null model estimates how genome-wide CRISPR screens differ from one another, accounting for the library gene-specific phenotypic variance observed under identical conditions.

To establish the first-generation human reference genetic interaction network (GIN), we conducted 363 genome-wide screens including 39 wild-type (wt) control screens, generating a dataset suitable for both accurate identification of GIs and systematic network analysis[61]. We selected the near-haploid human HAP1 cell line and derived isogenic clones harboring defined loss-of-function mutation in query genes. Genome-wide screening was performed using our TKOv3 gRNA library[62], which targets nearly 18,000 protein-coding genes with four gRNAs per gene, with Cas9 expressed from a gRNA-containing plasmid to ensure consistency across query backgrounds (see Methods for details). In total, this dataset comprised fitness effect measurements for approximately 23 million query gene-library gRNA double perturbations and 2.7 million single gRNA perturbations (Fig. 1b), providing the foundation for the development of the quantitative Genetic Interaction (qGI) score.

Robust identification of quantitative genetic interactions (qGI) from isogenic CRISPR-Cas9 screens

Distinguishing phenotypic variation caused by a query mutation from between-screen variation driven by incidental experimental factors, such as batch effects, clonal effects or gRNA sequencing data abnormalities, posed a central challenge to identify and quantify GIs. To address this, we developed the qGI score and its associated false discovery rate (FDR), which encompasses a series of normalization steps designed to account for both expected and previously unrecognized, yet highly impactful, artifacts in isogenic CRISPR screens. Normalization steps include fitting a robust null model, multi-level variance stabilization, correction of genome location-dependent artifacts, and unsupervised correction of previously unknown biologically driven screen-to-screen variation and covariation (Fig. 1b).

The qGI pipeline contrasts gRNA-level fitness effects in wt control and query screens by fitting a non-linear null model to M-A-transformed fitness effects of each query-wt screen pair (see Methods for details) (Fig. 1b, Fig. 2a). Fitness effects in each screen were quantified by computing the log2 fold-change (LFC) of normalized gRNA abundance between the initial (T0) and endpoint (T18; 10-12 population doublings) time points, followed by basic per-screen and per-gRNA quality assurance (see Methods for details; Fig. S1a-k). This null model substantially improved GI profile performance as measured by the ability of profile similarity on the resulting scores to predict known functional relationships (Fig. 2b).

To robustly quantify GIs and identify confounding patterns unrelated to the query gene, we computed all pairwise differential effects between the wt screens and each query screen. This resulted in a data cube comprising approximately half a billion measurements including more than 6000 query-wt screen pairs, each containing around 70,000 gRNA-level differential effects (Fig. 1b). Key corrections applied by the qGI pipeline included normalization of local chromosomal

shifts, SVD-based correction of genetic interaction-independent residual fitness effects, both of which are described below in detail, and exclusion of gRNAs whose effects were inconsistent with other gRNAs targeting the same gene (Fig. 1b). To quantify each query-library gene interaction, multiple query-wt screen pairs were contrasted, where each query ko screen was contrasted to several wt control screens, and gRNA-level values were summarized to estimate the effect size, the qGI score. The variance between sequence-independent gRNAs targeting the same gene and query-wt control contrasts was used to estimate the significance of the effect (see Methods for details) (Fig. 2c-f; Figure S2a, b).

Additionally, to improve the accuracy of GI scores, the qGI pipeline computed a gRNA quality score (QS) for each gene, measuring the consistency of differential effects among gRNAs targeting the same gene. Using this metric, we excluded the lowest-performing gRNA for 517 genes (Fig. 2e, g, h; Fig S2c-g). We observed that a two-gRNA-per-gene library, selected based on the QS, has the potential to perform as well as the full 70,000-gRNA TKOv3 library, but four gRNAs per gene were crucial for accurate scoring of this dataset since the gRNA QS requires a large collection of screens to be computed (Fig. 2g, h; Fig. S3a-e).

Altogether, the qGI score contrasts fitness effects from genome-wide CRISPR-Cas9 query and wt control screens and applies multiple corrections to isolate the differential effects attributable to the query mutation (Fig. 2i, j). Full details of the entire qGI pipeline are described in Methods, and in the following sections, we describe two key correction modules that account for unexpected patterns that would otherwise confound the genetic signal in differential CRISPR-Cas9 screens.

Genomic linkage of genetic interactions in isogenic cell lines

Perhaps the most well-characterized cause of spurious fitness effects in CRISPR screens is copy number (CN) variation. On-target CRISPR-Cas9-induced DNA double-strand breaks generate gene knockouts but also trigger DNA repair-associated toxicity at the target site. This toxicity scales with CN and, if uncorrected, can lead to false positive gene essentiality calls in CRISPR screens[54]. Consequently, unequal CN in different cell lines can produce cell line-specific fitness effects and introduce genomic linkage-driven coessentiality among neighboring genes[50,53,54,63]. Conceptually, a key advantage of isogenic CRISPR screens is the ability to trace back differential fitness effects to a defined mutation in an otherwise identical genomic background, free of structural differences such as CN variation between a control and query clone. However, cultured cell lines are often genomically unstable, and such instability may arise during the generation of query clones or over the course of a genome-wide screen.

To understand the impact of ploidy changes on GI identification, we tested two scenarios for HAP1 cells. First, HAP1 cells, which are haploid (1n), can undergo diploidization (2n) during query clone generation or screening. Given the CN-dependent toxicity of CRISPR-induced DNA damage, we reasoned that contrasting CRISPR screens performed in wildtype cells that remained haploid with those that became diploid during the experiment, or those that started diploid may reveal inflated numbers of differential effects (pseudo-GIs), if unaccounted for by our scoring pipeline. However, neither the number of gRNAs producing strong fitness defects nor the number

of detected pseudo-GIs differed between the three ploidy classes (Fig. S4a-d). Thus, we conclude that whole-genome ploidy differences at the start or end of a screen are unlikely to systematically generate false positive GIs.

Next, we examined a well-described case of local differential ploidy in the HAP1 cell line: a 30 Mbp region of chr15 that is heterozygous diploid. Specifically, we tested GIs in this region, comparing a HAP1 clone engineered to be fully haploid in this chr15 region[64] to standard HAP1 WT control screens that harbor the heterozygous diploid chr15 region (Fig. 3a). To identify CN-associated regional GI biases, we developed an unsupervised algorithm to detect genomic regions showing systematic shifts toward positive or negative interaction scores (see Methods). In the WDR73 clone (haploid on the entire chr15), we found that gRNAs targeting 222 genes within this region exhibited markedly positive shifts (Fig. S5a-c), consistent with reduced CN-dependent toxicity. As a consequence, 33% of genes in that region showed a positive GI, a 30-fold increase above the global density (Fig. 3b) (hypergeometric test; $p < 2.2 \times 10^{-16}$). Notably, even high-confidence interactions (more stringent effect size, FDR thresholds), which are typically robust to minor biases, showed a 100-fold enrichment of positive GIs within this region, suggesting that unequal local CN between a query and wt control screen pair can cause high rates of false positive GIs.

To systematically assess genomic linkage effects across the genome, we applied this algorithm to all query screens and all chromosomes. Overall, we detected GI score shifts affecting 75,739 gene pairs (1.3% of all pairs tested), involving 42% of all query screens and 56% of all library genes. These shifts fell into two categories: shifts that affected the full chromosome or shifts that affected a subchromosomal region only. The frequency of shifts varied by chromosome, with chrX most frequently affected (26.9% of query clones), followed by chr18, chr13, chr16, chr15, and others (4.9-2.2% of clones) (Fig. 3c). Some chromosomes exhibited a preference for either full-chromosome or subchromosomal shifts (Fig. 3c; Fig. S5d); for example, chr13, chr18, and chrX predominantly displayed whole-chromosome biases, while others showed mostly subchromosomal shifts, often clustering in specific regions prone to genetic or epigenetic instability (Fig. 3d; Fig. S5e, f). When subchromosomal shifts occurred, they clustered in specific regions, suggesting that those regions are prone to genetic or epigenetic instability in HAP1 clones (Fig. 3d). Some of those subchromosomal shifts, such as those on chr16 or chr10 affected one but not the other chromosomal arm (Fig. 3d). On chrX, which otherwise predominantly showed whole chromosome biases, subchromosomal frequent shifts occurred at p11 and q28 (Fig. 3d).

We also observed rare, local shifts only found in individual clones, such as a 64-gene region at the end of chr9 that encodes TUBB4B in the TUBB query clone (Fig. 3d, e). In HAP1 wildtype cells, TUBB4B is highly expressed but not essential when depleted. In the TUBB knockout clone, TUBB4B expression is upregulated (Fig. 3f), presumably to compensate the loss of TUBB, and TUBB4B becomes highly essential, which we measure as a strong negative GI (Fig. 3e, f). This buffering relationship was further supported by data from the CCLE cell line panel (Fig. S5g-i) and in an Perturb-Seq experiment in K562 cells[65]. Besides TUBB4B, 16 neighboring genes in this region exhibited negative GIs, a ~25-fold enrichment over background. Since these genes did not

show altered mRNA expression in the TUBB clone nor dependency on TUBB mRNA level in the Cancer Dependency Map (Fig. 3f, Fig. S5j), those 16 genes were unlikely to contribute to buffering of TUBB. We found that nearly all of these 16 genes were located to the centromere-side of TUBB4B (Fig. 3f), which instead suggests that their observed fitness defects may be due to occasional DNA double-strand break-mediated loss of the end of chr9, following a mechanism previously described by Lazar and colleagues[66]. Such a loss of the end of chr9 would knock out TUBB4B collaterally and produce many false positive negative GIs.

Since this putative chr9 mechanism differs from what we observed in the heterozygous diploid region of chr15, where we observed significant mRNA deregulation in the shifted region, we performed mRNA sequencing analyses in 42 of our HAP1 query mutation clones covering 77 shifted regions. Only about 20% of these regions showed differential mRNA expression in the shifted region and, of those, only shifts on chr15 and chr8 (Fig. 3d) consistently showed differential expression coherent with the CN hypothesis. Notably, only 1 of 26 tested shifts on chrX showed differential expression. Together, this suggests that the observed shifted regions may only be attributed to local CN differences between query and control screens in a small number of cases.

The qGI scoring pipeline mitigates these genomically linked potential false positives by mean-centering gRNA scores within shifted regions (Fig. 3g). For instance, correcting the shift on chr15 of the WDR73 KO screen retained only 7 of the 77 positive interactions (qGI > 0.3; FDR = 10%) while simultaneously identifying 11 new negative interactions (qGI < -0.3; FDR = 10%) none of which passed the threshold without this correction (Fig. 3h). The correction affected each chromosome differently by removing different fractions of interactions as well as different ratios of positive and negative GI calls (Fig. 3i, k). For instance, about 30% of all negative and 10% of all positive GIs were removed on chrX (Fig. 3i). Overall, the correction removed around 6,000 false positive GIs (~3%), which exceeded the number of true positive GIs in the shifted regions approximately three-fold (Fig. 3j; Fig. S5l).

Finally, to evaluate the effect of local interaction shifts on systematic interaction analysis, we tested whether genes co-localized in shifted genomic regions exhibited inflated GI profile correlations. Several frequently shifted regions showed substantially increased relative local GI profile correlation density (see Methods) before correction (Fig. 3k). For instance, the 222 genes in the heterozygous diploid chr15 region in HAP1 cells were nearly 50 times more likely to have highly correlated GI profiles with one another. This bias was strongly reduced after correction, which substantially improved the predictive performance of GI profiles (Fig. 3k, l). Indicative of a more general issue in cultured human cell lines, several regions with shifted HAP1 interaction scores also had strongly increased relative local co-essentiality density in the Cancer Dependency Map, despite sensitive corrections for proximity bias included in the current DepMap scoring pipelines[53] (Fig. 3m).

In conclusion, we identified stretches along the HAP1 genome prone to GI score shifts that can be caused by chromosomal abnormalities, which generate numerous putative false positive GIs and inflated profile similarities when uncorrected in isogenic CRISPR screens.

Genetic interaction-independent patterns in genetic interaction screens

Relative to other large-scale phenotype data, genetic interactions are rare, which makes their detection challenging. Previous work on yeast genetic interactions demonstrated that technical artifacts can generate non-random patterns that can frequently explain substantially more variation than the genetic interactions that are the target of detection[13]. Given there have been relatively few previous data capturing large-scale genetic interaction networks from isogenic CRISPR-Cas9 anchor screens from which to fully characterize technical artifacts, we generated a large collection of control screens for this purpose. The control screens were generated with two potential types of artifacts in mind (Fig. 4a): (1) between-screen variation due to differences in query mutant growth dynamics, and (2) between-screen variation due to random differences in when the screen was conducted and by whom. Regarding query mutant growth dynamics, one of the major factors that varies in human cell culture is the growth rate of a particular query mutant. Query mutations can have a range of impacts on cell fitness and growth rate in human cell culture. The number of doublings of a query mutant cell population is a parameter that is tracked and controlled in our screening protocol, but due to the multiple serial passages, restarting and stopping of growth in culture, precise control of this factor across hundreds of screens is not practically feasible. To understand this variation, in addition to single endpoints, for several control screens, we also collected multiple intermediate timepoints, sampled every 3-4 days, to simulate query screens that were unintentionally either undergrown (fewer cell doublings than average) or overgrown relative to the average control. Regarding general between-screen variation (2), as described in Methods, screens were conducted following a strictly defined protocol, controlling as many factors as practically possible. However, there are inherently uncontrollable elements that may drift over the course of a project (e.g., staff conducting the screens, virus batch). To model this variation, we collected HAP1 WT control screens spread across time. Several control screens (n=21) were periodically completed using the same screening protocol over several years' worth of screening, which would allow us to measure systematic between-screen variation.

Using this large collection of control screens, we performed a simulation to characterize random and systematic variation associated with our GI screening platforms. Specifically, we successively held out one of the control screens and used this "decoy" screen as a test query screen to score "pseudo-GIs", i.e. differential effects that resulted from contrasting that decoy to the average of all other control screens. This scoring process included all of the components of the qGI score except for the SVD-based genetic interaction-independent pattern removal components (Fig. 1b). This process produces a residual pseudo-GI score we then analyzed for random or systematic variation and used to quantify the expected number of false positive GIs. We first focused on using intermediate sampling points to characterize residual effects produced due to differences in the query growth dynamics and/or screens' sampling time. We found that especially early time points yielded a high density of pseudo-GIs, which even surpassed the average query screen GI density (Fig. 4b). Endpoint samples from this control set, which model screen-to-screen variation effects unrelated to sampling time effects, displayed fewer pseudo-GIs, but still resulted in a substantial number of false positive interactions (31 and 44% of the median positive and

negative interaction degree, respectively) relative to the average number of GIs observed across query mutant screens (Fig. 4b). Importantly, a majority of the library genes that were GI hubs in the query gene ko screens were also pseudo-GI hubs in this control screen simulation (Fig. 4c), suggesting that uncorrected between-screen systematic variation was producing many false positives in our query mutant GI screens.

To further characterize these between-screen variation effects, we contrasted each query mutant screen with the entire collection of WT screens (n=21), individually, which resulted in 21 differential effect measurements for each query x library gene combination. For this analysis, we focused on two query mutant genes with well-characterized functions, PTAR1 and FASN. Specifically, we wanted to understand whether genes that showed positive or negative qGI scores in one query-wt contrast also did so in the remaining 20 contrasts for the same query gene. As a positive control, strong negative, functionally coherent GIs with PTAR1 or FASN were found in every of the 21 contrasts, suggesting that a simple query-wt screen comparison can robustly identify true interactions (Fig. 4d, Fig. S6a). For instance, the true negative interaction between PTAR1 and a vesicle trafficking regulator VAMP2 showed a strong negative qGI score in each of the 21 contrasts (qGI scores -2.2 to -1.6) ranking among the top 4 most negative interactions in each case (Fig. S6b). This observation also held when we tested three independent FASN screens, each of which had been contrasted to the 21 WT control screens individually: the previously identified GI between FASN and LUR1[32] was among the most negative GIs in each of the 63 contrasts (Fig. 4e). Importantly, some of the highly variable genes across these contrasts resulted in scores that were similar in magnitude in specific contrasts to reproducible interactions. We wondered whether these highly variable genes across the control screens reflected something specific about our control screens or whether that set reflected genes likely to produce between-screen variation more generally, including for our query mutant screens. To explore this, for each of the three biologically independent query screens in the FASN KO background, we first mean-summarized the 21 KO-WT contrasts, which resulted in a single control profile robust against control screen variation. The three resulting BCL2-FASN GI scores ranged from -1.2 (very strong negative interaction) to 0.4 (moderate positive) (Fig. 4f). More broadly, several of the highly variable fitness genes in WT control screens exhibited increased variance across biological replicate screens (Fig. S6d). This suggests that highly variable genes across WT control screens also vary frequently in query mutant screens and indicates that between-WT control screen variation can potentially be used to correct between-query screen variation.

