

Interspecific hybridization in tomato influences endogenous viral sRNAs and alters gene expression

Sara Lopez-Gomollon, Sebastian Y. Müller & David C. Baulcombe*.

Additional file 1: Figures S1-S7

Fig. S1. Phenotyping analysis of the hybrid population along the generations. For the phenotypic study, 4 cuttings (obtained from axillary new leaves) per plant were grown simultaneously for 8 weeks. **a.** The 5th leaf from the bottom was used for the comparative phenotypic analysis. All pictures at the same scale being scale bar 10cm. **b.** Quantitative analysis of phenotypical traits (dry weight, height, and number of leaves). Parental lines in red, F1-F3 plants in light blue and F4 plants in dark blue. Dotted lines show the values for the parental lines for comparison to identify transgenic phenotypes. Box plots elements: box limits, upper and lower quartiles; center line, median; whiskers, 1.5x interquartile range; points, outliers.

Fig.S2. F4 genomes are chimeras of *S. lycopersicum* and *S. pennellii*. RNAseq-based SNP genotyping for each of the F4 plants (P1512, P2561, P2562, P3611, P3612, P4041, P4042). Top track (within each chromosome panel): For each tomato nuclear chromosome is it shown the genotypic results (red: region homozygous transcribed for *S. lycopersicum*, green: heterozygous transcribed, light blue: homozygous transcribed for *S. pennellii*, pink: could not be determined). Bottom track: chromatin status divided into euchromatin (light grey) or heterochromatin (dark grey) as described [14]. As expected, there were very few meiotic crossovers in the pericentromeric region. Preferential inheritance of some regions of the genomes including those that were homozygous transcribed *lyc* in all F4 plants (regions in chromosomes 2,8,10 and 11) may be due to selection against genes or combinations of genes that caused failure to produce progeny in some F2 and F3 progeny.

Fig.S3. Gene expression and sRNA analysis for the *S pennellii* homozygous transcribed regions for each F4 compared to *pen*. **a.** Percentage of genes in each group for the F4s compared to *pen* parent, calculated from the total number of genes in the *pen* homozygous transcribed region, being each bar an individual plant. The right-hand panel shows an expanded view of the +DEG (red) and -

DEG (blue). **b.** Percentage of sRNA loci in each group, calculated from the total number of sRNA loci in the *pen* homozygous transcribed region. Groups are sRNA loci upregulated (+DESL, red), downregulated (-DESL, blue), “in between” (lilac) and non-differentially expressed (non-DESL, yellow). The right-hand panel shows an expanded view of the +DESL (red) and -DESL (blue). Thresholds: DESL, FDR <0.05; “in between” DESL, 0.05 <FDR < 0.9; non-DESL, FDR > 0.9. **c-e.** Box plot showing the percentage of *pen* sRNA loci (**c**) genome wide, (**d**) to each genomic feature, (**e**) to TE order, that are +DESL or -DESL. Ratio of +DESL/-DESL is shown at the bottom. Each plant is represented by different dot shapes. Box plots elements: box limits, upper and lower quartiles; center line, median; whiskers, from each quartile to the minimum or maximum. For comparison, colored lines are the percentage values for +DESL (red) and -DESL (blue) mapping the whole genome (obtained from **c**).

Fig.S4. Phylogenetic relationship of EPRV envelope sequence. Polar Tree layout of the phylogenetic tree build based on envelope sequences found on *S. lycopersicum* by manual BLAT search using as original the envelope sequence from EPRV on GIRI, followed by RaxML tree construction. Node names consist of domain name (envelope), followed by coordinates on genome (chromosome name and start coordinate), overlapping transposon family according to REPET annotation (NA if no overlap), and overlapping type of sRNA loci highlighting +DESL in red and -DESL in blue.

Fig.S5. Size distribution of sRNA loci in the F4s. Name of plant shown at the top of each set of panels. Normalised sRNA counts plotted by size for -DESL, +DESL and non-DESL (**a**) genome wide and (**b**) mapping to each TE order. **c.** Normalised sRNA counts plotted by size in each F4 (-DESL, +DESL and non-DESL) overlapping with sRNA loci in *dc12* (-D2SL, +D2SL and non-D2SL). Red line- sRNA counts for *lyc*, blue line -sRNA counts for each F4. Solid line- counts for all sRNA loci in each group category for that plant. Dotted line – counts for sRNA in each category that map to EPRVs.

Fig.S6. Heatmap of the inheritance of DESLs in each family. Normalised reads for each generation for the sRNA loci identified as *lyc* DESL in one of the F4s per family (in bold letter on top), filtered for $|FC| > 9$ and $FDR < 1e-10$. The strict threshold was used to reduce the size of the plot by focusing on the most significant changes. Normalization was performed for each bin by centering and scaling (z-

score) across the lineages (rows). The names of bins (rows) correspond to the genomic coordinates of their first base. Annotation tracks were derived by overlapping the respective bins with genomic annotation features. Columns correspond to data for each plant (in black) and the generation that they belong to (in red on top of the name of the plant).

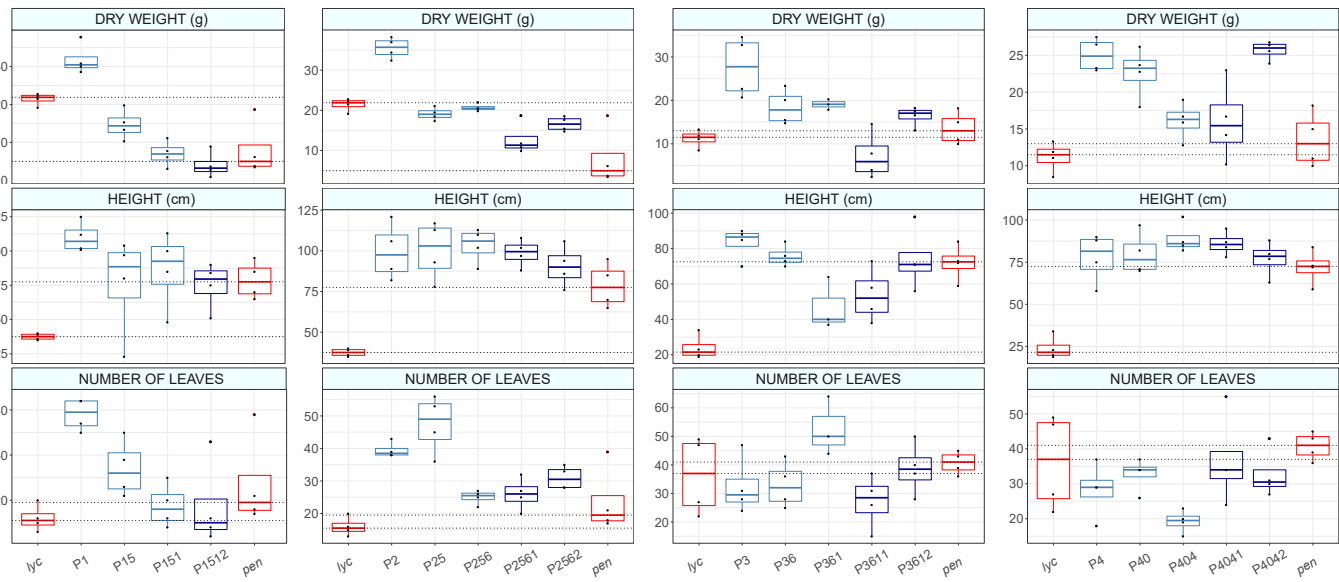
Fig.S7. Differentially expressed genes are shared between the F4s and *dc12* mutant. **a.** Venn diagram of genes upregulated (+DEG) and downregulated (-DEG) for an F4 plant compared to the genes upregulated (+D2G) or downregulated (-D2G) in the *dc12* mutant. Statistical association determined by Fisher's Exact Test, ** two-tailed $p < 10e-6$, *** two-tailed $p < 2.2e-16$. **b.** Scatter plot of the gene expression log fold change (FC) for each F4 compared to the *dc12* mutant. Grey: non-DEG or in between for any of the two plants; green: DEG for the F4 or *dc12*; purple: DEG for F4 and *dc12*. Pearson Correlation coefficient values (r) for the DEG and D2G genes (purple genes) for each scatter plot, $p = p$ value.

