Complementary human gene interaction maps from radiation hybrids and CRISPRi

Desmond J. Smith*,[1]

[1]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA, Los Angeles, USA

*Corresponding author: DSmith@mednet.ucla.edu

Abstract

The only comprehensive genetic interaction map of the human genome was constructed using increased gene copy number in radiation hybrid (RH) cells. More recently, interactions restricted to essential genes were identified using loss-of-function alleles made by CRISPRi technology. Here, the two maps are compared to understand their similarities and differences. Both maps overlapped significantly with a database of protein-protein interactions. However, interactions in the RH and CRISPRi datasets had no significant overlap, even though the participating genes overlapped significantly. In addition, the RH map showed highly significant similarity with an interaction map constructed from genome-wide association studies (GWASs), while the CRISPRi map did not. This study reveals the different aspects of the genetic interaction landscape illuminated by gain- and loss-of-function alleles.

Keywords: CRISPRi, Genetic Interactions; Radiation Hybrid Mapping; Genome-wide Association Studies

Introduction

Comprehensive genetic interaction maps of the human genome have been daunting to construct because of the large number of potential interactions. For the ∼20,000 protein coding genes, there are $\sim 2.0 \times 10^8$ possible interactions (Khan and Smith 2021). Including non-coding transcripts trebles the number of genes, and increases the number of potential interactions to $\sim 1.8 \times 10^9$.

Radiation hybrid (RH) mapping emerged in the 1990s as a statistical approach to high resolution mapping of the human genome (Goss and Harris 1975; Cox *et al.* 1990). Human cells are lethally irradiated, fragmenting the DNA into short pieces. A random sample of the fragments are then transferred to living cells by cell fusion (Figure 1A). The short fragments allow construction of high resolution genetic maps. Markers that lay close to each other in the genome are unlikely to be separated by the radiation and tend to be coinherited in the RH clones.

Usually distant genes are inherited independently. However, genes that interact will show coinheritance even if at a distance in the genome (Lin *et al.* 2010). For example, the poisonous effect of an extra gene copy on cell growth may be "blocked" by an extra copy of a distant gene. These two genes will show significant coinheritance in the RH clones, and constitute an "attractive" interaction. In contrast, two distant genes that together inhibit cell proliferation will show significant non-coinheritance, a "repulsive" interaction.

Each RH clone harbors ∼10-25% of the genome as an extra copy, corresponding to ∼2,000–5,000 protein-coding genes. This high degree of multiplexing gives the technique extraordinary efficiency and allows it to evaluate interactions for the entire genome. We used the RH approach to construct an interaction network for the entire human genome, including coding and non-coding genes, at single gene resolution (Lin *et al.* 2010). PCR genotyping data of 560 clones from human, mouse, rat and dog RH panels were employed, using the human genome as a scaffold.

Single gene resolution of the RH-PCR network was confirmed from its optimal overlap with a protein-protein interaction network when the RH-PCR network was restricted to nearest gene resolution (OR $> 1.0 \times 10^{-21}$), as well as the $-2\log_{10} P$ interval of interaction peaks (< 20 kb) (Lin *et al.* 2010). Further, the RH network specifically omitted gene pairs that resided within the maximum fragment length (10 Mb), thus excluding genomic linkage as a potential confound.

Unlike the RH network which used extra gene copies, CRISPR interference (CRISPRi) evaluates interactions using partial loss-of-function (hypomorphic) alleles (Horlbeck *et al.* 2018). Genes are silenced using a catalytically inactive version of Cas9 fused to a transcriptional repressor.

CRISPRi generally only targets one gene in each cell, limiting throughput and investigative scope. Methods to allow simultaneous targeting of multiple genes in one cell using CRISPR are being pursued (Fujii *et al.* 2023; Riesenberg *et al.* 2019). However, no more than ∼6-7 genes can be efficiently targeted at present, far short of the thousands of genes in RH cells.

Recently, CRISPRi was used to catalog interactions among pairs of protein coding genes with moderate individual loss-of-function growth effects in K562 and Jurkat cells (Horlbeck *et al.* 2018). The CRISPRi network assessed ∼0.05 % of all potential protein coding interactions, in contrast to the RH-PCR network which examined the entire genome.

Two types of genetic interaction were identified in the CRISPRi study. Buffering interactions occurred when the combination of two loss-of-function alleles resulted in better cell growth than the sum of their effects. In statistical terms, this non-additive phenomenon is a positive interaction. Synergistic interactions occurred when the combination of alleles resulted in poorer growth than expected, statistically a negative interaction. ("Buffering" can also refer to a gene deletion having a weaker than expected phenotype due to compensation by a paralog, De Kegel and Ryan 2019; Khan *et al.* 2020.)

In this study, the biological effects of genetic interactions due to extra gene copies and partial loss-of-function alleles are compared using the RH-PCR and CRISPRi networks. The similarities and differences of the two networks, the STRING protein-protein interaction network (Szklarczyk *et al.* 2019) and networks from genome-wide association studies (GWASs) (Sollis *et al.* 2023) are evaluated.

© The Author(s) 2024.

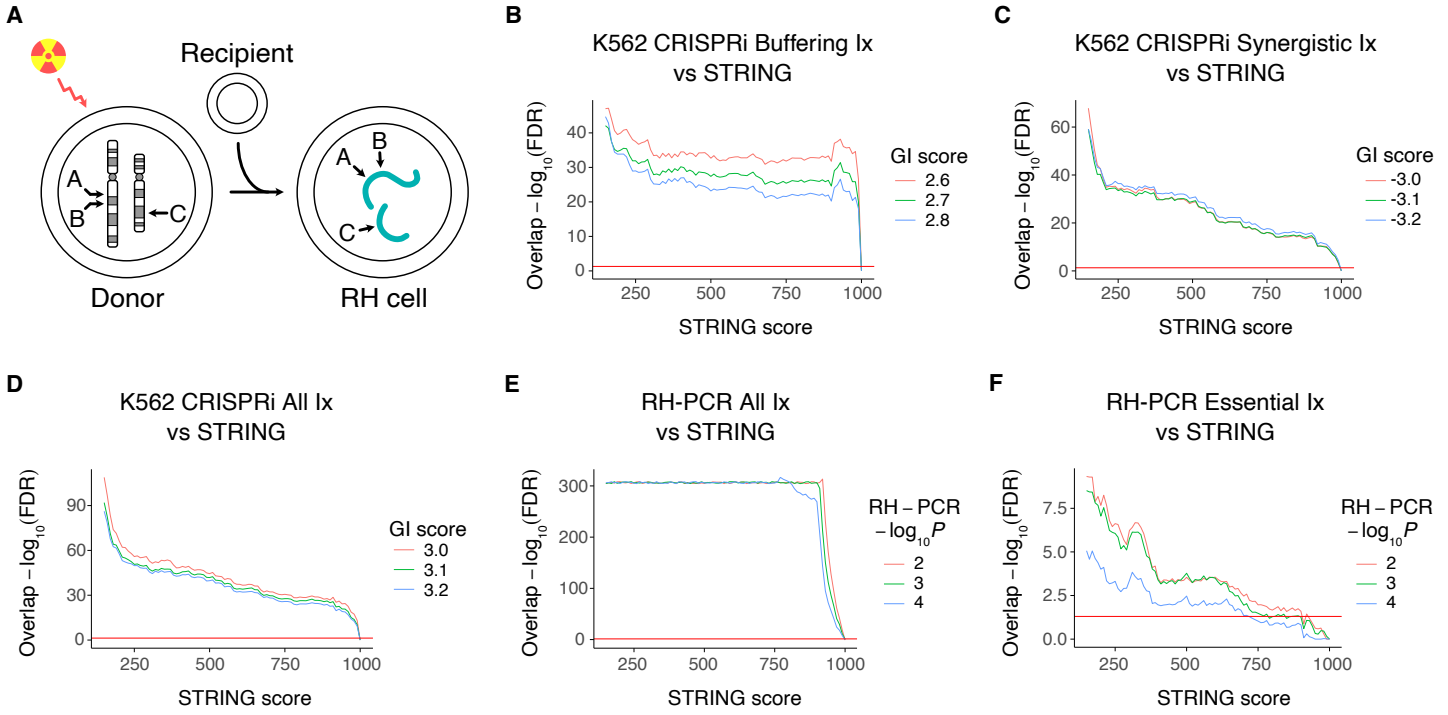

Figure 1 RH-PCR, CRISPRi and STRING interaction networks. **A.** Construction of RH cells. Genes A and B reside close to each other in the genome and tend to be co-inherited in the RH cells. Gene C is a long way from A and B, and is inherited independently unless it interacts with either A or B. Small DNA fragments, torquoise strands. Recipient cell chromosomes not shown. **B.** Overlap of K562 buffering CRISPRi interactions and STRING. **C.** Overlap of K562 synergistic CRISPRi interactions and STRING. **D.** Overlap of K562 combined buffering and synergistic CRISPRi interactions (All) and STRING. **E.** Overlap of protein coding RH-PCR network (All) and STRING. **F.** Overlap of RH-PCR interactions restricted to essential genes in the CRISPRi dataset and STRING. The higher the STRING score threshold, the greater the confidence in the protein-protein interaction. GI, CRISPRi genetic interaction. Positive GI scores, buffering interactions; negative, synergistic. Ix, interactions. Horizontal red line, FDR (false discovery rate) = 0.05.

Materials and methods

Data analysis

Statistical analyses employed R (R Core Team 2023). Graphs were plotted using `ggplot2` and `cowplot` (Wickham 2016; Wilke 2020). Receiver operating characteristic (ROC) curves and areas under the curve (AUCs) were obtained using `pROC`, `plotROC` and `caTools` (Robin *et al.* 2011; Sachs 2017; Tuszynski 2024). Precision recall (PRC) analyses employed `PRROC` (Grau *et al.* 2015). Networks were plotted using `Cytoscape` in the R environment implemented with `RCy3` (Gustavsen *et al.* 2019; Shannon *et al.* 2003).

CRISPRi network

Buffering or synergistic CRISPRi interactions were classified as significant if the genetic interaction (GI) score was greater or less than the 95% confidence intervals of the negative controls, respectively. For K562 cells, the GI threshold was 2.5 for buffering interactions and −2.6 for synergistic. For Jurkat cells, the thresholds were 1.9 and −2.0, respectively. Combined CRISPRi interactions were thresholded on GI score regardless of sign, with thresholds of 2.6 for K562 and 2.0 for Jurkat.

GWAS network

The GWAS network was constructed by linking all genes associated with the same phenotype. The $-\log_{10} P$ for the interactions was calculated as the mean value for each trait. Overlap comparisons of the essential RH-PCR and GWAS networks employed the full GWAS network. However, comparisons of the X chromosome-specific RH-PCR, essential RH-PCR plus one, and the full protein coding RH-PCR networks used the GWAS network downsampled 20-fold to allow manageable computation times.

In the GWAS dataset 70 % of entries were single gene, reflecting the high resolution of modern GWASs. Of the remaining entries, a single gene was chosen at the either beginning or end of the critical interval. Consequently, the GWAS networks are unlikely to be noticeably confounded by linkage.

Overlap analyses

Evaluation of network overlaps was performed as described (Ahn *et al.* 2009), except that instead of employing the χ^2 approximation when expected values in all table cells > 50, Fisher's Exact Test was used throughout. The numbers used for significance testing at optimal overlaps are shown in Supplementary Tables S1-S7. The total population size depended on the comparison.

Overlaps between the CRISPRi, STRING protein-protein, essential RH-PCR and GWAS networks employed the number of interactions tested in the CRISPRi study, depending on whether K562 or Jurkat cells were employed (Supplementary Tables S1, S2, S3, S5, S6). The protein coding RH-PCR and randomly downsampled RH-PCR networks were compared with the STRING and GWAS networks using all pairwise non-redundant combinations of protein coding genes in `GENCODE` v31 (Frankish *et al.* 2023). The total

interaction number was 199,490,325 (Supplementary Tables S2, S5, S7).

Comparison of the RH-PCR essential plus one and GWAS networks used all non-redundant pairwise combinations of unique genes in the RH-PCR essential plus one network and the analogous network based on the GWAS dataset (84,337,578 interactions) (Supplementary Table S5).

The RH-PCR networks with one gene restricted to the X or Y chromosomes and STRING were compared using a population equal to the non-redundant pairwise combination of protein coding genes on the relevant sex chromosome and the autosomes plus the pairwise combinations of genes on the relevant sex chromosome (RH-PCR chrX, 16,560,975 interactions; RH-PCR chrY, 1,260,039 interactions) (Supplementary Table S2). Overlap of the genes in the CRISPRi and RH-PCR networks used a population equal to the number of unique protein coding genes in `GENCODE` (19,942 genes) (Supplementary Table S4).

Clustering of RH-PCR network

Clustering of interactions restricted to CRISPRi essential genes in the RH-PCR network employed `heatmap.2` with Euclidean distance measures (Warnes *et al.* 2022). Clusters were defined using dendrogram cuts closest to origin. Genes were assigned to gene ontology (GO) categories using `ontologyIndex` (Gene Ontology Consortium *et al.* 2023; Greene *et al.* 2017). Significance values for GO enrichment of clusters were calculated using Fisher's Exact Test. Assignment of cell compartments in the essential RH-PCR network was performed as described (Horlbeck *et al.* 2018).

Results

Properties of RH-PCR and CRISPRi networks

The full RH-PCR network consisted of 7,248,479 significant interactions featuring 18,220 genes (false discovery rate, FDR < 0.05) (Benjamini and Hochberg 1995; Lin *et al.* 2010). When restricted to protein coding genes, the RH-PCR network had 2,705,945 significant interactions among 10,897 genes, or 1.4 % of all potential interactions. Of the significant protein coding interactions, 912 (0.03 %) were repulsive and the remainder attractive. None of the genes involved in repulsive interactions were in the CRISPRi dataset.

