Cell Systems, Volume *9*

Supplemental Information

Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors

Eva Yus, Verónica Lloréns-Rico, Sira Martínez, Carolina Gallo, Hinnerk Eilers, Cedric Blötz, Jörg Stülke, Maria Lluch-Senar, and Luis Serrano

Determination of the gene regulatory network of a genome-reduced bacterium highlights alternative regulation independent of transcription factors

SUPPLEMENTARY DATA

Figure S1. Supporting figure for Figure 1

A. ROC curves of the experiments that were performed to identify DNA/RNA-binding proteins (see Methods). We show the curves for those experiments with an area under the curve (AUC) higher than 0.7. These experiments distinguish better the gold set of proteins that bind either DNA or RNA, from the rest.

B. *M. pneumoniae* has six different groups of proteins that are fully duplicated or duplicated in part (DUF16, Lipoprotein3, Lipoprotein X, adhesins, hsdS and DUF237). The DUF16 group of duplicated proteins includes: MPN010, MPN013, MPN038, MPN094, MPN100, MPN104, MPN127, MPN130, MPN137, MPN138, MPN139, MPN145, MPN151, MPN204, MPN283, MPN287, MPN344, MPN368, MPN410, MPN484, MPN501, MPN504, MPN524, MPN655, and MPN675. Of these genes, the following were detected to bind to DNA: MPN013, MPN100, MPN127, MPN137, MPN138, MPN287, MPN344, MPN368, MPN484, MPN501, MPN655 and MPN675. These genes could be classified into four subgroups, all of which have the DUF16 domain (MPN010) in common and possibly other shared regions. MPN094, MPN100, MPN130, MPN138, MPN139, MPN204, MPN368, and MPN410 are full duplications containing the DUF16 domain (MPN010). From this subgroup we cloned MPN368 for further analysis. Another subgroup of duplicated proteins with DUF16 are MPN127, MPN151 and MPN484. From this group we cloned MPN127 for further analysis. The third subgroup comprises MPN013, MPN287, MPN344 and MPN655. MPN287 is lacking an N-terminal proline-rich fragment but all three of them bind DNA. For this subgroup we cloned MPN287 for further analysis. The fourth group comprises MPN038, MPN137 and MPN675, and we cloned MPN038 for further analysis.

Figure S2. Supporting figure for Figure 2

A. ChIP-seq reproducibility. The ChIP-seq profile for SigA and SpxA with biological replicas.

B. Correlation between gene expression and the extent of binding of the RNAP, as determined by ChIP-seq. Phantom peaks were discarded and peak height was normalized to the mean peak height of the experiment.
C. Changes in the ChIP-seq profile of the RNAP subunit alpha (RpoA) at the exponential (6 h) and stationary (96 h) phase in a selected region of the chromosome.
D. Novobiocin treatment leads to the release of RNAP binding as determined by ChIP-seq analysis, and subsequent RNA degradation.
E. ChIP-seq peaks of DnaA showing an enrichment of peaks in the first quarter of the chromosome

Figure S3. Supporting figure for Figure 3

To build dominant negative mutants we did multiple sequence alignment of the target *M. pneumoniae* proteins with orthologue proteins from other Mycoplasma species as well as other bacteria. The inverted triangles point to the mutated residues based on the structures determined for orthologous proteins, biochemical evidence or residue conservation. mge, *Mycoplasma genitalium*. mga, *Mycoplasma gallisepticum*. mpe, *Mycoplasma penetrans*. bsu, *Bacillus subtilis*. saa, *Staphyloccocus aureus*. mss, *Mycoplasma suis*. mpv, *Mycoplasma parvum*. lla, *Lactococcus lactis*. aad, *Alicyclobacillus acidocaldarius*. uue, *Ureaplasma parvum*. bca, *Bacillus cereus*.

