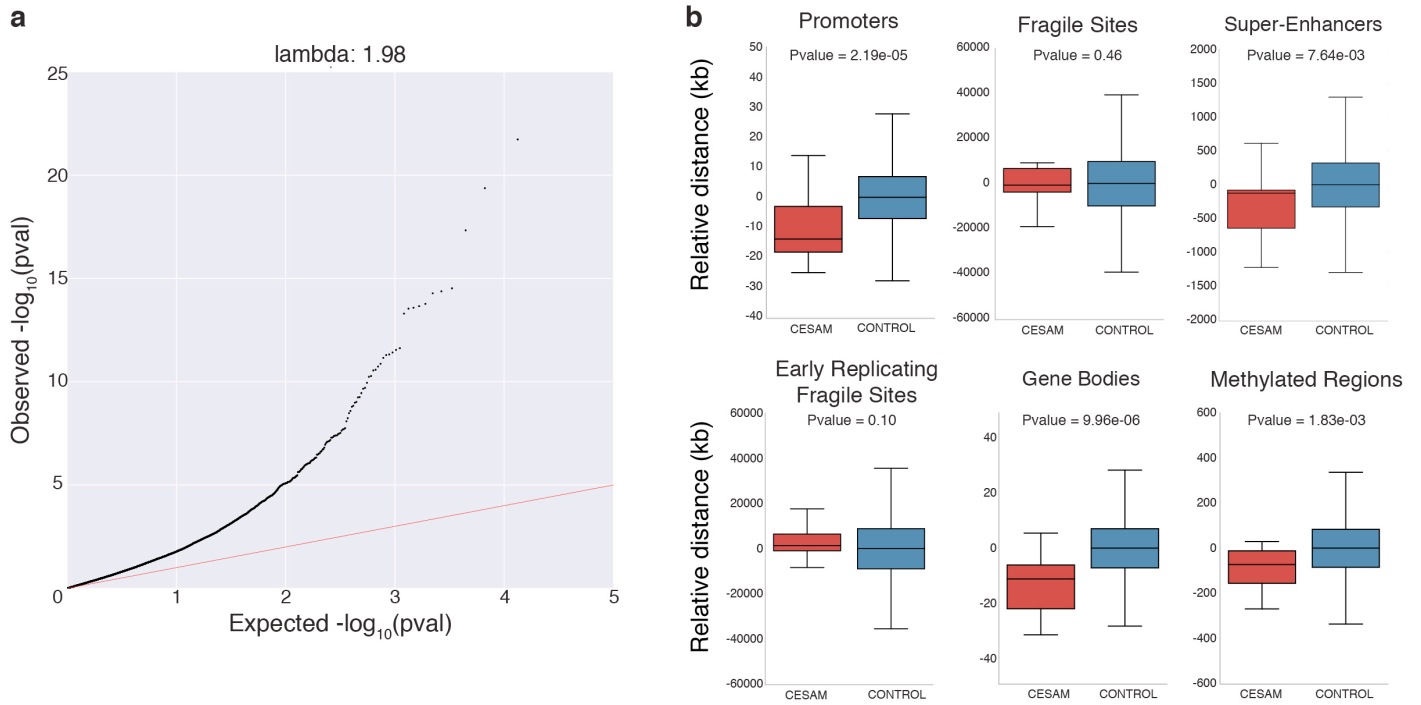


Supplementary Figure 1

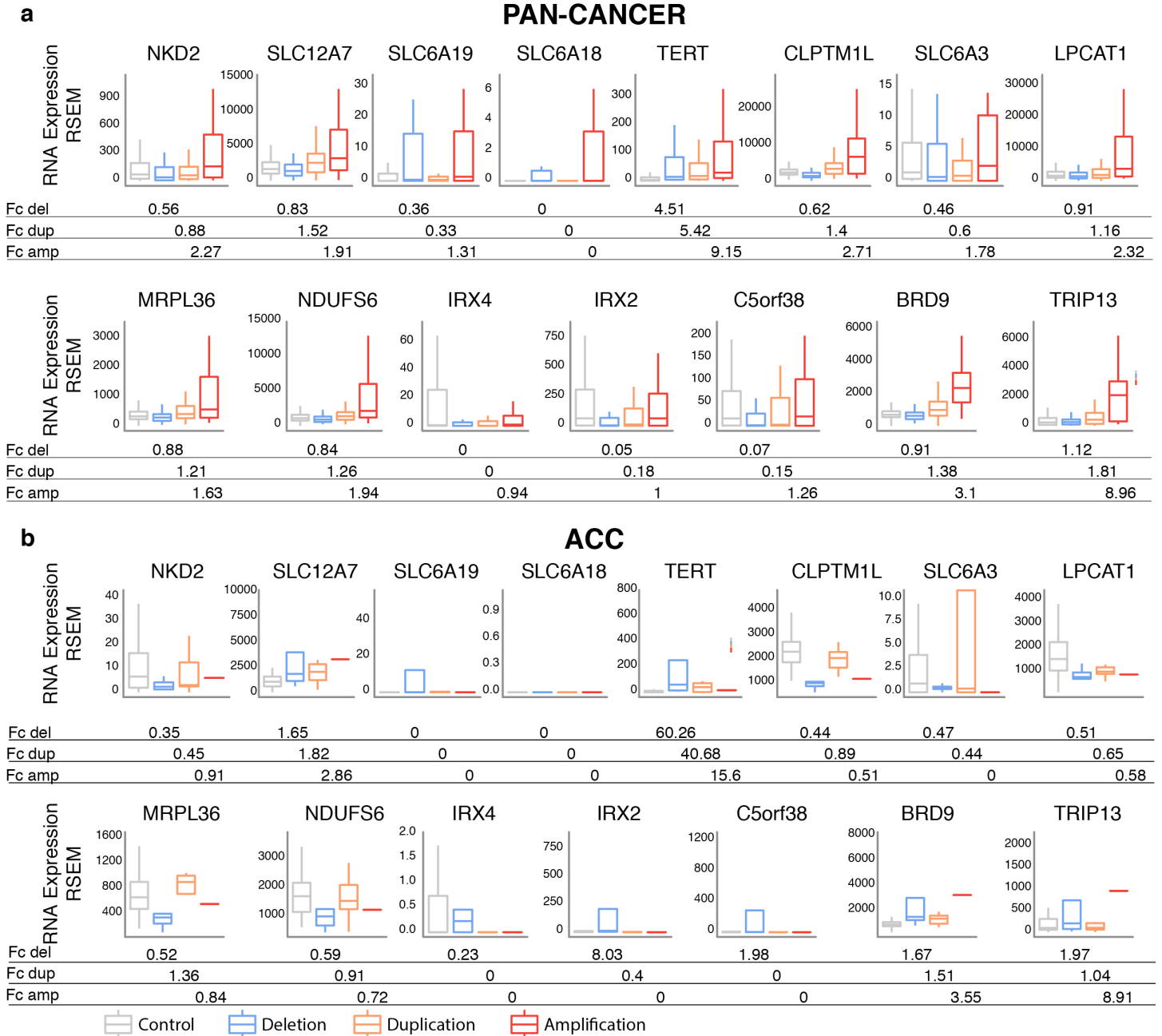


Supplementary Figure 1

p-value distribution and relative distance to genomic features of CESAM hits

(a) QQ plot depicting observed P -values ($-\log_{10}$) in comparison to expected P -values ($-\log_{10}$). (b) Shadow figure to **Figure 1d**. Relative proximity to genetic elements compared to background (mann-whitney U test) for CESAM hits ('CESAM') versus 'CONTROL'.