Figure S1

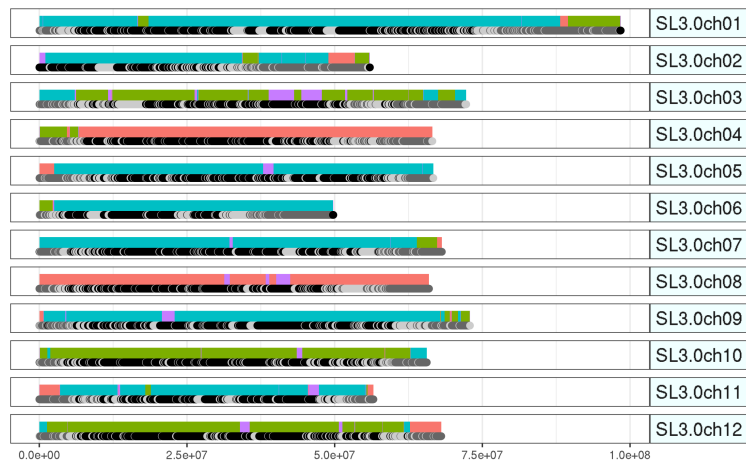
a



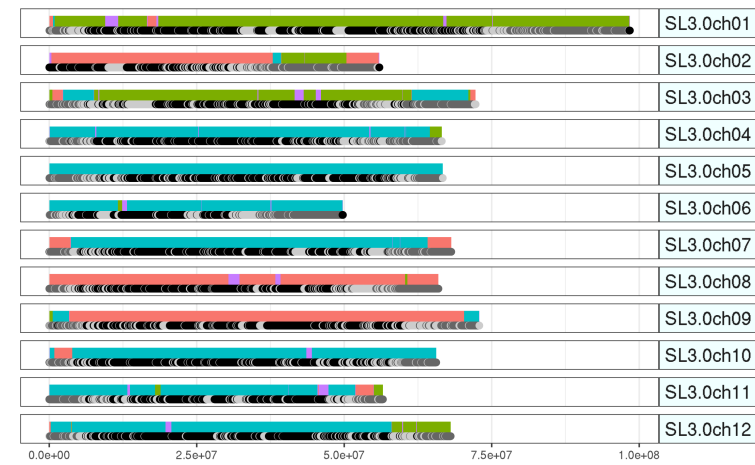
b



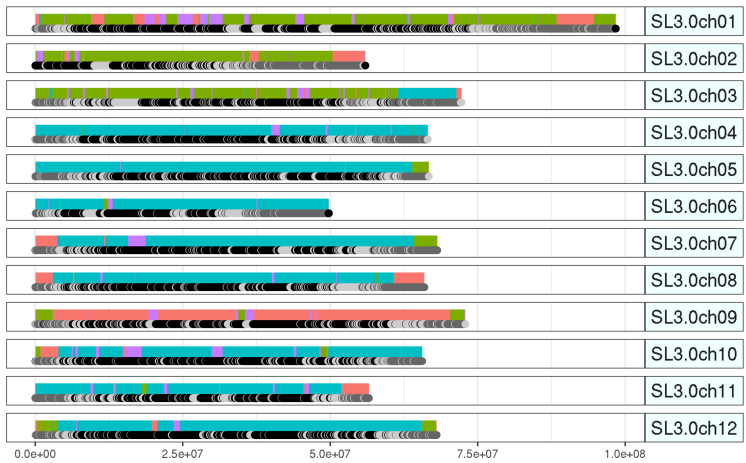
P1512



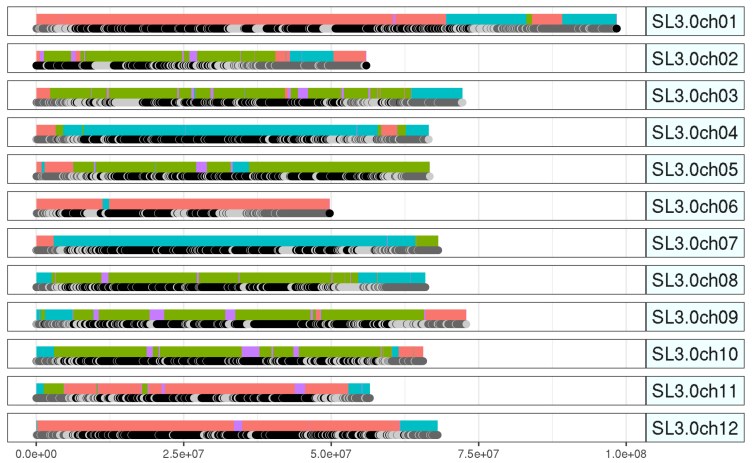
P2561



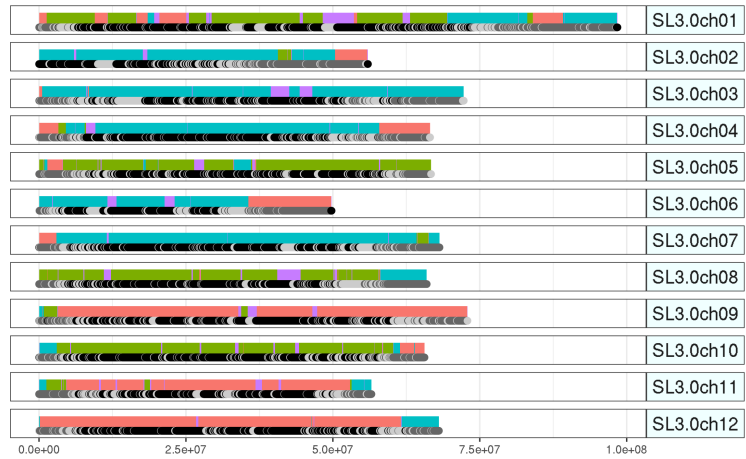
P2562



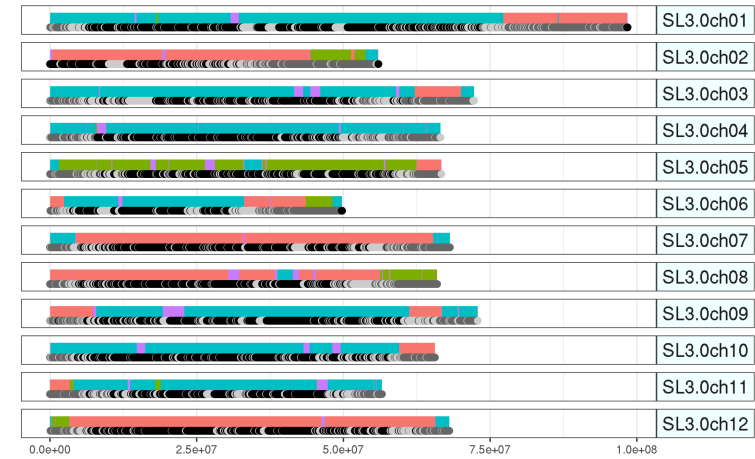