Our control screen analysis highlighted a set of highly variable fitness genes with the potential to produce strong effects that also exhibit variance in query mutant screens for GIs. To design a solution for controlling for this variation to accurately score GIs, we next asked the question of whether this variation was independent across library genes, or instead, if there was evidence of *covariation* between library genes in this control experiment. To look for evidence of covariation, we used the same dataset described above in which each query gene screen was individually contrasted against the 21 WT control screens, resulting in 21 differential effect measurements for each query x library gene combination. These data provided a collection of

profiles for each query mutant screen, which allowed us to measure covariance amongst the library genes targeted by our library. Surprisingly, we found very strong covariance between biologically coherent groups of genes with high between-screen variation. For instance, the anti-apoptotic regulator BCL2 showed a positive GI with FASN and the pro-apoptotic regulator BAX, which can form a dimer with BCL2 in response to stress to trigger apoptosis, shows a mild negative interaction in one FASN-WT contrast, while BCL2 shows a slightly negative trend and BAX a positive interaction in another FASN-WT contrast (Fig. 4g). Along the profiles of the 21 FASN-WT contrasts, the BCL2-BAX GI scores were strongly anti-correlated (Fig. 4h). In another example, TSC1 and TSC2, which form a dimer to inhibit nutrient-dependent growth regulation through mTORC1, both interact negatively with FASN in one WT contrast, but a genetic interaction is absent in another WT contrast, a residual genetic interaction pattern described by a strong positive correlation (Fig. 4g, h). We computed correlations across the entire set of highly variable genes to construct a global covariation network for the FASN query (Fig. 4i). Overall, this network was strongly structured and contained several biologically coherent functional modules, which are known to be central to cell growth and death (Fig. 4i). These findings suggest there is substantial random variation and covariation present in genome-wide CRISPR screens in biologically replicated conditions whose effect size is equal or greater in magnitude than the interaction effects being measured, and that furthermore, this covariation reflects biologically coherent modules. Thus, identifying true genetic interactions (i.e., differential effects that are specifically caused by the introduced query mutations) requires careful normalization of this systematic variation.

Correcting for genetic interaction-independent patterns in genetic interaction screens

To correct for the variation and covariation observed across control screens described above, we developed a two-step singular value decomposition (SVD)-based normalization module for the qGI pipeline that selectively removes covariation patterns indicative of non-query effects. First, this module removes pseudo-GI signal followed by normalization of the most dominant signal from the query x library gene matrix to improve GI scores in a semi-supervised fashion (Fig. 5a). The second step, the removal of the most dominant signal, is motivated by the assumption that GIs are rare, and thus, remaining patterns that explain substantial variation in the GI scores are likely technical artifacts and are removed.

For the first normalization step, we created two matrices: the raw GI matrix containing robust GI estimates for each query screen by contrasting it against a weighted panel of WT control screens, and a second matrix containing pseudo-genetic interaction scores for each WT control screen, which was also contrasted to the WT control screen panel (see Methods). This latter pseudo-GI matrix was decomposed using singular value decomposition (SVD) (Fig. 5a), and we then removed the resulting singular vectors from the query screen GI scores. Specifically, we computed the projection of the query screens' raw GI scores onto the pseudo-GI-derived singular vectors and subtracted the result from the raw GI scores to compute normalized GI scores (Fig. 5a) (see Methods for details).

To illustrate the effect of this correction, we focused on the 142 genes exhibiting the highest variance between control screens (Fig. 4d, i) and arranged them into a similarity network based on their covariance structure (Fig. 4i). We visualized the correction of individual query screen interaction profiles on this similarity network by plotting the delta qGI score on this network, the difference between normalized and unnormalized qGI scores. As described above, this similarity network groups genes into coherent functional neighborhoods. For each query screen, corrections showed coherent patterns on this similarity network but varied substantially in strength and magnitude from screen to screen (Fig. 5b, c). For example, GIs between FASN and genes in the cluster around INPP4A, INPPL1, FBXW7, KCTD5 and HDAC3 were corrected to become less negative (Fig. 5b). In another cluster around TSC1 and TSC2, the correction effect had the same direction but was less extreme (Fig. 5b). Interestingly, the third area in the covariation network, the cluster around TADA2B, TADA1, HK2, SLC16A1 and ATF4 was strongly corrected to show less positive genetic interactions with FASN (Fig. 5b). Other query screens showed distinctive correction patterns relative to FASN, each with a different combination of negatively and positively corrected library genes (Fig. 5c). This control screen-based correction is driven by the covariance observed across the controls, but which collection of library genes is corrected in any given screen can vary substantially, suggesting different genetic interaction-independent patterns emerge with varying strengths in each query screen.

A major determinant of how strongly a library gene is corrected is the combination of their between-screen variance, covariation patterns and the query-specific interaction effect. Notably, this allows this correction module to be both sensitive and specific. For instance, while scores of highly variable genes were substantially modified, functionally coherent negative GIs with FASN, such as with LDLR, LDLRAP1, SREBF2, LUR1/C12orf49, were not affected by this correction (Fig. 5d). This was also reflected by functional enrichment among the top negative GIs with FASN, which showed removal of GIs with genes involved in mRNA decay but retained GIs with genes in lipid metabolism-related pathways (Fig. 5e). The same trend was observed in another example, a PTAR1 query screen, where highly variable genes showed large corrections while functionally coherent negative GIs with VAMP2, VAMP3, EXOC8 or ARL1 remained unchanged (Fig. S6e). Interestingly, for the PTAR1 query, the correction removed a significant enrichment for MTOR signaling among negative interactions with PTAR1 while retaining enrichments for Golgi vesicle biogenesis reflecting PTAR1 function in vesicle trafficking (Fig. S6f). Overall, this correction module removed the majority (~90%) of pseudo interactions we observed in our simulation experiment (Fig. 5f). Overall, a surprising 45.2% of GIs in query screens were removed by this correction module, indicating a pervasive false positive signal in uncorrected GI calls (Fig. 5f). Among the GIs that were removed, positive GIs were more frequent (~57%), but the proportion of positive and negative GIs removed varied substantially between different library genes. Some genes, such as HDAC3, MORC2 and NF2, primarily lost negative GIs, while others, such as WSB2 or PTDSS1, lost more positive GIs.

In addition to the control screen SVD-based normalization, we applied an additional normalization step to correct additional technical artifacts in the genetic interaction data. Given

that genetic interactions are rare, any low-dimensional pattern that explains a substantial portion of variance in the data likely represents technical artifacts, not true genetic interactions. Previous studies on large-scale genetic interactions have demonstrated that normalization of such patterns (e.g., the strongest principal components) from interaction data can improve the quality and specificity of the data[13,53,67]. To test how top principal components contribute false positive signal in our context, we measured how well GI profiles reconstructed co-membership of gene pairs in orthogonal functional standards. We found that removing the first 4 principal components from the data, which account for ~11% of the total variance (Fig. S6g), increased the performance of qGI scores to detect co-complex and co-pathway membership (Fig. S6h, i). In contrast to the first SVD-based correction module, this second module, the top SV-removal correction, removed positive and negative GIs in a more balanced fashion. At the same time, the fraction of positive to negative GIs again was highly library gene-specific. Notably, this module removed many positive GIs involving components of the respiratory chain complex I, such as NDUFAF6, NDUFA6 or NDUFB8, which we previously suggested can dominate CRISPR screen data in part due to technical artifacts[68]. In contrast, a higher number of negative GIs involving SWI/SNF complex components such as SMARCC1, SMARCB1 or SS18 were removed.

Cumulative impact of qGI normalization on interactions

As described above, we characterized several biological and technical artifacts arising from our genome-wide CRISPR/Cas9 screening platform that obscure true genetic interaction and developed a collection of normalization procedures to address each one. We performed a series of ablation studies in which we removed components of the normalization pipeline and evaluated the impact on the quantity and quality of the resulting interactions. Specifically, we evaluated three variants of the qGI method relative to the complete approach, each with progressively more limited functionality: one without the removal of the top 4 SVs, one without any SVD-based correction, and one without the non-linear null model on M-A-transformed fitness effect data (referred to as simple dLFC). Each of those ablation experiments suggested that the qGI score correction modules we tested significantly improved the functional information derived from the resulting qGI scores, and cumulatively, they had a major impact on the qGI profiles' capacity to predict co-complex, co-pathway and co-GO BP memberships (Fig. 6a, Fig. S7a, b). In addition, each correction module reduced the number of strong GIs overall by more than 80% (Fig. 6b). This is important, because it contradicts the frequent assumption that technical artifacts in CRISPR screen data mostly affect moderate and weak phenotypes and that strong effects reflect the true signal of interest. In total, proportionally more positive than negative GIs were removed (Fig. 6b), and per screen, the positive-to-negative GIs ratio became more balanced upon normalization (Fig. 6c). For example, for the WNK1 query screen, a simple dLFC measurement generated large numbers of negative GIs due to a non-linear relation to the WT control reference set (Fig. 6d). Applying an accurate null model as well as the additional qGI correction modules removed many strong effects while retaining strong, expected GIs, such as those between the WNK1 paralogs WNK2 and 3. This trend for the WNK1 query, where a large fraction of genes with a specific fitness effect are called as interactions when no corrections are used due to the non-linear query-control relation of

genome-wide fitness effects was broadly representative of corrections for other queries as measured via a global correlation analysis (Fig. 6e).

Reproducibility analysis of qGI-derived interactions

In addition to evaluating the impact of the qGI normalization components, we also performed a comprehensive characterization of the reproducibility of the quantitative GIs measured by our pipeline. To do this, we generated a comprehensive dataset comprising 7 diverse query backgrounds and, in total, 30 biologically replicated, genome-wide screens. We applied a novel MCMC-based method, called JEDER [61,69,70], on this data to test precision and recall of genetic interactions identified in isogenic genome-wide CRISPR-Cas9 screens. From the four to five independent screens performed for each of the queries, JEDER generates a GI consensus, which can then be used to determine precision and recall at different combinations of thresholds based on interaction effect size, the qGI, and the associated FDR. To define standard and stringent GI thresholds, we selected two separate qGI/FDR threshold pairs to provide "standard" or "stringent" GI calls, which reflect different balances in the tradeoff between precision and recall (Fig. 6f, g). Standard GIs were defined at a JEDER-estimated precision of 61% and a recall of 37%, which was achieved at an absolute qGI score > 0.3 and an FDR < 10%. Stringent GIs were defined at a precision of 91% and a recall of 12%, which was achieved at |qGI| > 0.6 and FDR < 1%. Those two threshold pairs robustly identify GIs from genome-wide CRISPR-Cas9 screens in isogenic cell lines under the consideration of various experimental artifacts as well as empirically defined reproducibility metrics.

In addition to reproducibility of thresholded interactions, we evaluated the quantitative reproducibility of interaction profiles. Depending on the number of significant interactions, which varied broadly across the different query backgrounds we screened, between-replicate Pearson's correlation coefficients of genome-wide qGI scores ranged from approximately 0.2 for the PTAR1 query screens to 0.71 for the VPS52 screens (Fig. 6h, j; Fig. S7g). Using the Within-versus-Between-context replicate Correlation (WBC) score, a metric displaying how well a pair of replicates is distinguished from screens performed in another genetic background that can be interpreted as a z-score[71], we found that all replicates passed the threshold of WBC > 3. Between-replicate correlation coefficients consistently decreased at every correction step, but importantly, the distinction of replicates from different genetic backgrounds, as measured by the WBC score, increased substantially as steps of the qGI pipeline were applied (Fig. 6i, j; Fig. S7f). This suggests that experimental artifacts present within data produced by genome-wide CRISPR screens can contribute to non-specific background correlation between biologically unrelated screens, and that our qGI normalization pipeline can effectively remove those artifacts.

Discussion

We report the qGI score, a computational pipeline for accurate identification and quantification of genetic interactions from genome-wide CRISPR-Cas9 screens. Our framework incorporates a

series of data correction steps to control for experimental artifacts, including genomic linkage effects and biologically driven, genetic interaction-independent covariation between screens. The identification of these artifacts was unexpected given the use of isogenic cell lines and underscores the value of systematic evaluation of even controlled experimental systems. Additionally, we applied an MCMC-based approach to empirically characterize the precision and recall of qGI-derived interactions at different effect size and FDR thresholds, applied across on a large panel of replicated screens. Collectively, we provide a detailed scoring pipeline for and detailed analyses of GIs derived from genome-wide CRISPR screens. We anticipate that both the insights about technical artifacts and the correction strategies will be broadly applicable to other differential, or context-specific, CRISPR screening studies.

Context-specific CRISPR screens, where the context reflects a specific cell line, cell type or tissue, can be strongly affected by false positive differential fitness effects resulting from genomic linkage, via at least two different well-described mechanisms[50,51,53,54,66]. Most prominently, unequal DNA copy number (CN) between cell lines can exacerbate fitness effects for genes located within CN-altered regions[54]. Isogenic CRISPR screens are designed to avoid this by introducing a single, defined mutation in an otherwise identical genome. However, our analysis revealed chromosomal regions exhibiting systematic shifts in GI scores, resulting in numerous false positive, and to a lesser extent, false negative GIs, as well as linkage-driven biases in GI profile analyses.

We found evidence for multiple mechanisms for these genomic linkage artifacts. First, local CN alterations where a diploid region in a query clone is compared to a haploid region in the control produce spurious positive GIs to many genes in that region. This phenomenon parallels the CN effect previously described for context-specific essential genes, suggesting that CN effects should be considered when comparing clones to their parental cell line. In another scenario, strong negative GIs involving a library gene located near a chromosome end, such as TUBB4B in the TUBB knockout clone, produced numerous, spurious negative GIs. As described by Lazar and colleagues in the Cancer Dependency Map (DepMap)[66], sporadic loss of the chromosome arm may cause collateral loss of the strong interacting gene, such as TUBB4B, as well. Despite evidence we found supporting the presence of CN or chromosome arm loss effects in some cases, based on our RNAseq analysis of a large number of 42 query mutants, these mechanisms only explain a minority of the genomic linkage artifacts we observed, so more study of alternative mechanisms is worthwhile. Importantly, we observed that large-scale GI shifts affecting entire chromosomes, especially chrX, occurred in nearly one-third of query screens, suggesting that epigenetic regulation of sex chromosomes may require specific attention in future screens.

Context-specific functional screening datasets such as our HAP1 genetic interaction network and the DepMap are typically generated over years, during which experiments are conducted in different batches and under varying conditions. Batch effects and random noise are well-recognized challenges and can be addressed through established normalization strategies[9,13,53,54], and are implemented in our qGI pipeline. However, beyond technical variability, we identified pervasive, biologically coherent covariation across both control and replicated query

screens, with effect sizes comparable to the strongest GIs in a typical genome-wide query screen. Traditional genomics quality control approaches typically check for biological coherent signal among top effects for a given experimental pipeline using functional enrichment tests or related approaches. Such techniques would be blind to false positive GIs resulting from the between-screen covariation patterns we describe here, and may even identify spurious signal as evidence of strong screen quality. To correct for those effects, we developed an SVD-based method that identifies and removes shared patterns of variation present in HAP1 control screens from each query screen in a screen-specific manner. This correction will have a substantial impact on the analyses of context-specific screens, because it deprioritizes particularly many genes central to heavily investigated research areas such as cancer biology and aging. We found those genes to frequently and strongly vary between control screens in coherent patterns. We expect that this observation will further help to interpret context-specific functional screening data and improve the utility of such experiments to become more powerful discovery tools.

In the qGI scoring pipeline, correction modules sequentially reduced between- replicate screen correlation while progressively increasing query-specific signal reproducibility as measured by the within-vs-between-replicate correlation (WBC) score. These improvements were reflected in the improvement in functional information derived from GI profiles following each correction step. Together, the correction modules integrated into the qGI pipeline markedly improve the accuracy and specificity of GI quantification from genome-wide CRISPR screens and provide a means for accurate, comprehensive analysis of genetic interaction analysis in human cells.

Methods

Pooled genome-wide CRISPR-Cas9 dropout screening in HAP1 query mutant cell lines

Human HAP1 wild type cells were obtained from Horizon Genomics (clone C631, sex: male with lost Y chromosome, RRID: CVCL_Y019). HAP1 gene knockout cell lines were either obtained from Horizon or generated in house. See Billmann *et al.* for cell line annotation files and experimental details[61]. All gene knockout cell lines were confirmed to carry the expected out-of-frame insertions or deletions.

The experimental steps were previously described [32,72]. In brief, CRISPR library virus production was performed in HEK293T cells using the lentiviral lentiCRISPRv2 vector containing the TKOv3 gRNA library (Addgene #90294) [72]. Virus-containing media was harvested 48 hours after transfection and aliquoted and frozen at -80°C.

For determination of viral titers for pooled screens in HAP1 cells, multiplicity of infection (MOI) of the titrated virus was determined 72 hours post-infection by comparing percent survival of puromycin-selected cells to infected but non-selected control cells.

Pooled genome-wide screens were performed as previously described [32,34]. A total of 90 million HAP1 cells were transduced with lentivirus containing the TKOv3 gRNA library at a MOI of ~0.3, such that each gRNA was represented in ~200-300 cells. Transduced cells were allowed to recover for 24 h, selected using puromycin (1ug / mL) for 48 hours, and then a pellet was collected to represent the T0 starting population. The remaining cells were split into triplicates, and cultured for up to 20 population doublings while maintaining a library coverage of at least 200-fold throughout. Cells were passaged every 3-4 days and cell pellets, one from each triplicate, were collected to represent the midpoint and endpoint populations. Screens were performed in either DMEM with 10mM glucose and 1mM glutamine ('minimal medium'), or IMDM with 25mM glucose and 1mM pyruvate ('rich medium') (see Billmann *et al.* for data file[61]). All media was supplemented with 10% FBS and 1% penicillin–streptomycin, and cultures were grown at 37˚C with 5% CO2.

The qGI score was developed and characterized using an initial set of 363 genome-wide screens (324 query mutant + 39 wt controls) all of which passed basic quality control. For final interaction network analysis, described in [61], a subset of the query mutant screens was removed based on manual inspection and more detailed screen QC filters, which resulted in a total set of 298 query mutant screens used in that study.