Supplementary Figure 2

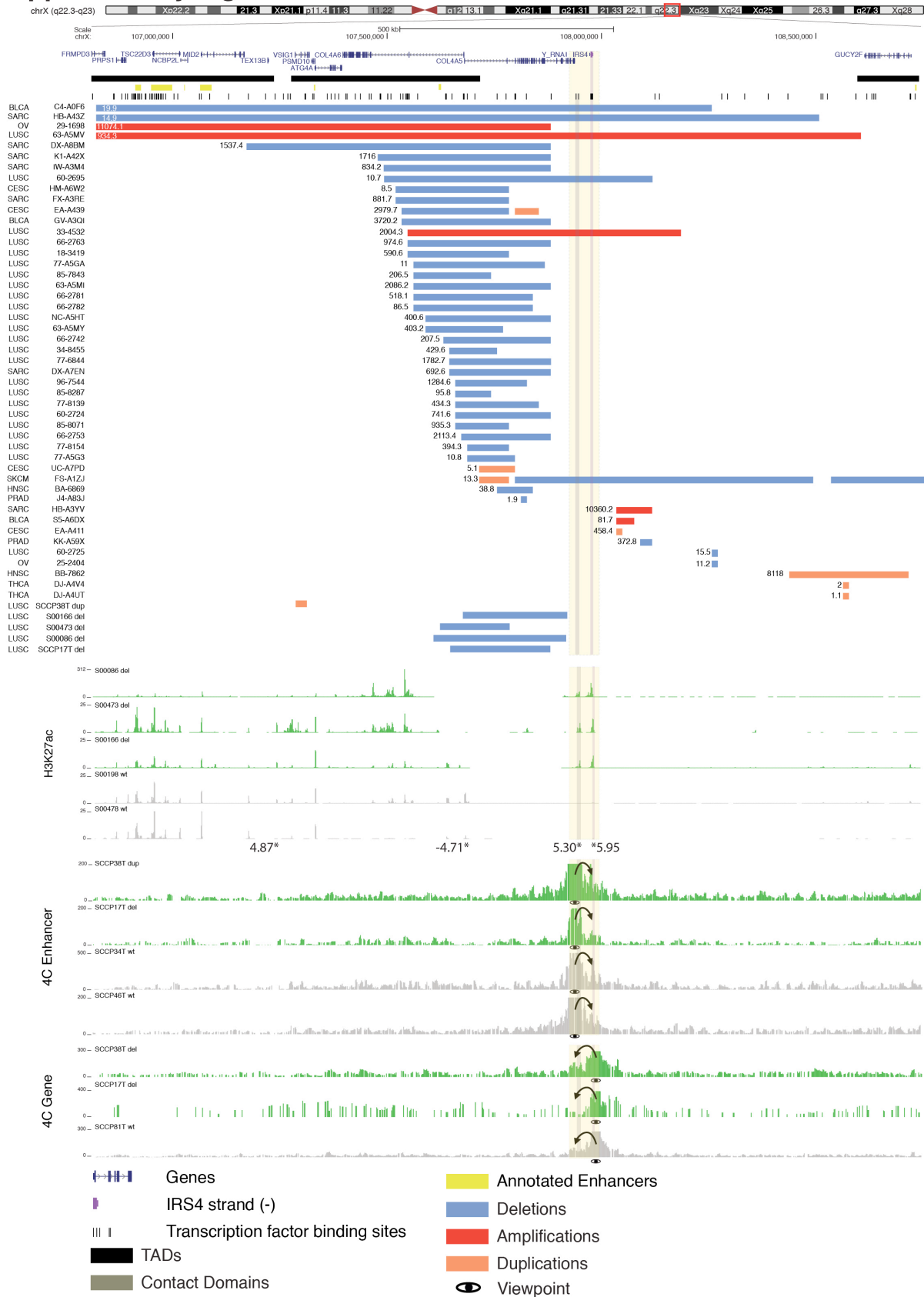


Supplementary Figure 2

Pan-cancer and ACC-specific CESAM analysis of gene expression changes for *TERT* locus-proximal genes

Related to main **Figure 2**. (a) Pan-cancer CESAM analysis of expression for genes in the vicinity of *TERT*, which demonstrates *TERT* as plausible target. (b) ACC-specific CESAM analysis of expression for genes in the vicinity of *TERT*, which demonstrates *TERT* as plausible target. For tumor type abbreviations, see **Supplementary Table 6**.