A. Multiple sequence alignment of the DNA-binding domain of MPN239 with orthologous genes from other Mycoplasma family members and the putative PLP-dependent transcriptional regulator YdfD from *B. subtilis*.
B. Multiple sequence alignment of MPN275 and ybaB-ebfC TF family members and dominant negative mutant. Conserved Glu,Asp was mutated to Lys, Ala according to (Jutras et al., 2012).
C. Multiple sequence alignment of MPN294 and Ara TF family members. Positively charged residues (in red) are the mutations (to Glu) introduced to prevent DNA binding.
D. Multiple sequence alignment of MPN295 and the LuxR family. The indicated positive residues were mutated to Ala to prevent dimerization.
E. Multiple sequence alignment of MPN329 and Fur family members. The indicated residues were mutated to Glu.
F. Multiple sequence alignment of MPN424 and YxlM members. Shown is a sequence alignment with the YlxM protein from *S. aureus*, whose 3D structure is available (1XSV). A structure for the *E. coli* TRFB transcriptional repressor protein has also been solved

(with bound DNA: 2W7N). Residues in red are the mutations (to Glu) introduced to prevent dimerization.

G. Multiple sequence alignment of MPN529 and HU family members. The indicated Arg residues were mutated to Glu and Ala, respectively.

H. Multiple sequence alignment of MPN608 and PhoU-family members. The indicated negatively charged residues were mutated to Ala.

Figure S4. Supporting figure for Figure 3

A. Correlation between the fold change in RNA levels (by transcriptomics) and the fold change in protein expression (based on proteomics) of the candidate genes for which proteomics data is available.

B. No relationship is observed between the level of RNA overexpression of a candidate gene of interest and the number of genes that change in the transcriptomic analysis.

C. Correlation between the transcriptional changes found in the Ldh KO and the MPN294 KO.

D. Correlation between the transcriptional changes found when inactivating (KO) CorB, MPN162 or RecA.

E. Expression of the *hit1* gene and the *serS* gene encoding the Ap_4A-producing serine-tRNA synthetase exhibited a significant anticorrelation (r=-0.37, p=7.45e-8) across all experiments in this study.

F. Negative correlation between the transcriptional changes found in the PrkC and PrpC inactivation (KO).

G. Example of two consecutive operons predicted to be regulated by the CorB protein. The first operon (*mpn647-mpn641*, minus strand) is identified as target of the CorB protein by the two independent approaches (the manual gene regulatory network reconstruction and the Inferelator network). However, for the next operon (*mpn640*) the fold change between the control and the CorB KO decreases, and is only detected by Inferelator but not in our analysis, since the fold change falls below our threshold for detection.

H. Venn Diagram showing the number of interactions in the expanded gene regulatory network, and the interactions found by Inferelator using the environmental perturbations set.

Figure S5. Supporting figure for Figure 4 and 5.

A. Negative correlation between cold-shock and heat-shock changes. Correlation was determined using RNA-seq datasets.

B. Negative correlation between the ratio of the stationary and exponential growth phase fold changes and the changes occurring in cold-shock. Correlation was determined using RNA-seq datasets.

C. Effect of changes in supercoiling on transcription, as induced by novobiocin inhibition of gyrase. All 689 genes in *M. pneumoniae* were clustered according to their profile upon addition of the novobiocin at different concentrations. Upper panel shows the general behaviour of the majority of the genes (degradation). The three different clusters (in different shades of gray) show different sensitivities to the drug, as genes take shorter or longer to decay. Lower panel shows a single cluster in which genes are upregulated upon changes in supercoiling.

D. Anticorrelation between consecutive genes in opposite strands (MPN244 in plus strand and MPN245 in the minus strand).

E. Upper panel: Alignment of *M. pneumoniae* ribosomal RNA and protein promoters and *B. subtilis* rRNA. The Pribnow box (green) and other functionally relevant elements are shown, including the TSS (blue) and the extended Pribnow box (TG, blue). Lower panel: Right. GLAM2 motif (Frith et al., 2008) of the ribosomal protein genes in *M. pneumoniae* and *M. genitalium*. Despite the variable spacer between the Pribnow and the TSS, the first two nucleotides of the RNA (GC) are highly conserved across the different ribosomal operons in these two species.