Illumina sequencing and mapping and quantification of gRNA reads

From the collected cell pellets, the library preparation for sequencing was performed as described previously [32]. In brief, sequencing libraries were prepared from 50 µg of the extracted genomic DNA in two PCR steps, the first to enrich guide RNA regions from the genome, and the second to amplify guide-RNA and attach Illumina TruSeq adapters with i5 and i7 indices as described previously using staggered primers aligning in both orientations to the guide-RNA region (see Aregger *et al.*[32] for details). Barcoded libraries were gel purified and final concentrations were

estimated by qRT-PCR. Sequencing libraries were sequenced on an Illumina HiSeq2500 using single read sequencing and completed with standard primers for dual indexing with HiSeq SBS Kit v4 reagents. The first 21 cycles of sequencing were dark cycles, or base additions without imaging. The actual 36-bases read begins after the dark cycles and contains two index reads, reading the i7 first, followed by i5 sequences. The T0 and T18 time point samples were sequenced at 400- and 200-fold library coverage, respectively (Fig. S1b, f).

FASTQ files from single read sequencing runs were first trimmed by locating constant sequence anchors and extracting the 20 bp gRNA sequence preceding the anchor sequence. Pre-processed paired reads were aligned to a FASTA file containing the TKOv3 library sequences using Bowtie (v0.12.8) allowing up to 2 mismatches and 1 exact alignment (specific parameters: -v2 -m1 -p4 --sam-nohead).

Estimation of fitness effects in HAP1 cells

To quantify genetic interactions, the fitness of HAP1 cells was recorded upon single and combinatorial perturbation of genes. In the following sections, CRISPR-Cas9-based gene perturbations are referred to as "mutants". Single mutant fitness (SMF) effects were estimated for each individual control and query screen and for each gRNA first. This was done by computing the $\log_2$ fold-changes (LFC) between read-depth normalized gRNA abundance of gRNAs in the TKOv3 library[72] in the starting population (T0) and the endpoint (10-12 population doublings) (T18) to quantify each gRNA's dropout from the pool at the end of the experiment:

$$SMF_i = Normcount_i^{T18} - Normcount_i^{hybridT0}$$

where hybrid T0 refers to a matched T0 for each control and query screen that is additionally stabilized (defined in detail in the next section). The read-depth of gRNAs in all samples, T0 and T18, was normalized using the following equation:

$$Normcount_i = log2\left(\frac{Rawcount_i}{\sum_{i=1}^{n} Rawcount_i} \cdot 1e6 + 0.5\right)$$

Where i represents a gRNA in the library, and n is the total number of gRNA in the library.

After read-depth normalization, gRNAs with a readcount before normalization below 40 or above 10,000 in a T0 sample were excluded from further analyses (Fig. S1c). The upper threshold was used to control extreme but very rare outliers. Of the 71,090 gRNAs in the library 2,092 (1,050 on the (-), 1042 on the (+) strand) gRNAs were flagged in at least one experiment. Of those, 985 gRNAs, including 367 gRNAs that were not in the library, were flagged in at least 20% of the T0 samples (Fig. S1e), and were removed from any sample.

Population doublings were determined by cell counting throughout the course of the experiment and depended on media conditions and query mutations. Throughout the process of completing the screens for this study, several plasmid and lentiviral packaged plasmids were required, which introduced dependencies between batches of screens. Moreover, we observed that gRNA abundance after library infection and puromycin selection differed from the abundance of the library, and that this difference was highly reproducible. Those T0 gRNA populations showed a tendency towards virus batch-driven patterns. This suggested that effects on the measurement of the starting population over the duration of the GIN would be manifested in the SMF effects if not accounted for during the LFC estimation. In such cases, query-specific effects could be driven by the abundance changes before and after library infection. The qGI score uses matched T0 measurement to assure that differences between screens during library infection and Puromycin selection would not result in false positive genetic interactions. In addition, we implemented a matched T0 stabilization by also using the median across all m T0 measurements of the same virus batch (common T0). Those two estimates were combined in a weighted fashion to minimize correlation between genetic interaction scores and the difference between matched and common T0 readcounts:

$$Normcount_i^{hybridT0} = Normcount_i^{matchedT0} \cdot \alpha + Normcount_i^{commonT0} \cdot (1-\alpha)$$

If the number m of genome-wide screens in a given batch is odd,

$$Normcount_{i,j}^{commonT0} = Normcount_{i,\frac{m+1}{2}}^{matchedT0}$$

If the number m of genome-wide screens in a given batch is even,

$$Normcount_{i,j}^{commonT0} = \frac{Normcount_{i,\frac{m}{2}}^{matchedT0} + Normcount_{i,\frac{m+2}{2}+1}^{matchedT0}}{2}$$

Where i represents a gRNA in a given genome-wide screen j, and m represents the number of genome-wide screens conducted in the same batch as j.
$Normcount^{matchedT0}$ and $Normcount^{commonT0}$ were balanced to compute the $Normcount^{hybridT0}$ by finding the optimal α, where $0 \leq \alpha \leq 1$:

$$Arg\ min_{\alpha j}\left(\left|corr(\overrightarrow{GI_{\alpha j}^{approx.}}, \overrightarrow{Normcount_j^{matchedT0} - Normcount_j^{commonT0}})\right|\right)$$

α was sampled 61 times in increments of $\frac{1}{60}$. To speed up the scoring process, gRNA-level GI scores were approximated by sampling gRNAs as well as estimating the GI Null model more efficiently.

Finally, for each control wildtype and query mutant genome-wide screen, the SMF effects were adjusted, by setting the median of non-essential genes, as defined in Hart *et al.* [20], to 0.

Estimation of residual fitness effects

To score genetic interactions, SMF effects estimated using the genome-wide TKOv3 gRNA library were compared between an isogenic HAP1 cell line harbouring a stable query mutation and the parental WT HAP1 cell line. To obtain robust estimates of SMF effects in the two media conditions used for screening, 21 and 18 genome-wide screens were performed in WT HAP1 cell lines cultured in minimal and rich media, respectively. In the following, we will refer to those 21 and 18 control screens as "control screens".

The TKOv3 library contains 71,090 guide (g)RNAs that target about 18k human protein-coding genes, most of them with four sequence-independent gRNAs. gRNA-level residual scores were derived for a given genetic background by estimating a non-interacting model between LFC values in this background and each control screen. For each WT-KO screen pair, the population of LFC values were M-A-transformed [73], which contrasts the per-gRNA LFC difference (M) with per-gRNA arithmetic mean (A). Since SMF values are represented by log2-foldchanges (LFC), we computed M and A as follows:

$$M_i = QuerySMF_i - ControlSMF_i$$

and

$$A_i = \frac{QuerySMF_i + ControlSMF_i}{2}$$

where *i* represents each gRNA, ranging from 1 to 71,090. To obtain the null model for each query screen, a second-degree Loess regression for each KO-WT screen pair was least squares-fit along *A* using the R stats function loess. A was additionally locally stabilized by binning the data along *A* with equal numbers of data points in every bin to result in the vectors $M_{balanced}$ and $A_{balanced}$.

$$\mu(A_{balanced}) = LOESS(M_{balanced} \sim A_{balanced})$$

Where $\mu(A)$ is the smooth regression function estimated from bin-balanced $M_{balanced}$ and $A_{balanced}$ and represents the genetic interaction null model. ΔSMF_i^{gRNA} represents the per-gRNA residual SMF effects, which are the gRNA-level interaction scores:

$$\Delta SMF_i^{gRNA} = M_i - \mu(A_i)$$

For each gRNA, this resulted in 21 (minimal media) or 18 (rich media) residual scores, which represent the contrasts of a given KO with the 21 or 18 WT HAP1 control screens, respectively.

Overall, the qGI score compares for each screening condition all query screens to all control (wildtype) screens. In addition, for each condition, all control screens are contrasted

amongst each other to identify pseudo- (false positive) genetic differential effects. For the first GIN cell culture condition in minimal media, we contrasted 21 control screens to each other and 147 query screens. For rich media conditions, we contrasted 18 control screens to each other and 177 query screens. The control screens were also contrasted against earlier sampling time points of the control screens to cover a wide spectrum of screen sampling time points when modeling and correcting for genetic interaction-independent screen-to-screen variation (see below). Together, this resulted in 7524 pairwise comparisons between genome-wide screens including 1251 comparisons representing pseudo differential effects. We noted that the range of residual effects differed substantially, and that this variation was largely driven by SMF dynamic ranges and not associated with query phenotypes. Therefore, the qGI score scales residual effects per screen pair by adjusting the standard deviation of the central 80% of gRNA residual values. Under the assumption that specific query screens can have extreme interaction effects, the top 10% of most extreme positive and negative gRNA residual scores are not considered for standard deviation estimation. Scaling of residual SMF effects globally improved the ability of gene-level qGI scores to recapitulate genes' co-complex or co-pathway membership.

Robust estimation of library gene fitness effects

Every query screen is contrasted to multiple control (HAP1 wildtype) screens, which provides a stable estimate of the differential effects caused by the query mutation. At the same time, control screens that share the same putative (unknown) batch effect as a given query would have fewer differential effects that those with a different set of batch effects. Since putative batch effects of screens are unknown, we applied an unsupervised weighting of all query-control screen pairs. The weights were assigned based on the assumption that genetic interactions are sparse[74]. We reasoned that experimental artifacts such as batch effects would introduce additional signal into the population of residual values and assigned a higher weight to WT HAP1 screen contrasts with lower sum of squared residuals mean across the 95% least extreme ~70,000 gRNAs in the library. Excluding the top 2.5% most negative and most positive residuals, respectively, prevented penalizing wt-query contrasts with a small number of very strong residuals, which could, in contrast to consistently large residuals, represent true GIs. Based on those weights, all 21 query-control pairs in minimal and all 18 query-control pairs in rich media were summarized, resulting in a single ~70,000 gRNA-long vector of stabilized residual fitness effects for each query. We refer to the resulting value for each gRNA as the "guide-level" genetic interaction (GI) score.

Correction of genomically linked residual fitness effects

On-target CRISPR-Cas9-introduced DNA cuts are accompanied by target gene-unspecific DNA repair-associated toxicity, which has been shown to associate with copy number and genomic position and can create false essential gene calls in CRISPR screens if uncorrected[54]. Supervised and unsupervised techniques to correct for falsely identified essential genes are routinely applied [51,53,54]. Those methods are applied on single mutant fitness (SMF) data, and unequal false positive fitness effects in control and query screens can be expected to cause false positive GIs. In our

context, given that we are performing isogenic screens in which one query gene is perturbed in a common HAP1 parental background, we reasoned that any copy number or other chromosomal artifacts should largely be shared between the WT and mutant HAP1 cells. Thus, any chromosomal artifacts present uniquely in the query would become evident in the residual fitness effects (e.g., contrasting the query mutant screen with the WT control screens). We designed our chromosomal artifact correction focusing on these residual effects and developed an algorithm that to find such genomically linked shifts of residual SMF effects.

Shifts are identified for residual SMF effects at the gRNA-level. As described above, those residual SMF effects were robustly estimated by taking the weighted consensus of a given query screen compared to all control HAP1 wildtype screens in the same cell culture condition. This assured that any detected shift can be attributed to the query screen (instead of a control screen).

First, the qGI pipeline tests for differences between mean residual SMF effects ($\overline{\Delta SMF_j}$) of adjacent groups of gRNA values while moving along genome coordinates, where j is in 1 to $71{,}090 - n$. The mean residual SMF effects are defined as:

$$\overline{\Delta SMF_j} = \frac{1}{n}\sum_{i=j+1}^{j+n} \Delta SMF_i$$

With this mean residual SMF effect, differences between groups are tested as:

$$\left| \overline{\Delta SMF_j} - \overline{\Delta SMF_{j+n}} \right| > c$$

The size of this group is defined per chromosome and is higher for chromosomes for a higher gRNA per bp ratio. The group size thus covers comparable distances of the genome while maintaining a number of measurements (gRNA values) that allows a stable estimate of the residual SMF effect. The group size depended on the number of gRNA on each chromosome k_{chr} (see R code implementation for details).

$$n \sim gRNA\ density_{chr}$$

If a difference between mean residual SMF effects (dLFC) of adjacent groups of gRNA values is detected, its genome coordinate (X) is saved. Shifts in residual SMF effects between X_i and X_{i+1} or, for the first and last X on a chromosome, between X_i and the start or end of the chromosome are tested. If the mean residual SMF effects in such a genomic region deviates from 0 (positive and negative deviation is possible) more than a defined threshold (flexible parameter, set at > 0.15 for this study), this region is saved as a putative shifted region.

Finally, to avoid affecting between-gRNA variance for genes at the boundaries of shifted regions, gRNAs were assigned to a shifted region if all gRNAs targeting the same gene were identified. Those shifted regions are sub-chromosomal. In addition, potential residual SMF shifts affecting each entire chromosome are identified for chromosomes without detectable local shifts. For this whole-chromosome test, we measured if the mean gRNA-level residual SMF effect for

the entire chromosome deviated from 0 (positively or negatively) more than a defined threshold (flexible parameter, set at >0.09 for this study). For any regions with identified chromosomal shifts, the residual SMF effects were corrected by subtracting the mean effect from all gRNA-level measurements in the affected region.

Unsupervised correction of genetic interaction-independent interaction patterns using SVD

The qGI score was designed to robustly identify genetic interactions, which are differential fitness effects between a query and wildtype control screen that can be attributed to the query mutation. A simple comparison between control and query CRISPR screens is obscured by a compendium of factors contributing to structured screen-to-screen variability, many of which are poorly understood. In our data, screen-to-screen variability affects both the control as well as the query screens (Fig. 4a-i; Fig. S6a-d). The qGI score implements a series of correction steps that estimate artifacts from the large collection of control and query mutant data and robustly correct for those artifacts. Unwanted screen-to-screen variance as well as genetic interaction-independent covariance patterns are removed via a two-step singular value decomposition (SVD)-based qGI correction module.

In the first step, the qGI scoring pipeline estimates and corrects screen-to-screen variation patterns in screen-to-screen differential fitness effects from wildtype control screens. Specifically, like a query screen, each wildtype control screen is contrasted to all remaining control screens. Again, as for the query screens, all resulting contrasts for a given control screen with all other control screens are summarized. This generates a matrix of control screens (n = 21 for minimal media; n = 18 for rich media) X gRNAs (~70,000) of residual fitness effects, only that residual effects are not the result of a query mutation. This captures all genetic interaction-independent (i.e. no query gene) screen-to-screen variation. In addition, the effective number of doublings past library infection when a screen is stopped can affect fitness phenotypes [60,68], which we refer to as temporal screen-to-screen variation, and this can occur when query cell lines grow substantially slower than HAP1 wildtype cells and the cell population goes through substantially fewer cell doublings. Therefore, additional intermediate time points between day 6 and day 17 were collected and like all individual endpoint control screens, are contrasted to the remaining endpoint control screens. This causes strong residual fitness effects (Fig. 4b) and captures temporal screen-to-screen variation. This extended the residual fitness effects matrix to 40 (21 of which are endpoint) and 25 (18 of which are endpoint) wildtype control screens times gRNAs (~70,000) for minimal and rich media, respectively. Utilizing these control screen data matrices, the query screen residual fitness effect matrices are corrected in multiple steps.

First, the genetic interaction-independent residual fitness effect variance in query data was estimated for each library gRNA. This was done by computing the per gRNA-variance across query screens as well as control screens. Here, intermediate time point data was not considered, because those were taken from a screen that was represented by an endpoint and samples taken from the same screen are highly correlated, which impacts variance estimates. For the entire gRNA population, the query variance was fitted via a first-degree loess model with a span of 0.5. This

per-gRNA estimate was then scaled to have a mean of 1 and query screen as well as all control screen residual fitness effects were divided by this variance estimate.

Second, genetic interaction-independent covariance patterns were removed. This was done by decomposing the 40 (min) or 25 (rich) control screens $c \times$ ~70,000 gRNA l variance-corrected residual fitness effect matrices via singular value decomposition (SVD) using the R function svd(). NAs were imputed using row mean values. The left singular vectors $L_{l\times c}$ were then multiplied to create a matrix $A_{l\times q}^{pseudoGI}$.

$$A_{l\times q}^{pseudoGI} = L_{l\times c-1} \cdot L_{l\times c-1}^{T} \cdot A_{l\times q}^{uncorrected}$$

The unit matrix was projected onto the variance-corrected query residual fitness effect matrix, and the result was then subtracted from the variance-corrected query residual fitness effect matrix.

$$A_{l\times q}^{ctrl_corrected} = A_{l\times q}^{uncorrected} - A_{l\times q}^{pseudoGI}$$

This removal of genetic interaction-independent covariation patterns was done for each media condition separately.