Supplementary Figure 3

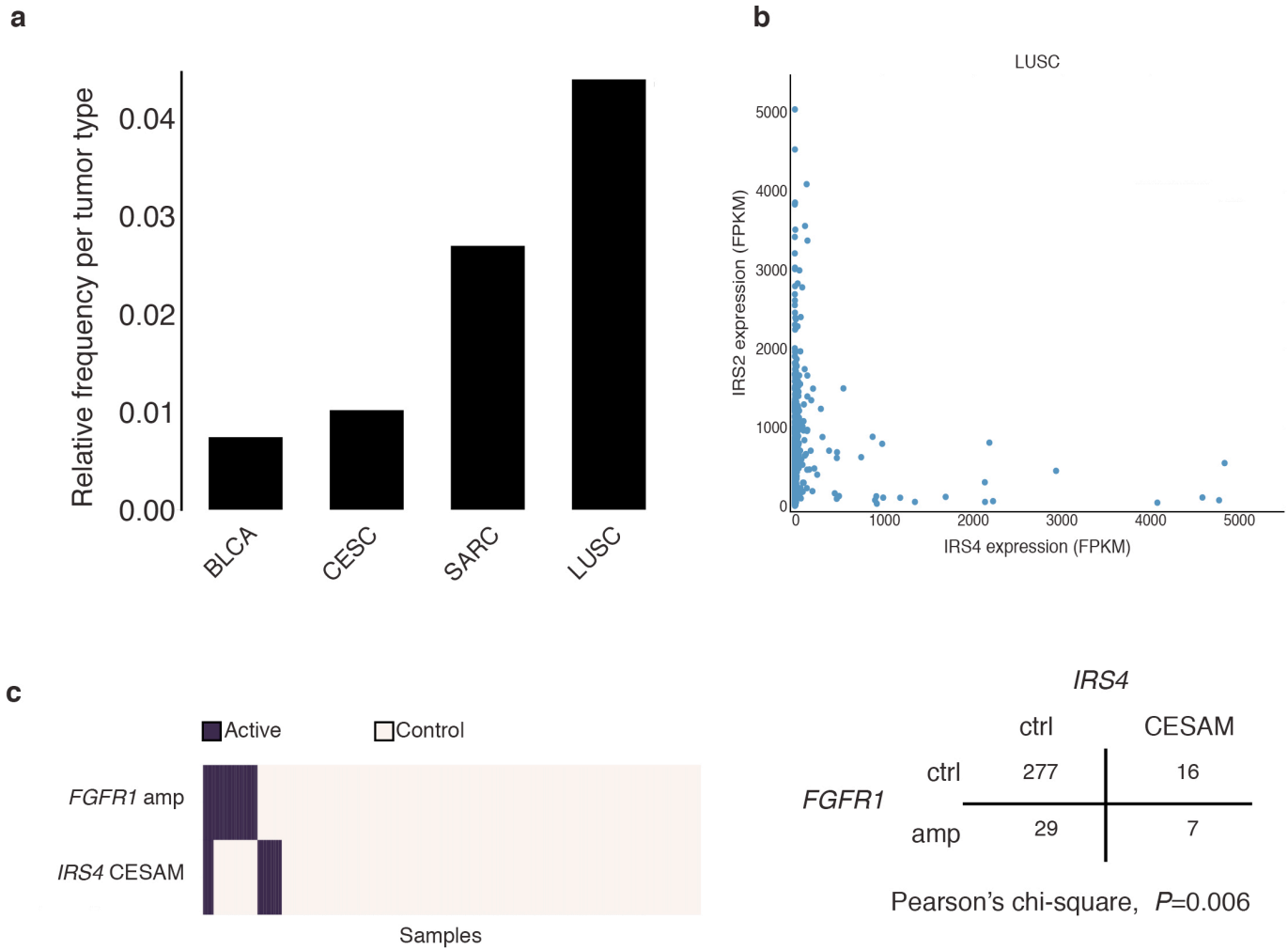


Supplementary Figure 3

Recurrent deletions at a TAD boundary on chromosome X associate with *IRS4* dysregulation in different cancer types

Related to main **Figure 3**. A genomic region near *IRS4* (a gene transcribed from the (-) strand of the reference genome), highlighted in yellow, exhibiting clustered transcription factor binding sites is highlighted in gray. Each SCNA is accompanied with the depicted expression fold change of *IRS4* (shown adjacent to each SCNA). Samples harboring the *cis* deletion, but not samples without the *cis* deletion, exhibit H3K27ac chromatin marks at *IRS4* as well as at the candidate CRE nearby (highlighted in grey; asterisks indicate significant differential H3K27ac marks between carriers and non-carriers). 4C-Seq experiments using the candidate CRE as a viewpoint clearly demonstrate physical interaction with *IRS4*; a similar intensity in interaction was observed in samples harboring the *cis* deletion as in samples lacking the deletion. And also 4C-Seq experiments using the gene as a viewpoint show interaction between *IRS4* and the putative CRE in both deletion carriers and non-carriers. LUSC samples S00086, S00473 and S00166 were verified, using semi-quantitative RT-PCR, to exhibit high-level upregulation of *IRS4*, whereas S00478 and S00198 showed merely baseline-level expression (data not shown). For all remaining carrier and non-carrier samples, we verified outlier and non-outlier expression, respectively, by qPCR (**Supplementary Table 3**).

Supplementary Figure 4

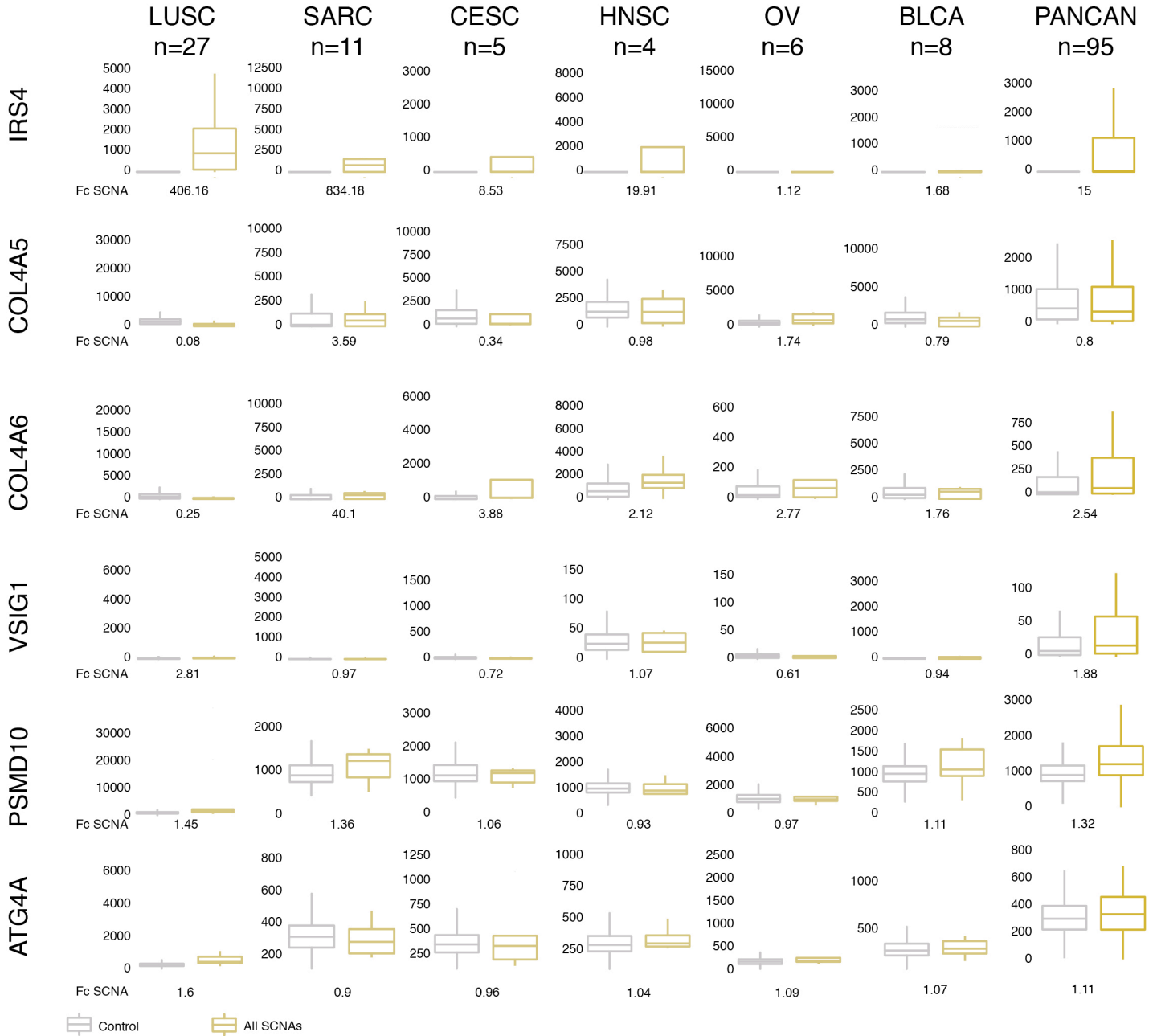


Supplementary Figure 4

IRS4* CESAM pan-cancer frequency and association with *IRS2* and *FGFR1

Related to main **Figure 3**. **(a)** Fraction of donors per cancer type for which CESAM inferred *IRS4* dysregulation in conjunction with SCNAs in *cis* in at least 3 donors. For tumor type abbreviations, see **Supplementary Table 6**. **(b)** *IRS4-IRS2* gene expression correlation plot, showing significantly anticorrelated expression ($P=0.008$, Pearson correlation; $r=-0.11$) **(c)** Co-occurrence heatmap of samples exhibiting *IRS4* CESAM hits versus *FGFR1* amplification in LUSC samples (ctrl: control). Significant co-occurrence was seen with $P=0.006$ (Pearson's chi-square test).