P3611



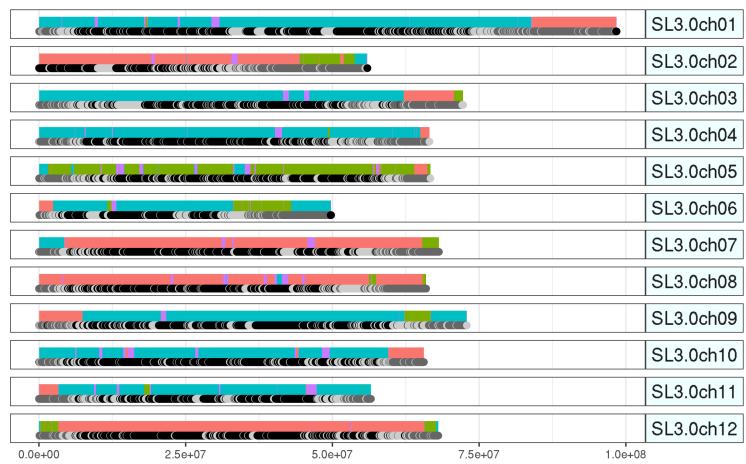
P3612



P4041



P4042



■ Slyc homozygous ■ heterozygous ■ Spenn homozygous
● Euchromatin ● Heterochromatin ●

Figure S3

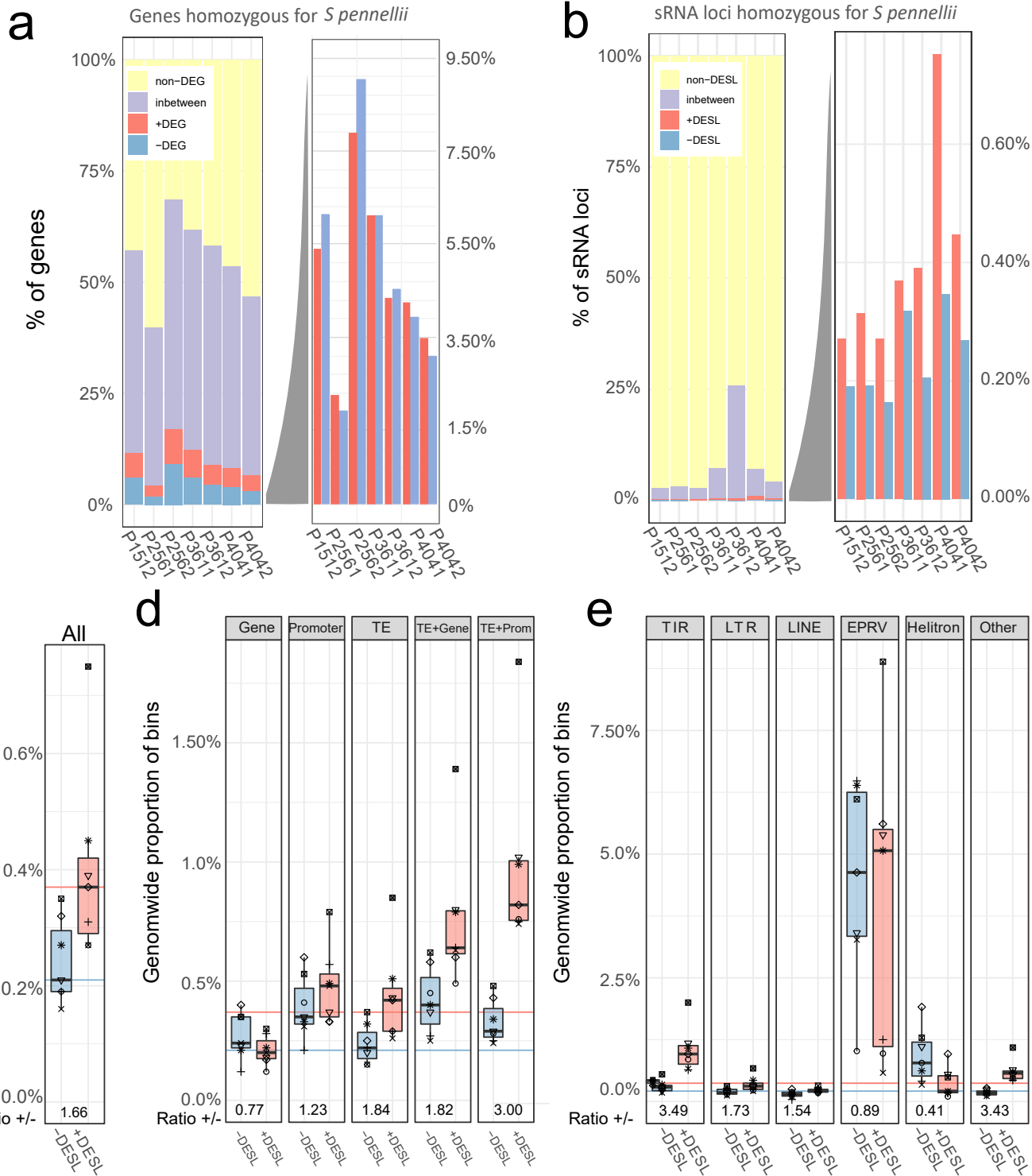
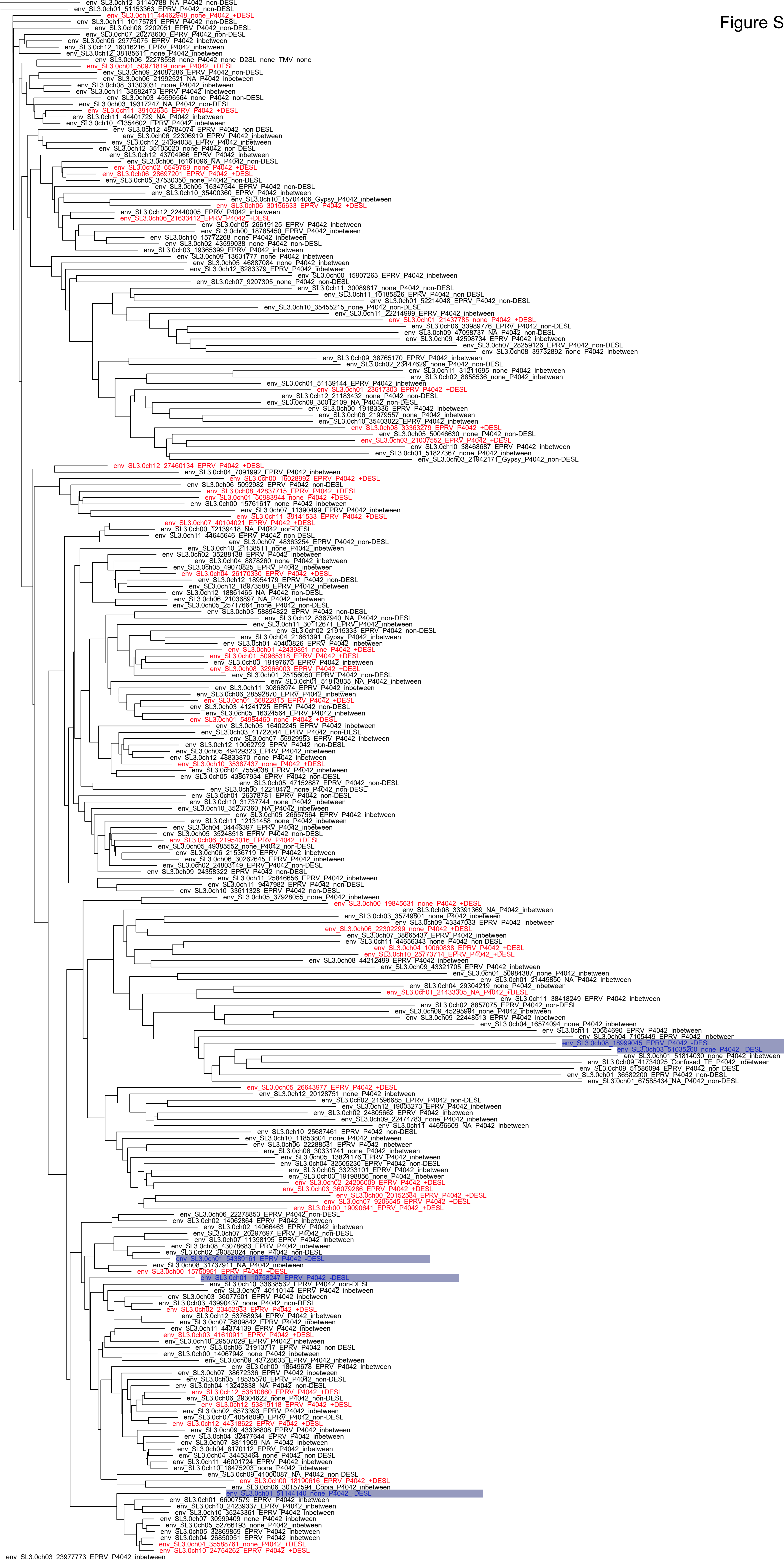