In the second genetic interaction-independent pattern normalization step, the qGI scoring pipeline removes dominant query screen covariation patterns. Previous studies involving large genetic screens have demonstrated the utility of removing dominant covariation patterns to purify underlying sparse, biological signal[6,13,53,67]. The qGI pipeline decomposes the corrected residual fitness effect matrix between all non-redundant query screens and the ~70,000 gRNAs via SVD. To avoid capturing latent variables that represent genetic interaction screens with a larger number of biological replicates, query genes with larger numbers of biological replicate screens were represented by one of those replicates. Just as for the wildtype control screens, NAs were imputed using row mean values. The left singular vectors $L_{l\times sv}$ were then multiplied to create a matrix $A_{l\times q}^{SVrm}$.

$$A_{l\times q}^{SVrm} = L_{l\times sv} \cdot L_{l\times sv}^{T} \cdot A_{l\times q}^{ctrl_corrected}$$

This was done by setting the first, the first two, first three, etc. singular vectors to 0. The different matrices were projected onto the query residual fitness effect matrix $A_{l\times q}^{ctrl_corrected}$ and the resulting representation $A_{l\times q}^{SVrm}$ was subtracted from the query residual fitness effect matrix.

$$A_{l\times q}^{corrected} = A_{l\times q}^{ctrl_corrected} - A_{l\times q}^{SVrm}$$

This gRNA-level genetic interaction score matrix was summarized at the gene level and the capacity of genetic interaction profiles to reconstruct CORUM 3.0 co-complex[75], co-pathway (MSigDB Bioconductor package) and co-GO biological process memberships was assessed using the FLEX pipeline [68]. Based on the precision-recall performance, the first 4 singular vectors were removed from the data presented in this study. The decision of how many singular vectors should be removed is a dataset-dependent decision that should be made with all available information about screens and an objective benchmarking pipeline (e.g. FLEX) [68].

Empirical TKOv3 gRNA quality score

Before summarizing gRNA-level into gene-level qGI scores and computing the associated FDR, gRNAs were tested for their gene-specific phenotype, and poorly performing gRNAs were excluded based on an empirical TKOv3 gRNA quality score (gRNA QS). This gRNA QS is based on the observation that interaction profiles are highly informative of gene function and the assumption that independent gRNA sequences targeting the same gene have similar interaction profiles. We further assumed that the majority of independent gRNA sequences for each gene are very unlikely to have off-target signal that is correlated, and that any observed strong GI profile correlation between gRNAs targeting the same gene is on target. Therefore, an outlier gRNA that does not correlate with the other gRNAs against the same gene, when others do correlate, is assumed to perform poorly and should not be used for computing gene-level scores.

To find and exclude those poorly performing gRNAs, for each gene targeted by the TKOv3 library, the pairwise Pearson's correlation coefficient (PCC) was computed between all gRNA targeting that gene along the corrected gRNA-level interaction scores across all 324 query screens. For all gRNAs targeting the same gene, the mean of PCCs was computed (excluding self-correlation), resulting in the gRNA QS. Based on this gRNA QS, each gene was tested for outlier gRNAs. First, gRNAs were only allowed to be flagged if other gRNAs targeting the same gene showed evidence of coherent interaction signal because we reasoned that reducing the number of gRNAs for genes with subtle interaction signal is likely to magnify spurious interaction signal driven by a single gRNA. This was judged based on the assumption that true and detectable interaction signal of at least two gRNAs targeting the same gene pass a minimum gRNA QS. The threshold was chosen such that 40% of the genes (7,129 of 17,804 genes) were considered to have sufficient interaction signal based on the mostly positive within-gene gRNA interaction score correlation (Fig. S2c). Selecting a higher threshold further increases this positive shift for within-gene correlations and would increase the confidence in gRNA QS-based calls but would also limit the sensitivity to flag low-quality gRNAs. The large majority of the selected genes show evidence of expression in HAP1 (Fig. S2d, e), but a substantial portion of these genes show no significant single mutant fitness effect (~40.7% are not detected at FDR < 20%) (Fig. S2f).

Second, for each of those 7,129 genes with reliable interaction signal, the worst gRNA per-gene based on its QS was flagged if it had a large negative differential gRNA QS to the per-gene median gRNA QS by a second threshold shortlisting the worst gRNAs for 777 (4.4% of 17804) genes. Amongst this set of implicated gRNAs, a subset was kept if their gRNA QS surpassed a

third threshold showing reasonable agreement with the other gRNA against the same gene, resulting in a reduced set of 517 genes (2.9% of 17804) flagged for low gRNA QS. For those 517 genes, the worst gRNA was removed before computing the SMF, the qGI score and the qGI-associated FDR (Fig. S2g).

Gene-level quantitative genetic interaction (qGI) scores

After the corrections described above were applied, we computed a quantitative, gene-level score summarizing the genetic interaction effect per query gene-library gene combination. The qGI score consists of an effect size measure (a differential log2-fold change) as well as an associated per-screen false discovery rate (FDR), representing the statistical confidence in a non-zero interaction for that gene pair, accounting for the multiple tests per screen. Gene-level genetic interaction scores are computed by mean-summarizing corrected gRNA values, with poor empirical TKOv3 gRNA library quality scores removed. Therefore, the resulting qGI score benefits from all previously described correction steps. The qGI-associated FDR is computed based on residual fitness effect scores that are corrected for coarse screen-specific effects (i.e., genomic linkage, residual-on-readcount abundance dependency, per-screen residual effect scaling, wt-query contrast weighting) but without between-screen normalization steps (i.e., SVD-based corrections for screen-to-screen variance and covariance). Also, the multiple wildtype control contrasts for each query screen were not summarized before estimating FDR. For each query-library gene pair, a p-value was computed using limma's moderated t-test, which considered independent gRNAs targeting the same library gene as biological replicates and values derived from different control vs. query screen comparisons as technical replicates[76]. Here, limma's duplicateCorrelation() function is used to estimate the inter-duplicate correlation for each query screen (Fig. S2a, b). For each query screen, p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method, resulting in the qGI-associated q-values, which we refer to hereafter as the FDR.

Evaluation of genetic interaction reproducibility

To quantify qGI score reproducibility, the Within-vs.-Between context replicate Correlation (WBC) score (described in [71]) was computed on all query screens with two or more biological replicates. To summarize briefly, first, the pairwise Pearson's correlation coefficient was computed between all biological replicate query screen pairs. As described in Billmann et al. [61,71], biological replicate correlations were then scaled to the expected background correlation distribution: the mean and standard deviation of its correlation with screens performed in another context (e.g., query mutation). This converted the correlation coefficients into a metric with an unambiguous statistical interpretation that can be interpreted as a z-score. Here, each query creates its own background correlation distribution. If all biological replicates for a given query were performed in the same media condition, only all other query screens in this media condition were used as background. If two biological replicates were performed in different media conditions, the background was only generated across media conditions (e.g., the minimal media replicate obtained its background correlation distribution from all rich media screens and vice versa). To

judge all biological replicates for a given query gene separately, as well as all screens sharing the same query gene as a whole, WBC scores were computed for all possible pairs as well as collectively for the entire set of replicates for a given query. Note that the selection of query screens used for creating each query screen's own background correlation distribution determines the exact numerical WBC score, and scores for an identical set of replicates for a given query will not be identical in a growing data set where the background correlation distribution is estimated from more queries.

Functional evaluation of genetic interactions

To functionally evaluate genetic interactions (GI), we tested how well GI profiles across all 324 query screens in our data set recapitulated known biology. Specifically, we assumed that a positive Pearson's correlation coefficient (PCC) between the qGI profiles along the 324 query screens of a gene pair would be indicative of functional similarity. We utilized the FLEX package to test how well our data reconstructed CORUM 3.0 co-complex[75], co-pathway (MSigDB Bioconductor package) and co-functional (GO biological process) annotations [68].

Analysis of genomic linkage of fitness effects in the DepMap dataset

We tested if genomically linked genetic interaction scores in the HAP1 isogenic cell lines set were indicative of a general trend beyond HAP1 cells. For chromosomal and sub-chromosomal regions with recurrent residual SMF effect shifts in HAP1 cells, we also quantified the density of strong co-essentiality in the Cancer Dependency Map (DepMap)[77]. Chronos gene fitness scores (version 23Q2) were downloaded from https://depmap.org/portal/. The data contained 1,095 genome-wide CRISPR-Cas9 gene knockout screens with scores corrected for proximity bias[50]. To test the Relative Local co-essentiality Density (RLD), co-essentiality was computed as the PCC along the Chronos score profiles across 1,095 cell lines. The ratio of the density of edges with a z-score larger than 4 between genes in a given region over the density among all genes was computed.

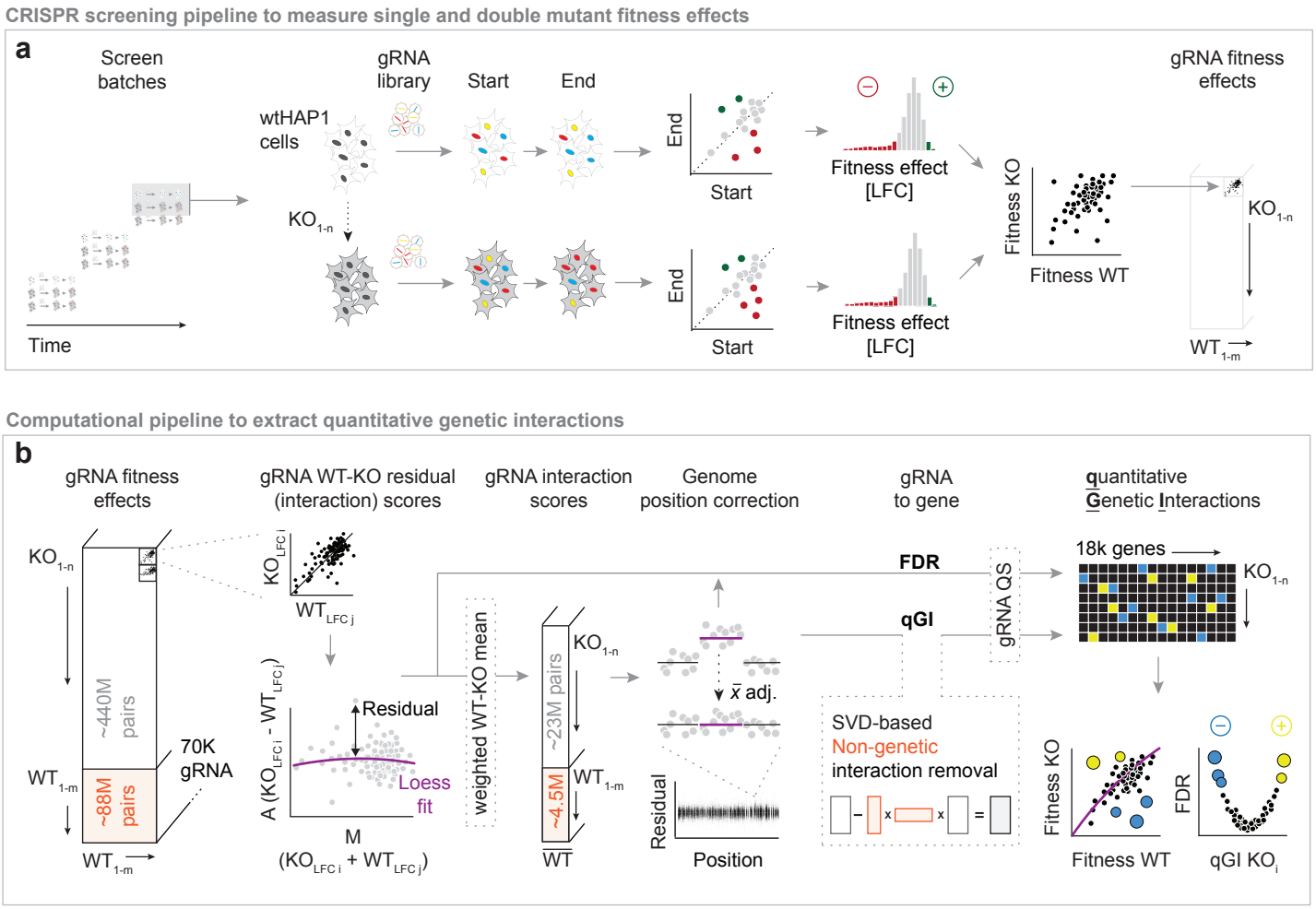

Figure 1: Isogenic genome-wide CRISPR/Cas9 screening to identify genetic interactions

Figure 1: Isogenic genome-wide CRISPR/Cas9 screening to identify genetic interactions. (a) Quantitative genetic interactions are identified by contrasting fitness effects measured in HAP1 cells harboring a specific KO and wildtype (WT) HAP1 screens as reference. Every KO-WT screen pair is compared at the gRNA level, where fitness scores are represented by the log2-foldchange (lfc) between starting and endpoint gRNA abundance, to obtain residual lfc values. **(b)** In a series of normalization steps, gRNA-level residual lfc values are processed to represent the effect of the specific KO background. Normalized residual lfc values are summarized at the gene level to compute the quantitative genetic interaction (qGI) score and an associated FDR.

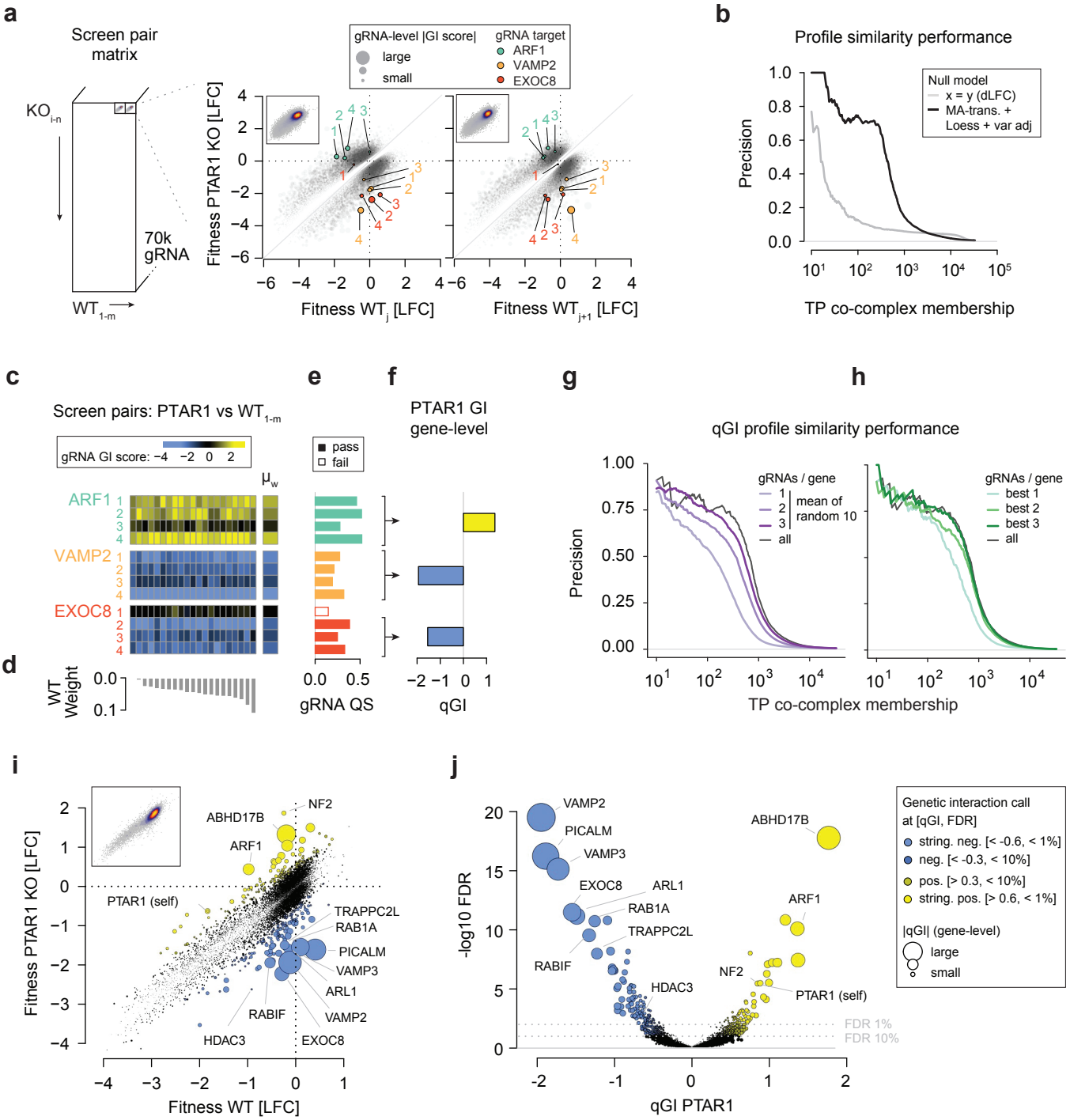

Figure 2: Robust identification of quantitative genetic interactions

Figure 2: Robust identification of quantitative genetic interactions. (a) Pairwise comparison of gRNA-level fitness effects measured in KO query and wildtype (WT) control genome-wide CRISPR-Cas9 screens in HAP1 cells. Shown is an example screen performed in a PTAR1 KO background and highlighted are gRNAs targeting ARF1, VAMP2 and EXOC8. Dot size corresponds to the fitted residual fitness effect (gRNA-level GI) between the PTAR1 KO and each WT control screen. **(b)** Performance of dLFC (grey) and raw qGI (black) profiles across 324 query screens to identify CORUM co-complex membership. The raw qGI scores represent differential effects between query and control screen using MA-transformation, a Loess-based null model, and variance adjustment removing T0 readcount dependencies. **(c)** Complete set of residual fitness effects (gRNA qGI) between the PTAR1 KO and all 21 reference WT screens for the three example genes. The weighted per-gRNA mean was taken across the 21 KO-WT screen contrasts. **(d)** Weights of each KO-WT screen are adjusted to prioritize residual effects derived from WT screens that fit the KO screen best (see Methods). **(e)** gRNA quality score (QS) for the gRNAs targeting the three example genes. The gRNA QS is derived from the sum of within-gene pairwise Pearson correlation coefficients of residual fitness effects across 324 query screens (see Methods). **(f)** Gene-level contrast of PTAR1 KO and the mean WT reference fitness effects. Dot size corresponds to the absolute qGI score. **(g, h)** Performance of qGI-based profiles across 324 query screens to identify CORUM co-complex membership. qGI scores are based on 1, 2 or 3 gRNAs per gene. Selected were either random gRNA targeting a specific gene (g; 10 iterations) or the gRNA per gene with the best gRNA QS (h). qGI profile Pearson correlation coefficients are used as similarity metric. **(i)** Comparison of genome-wide PTAR1 qGI scores with simple differential fitness effects between the PTAR1 KO and the WT reference screens. Dots are colored based on the qGI score (yellow: positive GI, blue: negative GI) and FDR thresholds. **(j)** Genome-wide PTAR1 qGI scores and the associated FDR. Dot color indicates significant genetic interactions of a library gene with PTAR1 by considering the qGI score and the associated FDR.