Supplementary Figure 5

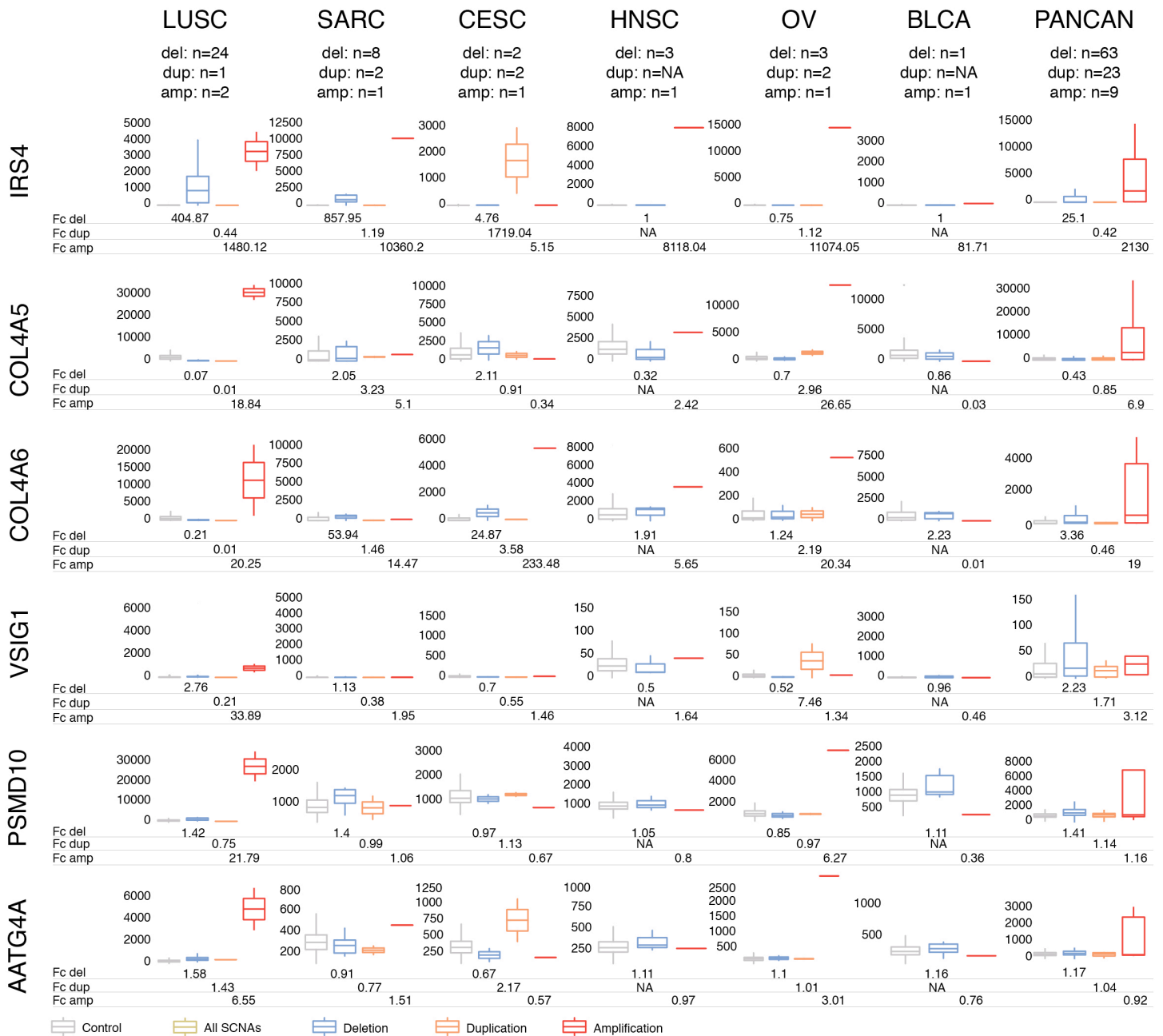


Supplementary Figure 5

Cancer-type specific CESAM analysis of gene expression changes for SVs vs control of *IRS4* locus-proximal genes

Related to main **Figure 3**. *IRS4*, *COL4A5*, *COL4A6*, *VSIG1*, *PSMD10* and *ATG4A* expression values (RSEM) for SVs and controls, shown for different cancer types as well as for the pan-cancer setting. *IRS4* represents the most consistently upregulated gene, and SCNAs clearly appear to converge on *IRS4* overexpression at this extended genomic locus, implicating *IRS4* as a plausible candidate gene.