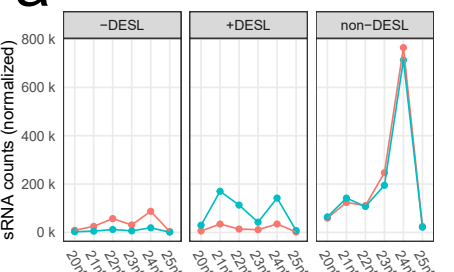
Figure S4



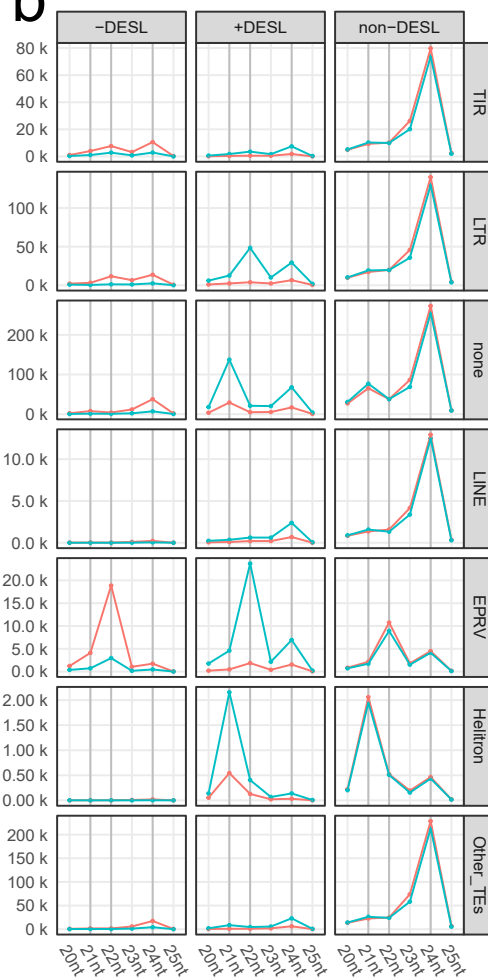
P1512

Figure S5

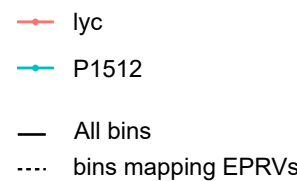
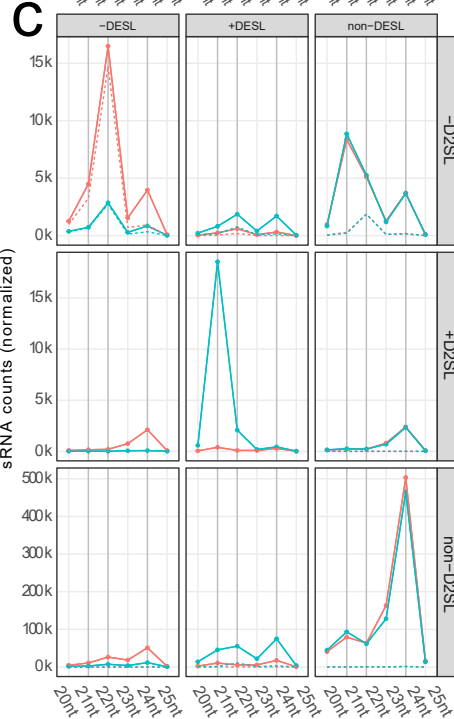
a