Interaction strengths in the CRISPRi networks were quantitated using a genetic interaction (GI) score, with positive GI scores indicating buffering interactions and negative indicating synergistic (Horlbeck *et al.* 2018). For example, a buffering interaction has a positive GI score if cells with two partial loss-of-function alleles grow better than expected from the additive inhibitory effects of the individual alleles. The higher the GI score, the stronger the buffering interaction.

In contrast, a synergistic interaction has a negative GI score and occurs when cells grow more poorly than expected from the additive inhibitory effects of the two alleles. The more negative the score, the stronger the synergistic interaction. In addition to buffering and synergistic CRISPRi interactions, the combined properties of both types of interaction together were examined using significant interactions regardless of sign.

Significant GI thresholds based on negative controls in K562 cells were > 2.5 and < -2.6, for buffering and synergistic interactions respectively, and > 1.9 and < -2.0 for Jurkat (Materials and methods). The thresholds for the combined buffering and synergistic networks were 2.6 for K562 and 2.0 for Jurkat.

Using CRISPRi, 100,128 potential interactions among 448 essential genes were tested in K562 cells. A total of 3,885 (3.9 %) interactions had significant growth effects, featuring 446 genes. Of the interactions, 1,493 were buffering and 2,392 synergistic. In Jurkat cells, 95,266 interactions were tested using 436 essential genes. A total of 2,552 (2.7 %) interactions featuring 424 genes were significant. Of the interactions, 1,318 were buffering and 1,234 synergistic.

The protein coding RH-PCR network featured $\sim$25 times more genes and $\sim$840 times more significant interactions than the CRISPRi network.

Network subsets

Six subsets of the full RH-PCR network were used in network comparisons.

1. The protein coding RH-PCR network ("All").

2. The protein coding RH-PCR network restricted to the interactions of essential genes evaluated in the CRISPRi study (279 genes, 1,778 interactions for K562; 273 genes, 1,686 interactions for Jurkat). In total, the "essential" RH-PCR network consisted of 285 genes and 1,839 interactions, allowing equally powered overlap comparisons with the CRISPRi networks.

3. Essential RH-PCR network plus genes linked by one interaction (10,497 genes, 140,308 interactions for K562; 10,497 genes, 140,216 interactions for Jurkat). The combined size of the "essential plus one" RH-PCR network is 10,497 genes and 140,369 interactions and is intermediate in size between first two subsets.

4. The protein coding RH-PCR network randomly downsampled to the same number of interactions evaluated in the CRISPRi study, whether from K562 or Jurkat cells. Overlaps of the "RH-PCR ds" network employed the mean results of 20 independent downsamples. The RH-PCR ds network consists of $8{,}265 \pm 20$ s.d. genes and 104,443 interactions, and is similar in size to the essential plus one RH-PCR network.

5. The protein coding RH-PCR network downsampled so that each interaction has at least one gene on the X but not Y chromosome. The RH-PCR chrX network has 10,071 genes and 151,980 interactions.

6. The protein coding RH-PCR network downsampled so that each interaction has at least one gene on the Y but not X chromosome. The RH-PCR chrY network has 3,614 genes and 8,440 interactions.

Four subsets of the CRISPRi networks were used.

1. Buffering, with significant positive GI scores.

2. Synergistic, with significant negative GI scores.

3. Combined, with all significant GI scores regardless of sign.

4. CRISPRi network downsampled so that each interaction has at least one gene on chromosome X. A total of 6,167 interactions involving 448 essential genes were tested in the K562 chrX network, of which 339 (5.5 %) interactions had significant growth effects, featuring 225 genes. Of the interactions, 88 were buffering and 251 synergistic. In Jurkat cells, a total of 5,590 interactions were tested involving 437 genes. A total of 252 (4.5 %) interactions featuring 167 genes were significant. Of the interactions, 109 were buffering and 143 synergistic. The CRISPRi dataset had no genes on chromosome Y.

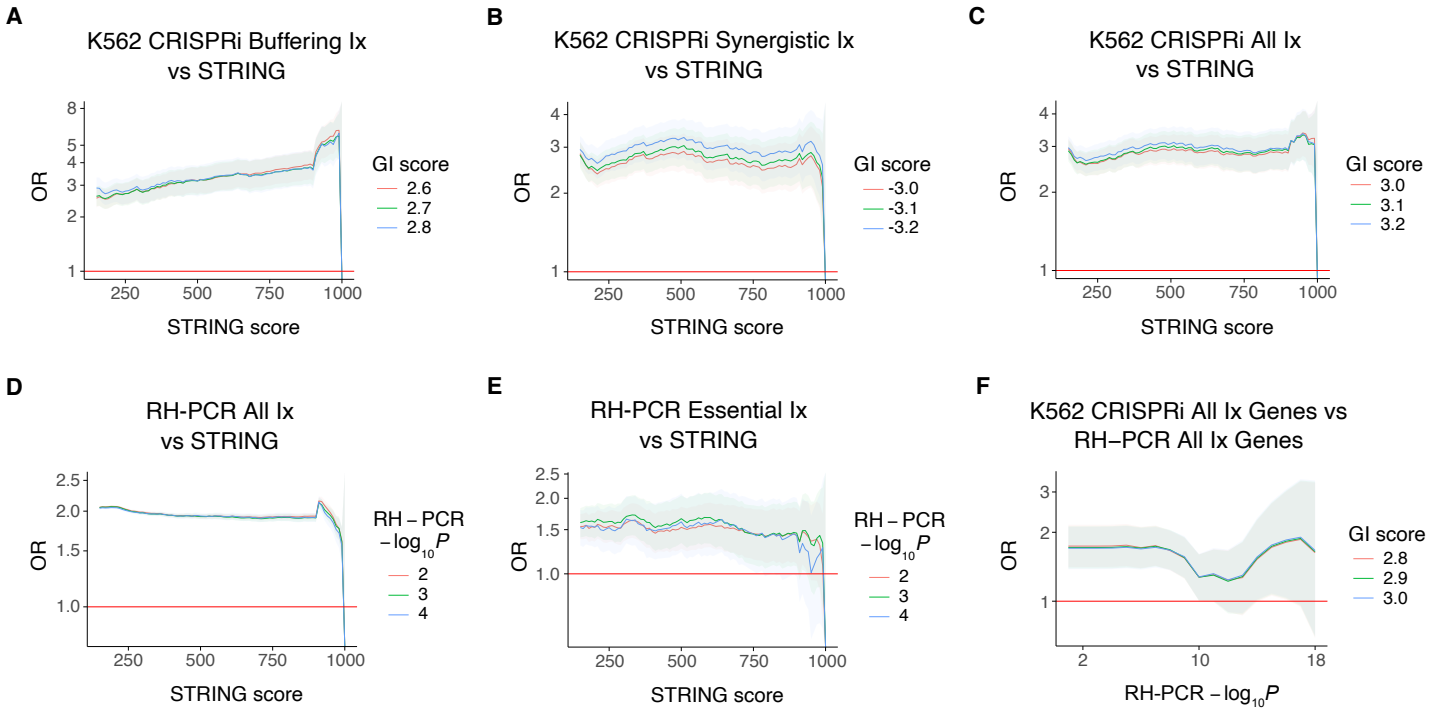

Figure 2 Odds ratios for overlaps of RH-PCR, CRISPRi and STRING networks. **A.** K562 buffering CRISPRi interactions and STRING. **B.** K562 synergistic CRISPRi interactions and STRING. **C.** K562 combined buffering and synergistic CRISPRi interactions and STRING. **D.** Protein coding RH-PCR network (All) and STRING. **E.** RH-PCR interactions restricted to essential genes in the CRISPRi dataset and STRING. **F.** Genes involved in K562 combined CRISPRi and protein coding RH-PCR networks. Odds ratios on $\log_{10}$ scale. Horizontal red line, odds ratio = 1. Grey areas, 95% confidence limits.

In network comparisons using the CRISPRi data, statistics from K562 cells are reported. Observations from the Jurkat cells are presented in the Supplementary material, and were essentially similar.

Overlap significance

The overlap significance of two genetic networks depends on the interaction thresholds used in each of them. Rather than simply using one threshold, the whole overlap terrain was surveyed by scanning threshold parameters from least to most stringent (Ahn *et al.* 2009). The overlap significance at each threshold was then calculated, correcting for multiple tests using false discovery rates (Benjamini and Hochberg 1995). This approach provides a survey of the overlap landscape instead of a single snapshot at one threshold parameter.

Overlaps of RH-PCR and CRISPRi networks with STRING

First to be examined was the overlap of the protein coding RH-PCR and CRISPRi interaction networks with the STRING repository of protein-protein interactions (Szklarczyk *et al.* 2019). STRING v11 contains 5,879,727 human protein-protein interactions, identified by collating information from multiple publicly available sources. Interactions are assigned a score ranging from 150 (lowest confidence) to 999 (highest).

The buffering, synergistic and combined CRISPRi interactions showed significant overlap with STRING in both K562 and Jurkat cells (K562 combined interactions, optimum FDR = 1.3×10^{-109}, odds ratio, OR = 2.9, Fisher's Exact Test) (Figures 1B-D, 2A-C, Supplementary Figures S1, S2, S3, Supplementary Table S1). The protein coding RH-PCR network also overlapped significantly with STRING (optimum FDR = 4.3×10^{-319}, OR = 1.9) (Figures 1E, 2D, Supplementary Figure S4, Supplementary Table S2), but with greater significance than CRISPRi (Figure 3A, Supplementary Figure S4), reflecting the more comprehensive nature of the RH-PCR network. As expected, overlap significance diminished as threshold stringency increased, due to diminishing interaction numbers (Supplementary Figure S5).

When restricted to the essential genes evaluated in the CRISPRi study, the RH-PCR interactions still showed significant overlap with STRING (optimum FDR = 4.7×10^{-10}, OR = 1.5) (Figures 1F, 2E, Supplementary Figure S4, Supplementary Table S2), but was less significant than CRISPRi (Figure 3B, Supplementary Figure S4). Random downsampling of the RH-PCR network to the same number of interactions tested in the CRISPRi study did not decrease the optimum overlap with STRING compared to the protein coding RH-PCR network (optimum FDR = 7.4×10^{-321}, OR = 2.1), and was more significant than the overlaps of the CRISPRi networks with STRING (Figure 3A, Supplementary Table S2). However, due to the decreased size of the downsampled RH-PCR network, the overlap decayed more quickly than the full protein coding network as stringency was increased,.

STRING scores for interactions specific to the K562 or Jurkat CRISPRi network were significantly greater than the null distribution of STRING scores (K562, $t[1,1256] = 5.2$, $P = 2.3 \times 10^{-7}$; Jurkat, $t[1,685] = 4.9$, $P = 1.5 \times 10^{-6}$; Welch Two Sample t-test) (Figure 3C). In contrast, interactions in the essential RH-PCR network showed no significant difference. The RH-PCR interactions appear to be more representative of protein-protein interactions than the

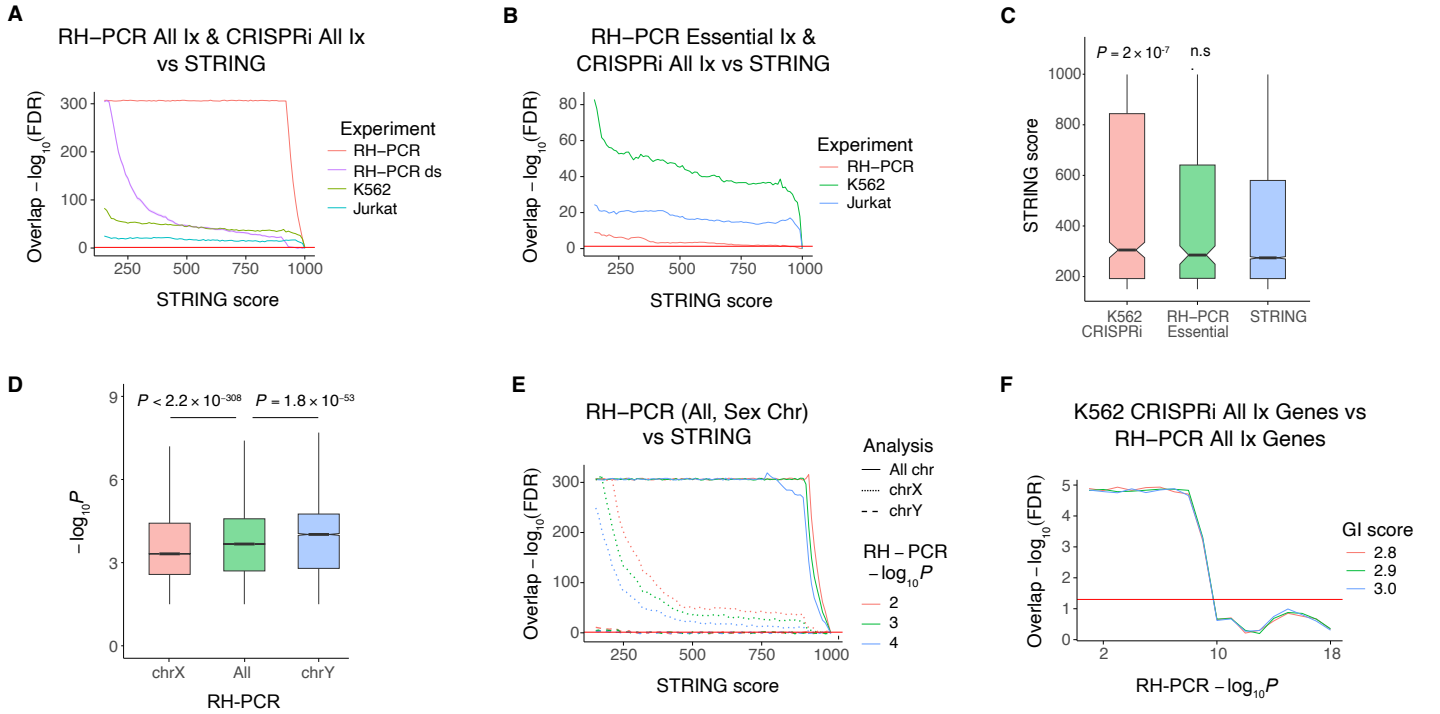

Figure 3 Overlaps of genetic interaction networks and their participating genes. **A.** Overlaps of protein coding RH-PCR network, downsampled RH-PCR network (RH-PCR ds), and K562 and Jurkat combined buffering and synergistic CRISPRi networks with STRING. RH-PCR downsampled network 95% confidence interval shown, but hard to see. **B.** Overlaps of essential RH-PCR as well as K562 and Jurkat combined CRISPRi networks with STRING. **C.** STRING scores for interactions specific to K562 combined CRISPRi, essential RH-PCR and STRING networks. **D.** Significance values for the chrX and chrY specific RH-PCR networks as well as the protein coding RH-PCR network. **E.** Overlaps of the protein coding RH-PCR network and chrX and chrY specific RH-PCR networks with STRING. **F.** Overlaps of genes involved in interactions in K562 combined CRISPRi and protein coding RH-PCR networks.