F. Conservation of promoter sequences as seen by alignment with *M. genitalium*. The conservation score for each position represents the fraction of promoters of *M. pneumoniae* and *M. genitalium* that have that position conserved. In general, the conservation at the level of promoters is low except for the Pribnow box and surrounding bases. The red arrow marks the theoretical TSS position. Despite the spacer between the Pribnow and the TSS being variable across different genes, the theoretical +1 position shows higher conservation than the previous positions.

G. Upper panel shows an example of the RNA-seq profiles of the 5'-UTR riboswitch of the *oppB* gene, which belongs to the peptide transporter operon. Upon exposure to glycerol, which induces peroxide production, there is read-through in the 5'-UTR of *oppB* in the corB MPN159 KO. Rather than the sequence of the 5'UTR, its secondary structure is conserved with *M. genitalium* (low panel). Additional examples can be found in Online Figure 3 and Online Table 13.

H. Higher correlation values between genes sharing the same Pribnow box in their promoter. The histogram shows an example of a Pribnow box motif (TAAAAT). Genes with this Pribnow motif are more correlated to each other (green bars) than to the genes with different Pribnow boxes (red bars).

Figure S6. Supporting figure for Figure 4 and 5

Examples of differential transcriptional read-through in different conditions.

Figure S7. Supporting figure for Figure 6

A. Quantitation of biological noise in our perturbations. Two perturbations were tested in five independent experiments (diamide and glycerol). The Random Forest was calculated in each of the independent experiments and tested both in the same ("self") and in the other experiments ("others"). The boxplots depict the variance explained in each case, showing a small but significant decrease when the prediction is made on the other independent experiments.

B. Pribnow and CID randomization. Variance explained by the Pribnow motif or CID in all perturbations compared to the variance explained randomizing the Pribnow or CID 100 times (see Methods). Operon structure was maintained in the randomizations in order to determine the effect that is not due to operons.

C. Gene regulation by different mechanisms. All 689 genes of *M. pneumoniae* have been classified into 36 functional categories (Table S1), and each gene has been assigned its corresponding regulatory mechanisms (transcription factors, regulators, supercoiling, etc.). Different cellular processes can preferentially use one regulatory mechanism over another to control the expression of the implicated genes.

Figure S8. Supporting figure for Figures 3D and 5

Relative importances of the different factors for each group of clustered experiments (following clusters of correlated experiments obtained in Figure 3D). We fitted a Random Forest model to each experiment in the correlated clusters and observed that the relative importances of each of the features studied were more similar in experiments within the same cluster than across clusters. Then, we calculated the average relative importance of each feature for each cluster. The size and opacity of the dots reflect the variance explained by each feature in each group of experiments.

Table S1. Putative DNA- and RNA-binding proteins (supporting Table for Figure 1)

Gold sets of DNA- and RNA-binding proteins were constructed using information from the literature. For the gold set, we did not consider components of large machines that need to be assembled in order to bind to DNA (i.e., the DNA replication complex), components of DNA-binding complexes that do not contact DNA, and metabolic enzymes involved in nucleotide metabolism. We did include all proteins that bind to DNA and/or RNA in a direct manner (information from Uniprot or from PDB structures). The negative gold set is composed of membrane proteins and lipoproteins exposed to the medium. ID: identifier. 1: Protein present in the gold set. 0: Protein not present in the gold set. We also indicate those proteins not detected by mass spectrometry. TRUE in columns I to L means that the protein passed the ROC curve cut-off as a DNA/RNA-binding protein. DNA binding probability: the score is calculated by dividing the number of times a protein is found in the experiments (TRUE) by 4. Criterion 1, not to be further characterized (applied to all proteins with a score equal to or greater than 0.5): RNA-binding proteins, membrane-associated ATPases, proteins involved in DNA replication and maintenance, RNases requiring other components, and chaperones that are usually found as contaminants. In the case of duplicated proteins, we only cloned some members of the group. Criterion 2, cloned (despite having a score below 0.5, or being a TRUE RNA-binding protein): we selected some key metabolic enzymes of the main processes (nucleotides, glycolysis, lipids and fermentation) to study the link between metabolism and transcription. We also included some proteins as controls, true RNA-binding proteins described in the literature that possibly regulate transcription, as well as RNases that can specifically degrade RNA transcripts.