b

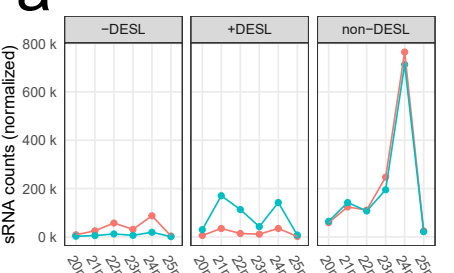


c

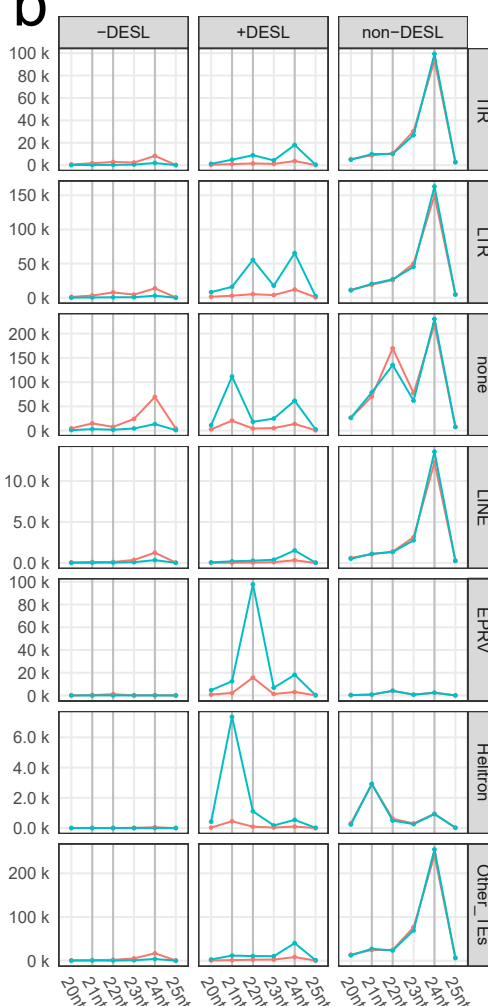


P2561

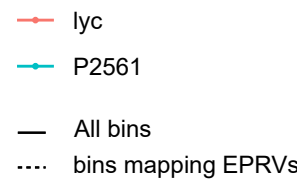
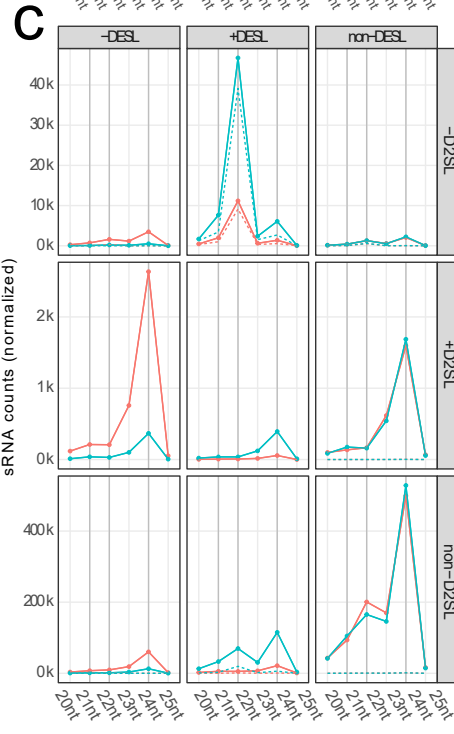
a