CRISPRi interactions, which preferentially recover interactions in the STRING dataset with higher scores.

Sex chromosome specific RH-PCR networks

Retained autosomal genes show a copy number change of 1.5-fold in RH cells (three copies vs. two). In contrast, X or Y chromosomal genes show a 2-fold copy number change (two copies vs. one), since the recipient cells used to create RH clones are male. The relative increase in copy number of sex chromosome genes might result in stronger genetic interactions. Unexpectedly however, retained X chromosome genes in RH cells show decreased expression increases compared to the autosomes (Park *et al.* 2008; Wang *et al.* 2011). This dosage compensation is not due to X chromosome inactivation and might result in weaker genetic interactions.

To investigate the competing effects of copy number change and dosage compensation, the interaction strength of genes on the X or Y chromosomes were compared to the complete protein-coding RH-PCR network (Figure 3D).

Genes on the X chromosome showed significantly weaker interactions than the complete network ($t[1{,}169496] = 47.6$, $P < 2.2 \times 10^{-308}$, Welch Two Sample t-test), suggesting that dosage compensation outweighs copy number effects. In contrast, genes on the Y chromosome showed stronger interactions than the complete network ($t[1{,}8478] = 15.5$, $P = 1.8 \times 10^{-53}$, Welch Two Sample t-test), showing that copy number effects are significant for this chromosome, which has no dosage compensation.

The CRISPRi networks showed no differences in interaction strength for genes on the X chromosome compared to the complete networks ($P > 0.1$), although statistical power was limited by size. The CRISPRi networks had no Y chromosome genes.

The overlaps between the X chromosome specific RH-PCR networks and STRING were undiminished compared to the protein coding RH-PCR network (optimum FDR = 3.7×10^{-316}, OR = 2.1), although the significance of overlap for the X chromosome network decayed more rapidly due to its smaller size (Figure 3E, Supplementary Table S2). Thus, the network overlaps for genes on the X chromosomes were not strongly affected by either the copy number or dosage effects.

Because of its smaller size the Y chromosome specific RH-PCR network showed decreased, but still significant, overlap significance with STRING compared to the protein coding RH-PCR network (optimum FDR = 6.8×10^{-12}, OR = 2.6).

Figures of merit for genetic interaction networks

Receiver operating characteristic (ROC) curves were used to evaluate the performance of the essential RH-PCR and CRISPRi networks in identifying interactions in STRING (Figure 4A). Although the areas under the curve (AUCs) were only slightly better than random (0.5), all were significant ($P < 2.2 \times 10^{-308}$). Consistent with the results from Fisher's Exact Test (Figure 3B), the K562 network showed better predictive power than Jurkat, which in turn was better than the essential RH-PCR network.

When events of interest (significant genetic interactions), are greatly outnumbered by null events (no interactions) the precision recall (PRC) plot is superior to the ROC curve (Saito and Rehmsmeier 2015). PRC plots comparing the essential RH-PCR

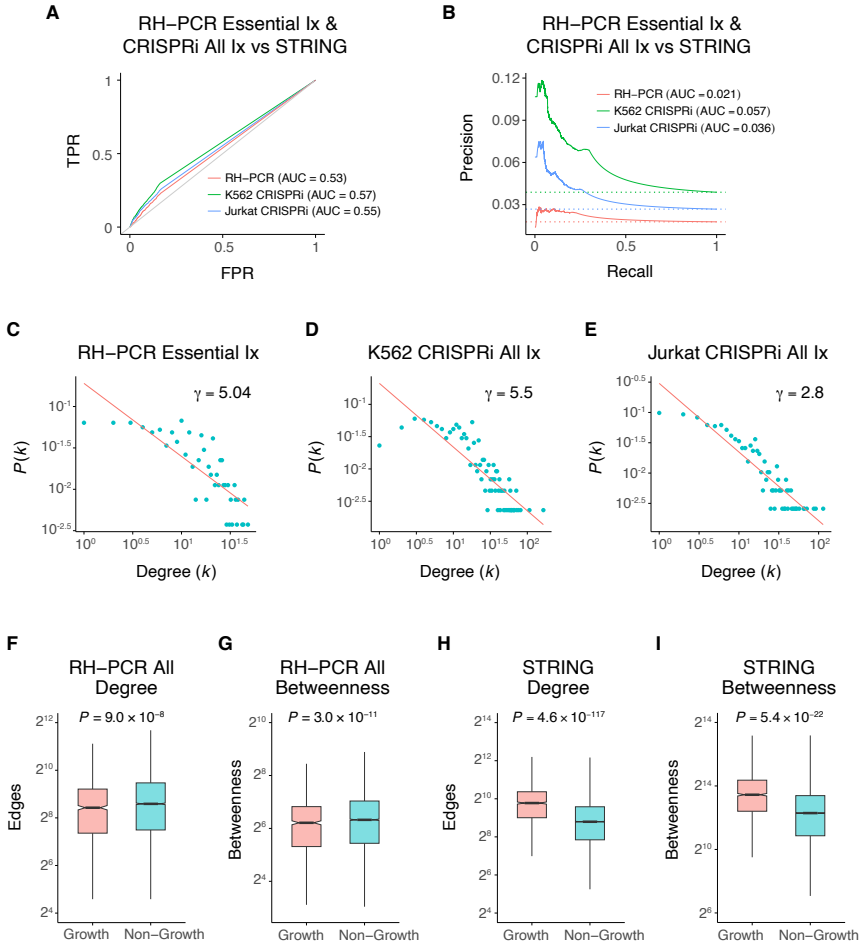

Figure 4 RH-PCR, CRISPRi and STRING interaction networks. **A.** ROC curves. Essential RH-PCR and combined buffering and synergistic K562 and Jurkat networks. TPR, true positive rate; FPR, false positive rate. Gray line, random. **B.** Precision recall curves. Dotted lines, random. **C.** Degree distribution for RH-PCR essential network. **D.** Degree distribution for K562 combined buffering and synergistic CRISPRi network. **E.** Degree distribution for combined Jurkat network. **F.** Degree centrality for protein coding RH-PCR network. **G.** Betweenness centrality for protein coding RH-PCR network. **H.** Degree centrality for STRING. **I.** Betweenness centrality for STRING. Growth, genes required for growth. Non-growth, not required.

and CRISPRi networks with STRING showed similar results to the ROC curve and Fisher's Exact Test (Figure 4B). The three networks had only modest power to identify interactions in STRING, though all were significantly better than random ($W > 8.2 \times 10^7$, $P < 3.8 \times 10^{-12}$, Wilcoxon Rank Sum Test).

Topology of RH-PCR and CRISPRi networks

Both the essential RH-PCR and CRISPRi networks were approximately scale-free (Figure 4C-E), a frequent property of biological networks (Albert 2005). The RH-PCR and K562 networks showed similar scale factors (γ) ~ 5, while the Jurkat network had a smaller scale factor of 2.8.

Degree and betweenness centralities in biological networks may be positively correlated with gene essentiality, but the finding remains controversial (Coulomb *et al.* 2005; Hahn and Kern 2005). To investigate these relationships in the protein coding RH-PCR AND STRING networks, a CRISPR knockout dataset (null alleles) in K562 cells was used (Wang *et al.* 2015).

Growth genes were significantly associated with decreased degree and betweenness centralities in the protein coding RH-PCR network (degree, $t[1,1341] = 5.4$, $P = 9.0 \times 10^{-8}$; betweenness, $t[1,1533] = 6.7$, $P = 3.0 \times 10^{-11}$; Welch Two Sample t-test) (Figures 4F,G). In contrast, there was increased degree and betweenness centralities for growth genes in STRING (degree, $t[1,1770] = 24.8$, $P = 4.6 \times 10^{-117}$; betweenness, $t[1,1692] = 9.8$, $P = 5.4 \times 10^{-22}$; Welch Two Sample t-test) (Figures 4H,I). Both the RH-PCR and STRING networks showed reversed relationships for growth suppressing genes (genes that increased growth when knocked out) (Supplementary Figure S6).

The RH-PCR network thus displays the opposite relationship between gene essentiality and centrality compared to STRING. Because the CRISPRi network was confined to essential genes, it was not possible to explore the relationship between centrality and essentiality for this network.

Gene expression and RH-PCR and CRISPRi networks

A microarray dataset from a human RH panel (Wang *et al.* 2011) was used to explore whether poorly expressed genes are less likely to be involved in genetic interactions. The essential RH-PCR network showed a weak but significant positive correlation ($R = 0.06$,

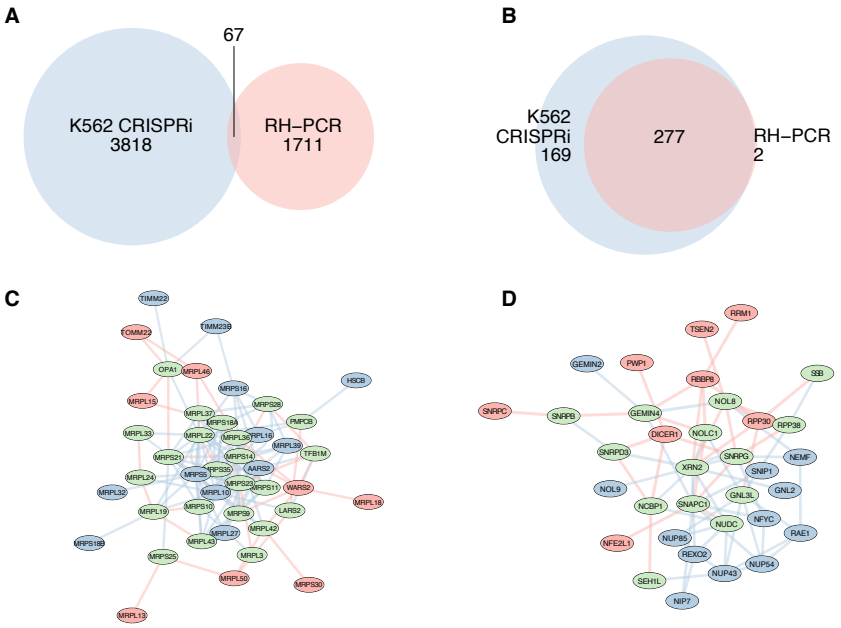

Figure 5 Overlaps of interactions and genes. **A.** Overlap of significant interactions in the essential RH-PCR and K562 combined buffering and synergistic CRISPRi networks. **B.** Overlap of genes involved in significant interactions in the two networks. **C.** Interactions between mitochondrial genes in amalgamated essential RH-PCR and K562 buffering and synergistic CRISPRi networks. Red nodes and edges, RH-PCR network only; blue nodes and edges, CRISPRi only; green nodes, both RH-PCR and CRISPRi. Edge width proportional to interaction strength. **D.** Interactions between genes that encode nuclear-and nucleolus-related proteins in amalgamated essential RH-PCR and K562 buffering and synergistic CRISPRi networks.

$t[1{,}1649] = 2.5$, $P = 0.013$) between genetic interaction strength ($-\log_{10}P$) and gene expression. The correlation for the full protein coding RH-PCR network was also significant and weakly positive ($R = 0.02$, $t[1{,}2320329] = 25.1$, $P = 3.3 \times 10^{-139}$).

Similarly, RNA-Seq data for K562 and Jurkat cells were used to examine the relationship between gene expression and the CRISPRi networks (ENCODE Project Consortium 2012; Liau *et al.* 2017). Both K562 ($R = 0.03$, $t[1{,}85076] = 8.7$, $P = 4.6 \times 10^{-18}$) and Jurkat ($R = 0.03$, $t[1{,}81001] = 9.9$, $P = 6.9 \times 10^{-23}$) cells showed a significant positive correlation between expression and GI score, though the magnitude of the correlation was small.

Genes involved in RH-PCR and CRISPRi interactions

The essential RH-PCR interaction network did not show significant FDR corrected overlap with the K562 or Jurkat CRISPRi networks (K562 combined interactions, optimum FDR = 1.0, OR = 0.21) (Supplementary Table S3). However, there were highly significant overlaps between the genes involved in the interactions of the essential RH-PCR and CRISPRi networks (K562 combined interactions, optimum FDR = 1.3×10^{-5}, OR = 1.8) (Figure 3F, Supplementary Figures S7, S8, Supplementary Table S4).

A random sample of genes the same size as the essential RH-PCR network showed no significant overlap with the CRISPRi networks at the parameters giving optimum overlap of the actual data (K562 combined interactions, most significant FDR = 0.4; Jurkat combined interactions, FDR = 0.9; 2,000 bootstrap samples, $P < 5.0 \times 10^{-4}$). This observation accords with the null expectations of Fisher's Exact Test.