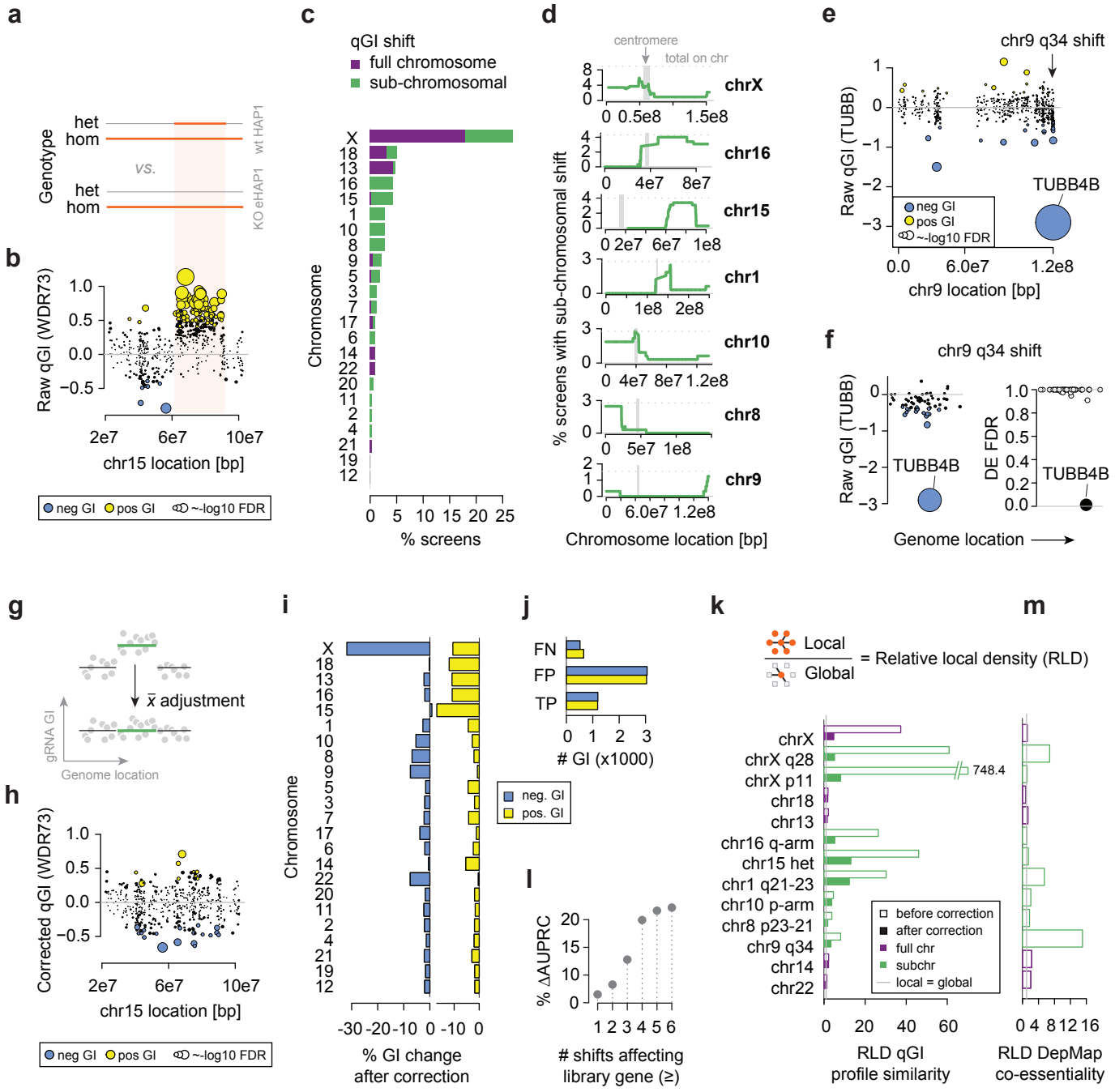

Figure 3: Correction of genomic linkage-based false positive genetic interactions

Figure 3: Correction of genomic linkage-based false positive genetic interactions. (a) CN differences between eHAP1-derived WDR73 query clone and HAP1 wt control cells on chr15. **(b)** Uncorrected (raw) WDR73 qGI scores on chr15. Dot size corresponds to -log10 FDR of the qGI score, and colors indicate significant GIs (|qGI| > 0.3, FDR < 0.1). **(c)** Percentage of HAP1 query gene ko screens with a residual lfc (raw qGI) shift on each chromosome (chr; purple) or a subchromosomal region (subchr; green). 100% corresponds to 324 screens. **(d)** Frequency of shifted regions on the seven chromosomes with the most subchromosomal shifts. The upper dotted line indicates the frequency of total subchromosomal shifts on that chromosome. **(e)** Uncorrected (raw) TUBB qGI scores on chr9. Dot size corresponds to -log10 FDR of the qGI score, and colors indicate significant GIs (|qGI| > 0.3, FDR < 0.1). **(f)** Raw TUBB qGI scores and differential expression (DE) in the shifted 64-gene chr9 q34 region. Blue (qGI; left) and black (DE; right) mark significant scores. Significant DE is called at FDR < 5%. **(g)** Schematic illustration of the correction of chromosomal GI shifts. **(h)** Corrected WDR73 qGI scores on chr15. Dot size corresponds to -log10 FDR of the qGI score, and colors indicate significant GIs (|qGI| > 0.3, FDR < 0.1). **(i)** Net percentage of GIs lost and gained after correction for local shifts in each chromosome across all 324 screens. **(j)** Number of positive (yellow) and negative (blue) GIs (|qGI| > 0.3, FDR < 0.1) across all shifted regions in the 324 screens that were gained (FN), lost (FP) and retained (TP) after correction. **(k)** Relative local density (RLD) of qGI profile similarity of genes in the genomic region with the most chromosomal and subchromosomal shifts before and after correction. A value above 1.0 indicates that strong (z > 4) profile similarities are observed more often between genes encoded in a shifted genomic region than between genes not encoded in the same region. **(l)** Change of the area under the PR curve when GO BP relationships are predicted through qGI profile correlation. The analyses were focused on all genes that occurred in at least a certain number (1-6) of shifts. **(m)** RLD of highly similar (z > 4) co-essentiality profiles across 1095 different cell lines (DepMap 23Q2 release) with the regions shifted in isogenic HAP1 screens.

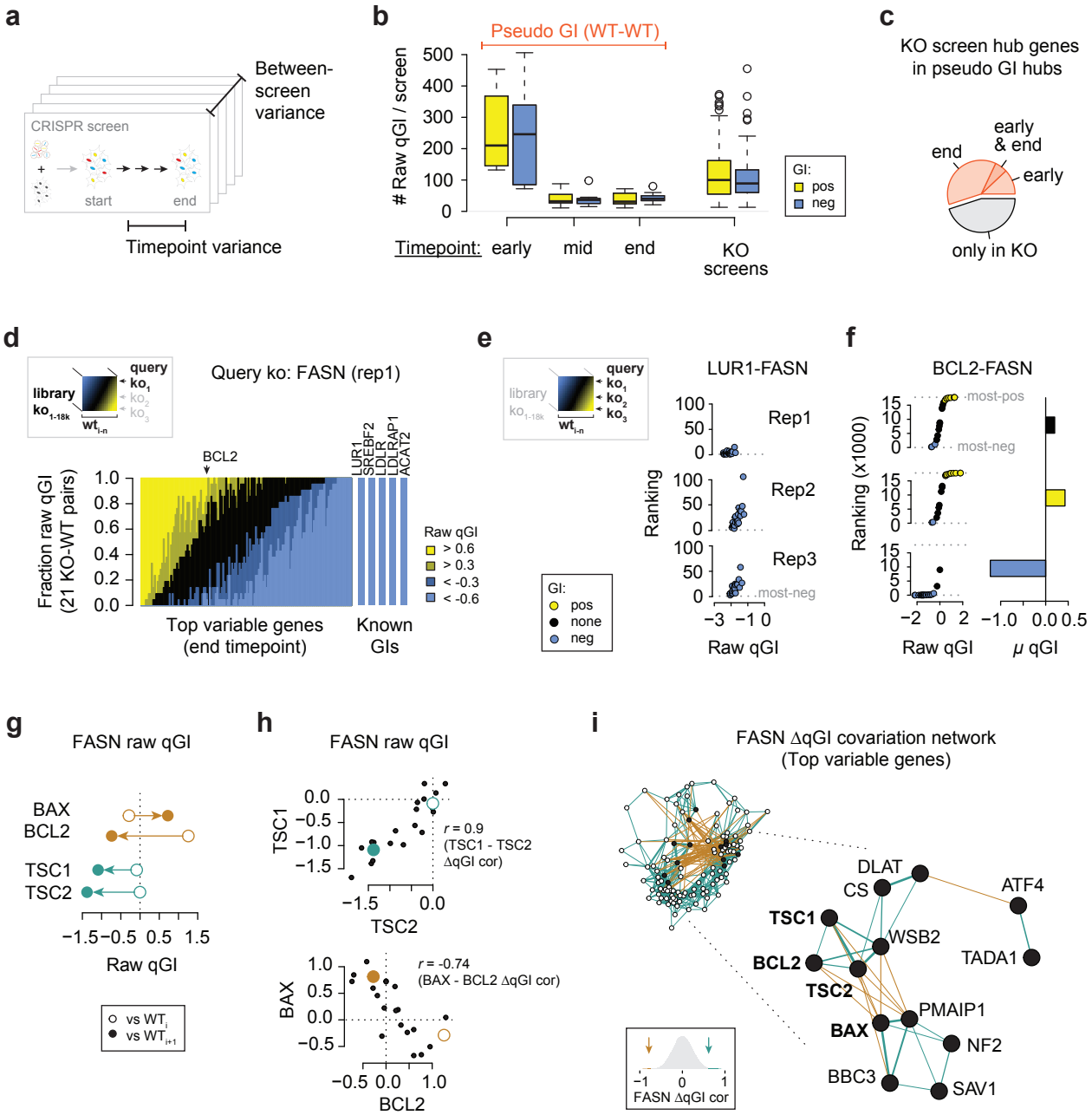

Figure 4: Genetic interaction-independent patterns in genetic interaction screens

Figure 4: Genetic interaction-independent patterns in genetic interaction screens. (a) Schematic illustration of sources of library gene-specific variance. **(b)** Number of GIs for early timepoint (T6 – T10), mid (T11 – T17) and end point (T18, T19) control simulations and query screens. The qGI are uncorrected and a GI is counted for |qGI| > 0.6. Simulations iteratively contrast a held-out control screen against the set of all other control screens (see Methods). **(c)** Hub GI genes overall between simulated "pseudo" GIs and query screens. In each data set, the top 100 genes with the highest absolute GI degree (|qGI| > 0.6) are tested for overlap. **(d)** Fraction of positive and negative uncorrected qGI scores of 142 highly variable genes and 5 high confidence negative genetic interactions with FASN when in all 21 KO-WT contrasts. The 142 were selected because they exhibited the highest fitness effect-adjusted variance between control screens or exhibited high covariance in the control screens. **(e, f)** Ranking and qGI scores (uncorrected) between FASN and LUR1 (known interactor) or the highly variable gene BCL2 in the 21 WT contrasts with three independent FASN KO screens. The GI spectrum thumbnail indicates that highly variable genes such BCL2 vary in their qGI score depending both on the WT reference as well as on the KO (here FASN) screen. Per FASN replicate mean BCL2 qGI scores are illustrated in the barcharts (f). **(g)** Uncorrected qGI scores between four highly variable genes and FASN. Shown are scores from the first FASN KO screen replicate contrasted to two different WT reference screens. **(h)** Uncorrected qGI scores between TSC1 and TSC2 (top) and BCL2 and BAX (bottom) with FASN when considering all 21 WT reference screens. The scores shown in (g) are highlighted. Pearson correlation coefficients indicate how raw qGI scores of two genes co-vary across the 21 WT reference contrasts. **(i)** Covariation network illustrating positive (green) and negative (brown) correlation between the 142 highly variable genes for the residual qGI scores between the replicate 1 FASN KO screen and the 21 WT reference screens. Edges are shown for Pearson correlations above 0.6 or below -0.75. The histogram shows the genome-wide distribution of the FASN-specific residual qGI covariation network. The zoom-in shows the closest partners of the genes illustrated in (g, h).

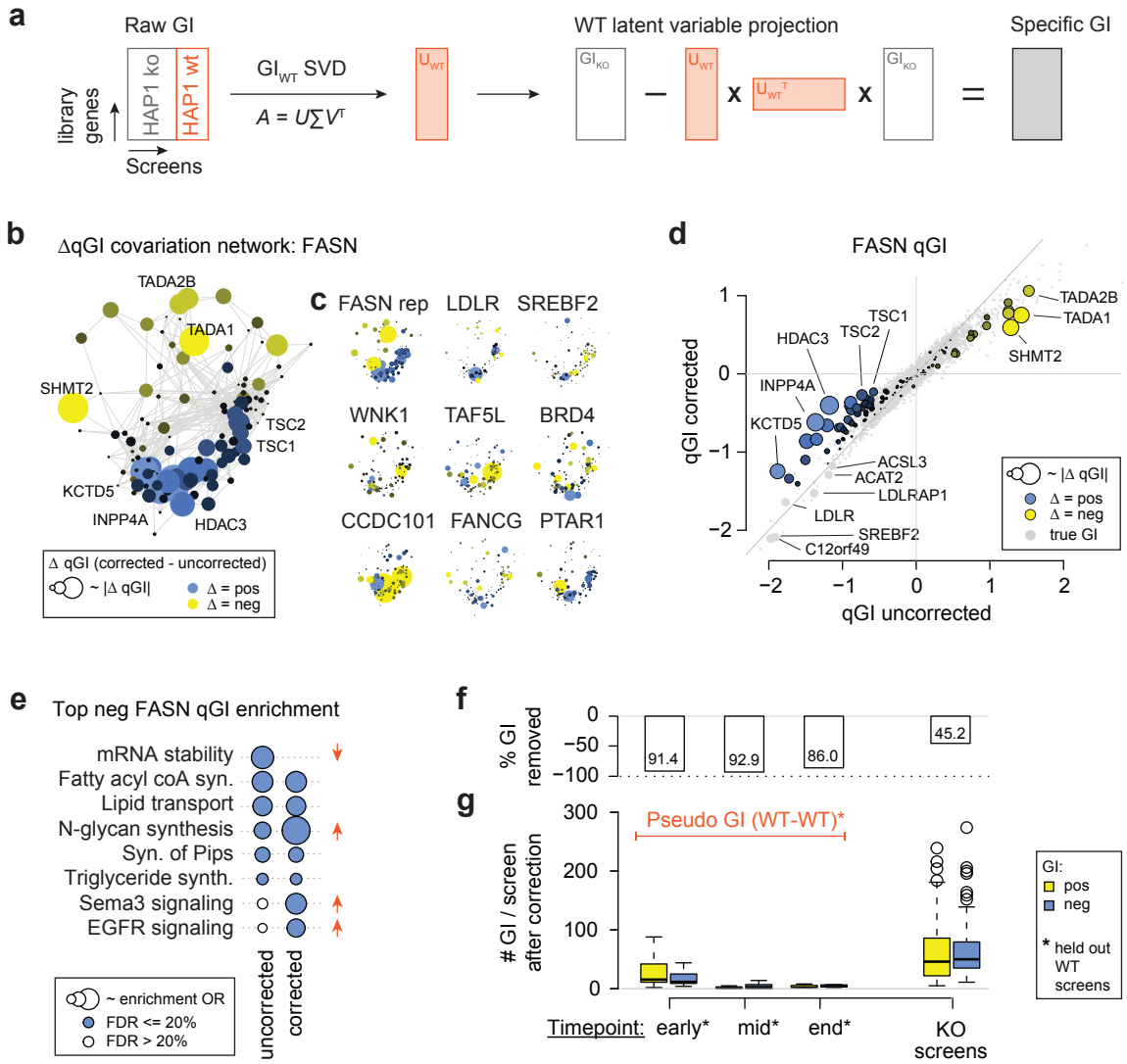

Figure 5: Correcting genetic interaction-independent patterns in genetic interaction screens

Figure 5: Correcting genetic interaction-independent patterns in genetic interaction screens. **(a)** Schematic illustration of the SVD-based removal of wt-derived variance on query screens: for each of the ~70k gRNAs, latent variables describing genetic interaction-independent variation are derived from the wt control screens and projected onto and removed from raw qGI scores measured on query screens. **(b)** FASN-specific covariation network for 142 highly variable genes (see Fig. 4i). Genes are connected by their residual qGI profile correlation across the 21 FASN-WT contrasts. Dot size corresponds to the gene-specific change of the qGI score. A correction towards a more positive value (less negative qGI score) is shown in blue, and correction toward more negative (less positive qGI score) in yellow. **(c)** Correction thumbnail laid out as in (b) for 9 different query screens. The gene-specific change of the qGI score represents the change in a given query screen. **(d)** FASN qGI scores before and after correction for genetic interaction-independent patterns. Shown are 17,804 genes, and the 142 highly variable genes are highlighted. Dot size corresponds to the gene-specific change of the qGI score. A correction towards a more positive value (less negative qGI score) is shown in blue, and correction toward more negative (less positive qGI score) in yellow. **(e)** REACTOME pathway enrichment of the top 100 negative and positive qGI scores in the FASN query screen before and after correction. All pathways enriched at an FDR of 20% before or after are shown. **(f)** Total net change in the number of GIs after correction (Fig. 4b vs Fig. 5g) **(g)** Number of GIs for early timepoint (T6 – T10), mid (T11 – T17) and end point (T18, T19) control simulations and query screens after correction (compare to Fig. 4b). The qGI are uncorrected and a GI is counted for $|qGI| > 0.6$. Simulations iteratively contrast a held-out control screen against the set of all other control screens.