Supplementary Figure 6

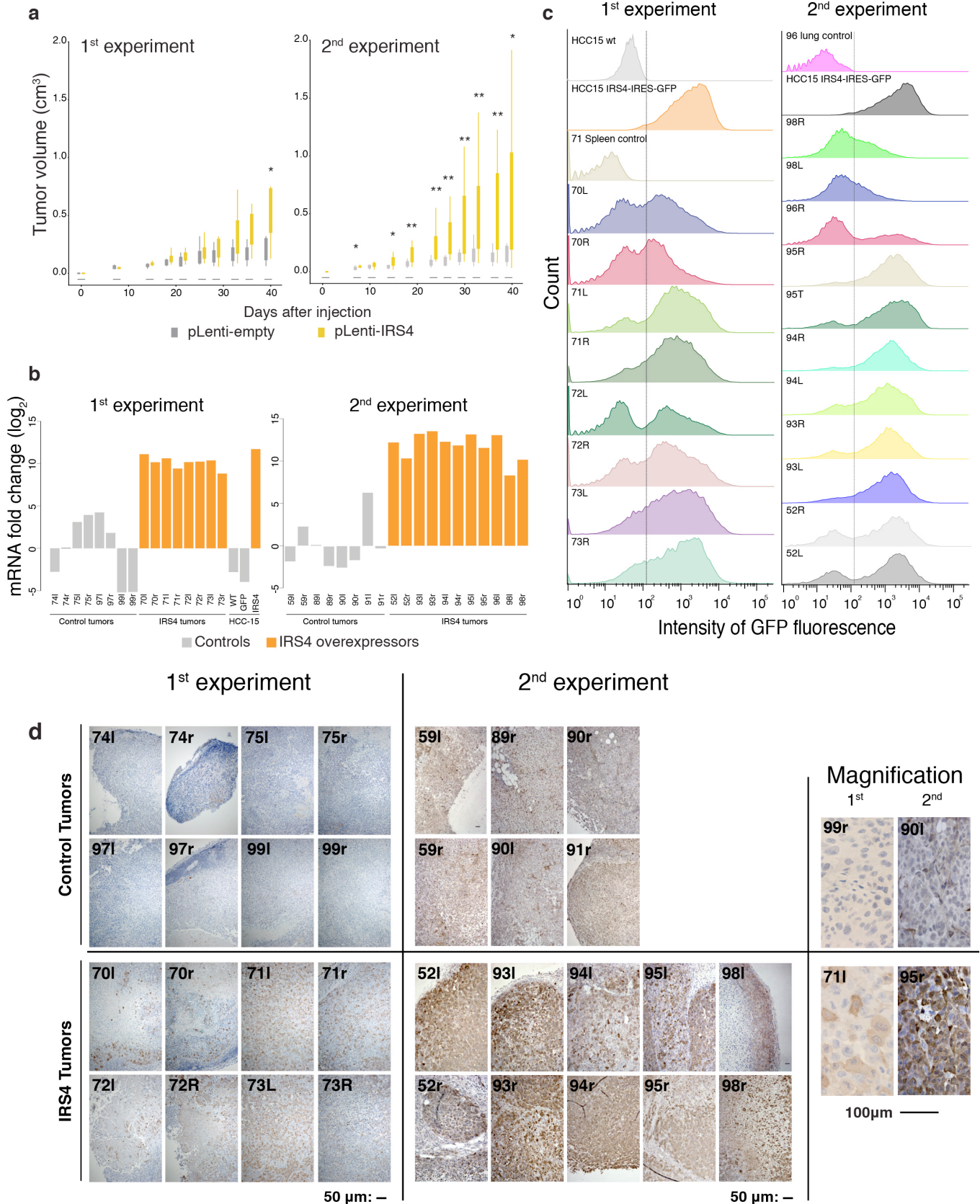


Supplementary Figure 6

Cancer-type specific CESAM analysis of gene expression changes for *IRS4* locus-proximal genes

Related to main **Figure 3** and **Supplementary Figure 5**. *IRS4*, *COL4A5*, *COL4A6*, *VSIG1*, *PSMD10* and *ATG4A* expression values (RSEM) for deletions (del), duplications (dup), amplifications (amp) and controls for different cancer types as well as in a pan-cancer setting. *IRS4* is the most consistently upregulated gene.

Supplementary Figure 7



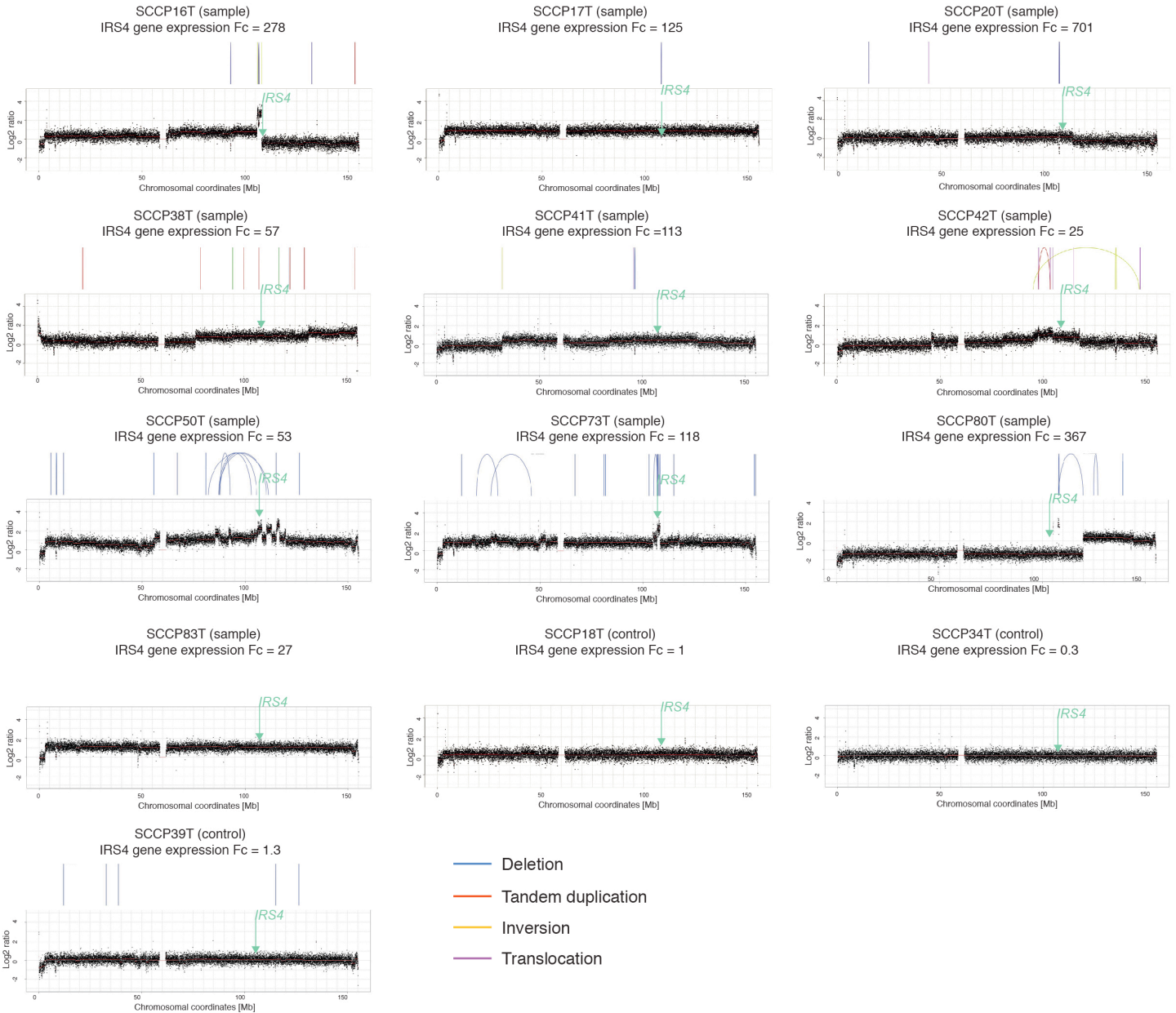
Supplementary Figure 7

Tumor progression of transplanted *IRS4* overexpressing HCC-15 cells and mock control in mouse model

Related to main **Figure 3**. **(a)** Boxplots depicting mouse tumor progression curves of HCC-15 cells containing *IRS4*-expressing lentiviral constructs (pLenti-*IRS4*) versus mock control (pLenti-empty HCC-15); 1×10^6 cells injected respectively; last time point: 1st $P=0.046$, 2nd $P=0.03$; two-tailed t-test; two-tailed t-test computed at last measured time point (day 39); $N=8$ for each group in first experiment, $N=9$ for control and $N=12$ for *IRS4* overexpressing sample in second experiment. **(b)** RT-qPCR of the control HCC-15 and *IRS4*-HCC-15 tumors confirming *IRS4* overexpression. **(c)** Flow cytometry of tumors injected with *IRS4-IRES-GFP* overexpressing vector HCC-15 cell line. **(d)** Representative immunohistochemistry experiment showing *IRS4* overexpression in the harvested tumors (bottom panel) but not in the control tumors (top panel).