Consistent with the non-significant overlap of interactions, only 67 were common between the 3,885 combined buffering and synergistic K562 CRISPRi interactions and 1,778 essential RH-PCR interactions (Figure 5A, Supplementary Figure S9A). In contrast, of the 446 and 279 genes involved in the K562 CRISPRi and essential RH-PCR interactions, respectively, a total of 277 were common (Figure 5B, Supplementary Figure S9B). The RH-PCR and CRISPRi approaches identify nearly the same genes involved in genetic interactions, but highlight different sets of interactions among these genes.

Examples of the complementary genetic interactions in the essential RH-PCR and K562 combined CRISPRi networks are shown for mitchondrial and nuclear-related genes in Figures 5C,D.

The mitochondrial interaction network obtained by amalgamating the essential RH-PCR and buffering and synergistic K562 CRISPRi datasets involved 43 genes, of which 23 were shared (Figure 5C). In contrast, of the 127 interactions in this network only 4 were shared between the essential RH-PCR and CRISPRi networks. Similarly, the network of essential nuclear- and nucleolus-related genes had 34 genes, of which 14 were shared between the RH-PCR and K562 CRISPRi networks, but of the corresponding 65 interactions only one was shared (Figure 5D).

Functional clustering of essential genes in the RH-PCR network

To further explore the essential RH-PCR network, unsupervised clustering of genes based on their interaction $-\log_{10}P$ value was employed (Figure 6A). Clusters were then assigned in principled fashion to gene ontology (GO) enrichment categories (FDR $< 10^{-5}$) (Gene Ontology Consortium *et al.* 2023).

Terms relating to mitochondria (mitochondrial inner membrane, mitochondrial fusion), the nucleus (nucleoplasm, nuclear envelope) and cytosol were prominent among the GO enrichments. For example, the mitochondrial inner membrane cluster contains 211

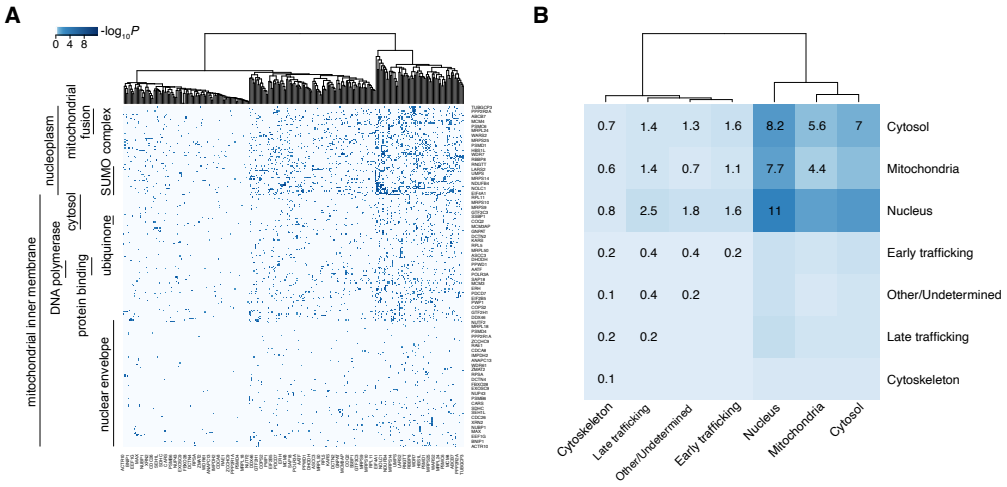

Figure 6 Functional clustering of essential RH-PCR network. **A.** Gene clusters based on $-\log_{10}P$ interaction values. Functional enrichments use gene ontology (GO) categories. For reasons of space, only 26 % of the genes in the cluster diagram are labeled. **B.** Clustering of shared cellular compartments based on percent interacting genes expressed in each compartment. Percent sharing indicated by color and number in each heatmap square. Dendrograms resulting from unsupervised clustering shown above heatmaps.

of the 241 genes related to mitochondria in the nuclear genome. Similarly, of the 308 nuclear-related genes in the genome, there were 105 in the nuclear envelope cluster .

Clustering of the CRISPRi interaction network gave similar enrichment patterns, with prominent enrichment of nucleus/cytoplasm, cytosol and mitochondria (Horlbeck *et al.* 2018). However, some GO terms such as mRNA stability and splicing that were prominent in the CRISPRi network did not appear in the essential RH-PCR network.

Cellular compartments of essential genes in the RH-PCR network

Gene interactions are reflected in the shared cellular compartments of the corresponding proteins. To better understand this aspect of the essential RH-PCR network, unsupervised clustering of cellular compartments was used based on the proportion of interacting genes expressed in each compartment (Figure 6B).

Reminscent of the GO enrichment clusters, the 3 compartments with the highest proportion of shared interactions were nucleus, cytosol and mitochondria. The essential RH-PCR network displayed some differences and similarities with the CRISPRi network, which showed strong clustering of mitochondria for buffering interactions and other/undetermined and early trafficking for synergistic interactions (Horlbeck *et al.* 2018).

Genetic interaction networks and GWAS

Genome-wide association studies (GWASs) identify individual variants that contribute to complex disorders. However, these single variants likely reflect underlying networks of genetic interactions, which are difficult to dissect because of a lack of statistical power (Fang *et al.* 2019; Ritchie and Van Steen 2018). The potential of the RH-PCR and CRISPRi networks to elucidate these interactions was investigated. A GWAS network was constructed by linking genes with the same phenotype from a database of GWAS results (Sollis *et al.* 2023). The overlap of the essential RH-PCR and CRISPRi networks with the GWAS network was then examined (Figure 7A, Supplementary Tables S5, S6).

The essential RH-PCR network showed significant overlap with the GWAS network (optimum FDR = 1.9×10^{-14}, OR = 2.0), while the K562 and Jurkat CRISPRi networks did not. Significant overlap with the GWAS network was also found for the entire protein coding RH-PCR network (optimum FDR = 7.2×10^{-319}, OR = 1.7) and a pruned network consisting of essential genes and genes one link away (optimum FDR = 4.5×10^{-250}, OR = 16.1) (Figure 7B).

The significant overlap of the RH-PCR and GWAS networks may reflect the less severe effect of an extra gene copy in the RH cells compared to the loss-of-function CRISPRi alleles. Most common alleles for GWAS affect gene regulation and have weaker effects than loss-of-function alleles (Nasser *et al.* 2021).

A significant overlap was found between the downsampled protein coding RH-PCR and GWAS networks in which at least one gene of each interaction was on the X chromosome (optimum FDR = 4.7×10^{-10}, OR = 2.6) (Supplementary Table S5). However, the overlap significance was not as strong as for the complete RH-PCR network, unlike the overlap between the X chromosome specific RH-PCR network and STRING (Figure 3E). There is a well documented paucity of GWASs that include the sex chromosomes (Sun *et al.* 2023). Consistent with this lack, X chromosome loci in the GWAS database accounted for only 0.5 % of all loci. This ascertainment bias may account for the weaker (but still significant) overlap of the X chromosome specific RH-PCR and GWAS networks.

The overlap of the complete protein coding RH-PCR and GWAS networks prompted an examination of whether loci from a higher powered GWAS could be predicted from a lower powered GWAS using the RH-PCR network. Two GWASs of colorectal cancer were employed. The earlier and smaller GWAS had ∼35,000 cases and ∼71,000 controls and identified 70 loci (Law *et al.* 2019). The later GWAS had greater power, with ∼100,000 cases and ∼155,000 controls, and identified 162 loci (Fernandez-Rozadilla *et al.* 2023). A total of 55 loci were shared between the two studies.

Interactions in the RH-PCR network that linked genes in the smaller GWAS with genes in the larger study by a single step were significantly enriched compared to background (OR = 1.5, $P = 6.2 \times 10^{-6}$). Of the 107 disease related genes (DRGs) for colorectal cancer found in the larger but not the smaller GWAS, 40 were linked by the RH-PCR network to the smaller study (Figure 7C, Supplementary Table S7). Examples of predicted DRGs in the larger study linked by the RH-PCR network to those in the smaller

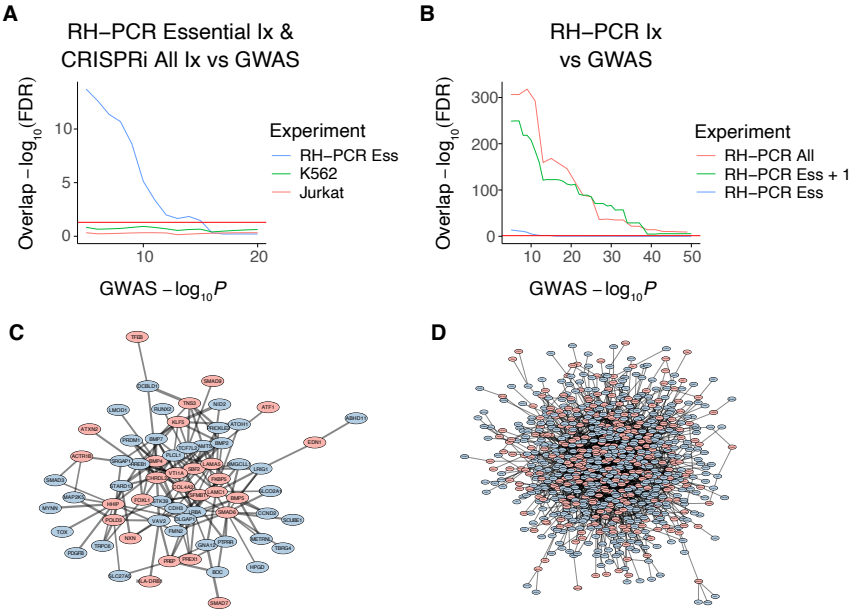

Figure 7 RH-PCR, CRISPRi and GWAS networks. **A.** Overlaps of essential RH-PCR and K562 and Jurkat combined buffering and synergistic CRISPRi networks with GWAS network. **B.** Overlaps of protein coding RH-PCR, essential RH-PCR plus genes one link away, and essential RH-PCR networks with GWAS network. **C.** Colorectal cancer GWAS genes from lower power study (blue nodes) and additional genes in the higher power study (pink) predicted by the RH-PCR network. Edge width reflects RH-PCR $-\log_{10}P$. **D.** Height GWAS.

study included genes with a role in oncogenic proliferation, such as *PDGFB* (linked to *HHIP* in the smaller study), *VAV2* (linked to 7 genes in the smaller study, including *BMP4*, *LAMC1*, *POLD3*) and *PTPRR* (*COL4A2*, *PREX1*, *SMAD6*).

In another example, two GWASs of height were evaluated. The lower powered study examined ∼628,000 individuals and identified 893 loci, while the higher powered study examined 5.4 million individuals and identified 2,483 loci (Sakaue *et al.* 2021; Yengo *et al.* 2022). A total of 529 loci were shared between the two studies. There was significant enrichment of interactions in the RH-PCR network linking genes in the smaller GWAS to genes in the larger (OR = 1.8, $P = 2.2 \times 10^{-308}$). Of the 1,954 genes specific to the higher powered GWAS, 895 were linked by the RH-PCR network to genes in the lower powered study (Figure 7D, Supplementary Table S7).

The RH-PCR network may thus help identify additional candidate genes using data from modestly powered GWASs. Possible scenarios include GWASs of less common or difficult to ascertain phenotypes.

Discussion

Deeper insights into cellular mechanisms can be obtained by charting biological networks from multiple perspectives. However, there is no reason *a priori* to expect that differing methods will give identical or even closely related networks. For example, two genes may strongly interact genetically even if their proteins do not bind. Understanding the differences between biological networks will require detailed and comprehensive studies. Comparisons between various networks have been possible in lower eukaryotes such as yeast (VanderSluis *et al.* 2018), but until recently have been daunting in mammalian cells.

This study represents the first investigation of the similarities and differences of two independently constructed genetic networks in mammalian cells. The RH-PCR network was constructed using extra gene copies and the CRISPRi network using partial loss-of-function alleles. The findings obtained from comparing these published datasets include the following. (1.) Both CRISPRi and RH-PCR networks overlap with a protein-protein interaction network. (2.) The RH-PCR and CRISPRi networks identify a core set of interacting genes, even though the actual interacting gene pairs differ. (3.) The RH-PCR network overlaps better than the CRISPRi network with GWAS networks, casting a new light on the relationship of copy number and loss-of-function variants to complex genetic traits. (4.) The ability to predict disease related genes using the RH-PCR network in combination with GWASs.

While experimental verification of selected interactions in the RH-PCR network will be useful, considering its size these studies can only explore a small corner of the network. Validation of a large network is best done at a genome-wide level. Comparison of the RH-PCR network with multiple interaction networks was highly significant, suggesting that the this approach can inform diverse biological contexts.

The relatively non-toxic effects of increased gene copy number in the RH approach allows mapping of interactions for all genes and is reflected in the fact that retention of donor DNA is uniform across the entire genome (Ahn *et al.* 2009; Khan *et al.* 2016, 2020; Khan and Smith 2021; Lin *et al.* 2010; Lin and Smith 2015; Park *et al.* 2008; Wang *et al.* 2011).

In contrast, the essential genes in the CRISPRi study were chosen for their moderate growth phenotype. The selection was necessary because interactions are difficult to assess when loss of one gene partner abolishes cell growth. Further, the relatively low throughput of the CRISPRi approach makes an evaluation of all interactions prohibitive.

The restriction of the CRISPRi network to genes with mild growth effects may introduce selection bias, even though comparisons included those in which the RH-PCR network was downsampled to the same genes as the CRISPRi network. Further insights

into the two types of interaction network may emerge with higher extra copy numbers or stronger loss-of-function alleles.