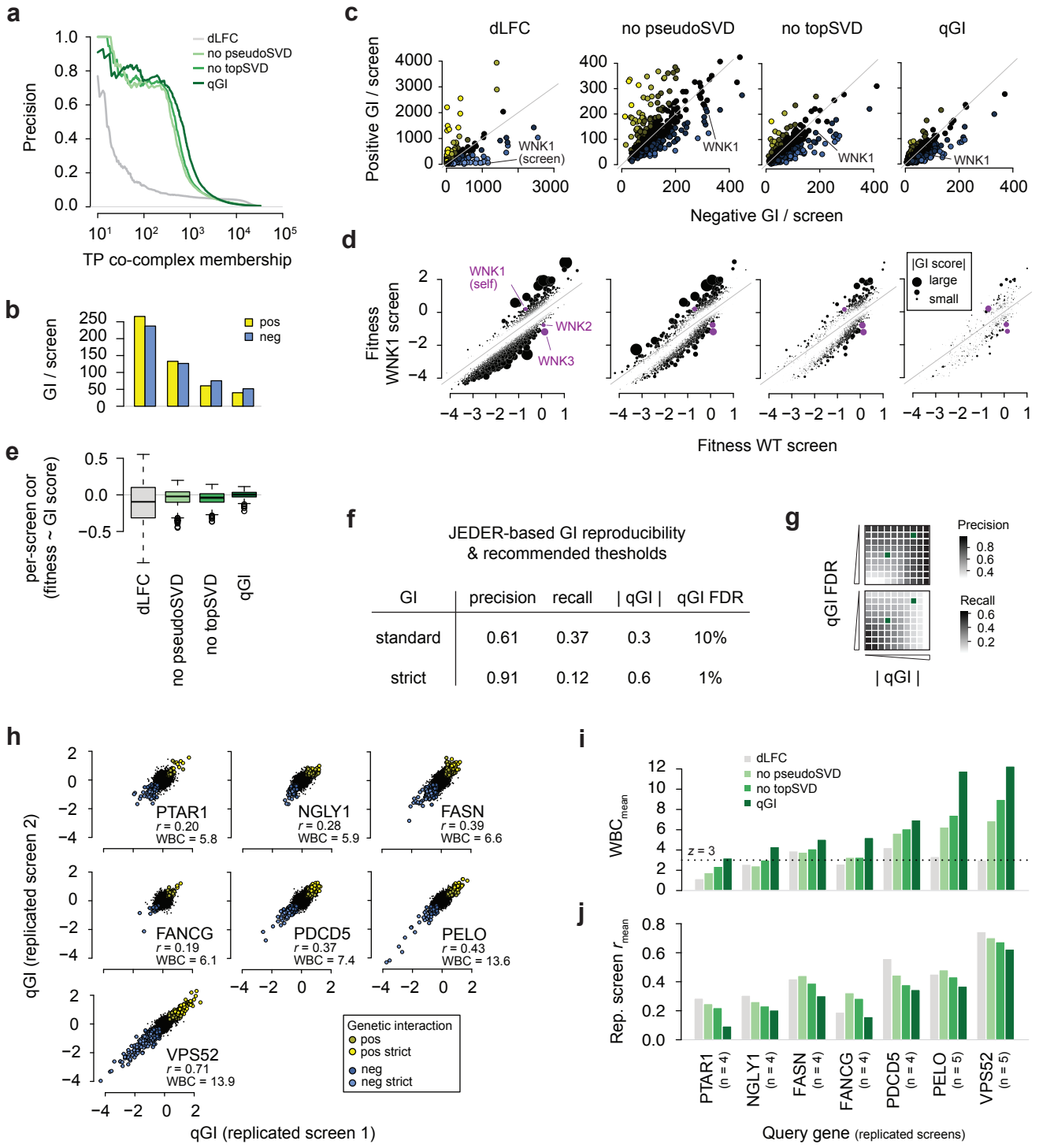

Figure 6: Cumulative impact of qGI normalization on genetic interactions and reproducibility analysis

Figure 6: Cumulative impact of qGI normalization on genetic interactions and reproducibility analysis. (a) Increasing precision-recall performance of qGI profiles at different normalization steps of the qGI pipeline to reconstruct known CORUM co-complex memberships. Shown is the simple differential (d)LFC, a fitted raw qGI version without removal of genetic interaction-independent patterns including the top SV correction, a version without removal of top SVs, and the completely corrected qGI score. **(b)** Mean number of GIs per query screen at different stages of normalization of the qGI pipeline. A GI is called at |qGI| > 0.6. **(c)** Per-query screen comparison of the number of positive (qGI > 0.6) or negative (qGI < -0.6) GIs at different normalization stages of the qGI pipeline. Blue indicates a higher ratio of negative, yellow a higher ratio of positive GIs. **(d)** Differential fitness effects of the library genes in the WNK1 query screen at different normalization of the qGI pipeline. Fitness effects in wt HAP1 control are plotted against fitness effects in the query screen. Dot size is scaled to the absolute interaction score. The paralogs of WNK1 are labeled. **(e)** Per-query screen correlation of the fitness effect and interaction scores at different normalization stages of the qGI pipeline. **(f)** Selected thresholds for the GI effect size, the qGI score, and significance, the qGI-associated FDR, for standard and strict GIs. The JEDER-derived precision and recall are highlighted in (g). **(g)** Precision and recall of positive and negative GIs of the 7 sets of replicated query screens (see i, j) at different qGI and FDR thresholds. The precision was estimated using a standard-free MCMC-based method to determine FP and FN rates. The set of qGI and FDR thresholds for standard (|qGI| > 0.3, FDR < 10%) and stringent (|qGI| > 0.6, FDR < 1%) GIs is indicated in each precision and recall heatmap. **(h)** Replicate agreement of genome-wide qGI scores (n = 17,804) of the a selected pairs of the 7 query genes that were replicated multiple times. The Pearson's correlation coefficient (PCC) r, and the WBC score (see Methods) quantify qGI reproducibility. The WBC score is a z-score-based metric that indicates how many standard deviations of PCCs among qGI scores measured in different genetic backgrounds the PCC between a pair of interest is away from the mean PCC**.** The WBC score is a z-score-based metric that indicates how much the Pearson correlation coefficient (PCC) between a pair of qGI profiles differs from the average PCC observed across different genetic backgrounds (see Methods). **(i)** Mean WBC score for 7 genetic backgrounds at different normalization stages of the qGI pipeline. The mean per genetic background was taken across all within-background pairs for FASN (n = 4), PDCD5 (n = 4), FANCG (n = 4), PTAR1 (n = 4), NGLY1 (n = 4), PELO (n = 5) and VPS52 (n = 5). A WBC score is a z-score-based metric, and the mean WBC for each genetic background passes the threshold of 3 only after complete normalization. **(j)** Mean replicate correlation coefficient for 7 genetic backgrounds at different normalization stages of the qGI pipeline.

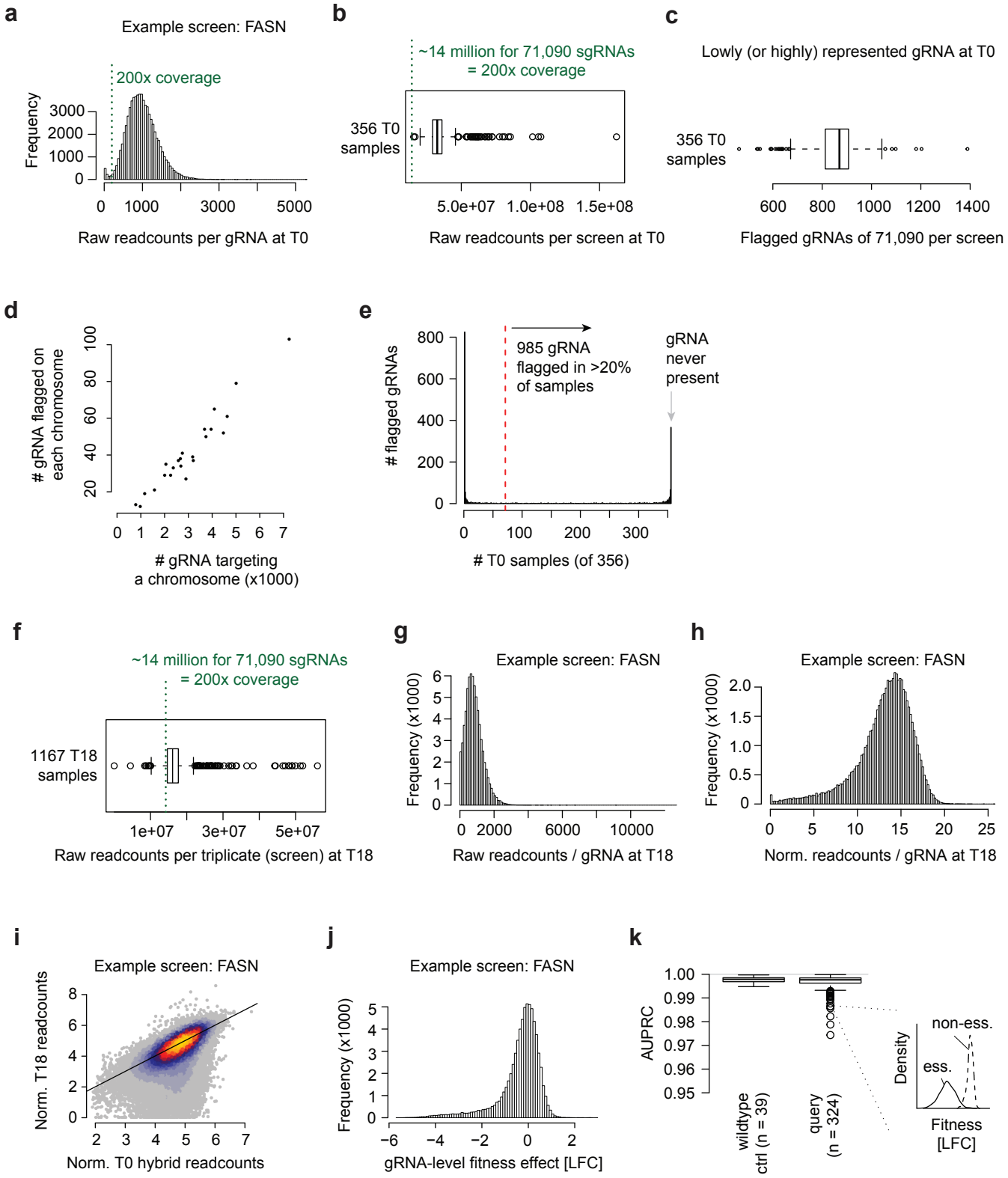

Figure S1: Initial quality control of the 363 genome-wide CRISPR-Cas9 screens

Figure S1: Initial quality control of genome-wide CRISPR-Cas9 screens. (a) Raw readcounts before sequencing depth normalization per gRNA for all 71,090 gRNAs in the TKOv3 library at T0. Shown is a FASN query screen as a representative example. **(b)** Total number of raw readcounts per screen at T0 for each query and control screen. Overall, 356 unique T0 samples from the 363 screens existed in the initial dataset, since several control screens that were performed in different media conditions shared a common T0. **(c)** Number of gRNAs with readcounts before read-depth normalization below 40 for each screen. **(d)** Mean number of gRNAs flagged on each chromosome compared to the number of gRNAs targeting genes on those chromosomes. **(e)** Number of screens a gRNA is flagged in. Only the 2092 gRNAs that are flagged in at least one T0 sample (screen) are shown. **(f)** Total number of raw readcounts per screen at T18 (screen endpoint) for each query and control screen. Overall, 1167 unique end and intermediate time point samples (technical triplicates) were sequenced from the 363 unique screens. **(g, h)** Raw (g) and normalized log2 (h) readcounts per gRNA for all 71,090 gRNAs in the TKOv3 library at T18 (endpoint). Shown is a FASN query screen as a representative example. **(i)** Normalized log2 readcount T0 hybrid versus T18 (endpoint) comparison of all 71,090 gRNAs in the TKOv3 library. Shown is a FASN query screen as a representative example. **(j)** T18 over T0 hybrid foldchange (lfc) of the FASN screen. **(k)** Area under the precision-recall curve (AUPRC) distinguishing the Hart *et al.* core essential genes from non-essential genes as a basic per-screen quality control statistic.

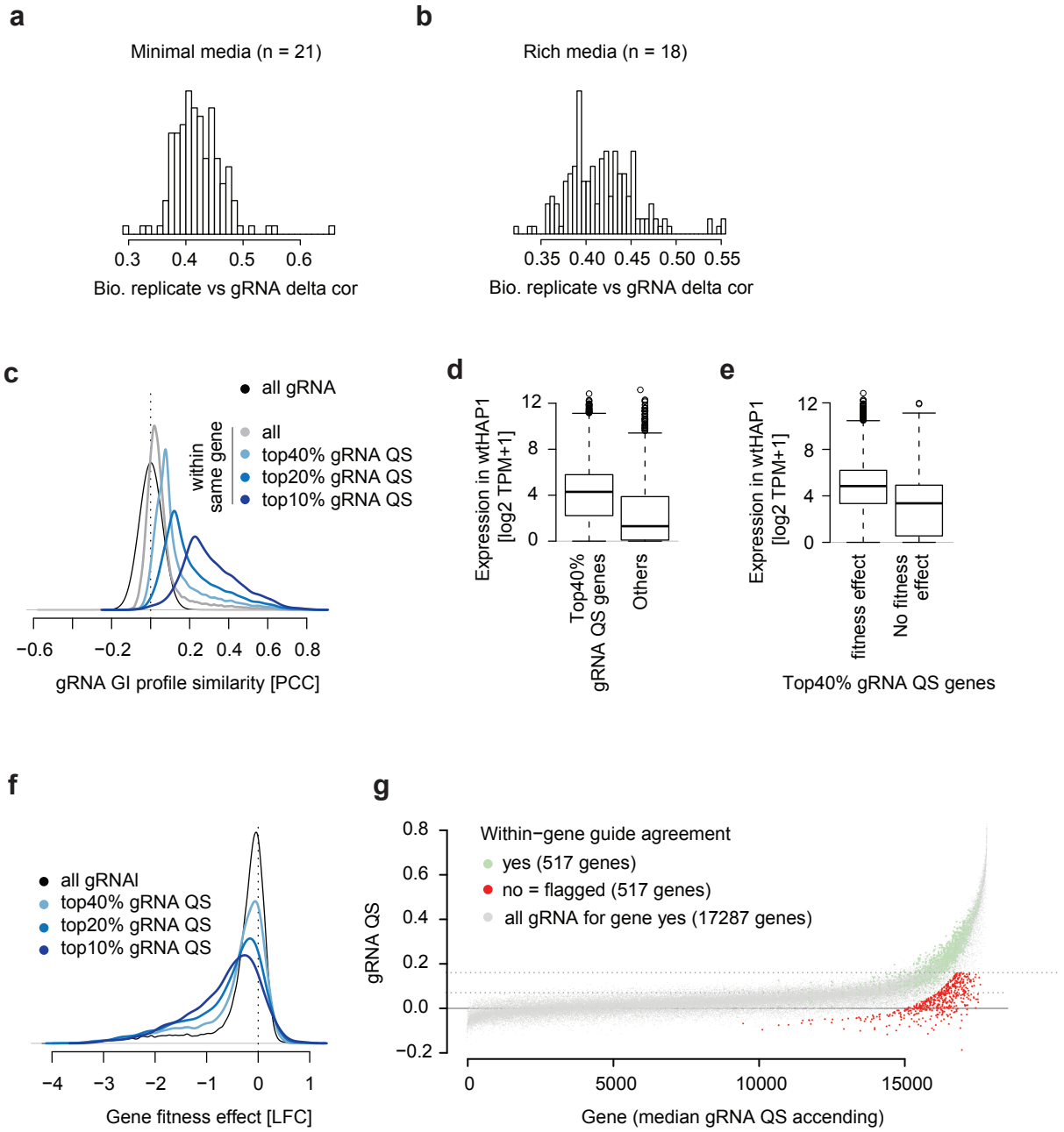

Figure S2: Agreement and quality control of gRNAs targeting the same gene

Figure S2: Agreement and quality control of gRNAs targeting the same gene. (a, b) Relative replicate correlation of gRNA-level GI scores to determine the setup of technical and biological replicates in limma's moderated t-test. This is computed independently for the minimal and rich media screening conditions. **(c)** GI profile similarity of all pairwise comparisons of gRNAs targeting the same gene. **(d)** Expression of the genes with the top 40% gRNA QS in wildtype HAP1 cells. **(e)** Expression of the genes with the top 40% gRNA QS separated by whether they showed a single mutant fitness effect at a lenient FDR of 20% in wildtype HAP1 cells. **(f)** single mutant fitness (LFC) for different groups of top gRNA QS genes. **(g)** The gRNA QS for all ~70,000 gRNAs grouped by targeted gene. Flagged gRNAs are highlighted in red.