b

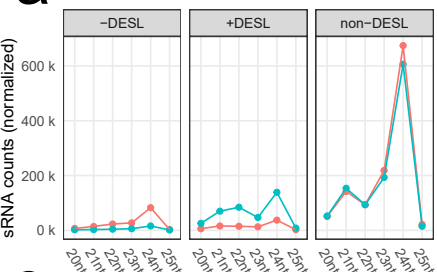


c

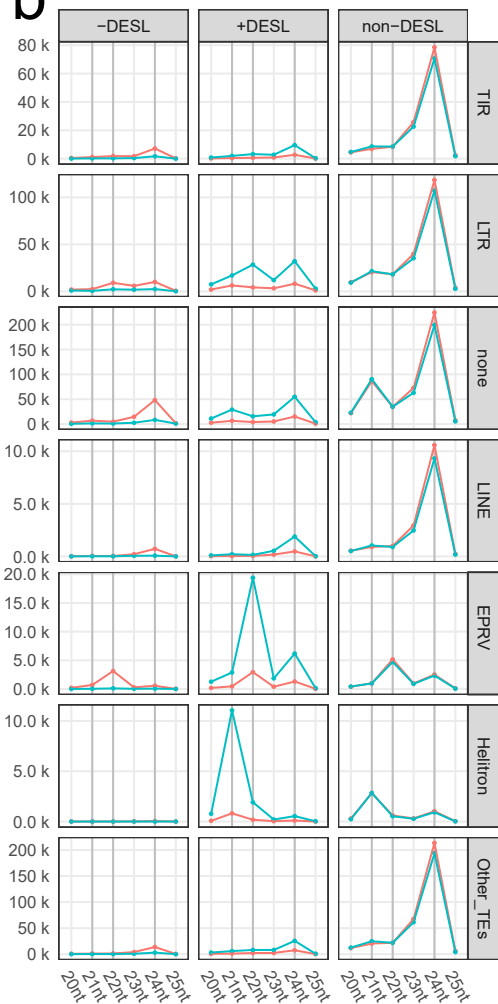


P2562

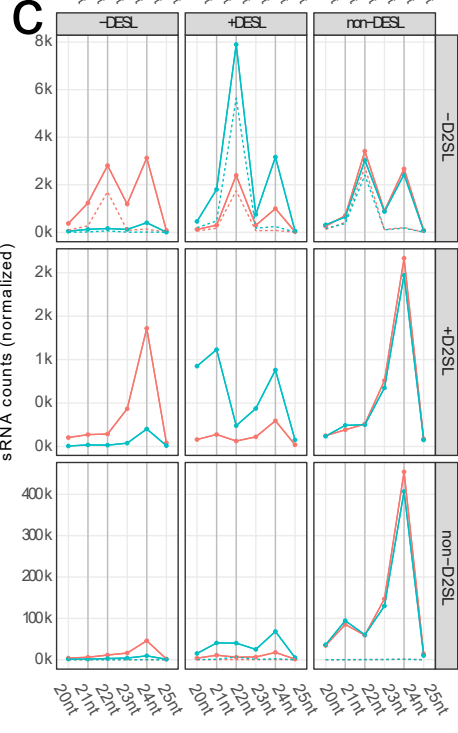
a



b

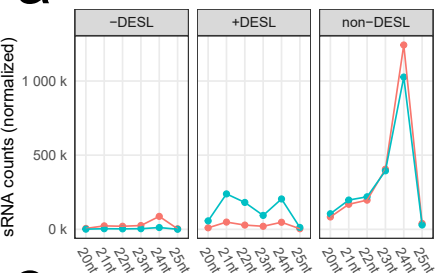


c

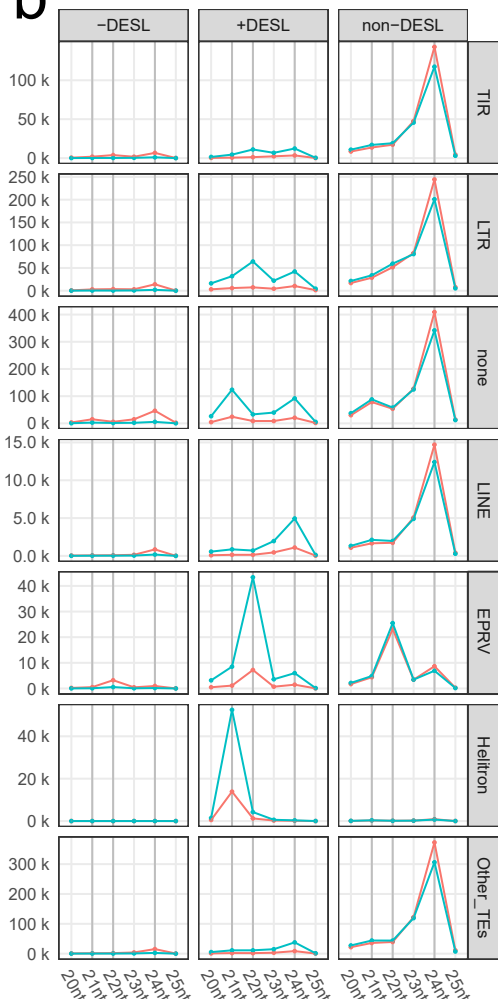


P3611

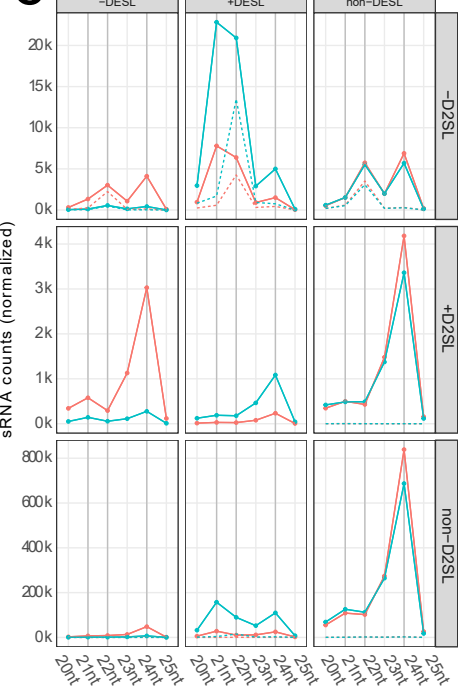
a



b

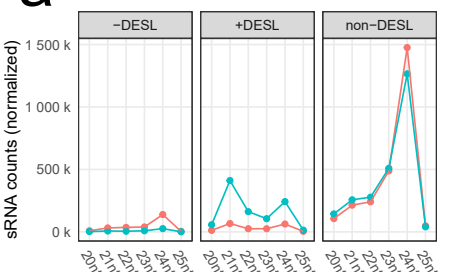