Both the CRISPRi and protein coding RH-PCR maps showed significant overlap with the STRING repository of protein-protein interactions. However, the CRISPRi interaction map showed greater overlap than the RH-PCR network restricted to essential genes. The protein coding RH-PCR network showed superior overlap, probably because of its much larger size.

There was no significant overlap between the RH-PCR and CRISPRi interaction networks, reflecting their different mechanisms. This finding is consistent with a study showing that individual growth genes identified by partial inactivation using CRISPRi were different from those identified by overexpression using CRISPR activation (CRISPRa) (Gilbert *et al.* 2014).

In contrast, the CRISPRi and RH-PCR networks showed significant overlap when analyzing the participating genes rather than the interactions. Partial loss-of-function and extra gene copies thus implicate similar genes in a biological process, but highlight distinct interaction partners. These similarities and differences could reflect hidden commonalities that may become manifest when higher order interactions are evaluated (Kuzmin *et al.* 2018).

Genome-wide mapping of three-way interactions is difficult using CRISPR, but is feasible using RH technology because of the large number of extra gene copies in each cell. Nevertheless, even in RH cells, three-way interactions occur at a lower frequency than two-way ($\sim$cube of the retention probability vs square). Large datasets will be required for adequate statistical power.

The essential RH-PCR network showed similar functional enrichments to that found in the CRISPRi network, consistent with their shared genes. GO analysis highlighted mitochondria and nucleus, while compartment clustering emphasized nucleus, cytoplasm and mitochondria, in that order. Other compartment classifications present in the CRISPRi network, such as early trafficking and cytoskeleton, were not clearly delineated in the RH-PCR network.

The RH-PCR network showed significant overlap with a network constructed from human GWAS data. In contrast, the CRISPRi network showed no overlap. The lack of overlap of the CRISPRi and GWAS networks was not due to insufficient power, since an RH-PCR network constructed from the same genes as the CRISPRi network did show significant overlap. Further, the overlap between the RH-PCR and GWAS networks is unlikely to be appreciably inflated by linkage, as both approaches have close to single gene resolution.

The overlap of the RH-PCR and the GWAS networks suggests that the extra gene copies used in the RH-PCR network are similar in their effects to common GWAS variants. This observation mirrors the finding that growth genes identified using bulk segregant analysis of RH cells showed significant overlap with GWAS genes, while growth genes identified using CRISPR did not (Khan *et al.* 2020).

The RH approach uses natural gene promoters, while CRISPRi employs partial loss-of-function alleles. These technologies will only result in a phenotype if the targeted gene is normally transcriptionally active, and is consistent with the weak but statistically significant correlation between interaction strength and gene expression in both networks.

In contrast, CRISPRa employs a catalytically dead version of Cas9 fused to a transcriptional activator to constitutively drive gene transcription (Gilbert *et al.* 2014). The RH-PCR and CRISPRi approaches may hence provide more physiologically relevant genotype/phenotype information. Nevertheless, it will be valuable to compare the RH-PCR and CRISPRi networks with genetic interaction maps constructed using CRISPRa when they become available.

The great preponderance of attractive to repulsive interactions in the RH-PCR network is not mirrored by the roughly equal abundance of buffering and synergistic interactions in the CRISPRi networks. In addition, the buffering and synergistic CRISPRi interactions showed similar overlaps with the unsigned STRING network. A direct mapping of the two types of RH-PCR interaction to the two types of CRISPRi interaction is thus unlikely.

Any parallels between the RH-PCR and CRISPRi networks are further complicated by the fact that the first depends on RH clone viability, while the second is based on promotion or inhibition of cell growth. An understanding of how genetic interactions contribute to cell growth and physiology will be deepened by developing genome-wide approaches that can evaluate the phenotypic consequences of a variety of loss- and gain-of-function alleles.

Data availability

All data and code is available from https://doi.org/10.6084/m9.figshare.27106078. Supplementary material is available.

Acknowledgments

I thank Jake Lusis for helpful comments on the manuscript.

Funding

This work was supported by the University of California Cancer Research Coordinating Committee grant C25CR8562.

Conflicts of interest

The author has no conflict of interest.

Literature cited

Ahn S, Wang RT, Park CC, Lin A, Leahy RM, Lange K, Smith DJ. 2009. Directed mammalian gene regulatory networks using expression and comparative genomic hybridization microarray data from radiation hybrids. PLoS Comput Biol. 5:e1000407.

Albert R. 2005. Scale-free networks in cell biology. J Cell Sci. 118:4947–57.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B-Methodological. pp. 289–300.

Coulomb S, Bauer M, Bernard D, Marsolier-Kergoat MC. 2005. Gene essentiality and the topology of protein interaction networks. Proc Biol Sci. 272:1721–5.

Cox DR, Burmeister M, Price ER, Kim S, Myers RM. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science. 250:245–250.

De Kegel B, Ryan CJ. 2019. Paralog buffering contributes to the variable essentiality of genes in cancer cell lines. PLoS Genet. 15:e1008466.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature. 489:57–74.

Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, Liu X, VanderSluis B, Oately B, Steinbach M *et al*. 2019. Discovering genetic interactions bridging pathways in genome-wide association studies. Nat Commun. 10:4274.

Fernandez-Rozadilla C, Timofeeva M, Chen Z, Law P, Thomas M, Schmit S, Díez-Obrero V, Hsu L, Fernandez-Tajes J, Palles C *et al*. 2023. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of european and east asian ancestries. Nat Genet. 55:89–99.

Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, Sisu C, Wright JC, Arnan C, Barnes I *et al*. 2023. Gencode: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res. 51:D942–D949.

Fujii T, Inoue N, Nobeyama T, Inoue J, Ogasawara S, Otani Y, Fujii S, Ito C, Fukunaga R. 2023. A simple, rapid, and efficient method for generating multigene-knockout culture cells by the CRISPR/Cas9 system. Genes Cells. 28:390–397.

Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL *et al*. 2023. The Gene Ontology knowledgebase in 2023. Genetics. 224:iyad031.

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC *et al*. 2014. Genome-scale CRISPR-mediated control of gene repression and activation. Cell. 159:647–661.

Goss SJ, Harris H. 1975. New method for mapping genes in human chromosomes. Nature. 255:680–684.

Grau J, Grosse I, Keilwagen J. 2015. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics. 31:2595–2597.

Greene D, Richardson S, Turro E. 2017. `ontologyX`: a suite of R packages for working with ontological data. Bioinformatics. 33:1104–1106.

Gustavsen JA, Pai S, Isserlin R, Demchak B, Pico AR. 2019. `RCy3`: Network biology using `Cytoscape` from within R. F1000Research. 8:1774.

Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 22:803–6.

Horlbeck MA, Xu A, Wang M, Bennett NK, Park CY, Bogdanoff D, Adamson B, Chow ED, Kampmann M, Peterson TR *et al*. 2018. Mapping the genetic landscape of human cells. Cell. 174:953–967.e22.

Khan AH, Bloom JS, Faridmoayer E, Smith DJ. 2016. Genetic screening reveals a link between Wnt signaling and antitubulin drugs. Pharmacogenomics J. 16:164–172.

Khan AH, Lin A, Wang RT, Bloom JS, Lange K, Smith DJ. 2020. Pooled analysis of radiation hybrids identifies loci for growth and drug action in mammalian cells. Genome Res. 30:1458–1467.

Khan AH, Smith DJ. 2021. Cost-effective mapping of genetic interactions in mammalian cells. Front Genet. 12:703738.

Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, Usaj M, Balint A, Mattiazzi Usaj M, van Leeuwen J *et al*. 2018. Systematic analysis of complex genetic interactions. Science. 360:eaao1729.

Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, Farrington S, Svinti V, Palles C, Orlando G *et al*. 2019. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun. 10:2154.

Liau WS, Tan SH, Ngoc PCT, Wang CQ, Tergaonkar V, Feng H, Gong Z, Osato M, Look AT, Sanda T. 2017. Aberrant activation of the GIMAP enhancer by oncogenic transcription factors in T-cell acute lymphoblastic leukemia. Leukemia. 31:1798–1807.

Lin A, Smith DJ. 2015. Copy number networks to guide combinatorial therapy of cancer and proliferative disorders. In: Tran QN, Arabnia HR, editors, *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology - Algorithms and Software Tools*. pp. 401–419. Elsevier Morgan Kaufmann.

Lin A, Wang RT, Ahn S, Park CC, Smith DJ. 2010. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. Genome Res. 20:1122–1132.

Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F *et al*. 2021. Genome-wide enhancer maps link risk variants to disease genes. Nature. 593:238–243.

Park CC, Ahn S, Bloom JS, Lin A, Wang RT, Wu T, Sekar A, Khan AH, Farr CJ, Lusis AJ *et al*. 2008. Fine mapping of regulatory loci for mammalian gene expression using radiation hybrids. Nat Genet. 40:421–429.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Riesenberg S, Chintalapati M, Macak D, Kanis P, Maricic T, Pääbo S. 2019. Simultaneous precise editing of multiple genes in human cells. Nucleic Acids Res. 47:e116.

Ritchie MD, Van Steen K. 2018. The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. Ann Transl Med. 6:157.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. 2011. `pROC`: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 12:77.

Sachs MC. 2017. `plotROC`: a tool for plotting ROC curves. J Stat Softw. 79:2.

Saito T, Rehmsmeier M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 10:e0118432.

Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M *et al*. 2021. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 53:1415–1424.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. `Cytoscape`: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504.

Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J *et al*. 2023. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 51:D977–D985.

Sun L, Wang Z, Lu T, Manolio TA, Paterson AD. 2023. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? Am J Hum Genet. 110:903–912.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*. 2019. String v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47:D607–D613.

Tuszynski J. 2024. *caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc*. R package version 1.18.3.

VanderSluis B, Costanzo M, Billmann M, Ward HN, Myers CL, Andrews BJ, Boone C. 2018. Integrating genetic and protein-protein interaction networks maps a functional wiring diagram of a cell. Curr Opin Microbiol. 45:170–179.

Wang RT, Ahn S, Park CC, Khan AH, Lange K, Smith DJ. 2011. Effects of genome-wide copy number variation on expression in mammalian cells. BMC Genomics. 12:562.

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015. Identification and characterization of essential genes in the human genome. Science. 350:1096–101.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S *et al*. 2022. *gplots: Various R programming tools for plotting data*. R package version 3.1.3.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilke CO. 2020. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.1.1.

Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M, Eliasen AU, Jiang Y, Raghavan S *et al*. 2022. A saturated map of common genetic variants associated with human height. Nature. 610:704–712.

Supplementary material: Complementary human gene interaction maps from radiation hybrids and CRISPRi

Desmond J. Smith*,[1]

[1]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA, Los Angeles, USA

*Corresponding author: DSmith@mednet.ucla.edu

Table S1 CRISPRi and STRING network overlaps

Comparison	GI	STRING	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	*P*	FDR
K562 Buffer	3	150	279	112	398	15835	81224	3.6	7.8E-53	4.7E-49
K562 Synergistic	-3	150	540	252	983	15777	81335	2.8	1.4E-72	1.4E-68
K562 All	$\|\pm 3\|$	150	819	379	1381	16794	83334	2.9	1.3E-113	1.3E-109
Jurkat Buffer	2	150	289	201	890	15704	77149	1.6	4.7E-11	2.0E-07
Jurkat Synergistic	-2	420	158	66	1025	5102	87663	2.6	9.4E-24	7.7E-20
Jurkat All	$\|\pm 2\|$	150	626	410	1736	16338	78928	1.7	3.8E-29	3.9E-25

Notes: Table shows optimal overlaps. GI, CRISPRi genetic interaction threshold; STRING, STRING threshold; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, CRISPRi specific interactions; n_{10}, STRING specific; Non-olap, neither CRISPRi nor STRING interactions; OR, odds ratio; FDR, false discovery rate; All, combined buffering and synergistic interactions.

Table S2 RH-PCR and STRING network overlaps

Comparison	$-\log_{10}P$	STRING	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	*P*	FDR
RH-PCR Ess	2	150	400	285	1286	17229	85528	1.5	4.6E-13	4.7E-10
RH-PCR ds	1.5	160	3537	1738	62583	5239410	194184794	2.1	1.5E-323	7.4E-321
RH-PCR chrX	3	170	3646	1822	83528	342478	16131323	2.1	2.1E-319	3.7E-316
RH-PCR chrY	1.5	160	91	36	5527	7874	1246547	2.6	3.8E-15	6.8E-12
RH-PCR All	4	770	4299	2255	1190102	372259	197923665	1.9	8.4E-323	4.3E-319

Notes: Table shows optimal overlaps. $-\log_{10} P$, RH-PCR significance threshold; STRING, STRING threshold score; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, RH-PCR specific interactions; n_{10}, STRING specific; Non-olap, neither RH-PCR nor STRING interactions; OR, odds ratio; FDR, false discovery rate; RH-PCR Ess, RH-PCR network restricted to essential genes; RH-PCR ds, RH-PCR network randomly downsampled to same node number as essential genes; RH-PCR chrX, RH-PCR network downsampled so that each interaction has at least one gene on the X but not Y chromosome; RH-PCR chrY, RH-PCR network downsampled so that each interaction has at least one gene on the Y but not X chromosome; RH-PCR All, protein coding RH-PCR network.

Table S3 Essential RH-PCR network overlaps and CRISPRi

Comparison	$-\log_{10}P$	GI	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	P	FDR
K562 Buffer	4	4.45	3	1	796	61	96876	6.0	1.6E-02	1.0
K562 Synergistic	2	-3.65	2	12	1595	768	96270	0.2	7.6E-04	0.7
K562 All	1	$\lvert\pm5\rvert$	1	5	1778	261	98350	0.2	0.1	1.0
Jurkat Buffer	5	2.55	3	1	293	430	93306	2.2	0.2	1.0
Jurkat Synergistic	1	-2.15	8	17	1654	940	91346	0.5	2.5E-02	1.0
Jurkat All	2	$\lvert\pm3\rvert$	3	7	1549	452	93717	0.4	0.1	1.0

Notes: Table shows optimal overlaps. $-\log_{10} P$, RH-PCR significance threshold; GI, CRISPRi genetic interaction threshold; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, RH-PCR specific interactions; n_{10}, CRISPRi specific; Non-olap, neither RH-PCR nor CRISPRi interactions; OR, odds ratio; FDR, false discovery rate; All, combined buffering and synergistic interactions.