Supplementary Figure 8

IRS4 locus structural variant analysis

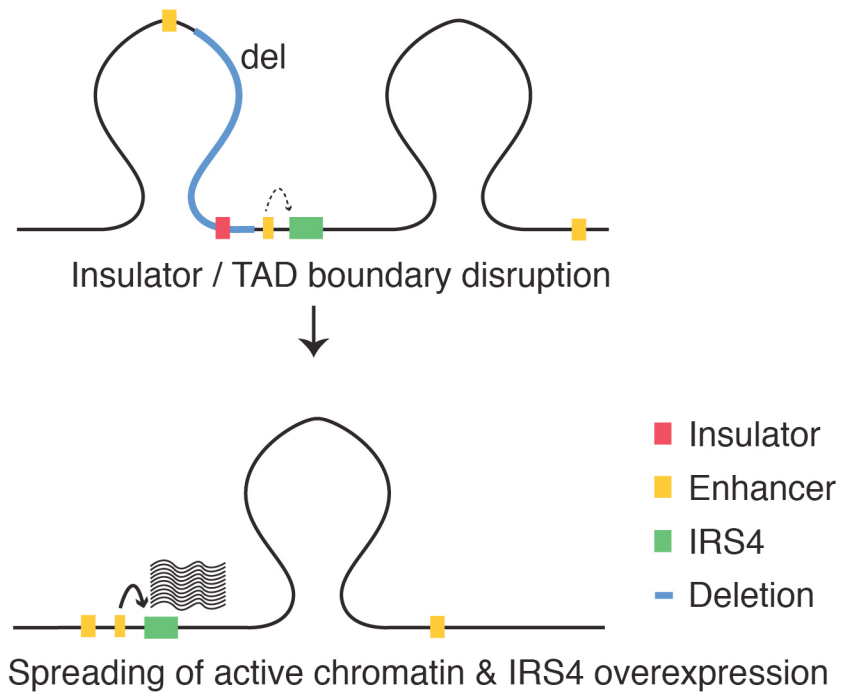


Supplementary Figure 8

SV analysis of chromosome X from LUSC samples associated with *IRS4* gene overexpression and controls

Related to main **Figure 3**. Read depth plot and somatic SVs from mate-pair sequencing data of LUSC samples exhibiting *IRS4* overexpression as well as LUSC controls samples with normal *IRS4* expression level. The location of the *IRS4* gene is indicated with a green arrow. Read depth changes and SV types are shown as colored graphs. *IRS4* alterations included simple as well as more complex SV events including such showing evidence for chromosome shattering (also known as chromothripsis). *IRS4* expression fold-change is shown for each sample.

Supplementary Figure 9

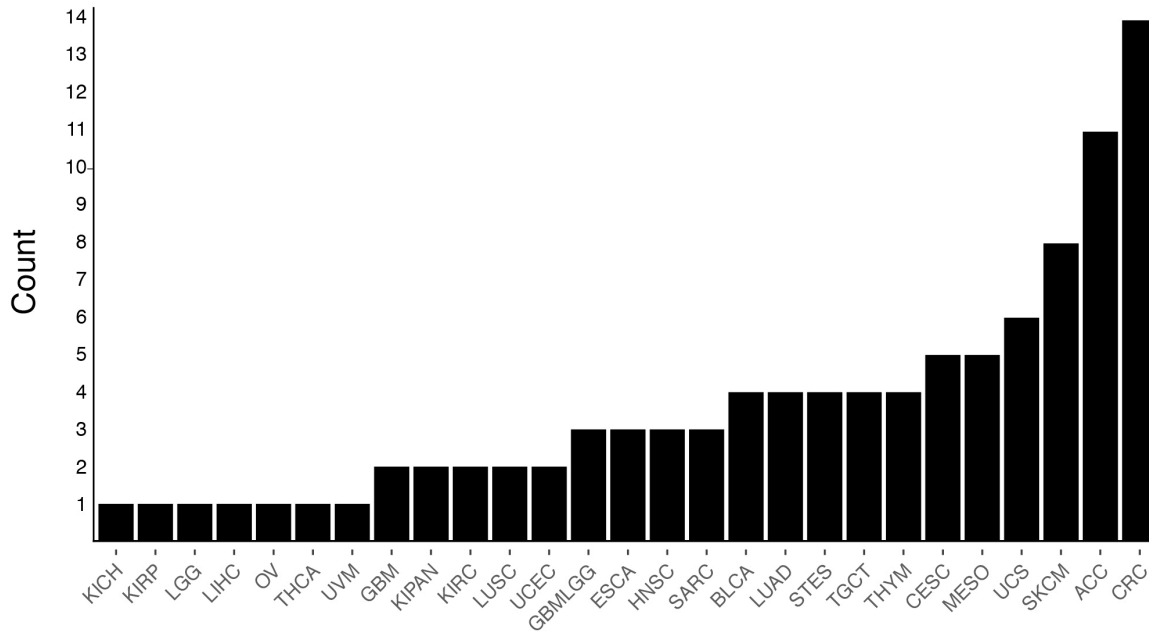


Supplementary Figure 9

Proposed model for *IRS4* overexpression by active chromatin spreading

Related to main **Figure 3**. Proposed model for *IRS4* overexpression by active chromatin spreading, resulting from clustered deletions at a TAD boundary *in cis* of *IRS4*.