c

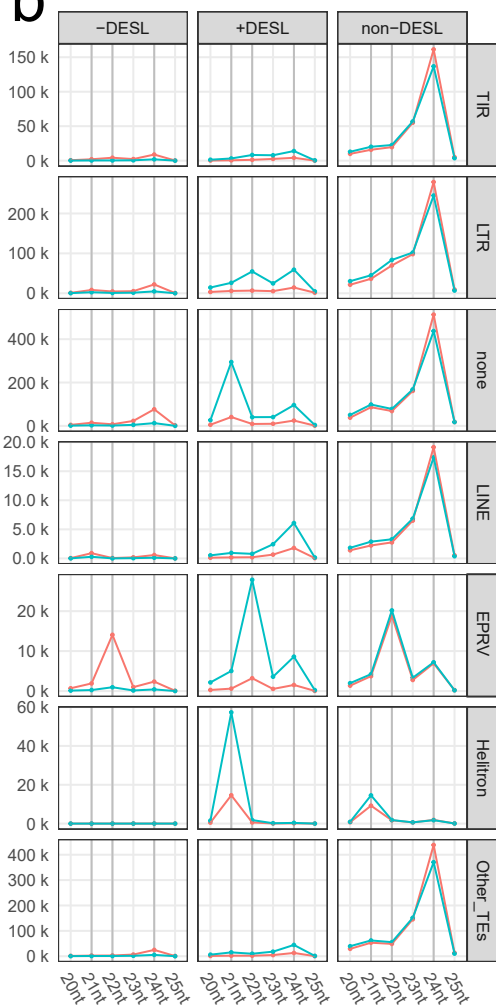


P3612

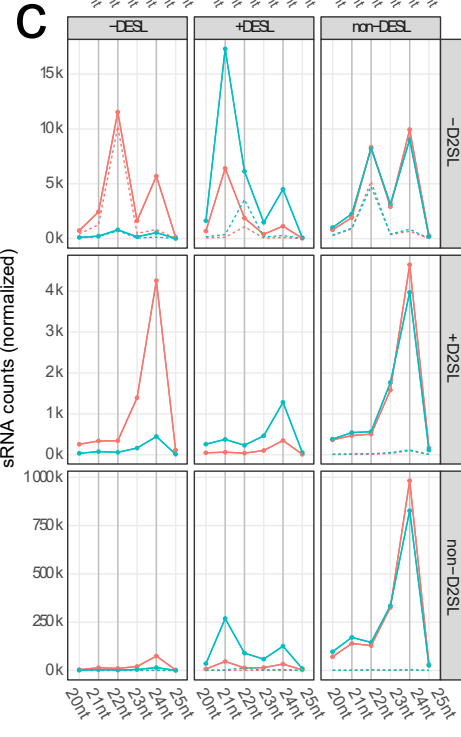
a



b

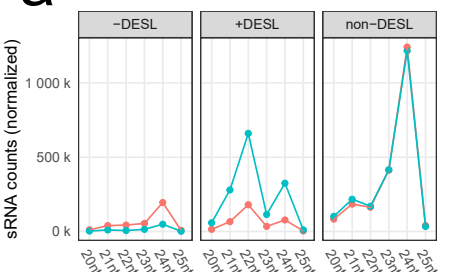


c

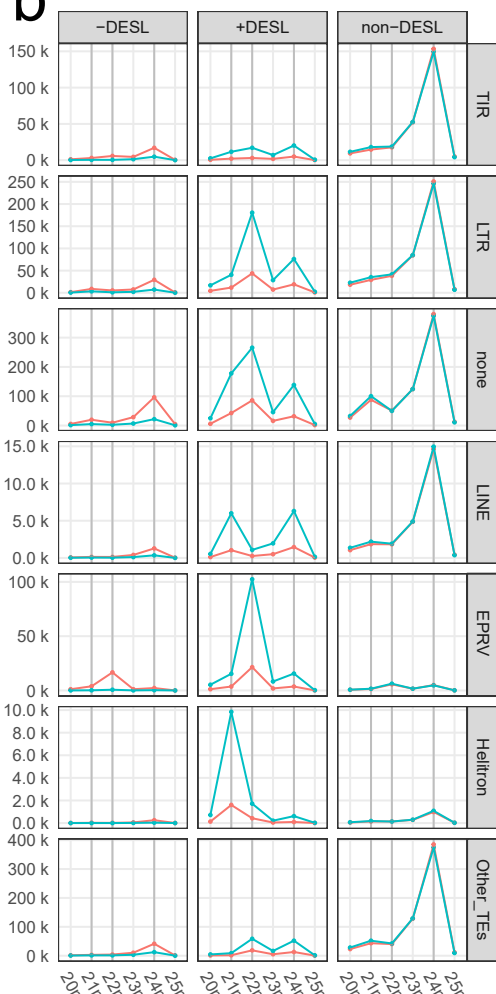


P4041

a



b



c

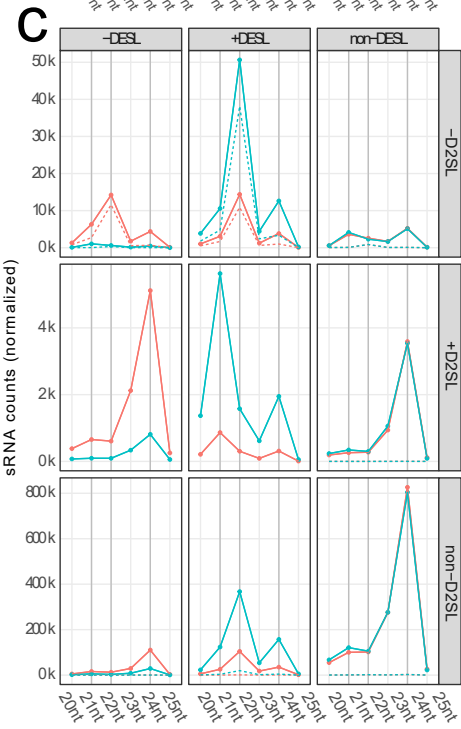
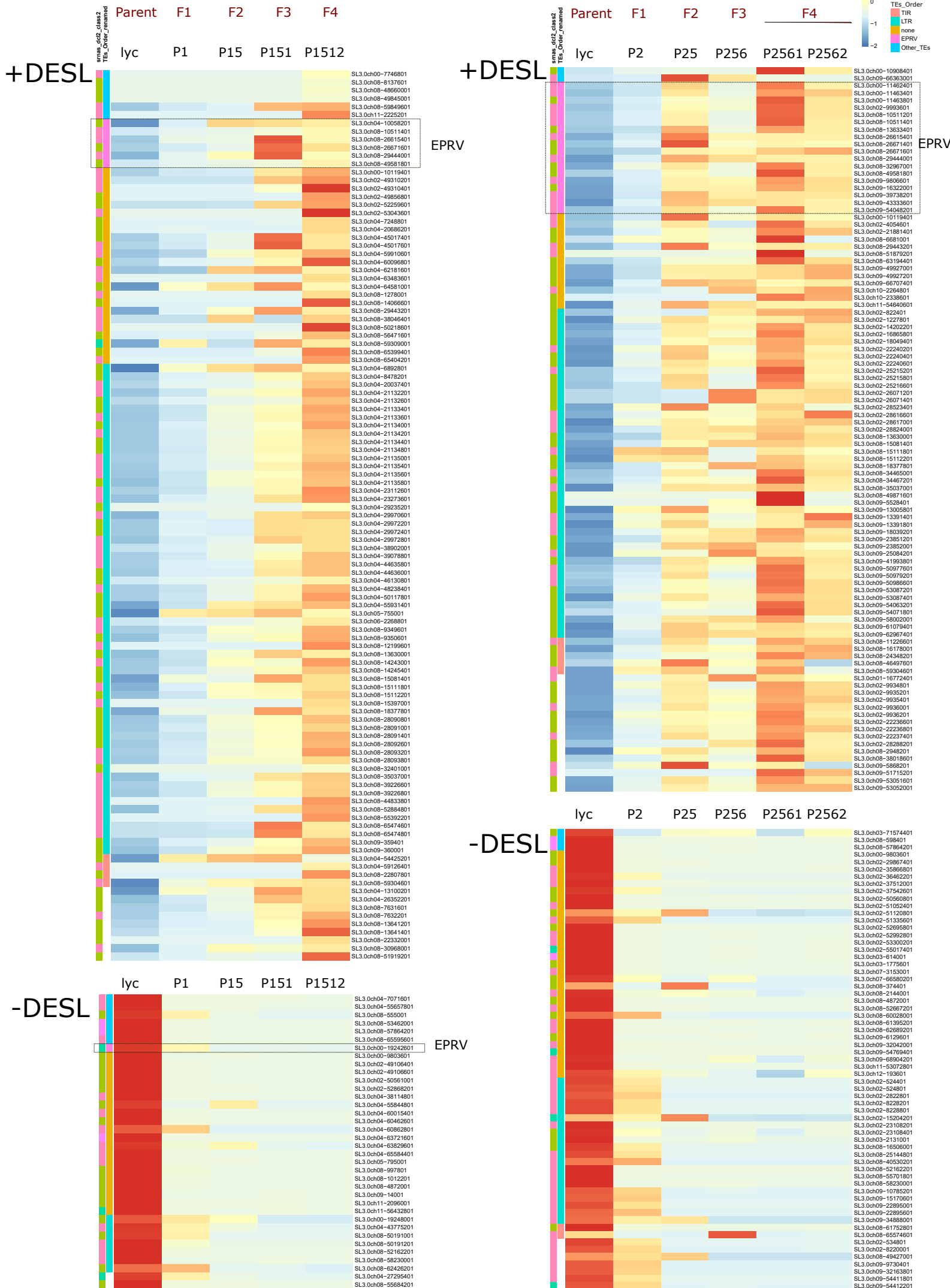