Table S4 RH-PCR and CRISPRi overlaps of interacting genes

Comparison	$-\log_{10}P$	GI	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	P	FDR
K562 Buffer	7	2.9	202	151	8778	134	10828	1.9	2.6E-08	2.7E-05
K562 Synergistic	7	-2.8	231	178	8749	164	10798	1.7	7.0E-08	5.6E-05
K562 All	7	$\lvert\pm2.8\rvert$	245	188	8732	173	10792	1.8	2.2E-08	1.3E-05
Jurkat Buffer	5	2	242	192	10556	113	9031	1.8	6.5E-08	1.7E-05
Jurkat Synergistic	6	-2.2	214	167	10072	109	9547	1.9	8.2E-08	4.0E-05
Jurkat All	5	$\lvert\pm2.1\rvert$	261	209	10519	125	9037	1.8	5.2E-08	1.4E-05

Notes: Table shows optimal overlaps. $-\log_{10} P$, RH-PCR significance threshold; GI, CRISPRi genetic interaction threshold; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, RH-PCR specific interactions; n_{10}, CRISPRi specific; Non-olap, neither RH-PCR nor CRISPRi interactions; OR, odds ratio; FDR, false discovery rate; All, combined buffering and synergistic interactions.

Table S5 RH-PCR and GWAS network overlaps

Comparison	$-\log_{10}P$	GWAS	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	P	FDR
RH-PCR chrX	1.5	7	78	30	148846	3295	16408756	2.6	3.2E-13	4.7E-10
RH-PCR Essential	1.5	5	210	114	1629	6260	96344	2.0	2.1E-17	1.9E-14
RH-PCR Ess + 1	1.5	7	315	20	139866	11751	84185646	16.1	9.7E-253	4.5E-250
RH-PCR All	4	9	6021	3525	1188380	582794	197713130	1.7	1.1E-321	7.2E-319

Notes: Table shows optimal overlaps. $-\log_{10} P$, RH-PCR significance value; GWAS, genome-wide association study $-\log_{10} P$; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, RH-PCR specific interactions; n_{10}, GWAS specific; Non-olap, neither RH-PCR nor GWAS interactions; OR, odds ratio; FDR, false discovery rate; RH-PCR chrX, RH-PCR network downsampled so that each interaction has at least one gene on the X but not Y chromosome; RH-PCR Essential, RH-PCR network restricted to essential genes; Ess + 1, essential RH-PCR network plus genes linked by one interaction; RH-PCR All, protein coding RH-PCR network.

Table S6 CRISPRi and GWAS network overlaps

Comparison	GI	GWAS	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	*P*	FDR
K562 Buffer	2.55	5	65	90	1410	5881	90380	0.7	6.0E-03	1.0
K562 Synergistic	-8.7	11	5	0.7	16	3317	95297	9.0	5.6E-04	8.1E-02
K562 All	$\lvert\pm 8.8\rvert$	11	5	0.7	16	3368	96760	9.0	5.6E-04	0.1
Jurkat Buffer	2.2	47	8	3	799	308	92917	3.0	6.4E-03	1.0
Jurkat Synergistic	-5.8	5	6	2	21	6001	87920	4.2	6.3E-03	0.5
Jurkat All	$\lvert\pm 5.8\rvert$	5	6	2	24	6091	89175	3.7	1.1E-02	0.5

Notes: Table shows optimal overlaps. GI, CRISPRi genetic interaction threshold; GWAS, genome-wide association study $-\log_{10} P$ threshold; Obs, observed interaction overlap; Exp, expected overlap; n_{01}, CRISPRi specific interactions; n_{10}, GWAS specific; Non-olap, neither CRISPRi nor GWAS interactions, OR, odds ratio; FDR, false discovery rate; All, combined buffering and synergistic interactions.

Table S7 GWAS overlaps based on RH-PCR network

Trait	Superset	Subset	Obs	Exp	n_{01}	n_{10}	Non-olap	OR	*P*
Colorectal cancer	36539618	31089142	123	80	2712281	5762	196772159	1.5	6.2E-06
Height	36224396	34594039	25021	14054	2687383	1008645	195769276	1.8	2.2E-308

Notes: Superset, PubMed ID of larger GWAS study; Subset, PubMed ID of smaller GWAS study, both studies thresholded at $P = 5.0 \times 10^{-8}$; Obs, observed RH-PCR links from subset to superset; Exp, expected links; n_{01}, RH-PCR interactions not included in superset-subset links; n_{10}, superset-subset links not included in RH-PCR interactions; Non-olap, links in neither RH-PCR nor GWASs; OR, odds ratio; FDR, false discovery rate.

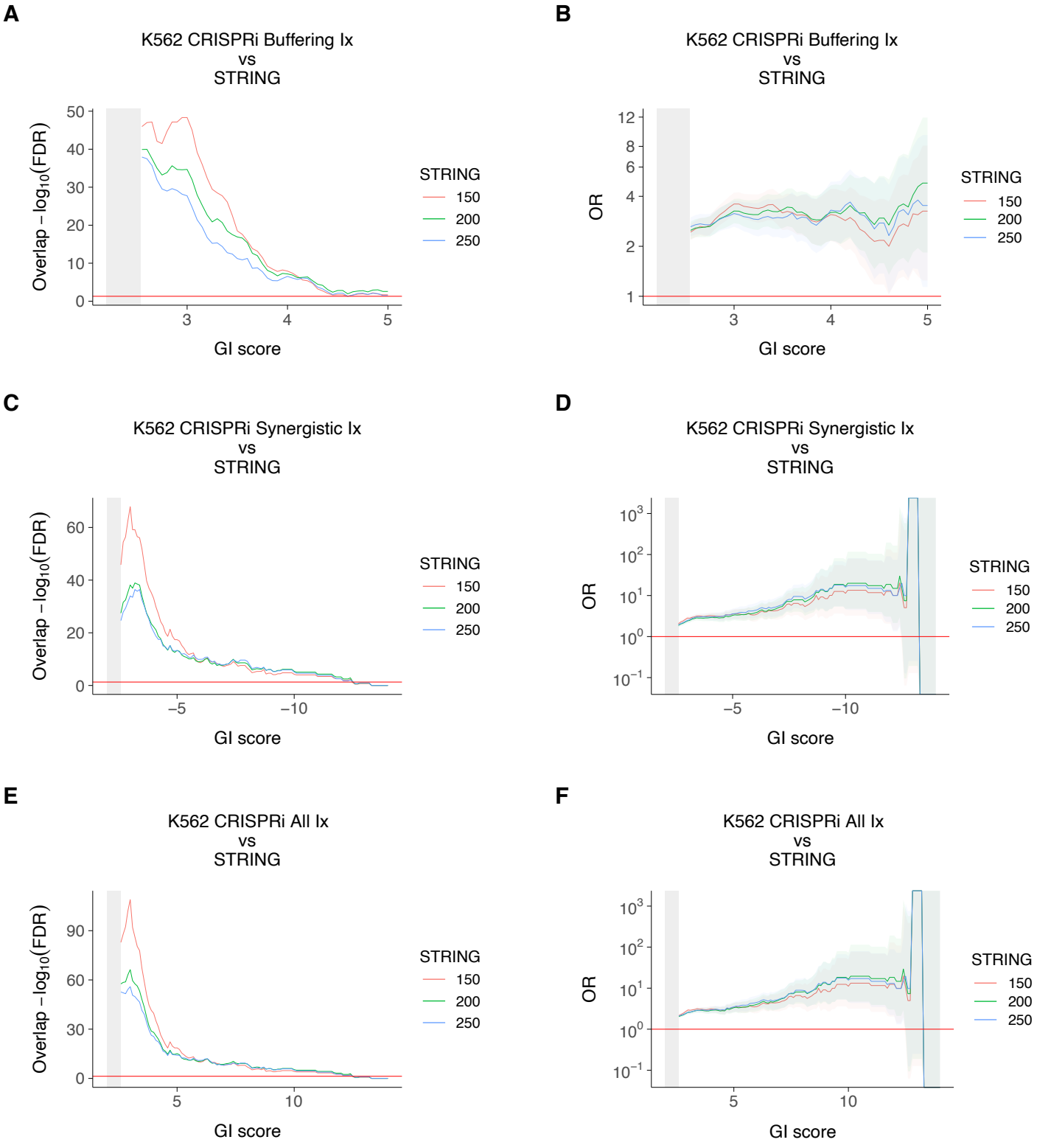

Figure S1 CRISPRi and STRING networks in K562 cells. **A.** Overlap of K562 buffering CRISPRi interactions (Ix) and STRING. **B.** Odds ratios (ORs) for overlap of K562 buffering CRISPRi interactions and STRING. Confidence intervals shown (95%). **C.** Overlap for K562 synergistic CRISPRi interactions and STRING. **D.** ORs for K562 synergistic CRISPRi interactions and STRING. **E.** Overlap of K562 combined buffering and synergistic CRISPRi interactions and STRING. **F.** ORs for combined K562 CRISPRi interactions and STRING. FDR, false discovery rate. Gray rectangles at left of each graph indicate insignificant CRISPRi genetic interaction (GI) scores. ORs on a $\log_{10}$ scale.

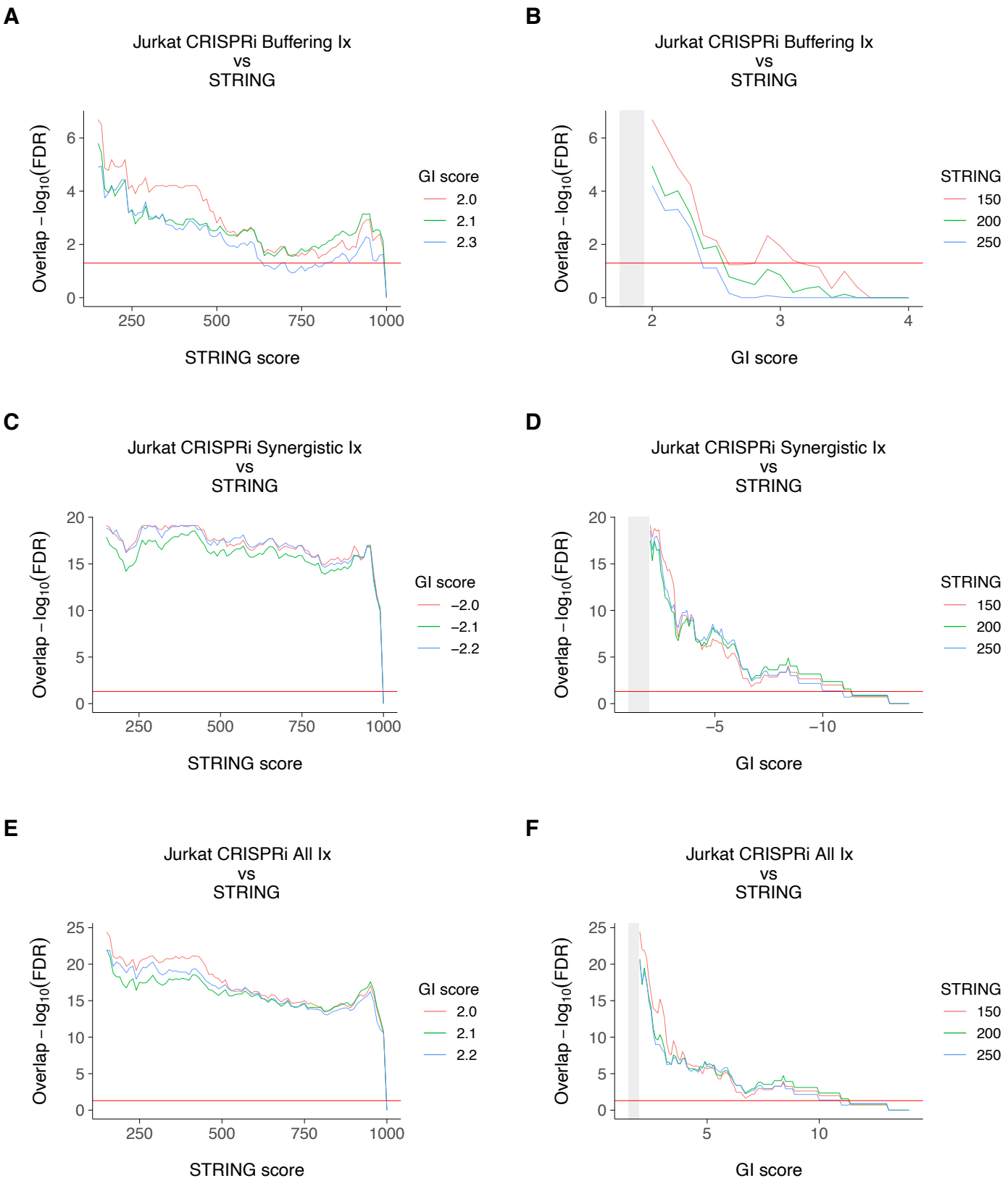

Figure S2 CRISPRi and STRING networks in Jurkat cells. **A,B.** Overlap of Jurkat buffering CRISPRi interactions and STRING. **C,D.** Overlap of Jurkat synergistic CRISPRi interactions and STRING. **E,F.** Overlap of Jurkat combined buffering and synergistic CRISPRi interactions and STRING. FDR, false discovery rate. Gray rectangles at left of graphs in B, D and F indicate insignificant CRISPRi genetic interaction (GI) scores.