Table S2. Experiments for each gene and mutant (supporting Table for Figures 1, 2, 3)

Genes and constructs used in this study. ID is the MPN identifier and Name the biological name of the gene. Construct indicates which vector was used. pMT85 is the standard transposon vector used. Normally we use the Ef-tu promoter (tuf) to drive gene expression. The relative position of the FLAG-tag is indicated in the construct name (i.e., FLAG-XXX means that the flag is located at the N-terminus, XXX-FLAG means that the flag is located at the C-terminus). Tn indicates a knockout strain wehre the transposon

position has been identified and Tc is resistance to tetracycline. Type of mutant: OE overexpression, KO knockout mutant, Mut means point or deletion mutant (the number indicates the position of the mutated amino acid and the letters are the wildtype and mutated amino acid in one letter codes; if the position is not indicated, it is because there were multiple mutations), Fus means fusion and DN: dominant negative mutant (see Figure S3, for the description of the mutations). For microarrays, RNA-seq, Proteomics, and ChIP-seq conditions we show the different experiments performed: Growth in standard medium collected at different times (6 h, 24 h, 48 h, 72 h, 96 h). Addition of compounds to the standard medium (Gly: Glycerol; TL: thiolutin; DA: diamide; Px: H_20_2; Nvb: novobiocin, Shx: serine hydroxamate). The number in parenthesis shows biological replicas (all experiments have at least one technical replica). Growth Curve indicates that the growth curve of the mutant was determined. Log2 fold RNA is the fold change in RNA expression with respect to the WT strain upon OE. Estimated log2 fold protein: fold increase calculated from the mRNA fold change (ND, not determined) with a linear interpolation of the existing data. Essentiality (Lluch-Senar et al., 2015): E, essential; NE, non-essential; NE*, non-essential because they are duplicated; F, fitness (Lluch-Senar et al., 2015). NOTE: TF149 had a mutation in the FLAG-tag. See Online Table 1 for the used constructs.

Table S3. ChIP-seq specific peaks (supporting Table for Figures 2 and 3)

Consensus ChIP-seq peaks of the RNAP subunits and associated proteins, as well as of proteins with specific peaks (either structural or mapping to promoters). In total, the profiles of the POD (combined results of 6 and 24hours) and 23 DNA-binding proteins are shown. In the case of multiple biological replicates, peaks appearing in at least two of them are marked with an asterisk (*), and the largest peak height is shown. Different peaks located less than 100 bases apart were considered to bind the same region.

Table S4. TSS annotation by 5'-mapping (supporting Table for Figure 2)

Annotation of TSSs from 5'-mapping data (TSS column numbers indicate the genome position). Alternative TSSs are included (TSS2, etc.), as well as the relative position of the TSS to the expected canonical +1 base (Start), the ID of the first base, and the sequence (including the natural="theoretical" +1 base in red, the identified +1 base in bold (=TSS1), and the Pribnow box underlined). The “RNAP/SigA ChIP-seq peak"

column indicates the binding by Chip-seq of at least one core component of the RNAP subunits.

Table S5. Summary of gene changes per regulator or TF (supporting Table for Figures 3 and 4)

Gene changes produced by the OE/KOs/Mut/DN of the TFs and regulators described in this study, as well as their adjusted p-values are shown (only significant changes, see Methods). In those cases, where the OE and KO or any combination with DN and Mut, resulted in similar or mirror results, we determined the consensus (using the reverse values of the KO, DN and Mut when they mirrored the OE). For some genes, we could only see a phenotype upon addition of drugs (Shx is serine hydroxamate; DA is diamide) or after perturbations (glucose starvation). Online Table 6 shows all unfiltered changes and p-values for all the candidates tested in this study.

Table S6. Major perturbations and their effect on the transcriptome (supporting Table for Figures 3 and 4)

Thirty-seven major perturbations were defined by their effect (i.e., antibiotics were grouped by their type, such as macrolides or tetracyclines), given that they correlate well. For each of these major perturbations, we determined a consensus of the changes (see Methods). Significant changes are those marked with an asterisk (*).