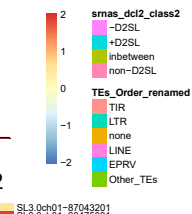
Figure S6 P1512

P2562



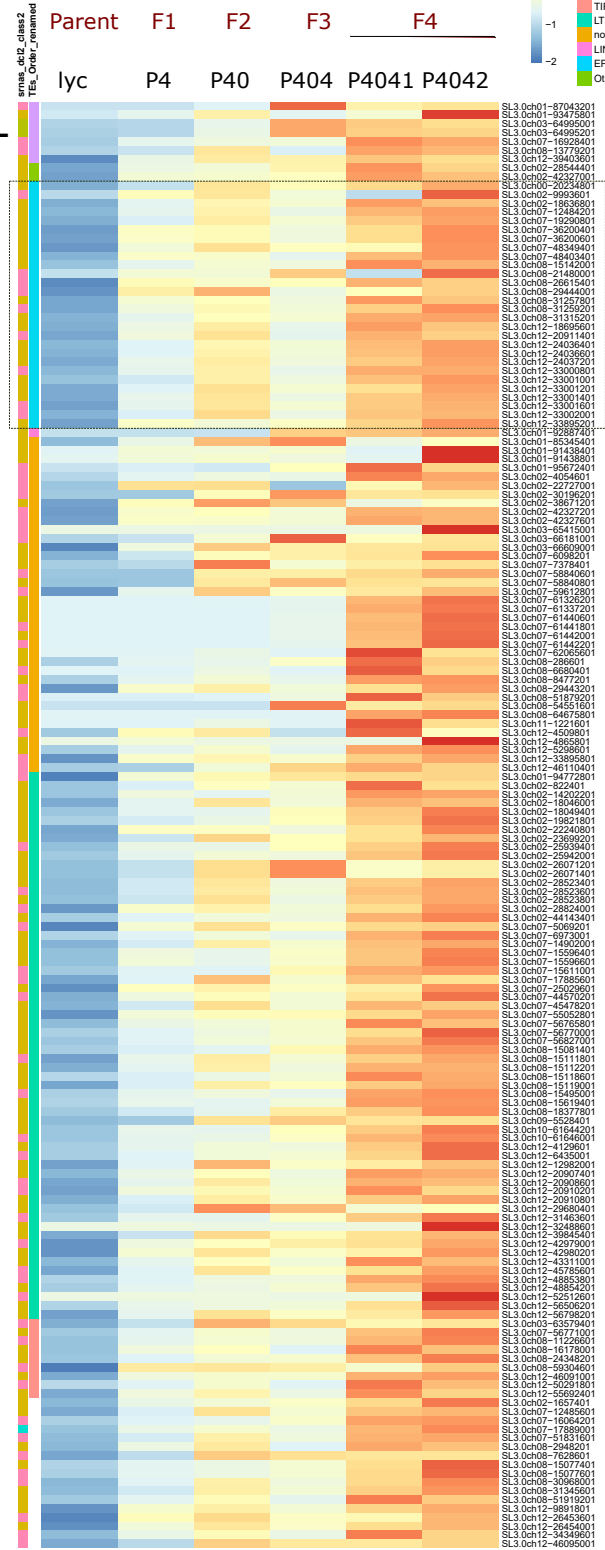
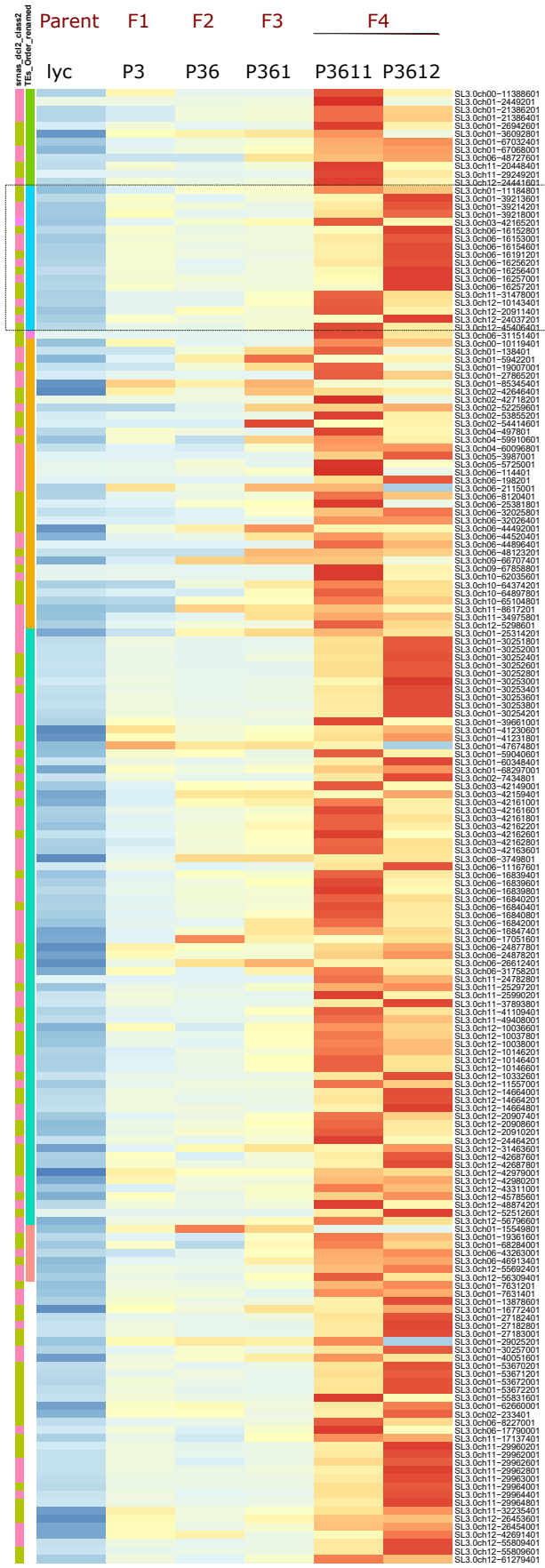
P3611

P4042



+DES�

+DES�



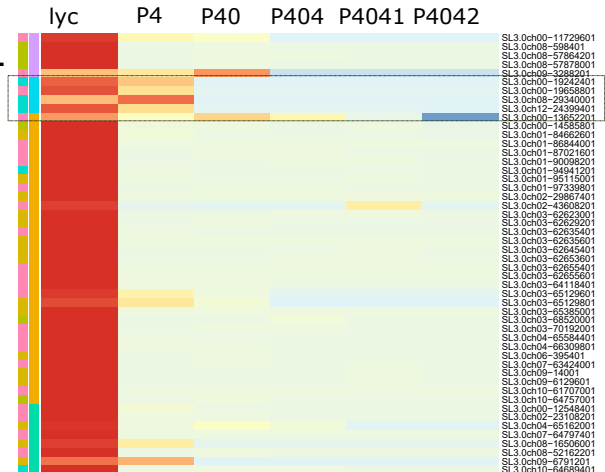
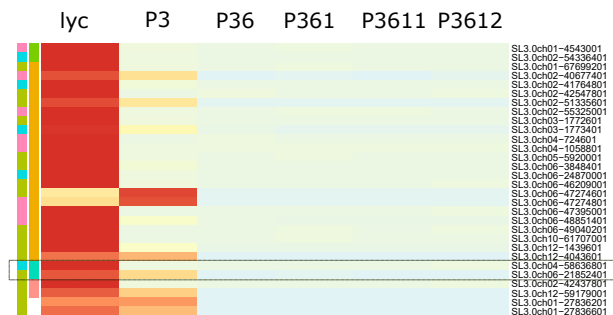
EPRV

EPRV

-DES�

EPRV

-DES�



EPRV

Figure S7