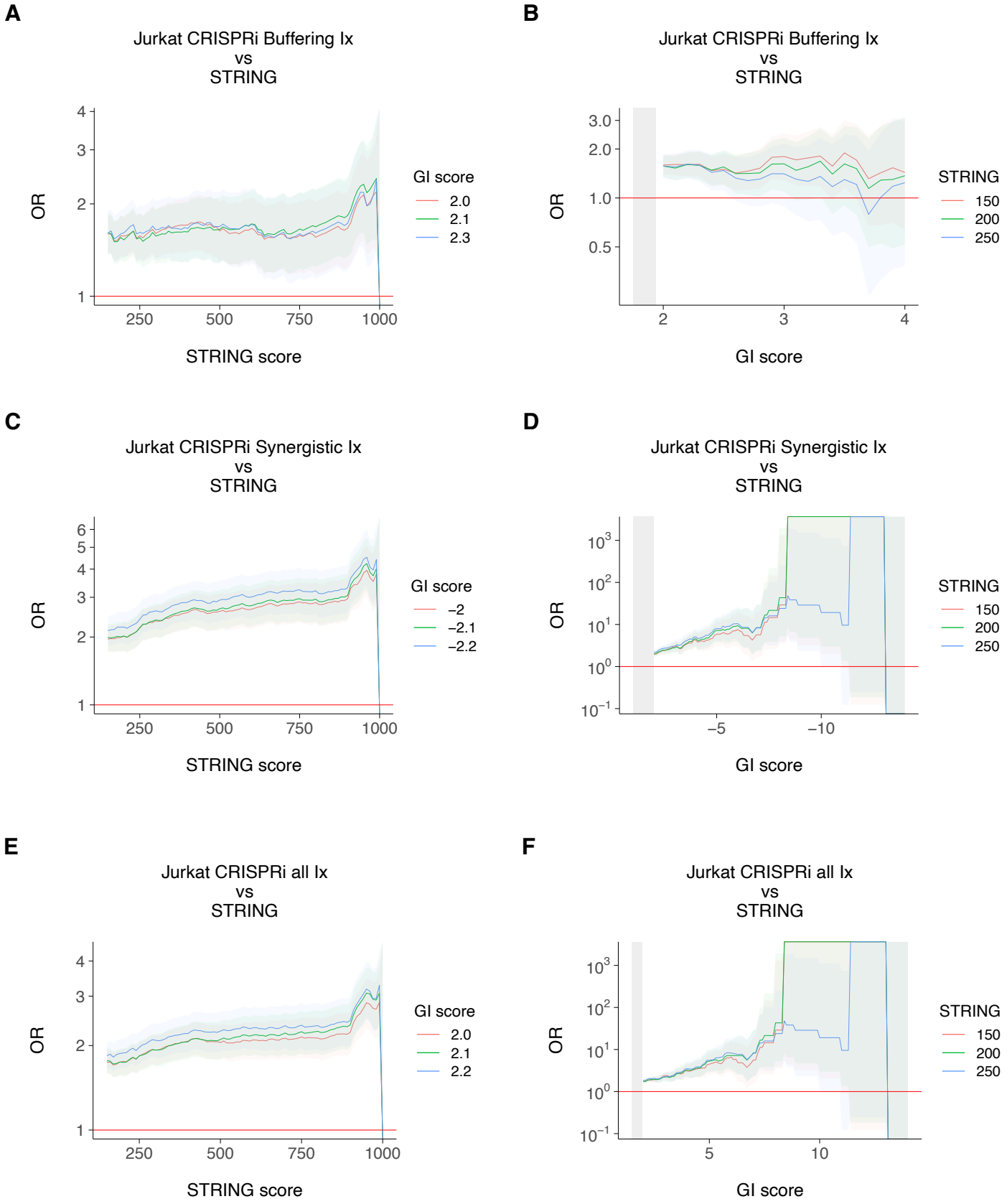

Figure S3 CRISPRi and STRING network odds ratios in Jurkat cells. **A,B.** Odds ratios (ORs) for overlap of Jurkat buffering CRISPRi interactions and STRING. **C,D.** ORs for overlap of Jurkat synergistic CRISPRi interactions and STRING. **E,F.** ORs for overlap of Jurkat combined buffering and synergistic CRISPRi interactions and STRING. FDR, false discovery rate. Gray rectangles at left of graphs in B, D and F indicate insignificant CRISPRi genetic interaction (GI) scores. ORs on a $\log_{10}$ scale.

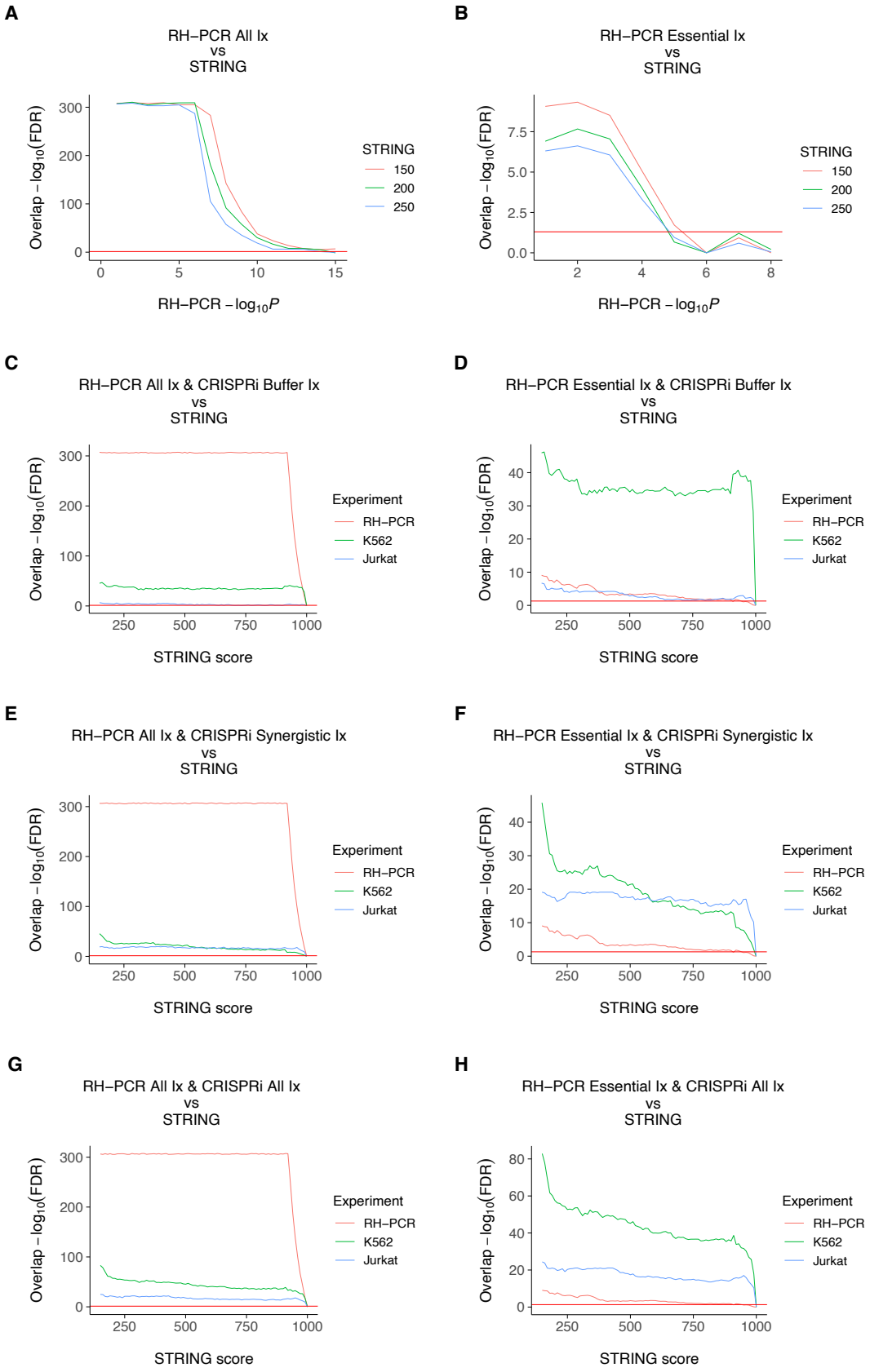

Figure S4 Overlaps of RH-PCR, CRISPRi and STRING networks. **A.** Overlap of protein coding RH-PCR network and STRING. **B.** Overlap of RH-PCR interactions restricted to CRISPRi essential genes and STRING. **C.** Overlap of protein coding RH-PCR, buffering CRISPRi, and STRING networks. **D.** Overlap of essential RH-PCR, buffering CRISPRi, and STRING networks. **E.** Overlap of protein coding RH-PCR, synergistic CRISPRi, and STRING networks. **F.** Overlap of essential RH-PCR, synergistic CRISPRi, and STRING networks. **G.** Overlap of protein coding RH-PCR, combined buffering and synergistic CRISPRi, and STRING networks. **H.** Overlap of essential RH-PCR, combined CRISPRi, and STRING networks. FDR, false discovery rate.

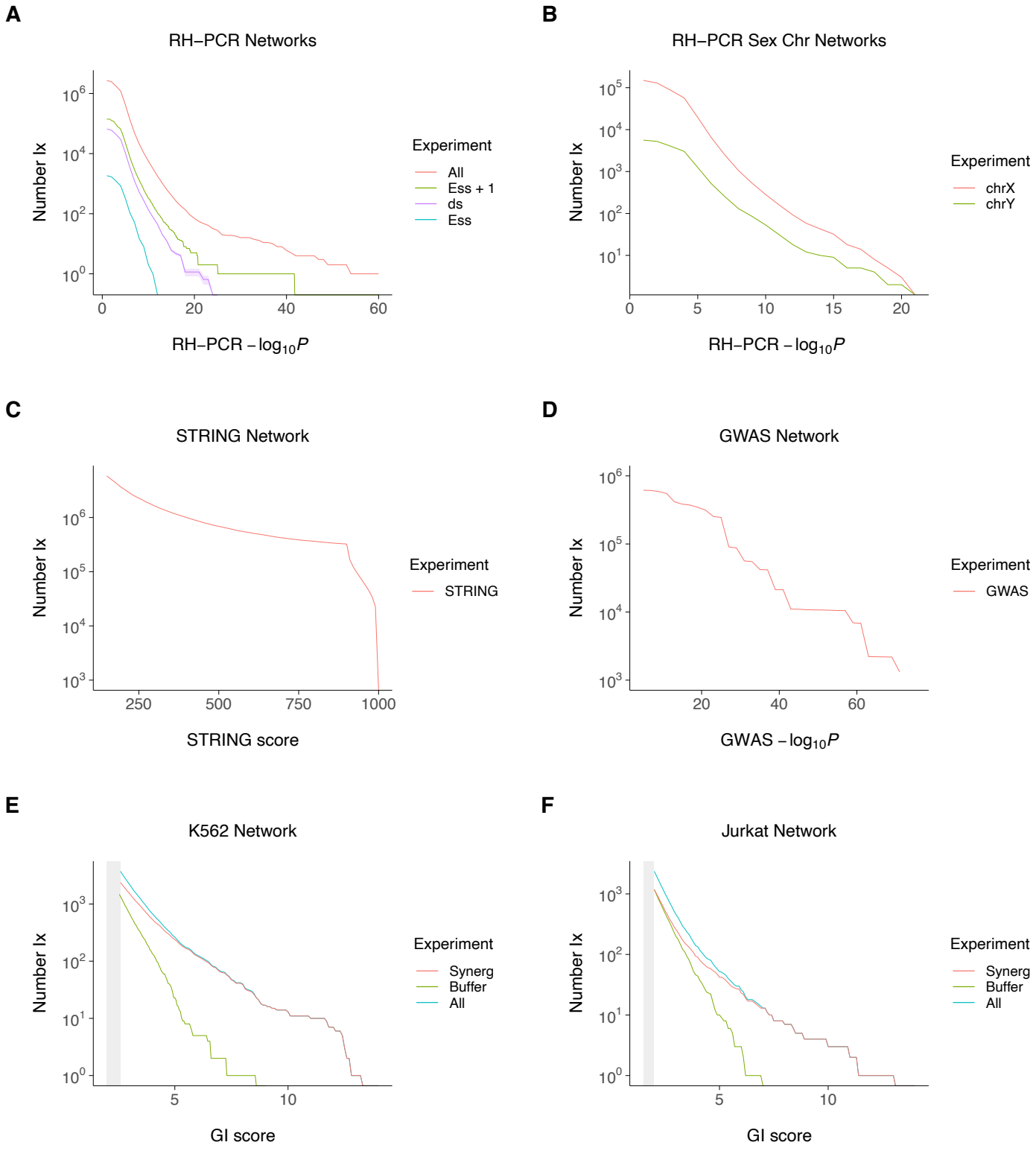

Figure S5 Network numbers. **A.** RH-PCR networks. All, protein-coding RH-PCR network. Ess + 1, essential RH-PCR network plus genes linked by one interaction. ds, downsampled RH-PCR network, 95% confidence limits shown. Ess, essential RH-PCR network. **B.** Protein coding RH-PCR network downsampled so that at least one gene in each interaction is on either the X or Y chromosome. **C.** STRING network. **D.** Geneome-wide association study (GWAS) network. **E.** K562 CRISPRi network. Synerg, synergistic interactions. Buffer, buffering interactions. All, combined synergistic and buffering interactions. **F.** Jurkat CRISPRi network. Gray rectangles at left of graphs in E and F indicate insignificant CRISPRi genetic interaction (GI) scores. Number Ix, number of interactions in each network as stringency increased.