Supplementary Figure 10

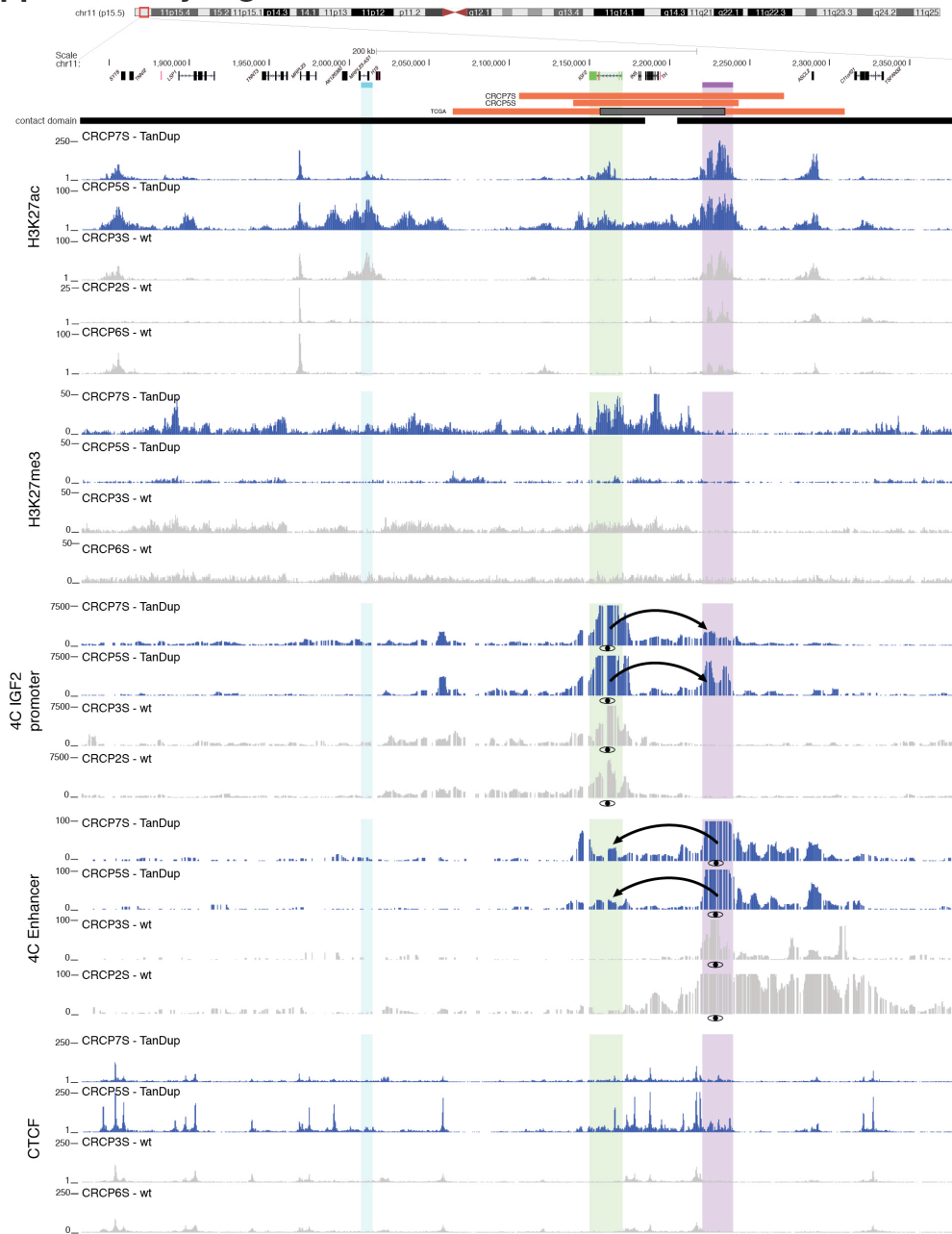


Supplementary Figure 10

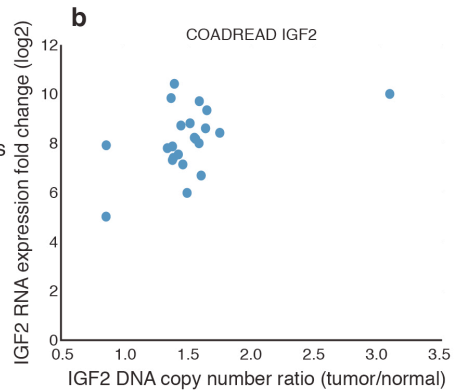
Tumor-type specific analysis and frequency of CESAM hits per tumor type

Related to main **Table 1**

Supplementary Figure 11



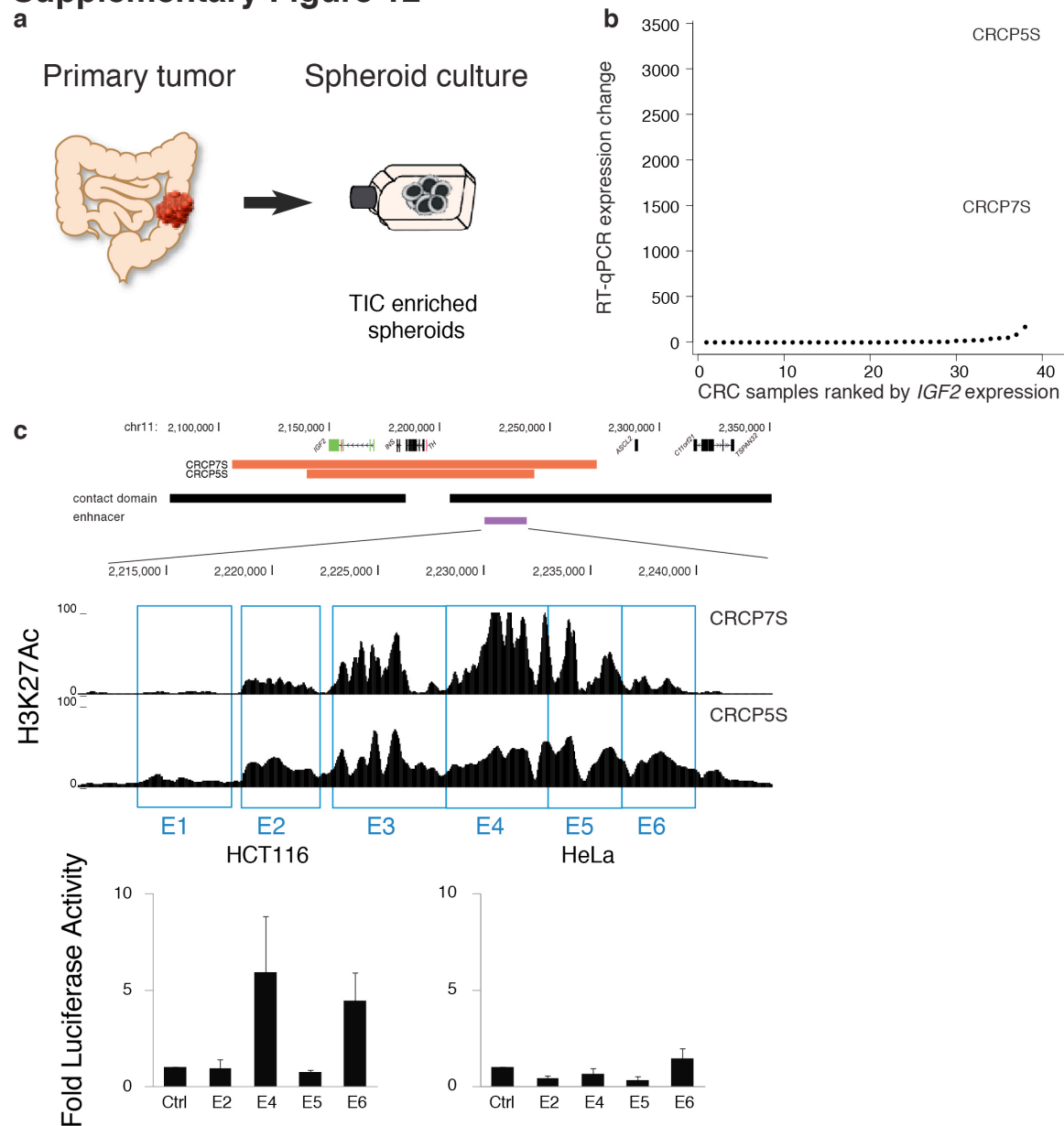
- Samples CRCP7S, CRCP5S
 Controls CRCP2S, CRCP6S
- genes
 - miRNAs
 - Tandem duplications detected based on mate-pairs
 - Inner / outer coordinates of TCGA focal duplication
 - IGF2 cognate enhancer
 - Super-enhancer in adjacent contact domain
 - Viewpoint



Recurrent somatic duplications at the *IGF2* locus associating with *IGF2* overexpression

Related to main **Figure 5**. Recurrent somatic duplications at the *IGF2* locus associating with *IGF2* overexpression encompass a contact domain boundary and a non-cognate annotated super enhancer in the adjacent contact domain, but do *not* encompass the known *IGF2* cognate enhancer (light blue). H3K27ac peaks show the presence of a non-cognate enhancer in the contact domain adjacent to the *IGF2* locus. H2K27me3 marks for WT samples and samples harboring the duplication show absence of repressive chromatin at the adjacent non-cognate enhancer. 4C-Seq experiments using *IGF2* as the viewpoint demonstrate a marked physical interaction between the *IGF2* locus and the non-cognate enhancer in samples with the recurrent tandem duplication, but not in samples lacking the tandem duplication (WT). By comparison, no physical interaction is seen between *IGF2* and its known cognate enhancer, neither in tandem duplication nor in WT samples. 4C-Seq experiments using the non-cognate enhancer as viewpoint verify the marked physical interaction with *IGF2* in tandem duplication carriers and not in WT samples. Collectively, these results demonstrate that hijacking of a non-cognate super-enhancer, mediated by a contact domain spanning recurrent SCNA, drive overexpression of the *IGF2* locus (see also **Figure 5**). CTCF marks for WT samples and samples harboring the duplication are consistent with the presence of TAD boundaries normally separating *IGF2* and the non-cognate (super) enhancer. **(b)** *IGF2* expression fold change versus copy number ratio (tumor/normal).