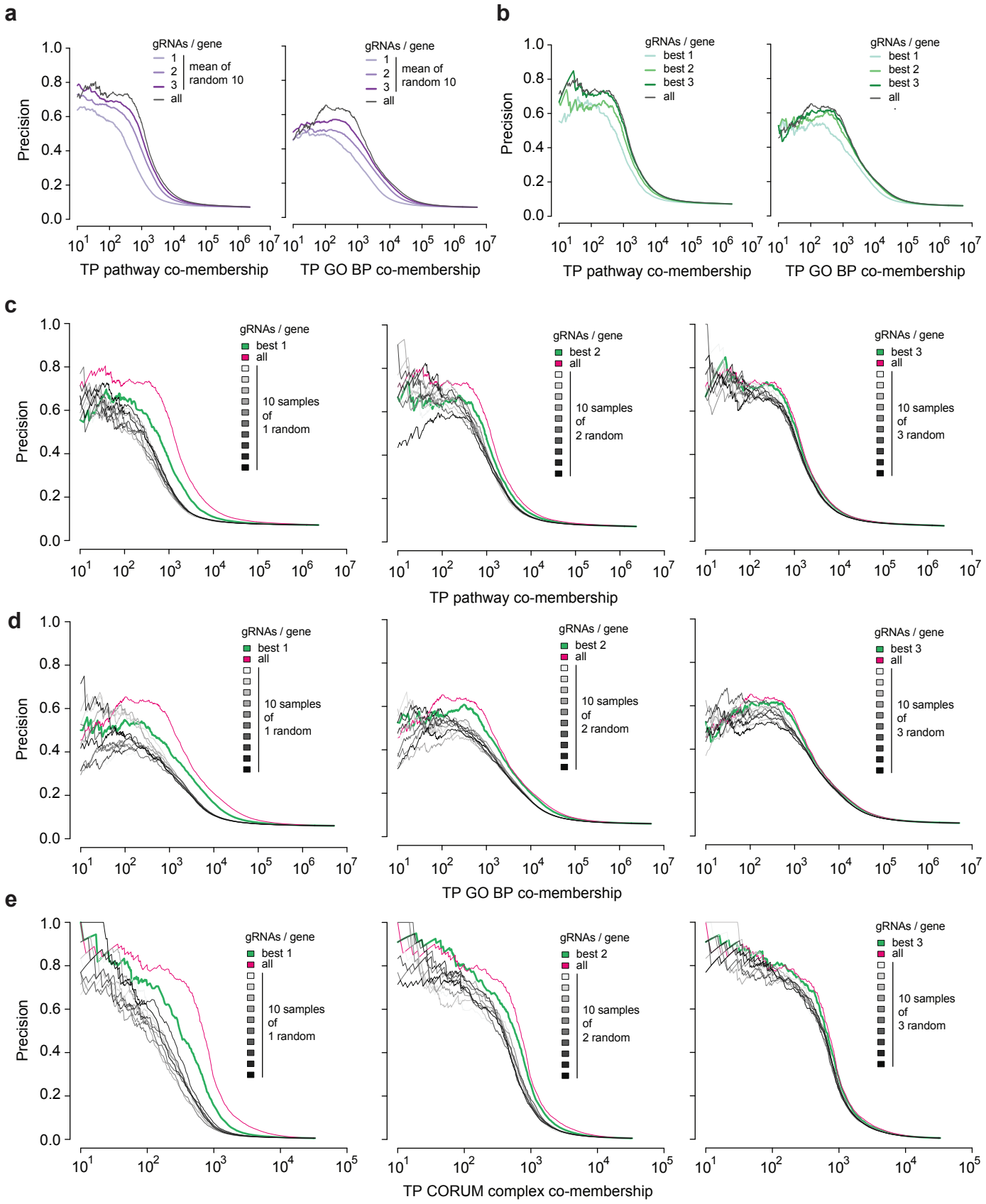

Figure S3: Effect of gRNAs-per gene on performance of qGI-based profiles to identify complex, pathway or GO-BP co-membership.

Figure S3: Effect of gRNAs-per gene on performance of qGI-based profiles to identify complex, pathway or GO-BP co-membership. qGI scores are based on 1, 2, 3 or all up to 4 gRNAs per gene. qGI profile Pearson's correlation coefficients across 324 query screens are used as a similarity metric. **(a)** Mean precision of 10 iterations of 1, 2 or 3 randomly selected gRNAs targeting a specific gene compared to all up to 4 gRNAs per gene. Shown is the performance of for identifying pathway or GO BO co-membership. Corresponds to Fig. 2g. **(b)** Top 1, 2 or 3 gRNAs per gene with the best gRNA QS score compared to all up to 4 gRNAs per gene. Shown is the performance of for identifying pathway or GO BO co-membership. Corresponds to Fig. 2h. **(c-e)** Comparison of the best 1, 2 or 3 gRNAs targeting a specific with 10 iterations of 1, 2 or 3 randomly selected gRNAs.

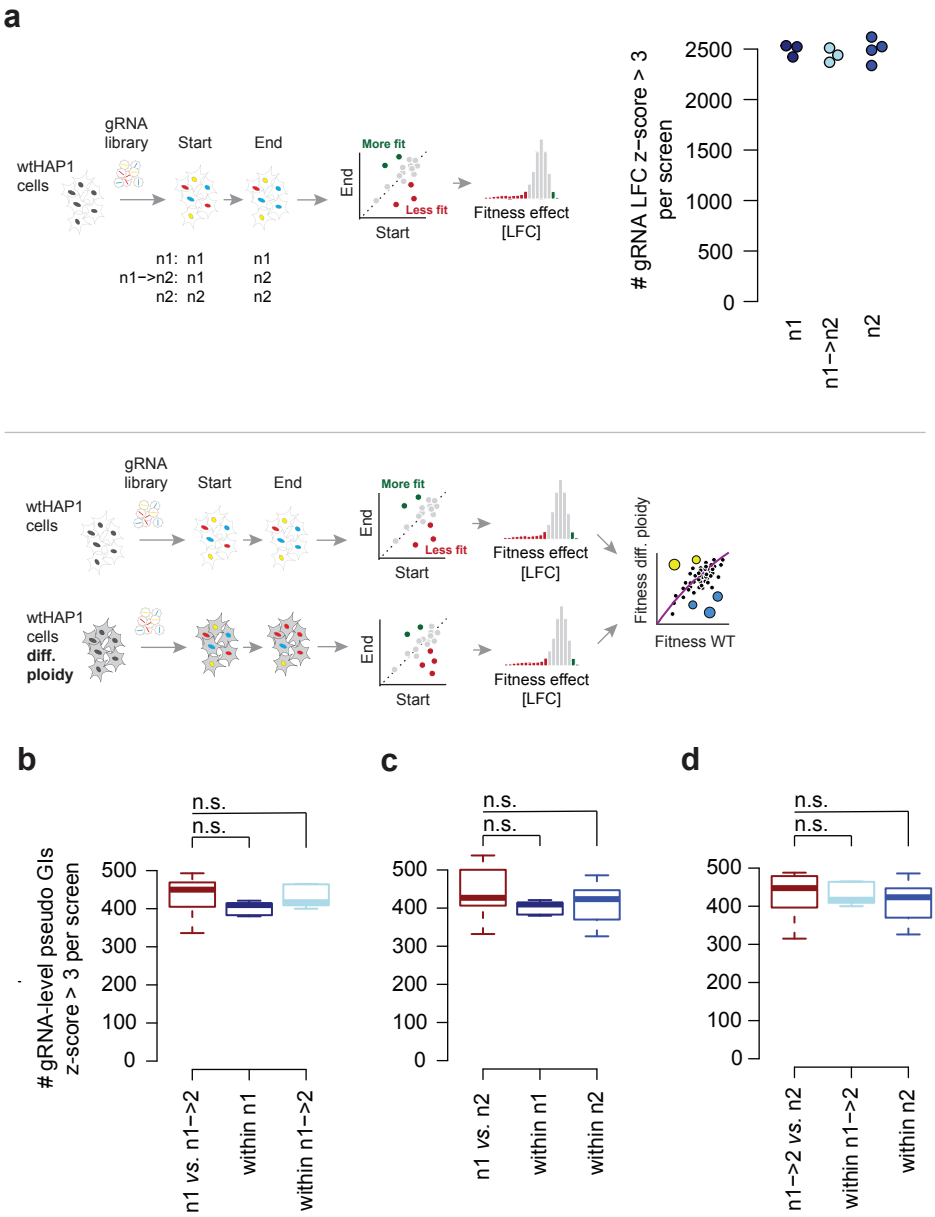

Figure S4: Number of differential fitness effects between wt HAP1 screens with different ploidy status.

Figure S4: Number of differential fitness effects between wt HAP1 screens with different ploidy status. (a) Number of strong fitness effects (gRNA-level LFC z-score < -3) in screens with 1n at both start and end of the experiment, screens that developed from 1n to 2n, and those that were 2n from the start. Each dot represents a single genome-wide screen. **(b-d)** Number of strong differential effects (gRNA-level uncorrected interaction z-score < -3) when screens with different ploidy status were contrasted (red). No statistical difference was observed when comparing to the number of strong differential effects is between screens with the same ploidy (Wilcoxon's rank sum test).

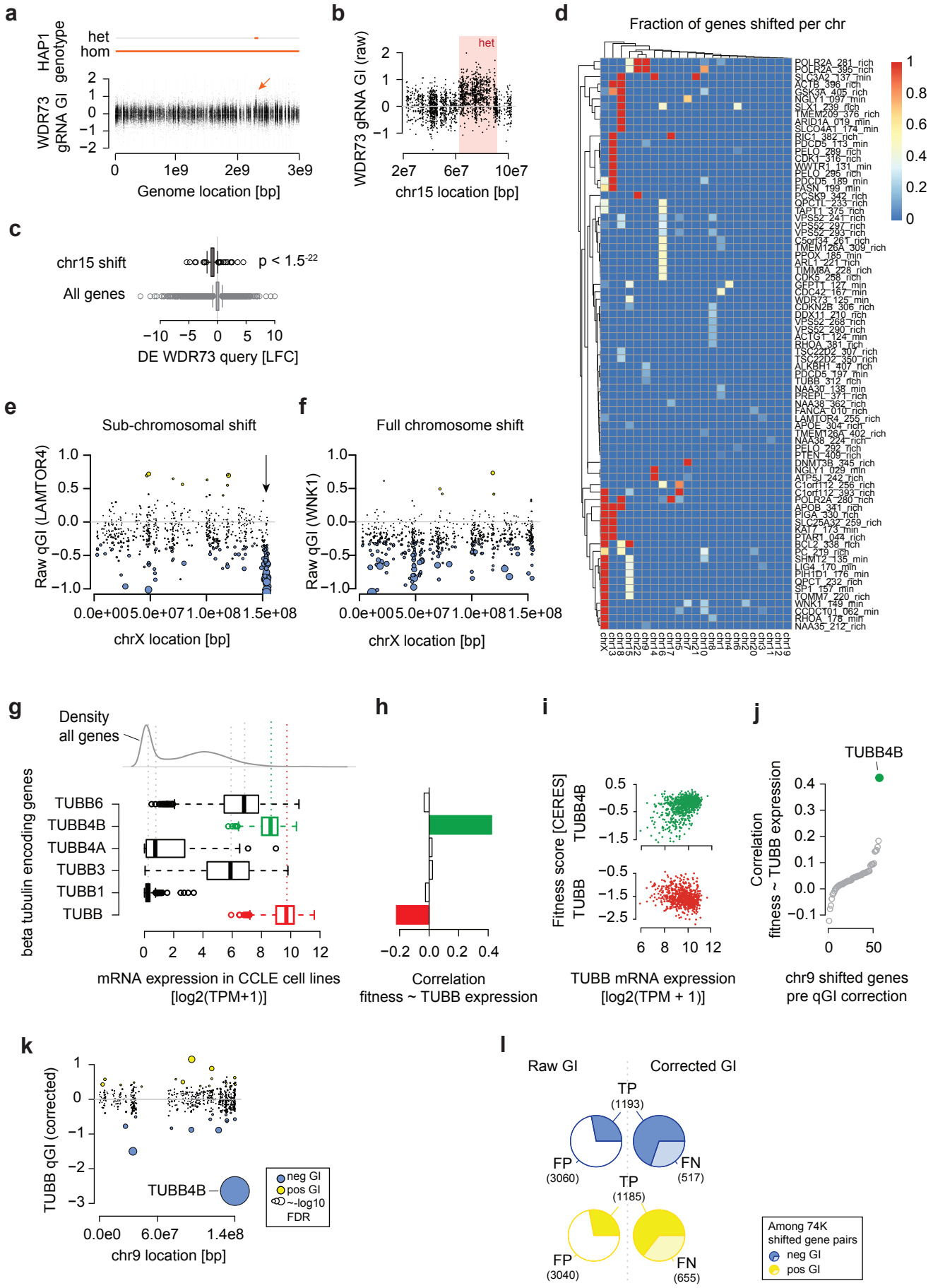

Figure S5: Exploration of genomic linkage-based false positive genetic interactions.

Figure S5: Exploration of genomic linkage-based false positive genetic interactions. (a) gRNA-level, uncorrected qGI scores showing interactions with a WDR73 mutation in the HAP1 cell line are plotted against the genomic location of their target sequence. The heterozygous 30 Mbp region on chr15 in parental HAP1 cells is shown and the arrow points at uncorrected gRNA-level GI scores in this region. **(b)** Uncorrected WDR73 gRNA-level qGI scores on chr15 (zoomed in view of region highlighted in (a)). **(c)** Differential expression distributions in the WDR73 KO clone of all genes in the genome compared to the genes on in the heterozygous chr15 region. **(d)** Fraction of genes per chromosome with detected shifts. Shown are query screens with at least one shift on chromosomes other than chrX. **(e, f)** Uncorrected (raw) LAMTOR4 and WNK1 qGI scores on chrX. Dot size corresponds to -log10 FDR of the qGI score, and colors indicate significant GIs ($|qGI| > 0.3$, FDR < 0.1). **(g)** Expression of beta tubulin encoding genes in the CCLE. **(h, i)** Correlation between TUBB expression and fitness effects upon perturbation of beta tubulin encoding genes in the Cancer Dependency Map (DepMap). **(j)** Correlation between TUBB expression and DepMap fitness effects of the shifted genes on chr9 q34. **(k)** Corrected TUBB qGI scores on chr9. Dot size corresponds to -log10 FDR of the qGI score, and colors indicate significant GIs ($|qGI| > 0.3$, FDR < 0.1). **(l)** Number of positive (yellow) and negative (blue) GIs ($|qGI| > 0.3$, FDR < 0.1) across all shifted regions in the 324 screens that were gained (FN), lost (FP) and retained (TP) after correction.

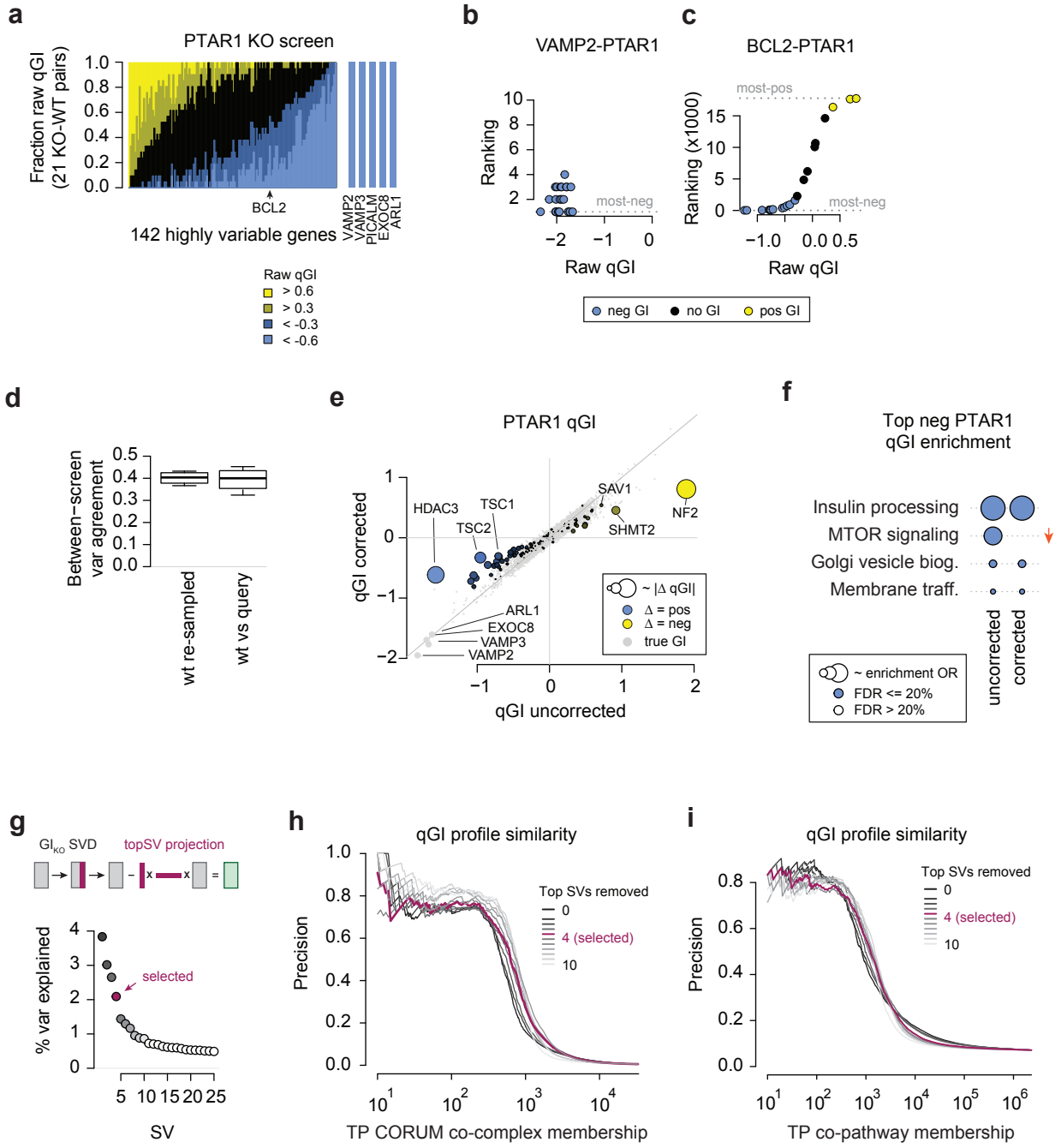

Figure S6: Unsupervised correction of genetic interaction-independent patterns in genetic interaction screens.