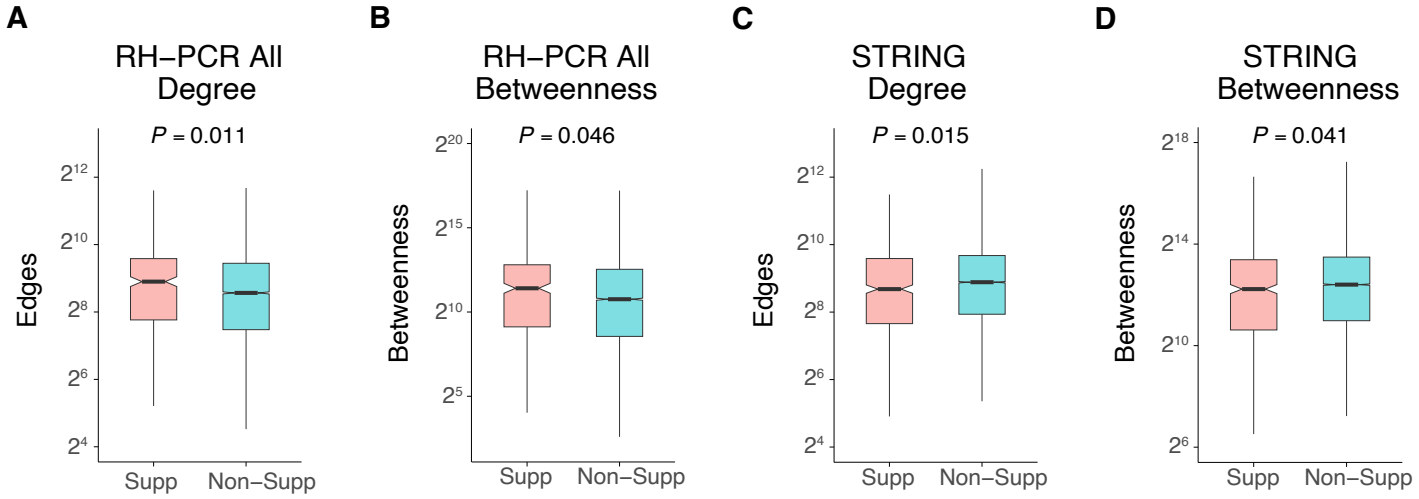

Figure S6 Centrality for growth suppressing genes in protein coding RH-PCR and STRING networks. **A.** Degree centrality for RH-PCR network, $t[1,415] = 2.6$, $P = 0.011$. **B.** Betweenness centrality for RH-PCR network, $t[1,408] = 2.0$, $P = 0.046$. **C.** Degree centrality for STRING network. $t[1,649] = 2.5$, $P = 0.015$. **D.** Betweenness centrality for STRING network. $t[1,711] = 2.0$, $P = 0.041$. Supp, genes that suppress growth. Non-Supp, non-suppressing. Welch Two Sample t-tests used.

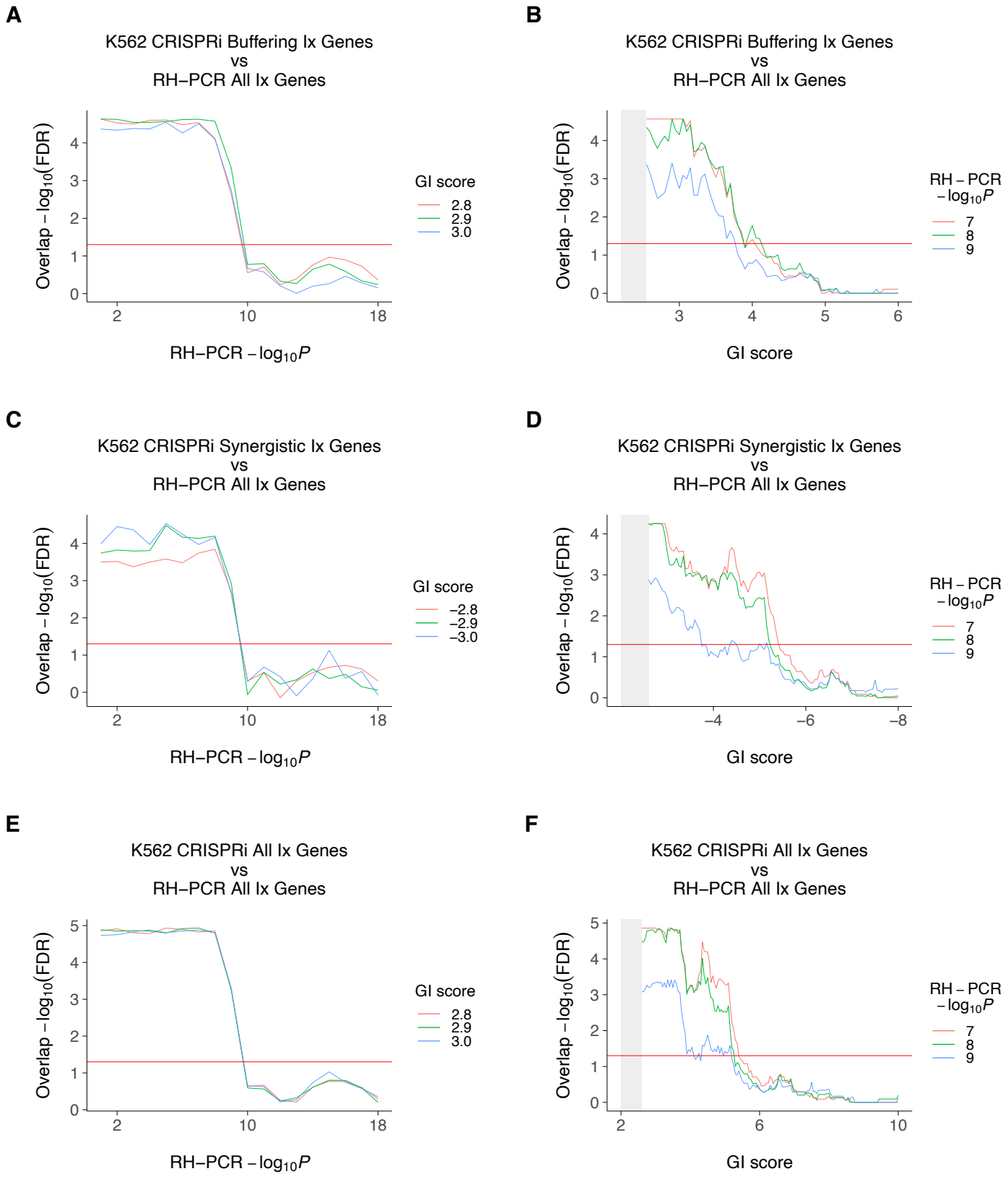

Figure S7 Genes in RH-PCR and K562 CRISPRi networks. **A,B.** Overlap of genes in K562 buffering CRISPRi and protein coding RH-PCR networks. **C,D.** Overlap of genes in K562 synergistic CRISPRi and protein coding RH-PCR networks. **E,F.** Overlap of genes in K562 combined buffering and synergistic CRISPRi and protein coding RH-PCR networks. FDR, false discovery rate. Gray rectangles at left of graphs in B, D and F indicate insignificant CRISPRi genetic interaction (GI) scores.

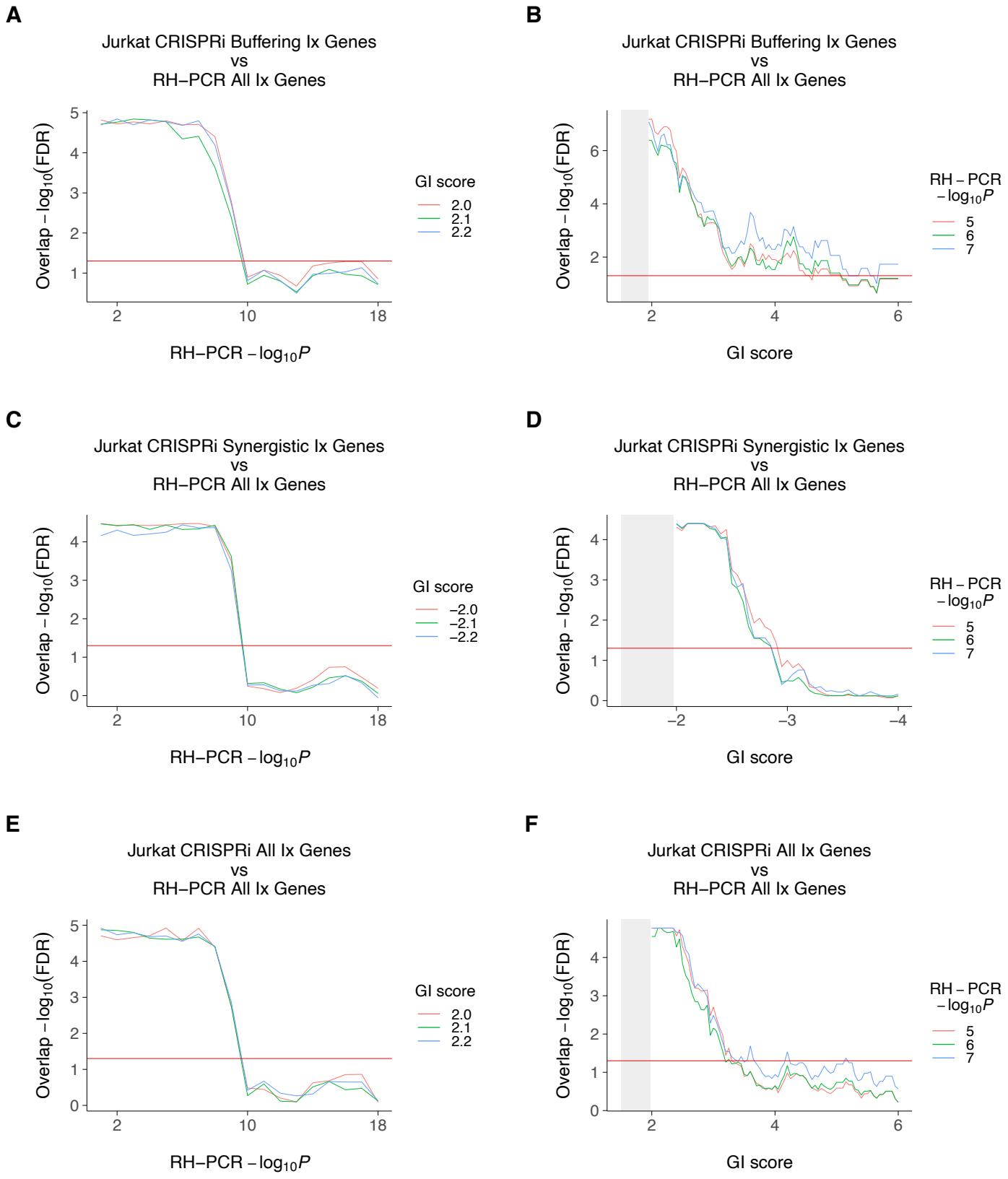

Figure S8 Genes in RH-PCR and Jurkat CRISPRi networks. **A,B.** Overlap of genes in Jurkat buffering CRISPRi and protein coding RH-PCR networks. **C,D.** Overlap of genes in Jurkat synergistic CRISPRi and protein coding RH-PCR networks. **E,F.** Overlap of genes in Jurkat combined buffering and synergistic CRISPRi and protein coding RH-PCR networks. FDR, false discovery rate. Gray rectangles at left of graphs in B, D and F indicate insignificant CRISPRi genetic interaction (GI) scores.

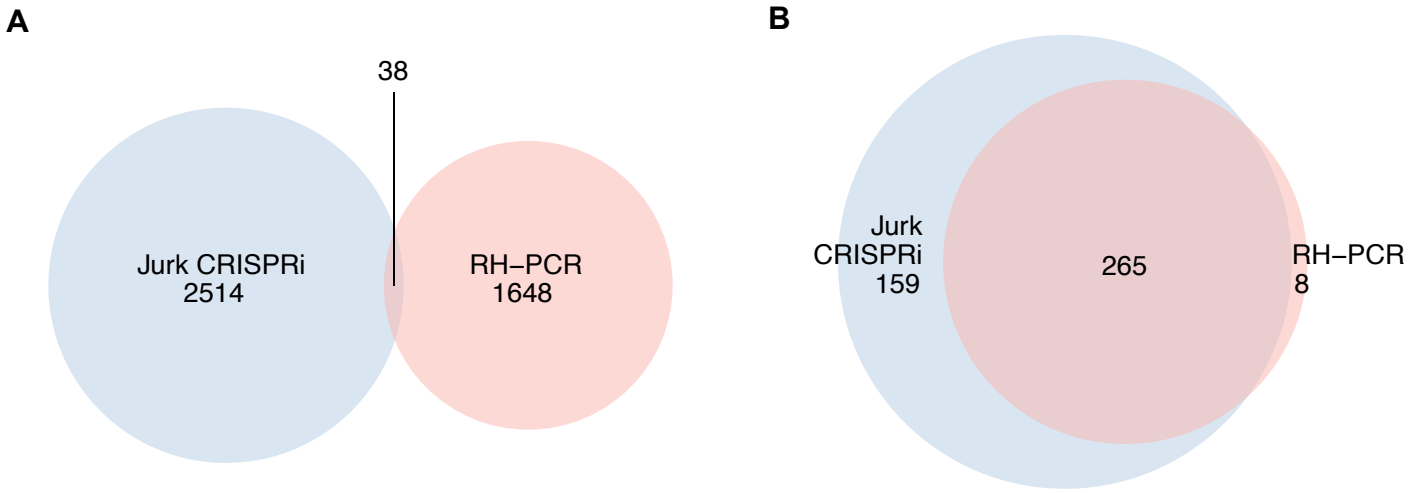

Figure S9 Interactions and genes in RH-PCR and Jurkat CRISPRi networks. **A.** Overlap of significant interactions in the essential RH-PCR and Jurkat combined buffering and synergistic CRISPRi networks. **B.** Overlap of genes participating in significant interactions in the two networks.