Supplementary Figure 12

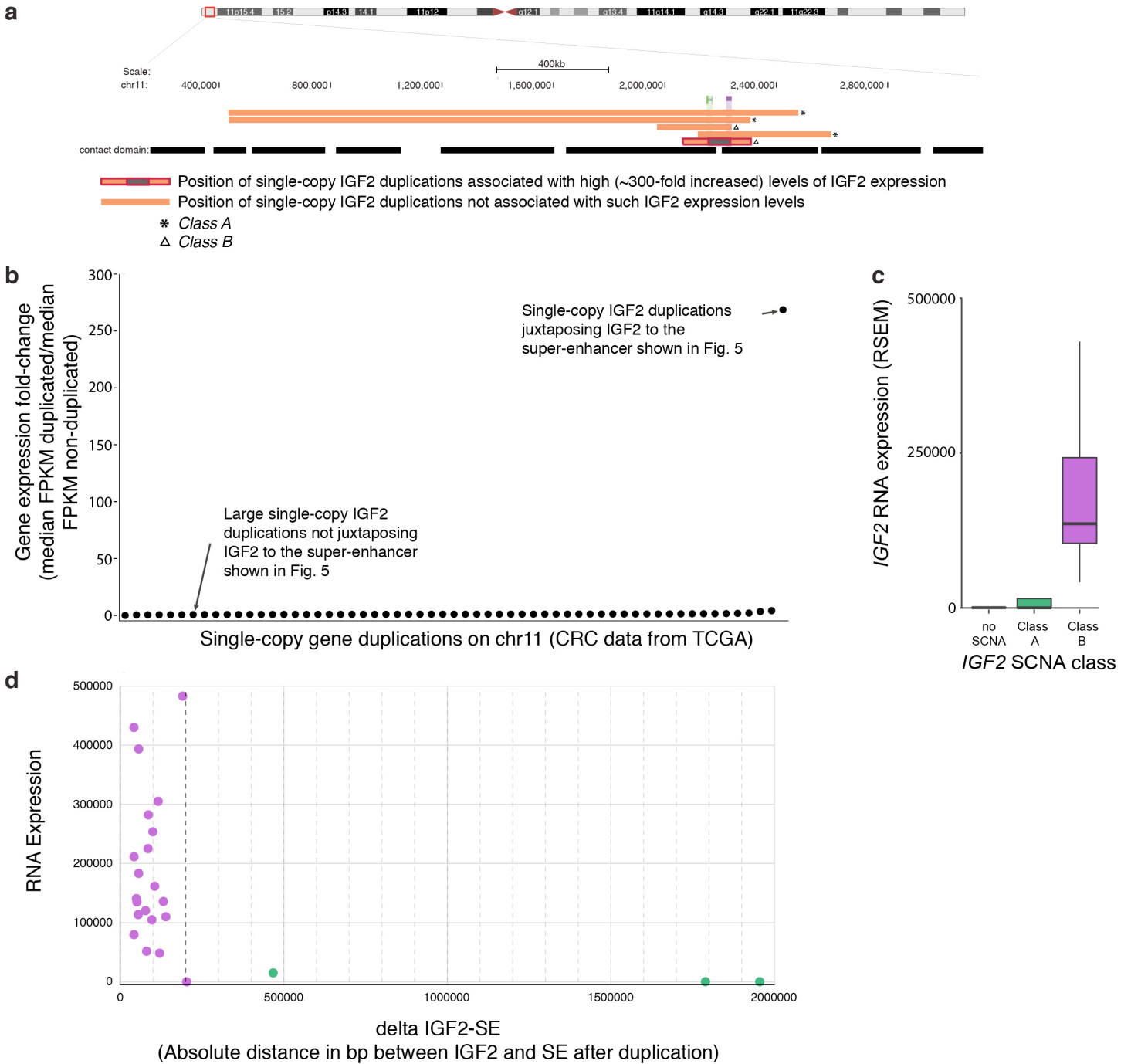


Supplementary Figure 12

Enhancer validation in primary CRC-derived spheroid cultures

Related to main **Figure 5**. (a) Additional experiments to characterize the mechanism of gene dysregulation were pursued in (non-TCGA based) spheroid cultures obtained from primary CRC. (b) Distribution of *IGF2* expression measurements in additional primary samples and spheroids. Two spheroids with marked overexpression, denoted CRCP5S and CRCP7S, were used for further characterization along with controls lacking *IGF2* expression. Expression values are based on qPCR. (c) Functional activity of enhancer regions spanning the *IGF2*-interacting distal super-enhancer locus. Selected parts of the enhancer regions found to be interacting with the *IGF2* locus in tandem-duplicated spheroid cells were amplified by PCR, cloned into a firefly luciferase reporter construct and co-transfected with a renilla luciferase control construct into the colon cancer cell line HCT116 as well as HeLa (cervical cancer cell line) cells used as a control. 48 h post transfection, cells were lysed and luciferase activities were measured. Firefly luciferase signal was normalized to renilla luciferase signal and displayed as fold activity normalized to empty vector control (Ctrl) (Mean \pm SEM in triplicate).

Supplementary Figure 13



Supplementary Figure 13

Single-copy super-enhancer juxtaposing tandem duplications result in *IGF2* overexpression

Related to main **Figure 5**. (a) Four *IGF2* single copy gene duplications not associated with stark levels of *IGF2* overexpression (orange bars) are shown next to *IGF2* single-copy tandem duplications leading to ~300-fold *IGF2* upregulation. Based on analyzing their relative position with respect to preexisting contact domains/TADs (black bars), 3/4 of these larger duplications are not able to bring *IGF2* and the super-enhancer (SE) depicted in Fig. 5 into a single *de*

novo contact domain – the lack of associated *IGF2* overexpression is hence consistent with our new model of *IGF2* dysregulation by enhancer hijacking (**Fig. 5d**). For the fourth duplication, juxtaposition of *IGF2* and the SE is inferred to occur at a larger distance. **(b)** Gene expression fold-change for single-copy duplicated genes on chromosome 11. Single-copy *IGF2* duplications not inferred to bring the SE and *IGF2* into a *de novo* contact domain (denoted **class A**) and duplications inferred to juxtapose the SE and *IGF2* bringing both elements into a *de novo* contact domain (**class B**) are depicted separately. For the observed 58 single-copy duplicated genes on chromosome 11 in CRC samples, the median gene expression increase was ~1.4-fold. **(c)** Panel supporting **(b)**, showing significant difference in *IGF2* expression between **class A** and **class B** single-copy *IGF2* duplications ($P=0.01$; Wilcoxon rank-sum test). **(d)** RNA expression of *IGF2* (y-axis) for **class A** (green dots) and **class B** (purple dots) versus the estimated distance in basepairs (bp) between *IGF2* and the SE upon tandem duplication (x-axis).