Figure S6: Unsupervised correction of genetic interaction-independent patterns in genetic interaction screens. (a) Fraction of positive and negative uncorrected qGI scores of 142 highly variable genes and 5 high confidence negative genetic interactions with FASN in all 21 KO-WT contrasts. The 142 were selected because they exhibited the highest fitness effect-adjusted variance between control screens or exhibited high covariance in the control screens. **(b, c)** Ranking and qGI scores (uncorrected) between PTAR1 and VAMP2 (known) or the highly variable gene BCL2 in the 21 WT contrasts with three independent FASN KO screens. **(d)** Between-biological replicate variance agreement of wt control and query screens. The standard deviation across the residual fitness scores of biological replicated screens was computed for each library gene. This scaled representation of the between-screen variance was compared between FASN, FANCG, NGLY1 and PDCD5 query screen sets and the wt control. **(e)** PTAR1 qGI scores before and after correction for genetic interaction-independent patterns. Shown are 17,804 genes, and the 142 highly variable genes are highlighted. Dot size corresponds to the gene-specific change of the qGI score. A correction towards a more positive value (less negative qGI score) is shown in blue, and correction toward more negative (less positive qGI score) in yellow. **(f)** REACTOME pathway enrichment of the top 100 negative and positive qGI scores in the PTAR1 query screen before and after correction. All pathways enriched at an FDR of 20% before or after are shown. **(g)** Variance explained by the first 25 singular vectors (SV). Singular vector decomposition (SVD) was performed on the gRNA-level qGI matrix and considering every unique query genetic background (see Methods). **(h, i)** Precision-recall performance of qGI profiles with increased number of SVs removed to reconstruct known co-complex (h) or co-pathway (i) memberships. Co-complex memberships were defined by CORUM 3.0, co-pathway memberships were defined by MSigDB. Top SVs were cumulatively removed from the data. Based on these performance evaluations, the version removing SV1-4 was selected.

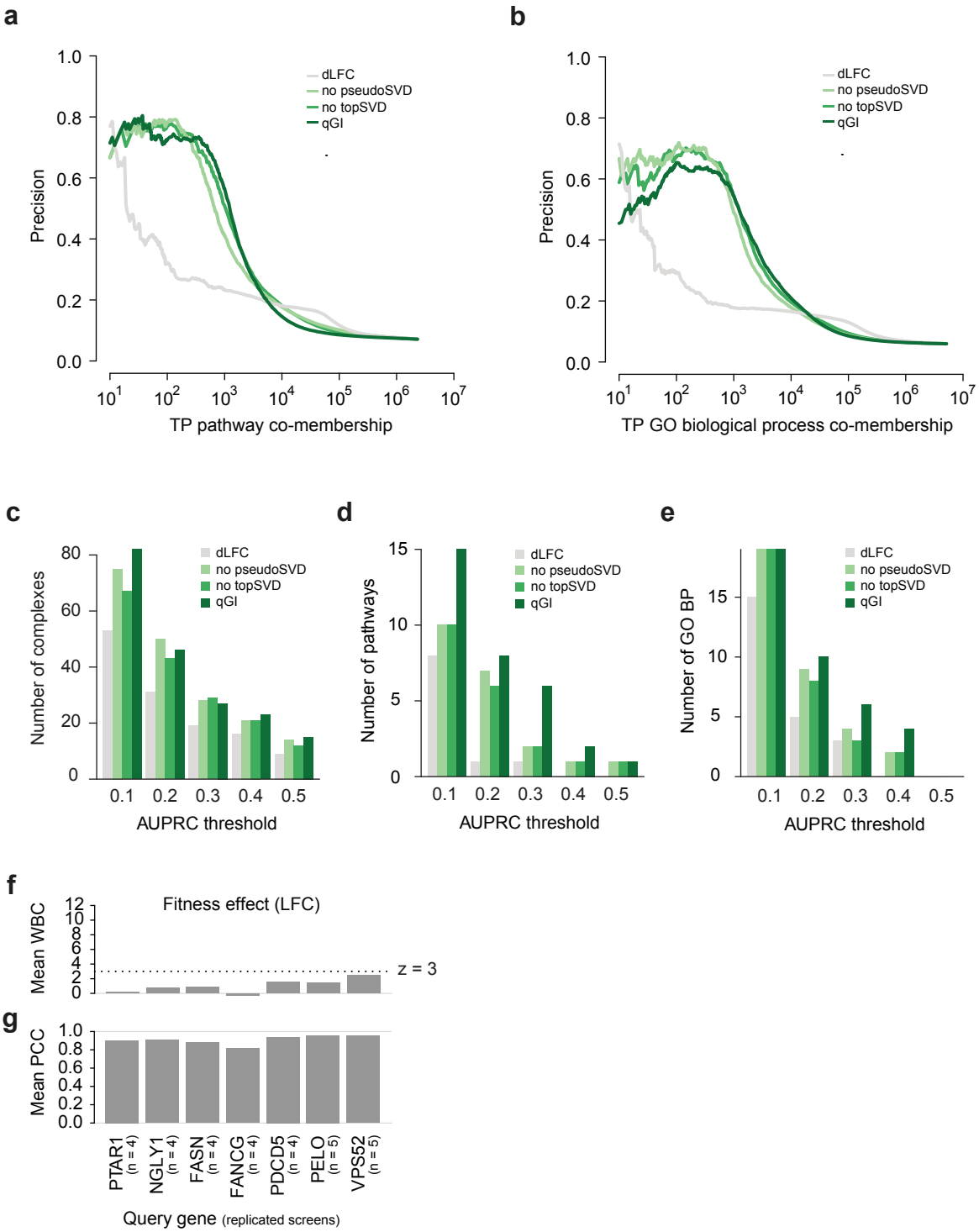

Figure S7: Performance of qGI profiles at different normalization stages of the qGI pipeline to reconstruct known complex, pathway and GO-BP co-memberships.

Figure S7: Performance of qGI profiles at different normalization stages of the qGI pipeline to reconstruct known complex, pathway and GO-BP co-memberships. Shown is the simple differential (d)LFC, a fitted raw qGI version without removal of genetic interaction-independent patterns and the overall top 4 SVs, a version without removal of the overall top 4 SVs, and the completely corrected qGI score. qGI profile Pearson's correlation coefficients across 324 query screens are used as a similarity metric. **(a, b)** Precision-TP curve for pathway (a) and GO BP (b) as functional standard. This figure corresponds to Fig. 6a. **(c-e)** Number of CORUM complexes (c), pathways (d) and GO BPs (e) different module-level AUPRC thresholds. Redundancy of modules (complexes, pathways, GO BPs) is controlled at a Jaccard index of 0.8. **(f)** Mean WBC score for 7 genetic backgrounds of the fitness scores (LFC). The mean per genetic background was taken across all within-background pairs for FASN (n = 4), PDCD5 (n = 4), FANCG (n = 4), PTAR1 (n = 4), NGLY1 (n = 4), PELO (n = 5) and VPS52 (n = 5). A WBC score is a z-score-based metric. **(g)** Mean replicate correlation coefficient for 7 genetic backgrounds of the fitness scores (LFC).

References

1. Costanzo, M. *et al.* Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell* **177**, 85–100 (2019).
2. Ryan, C. J., Devakumar, L. P. S., Pettitt, S. J. & Lord, C. J. Complex synthetic lethality in cancer. *Nat Genet* **55**, 2039–2048 (2023).
3. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
4. Ünlü, B. *et al.* Global analysis of suppressor mutations that rescue human genetic defects. *Genome Med* **15**, 78 (2023).
5. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).
6. Costanzo, M. *et al.* The Genetic Landscape of a Cell. *Science* **327**, 425–431 (2010).
7. Tong, A. H. Y. *et al.* Global Mapping of the Yeast Genetic Interaction Network. *Science* **303**, 808–813 (2004).
8. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet* **38**, 896–903 (2006).
9. Fischer, B. *et al.* A map of directional genetic interactions in a metazoan cell. *eLife* **4**, e05464 (2015).
10. Heigwer, F. *et al.* A global genetic interaction network by single-cell imaging and machine learning. *Cell Systems* **14**, 346-362.e6 (2023).
11. Horn, T. *et al.* Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods* **8**, 341–346 (2011).
12. Tong, A. H. Y. *et al.* Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science* **294**, 2364–2368 (2001).
13. Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods* **7**, 1017–1024 (2010).
14. Laufer, C., Fischer, B., Billmann, M., Huber, W. & Boutros, M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods* **10**, 427–431 (2013).
15. Roguev, A. *et al.* Quantitative genetic-interaction mapping in mammalian cells. *Nat Methods* **10**, 432–437 (2013).
16. Bassik, M. C. *et al.* A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility. *Cell* **152**, 909–922 (2013).
17. Billmann, M., Chaudhary, V., ElMaghraby, M. F., Fischer, B. & Boutros, M. Widespread Rewiring of Genetic Networks upon Cancer Signaling Pathway Activation. *Cell Systems* **6**, 52-64.e4 (2018).
18. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).

19. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87 (2014).
20. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515–1526 (2015).
21. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
22. Boettcher, M. *et al.* Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat Biotechnol* **36**, 170–178 (2018).
23. Ito, T. *et al.* Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. *Nat Genet* **53**, 1664–1672 (2021).
24. Gonatopoulos-Pournatzis, T. *et al.* Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9–Cas12a platform. *Nat Biotechnol* **38**, 638–648 (2020).
25. Esmaeili Anvar, N. *et al.* Efficient gene knockout and genetic interaction screening using the in4mer CRISPR/Cas12a multiplex knockout platform. *Nat Commun* **15**, 3577 (2024).
26. Thompson, N. A. *et al.* Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat Commun* **12**, 1302 (2021).
27. Shen, J. P. *et al.* Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat Methods* **14**, 573–576 (2017).
28. Najm, F. J. *et al.* Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat Biotechnol* **36**, 179–189 (2018).
29. Wolf, G. *et al.* The genetic interaction map of the human solute carrier superfamily. *Mol Syst Biol* **21**, 531–559 (2025).
30. Parrish, P. C. R. *et al.* Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports* **36**, 109597 (2021).
31. Burgold, T. *et al.* Genetic interaction library screening with a next-generation dual guide CRISPR system. Preprint at https://doi.org/10.1101/2024.03.28.587052 (2024).
32. Aregger, M. *et al.* Systematic mapping of genetic interactions for de novo fatty acid synthesis identifies C12orf49 as a regulator of lipid metabolism. *Nat Metab* **2**, 499–513 (2020).
33. DeWeirdt, P. C. *et al.* Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat Biotechnol* **39**, 94–104 (2021).
34. Varland, S. *et al.* N-terminal acetylation shields proteins from degradation and promotes age-dependent motility and longevity. *Nat Commun* **14**, 6774 (2023).
35. Xiao, Y.-X. *et al.* The TSC22D, WNK, and NRBP gene families exhibit functional buffering and evolved with Metazoa for cell volume regulation. *Cell Reports* **43**, 114417 (2024).
36. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174**, 953-967.e22 (2018).
37. Fielden, J. *et al.* Comprehensive interrogation of synthetic lethality in the DNA damage response. *Nature* **640**, 1093–1102 (2025).

38. DeWeirdt, P. C. *et al.* Genetic screens in isogenic mammalian cell lines without single cell cloning. *Nat Commun* **11**, 752 (2020).
39. Rohde, T. *et al.* BaCoN (Balanced Correlation Network) improves prediction of gene buffering. *Mol Syst Biol* (2025) doi:10.1038/s44320-025-00103-7.
40. De Kegel, B., Quinn, N., Thompson, N. A., Adams, D. J. & Ryan, C. J. Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. *Cell Systems* **12**, 1144-1159.e6 (2021).
41. Köferle, A. *et al.* Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Reports* **39**, 110636 (2022).
42. Krieg, S. *et al.* Mitoferrin2 is a synthetic lethal target for chromosome 8p deleted cancers. *Genome Med* **16**, 83 (2024).
43. Lord, C. J., Quinn, N. & Ryan, C. J. Integrative analysis of large-scale loss-of-function screens identifies robust cancer-associated genetic interactions. *eLife* **9**, e58925 (2020).
44. Ward, H. N. *et al.* Analysis of combinatorial CRISPR screens with the Orthrus scoring pipeline. *Nat Protoc* **16**, 4766–4798 (2021).
45. Dede, M., McLaughlin, M., Kim, E. & Hart, T. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol* **21**, 262 (2020).
46. Ajmal, H., Nandi, S., Kebabci, N. & Ryan, C. J. Benchmarking Genetic Interaction Scoring Methods for Identifying Synthetic Lethality from Combinatorial CRISPR Screens. Preprint at https://doi.org/10.1101/2025.03.31.645224 (2025).
47. Li, R. *et al.* Comparative optimization of combinatorial CRISPR screens. *Nat Commun* **13**, 2469 (2022).
48. Chou, J., Anvar, N. E., Elghaish, R., Chen, J. & Hart, T. Optimal methods for analyzing targeted pairwise knockout screens. Preprint at https://doi.org/10.1101/2024.08.19.608665 (2024).
49. Zamanighomi, M. *et al.* GEMINI: a variational Bayesian approach to identify genetic interactions from combinatorial CRISPR screens. *Genome Biol* **20**, 137 (2019).
50. Vinceti, A. *et al.* A benchmark of computational methods for correcting biases of established and unknown origin in CRISPR-Cas9 screening data. *Genome Biol* **25**, 192 (2024).
51. Iorio, F. *et al.* Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. *BMC Genomics* **19**, 604 (2018).
52. Allen, F. *et al.* JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome Res.* **29**, 464–471 (2019).
53. Dempster, J. M. *et al.* Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol* **22**, 343 (2021).
54. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779–1784 (2017).
55. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).

56. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756–780 (2019).
57. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* **34**, 695–697 (2016).
58. Pacini, C. *et al.* Integrated cross-study datasets of genetic dependencies in cancer. *Nat Commun* **12**, 1661 (2021).
59. Pacini, C. *et al.* A comprehensive clinically informed map of dependencies in cancer cells and framework for target prioritization. *Cancer Cell* S1535610823004440 (2024) doi:10.1016/j.ccell.2023.12.016.
60. Dempster, J. M. *et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nat Commun* **10**, 5817 (2019).
61. Billmann, M *et al.* A global genetic interaction map of a human cell reveals conserved principles of genetic networks. (2025).
62. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Genes|Genomes|Genetics* **7**, 2719–2727 (2017).
63. Wainberg, M. *et al.* A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat Genet* **53**, 638–649 (2021).
64. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).
65. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559-2575.e28 (2022).
66. Lazar, N. H. *et al.* High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR-Cas9 editing. *Nat Genet* **56**, 1482–1493 (2024).
67. Simpkins, S. W. *et al.* Using BEAN-counter to quantify genetic interactions from multiplexed barcode sequencing experiments. *Nat Protoc* **14**, 415–440 (2019).
68. Rahman, M. *et al.* A method for benchmarking genetic screens reveals a predominant mitochondrial bias. *Mol Syst Biol* **17**, (2021).
69. Costanzo, M. *et al.* Environmental robustness of the global yeast genetic interaction network. *Science* **372**, eabf8424 (2021).
70. Rahman, A. M. Computational methods to evaluate and interpret genome-wide perturbation screens. (2021).
71. Billmann, M. *et al.* Reproducibility metrics for context-specific CRISPR screens. *Cell Systems* **14**, 418-422.e2 (2023).
72. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3: Genes, Genomes, Genetics* **7**, g3.117.041277 (2017).
73. Dudoit, S. & Fridlyand, J. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol* **3**, research0036.1 (2002).
74. Mani, R., St.Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3461–3466 (2008).

75. Giurgiu, M. *et al.* CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559–D563 (2019).
76. Smyth, G. K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, 1–25 (2004).
77. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).