Table S7. Phenotypic analysis (supporting Table for Figure 3)

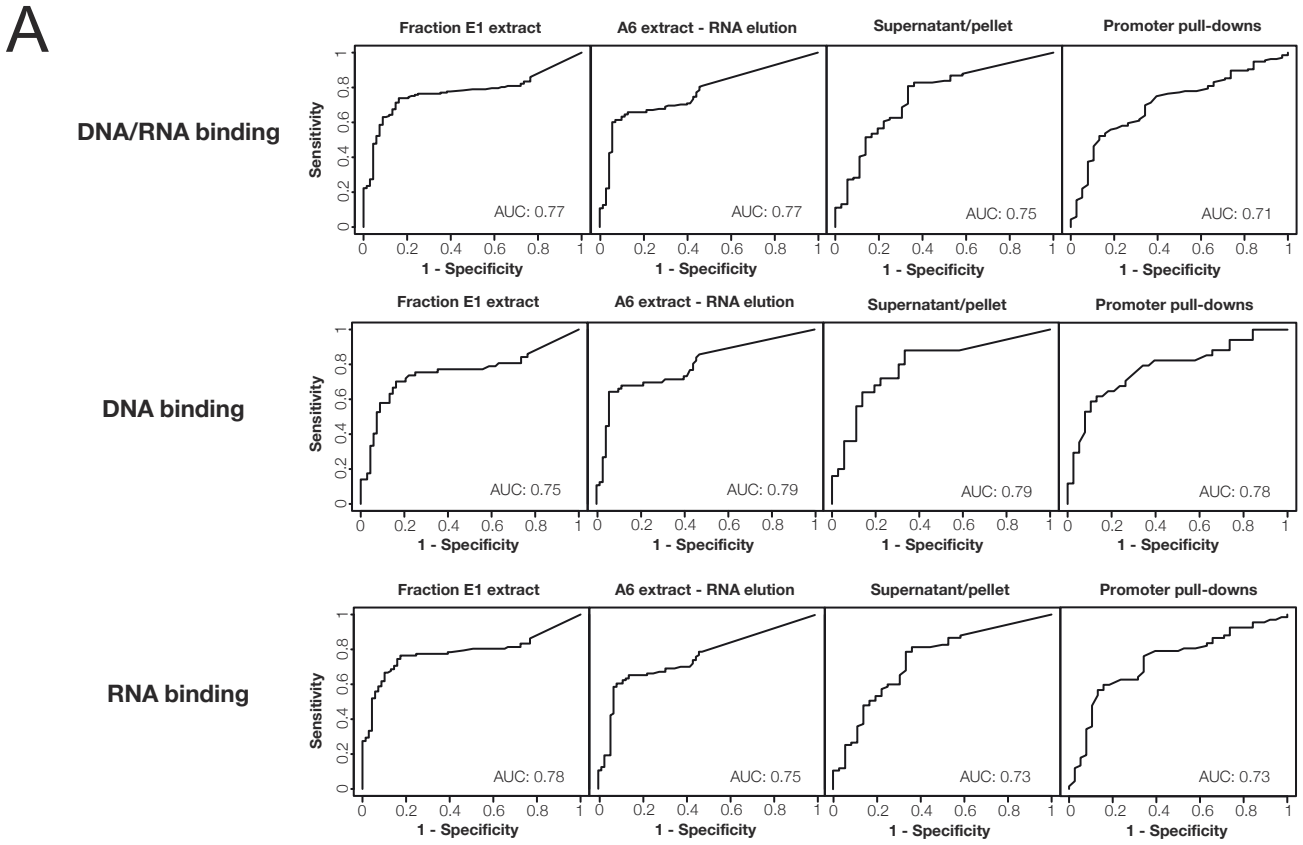
Qualitative phenotype of all generated strains as determined by growth curves (quantitative data can be found in Online Table 7). Two indicators were extracted from the pH-growth curves: growth (based on the protein biomass and the early slope of the growth curve) and metabolism (based the color of the medium at the highest point of the curve and the late slope). We also indicate the number of changes in each of the strains in transcriptomics experiments.

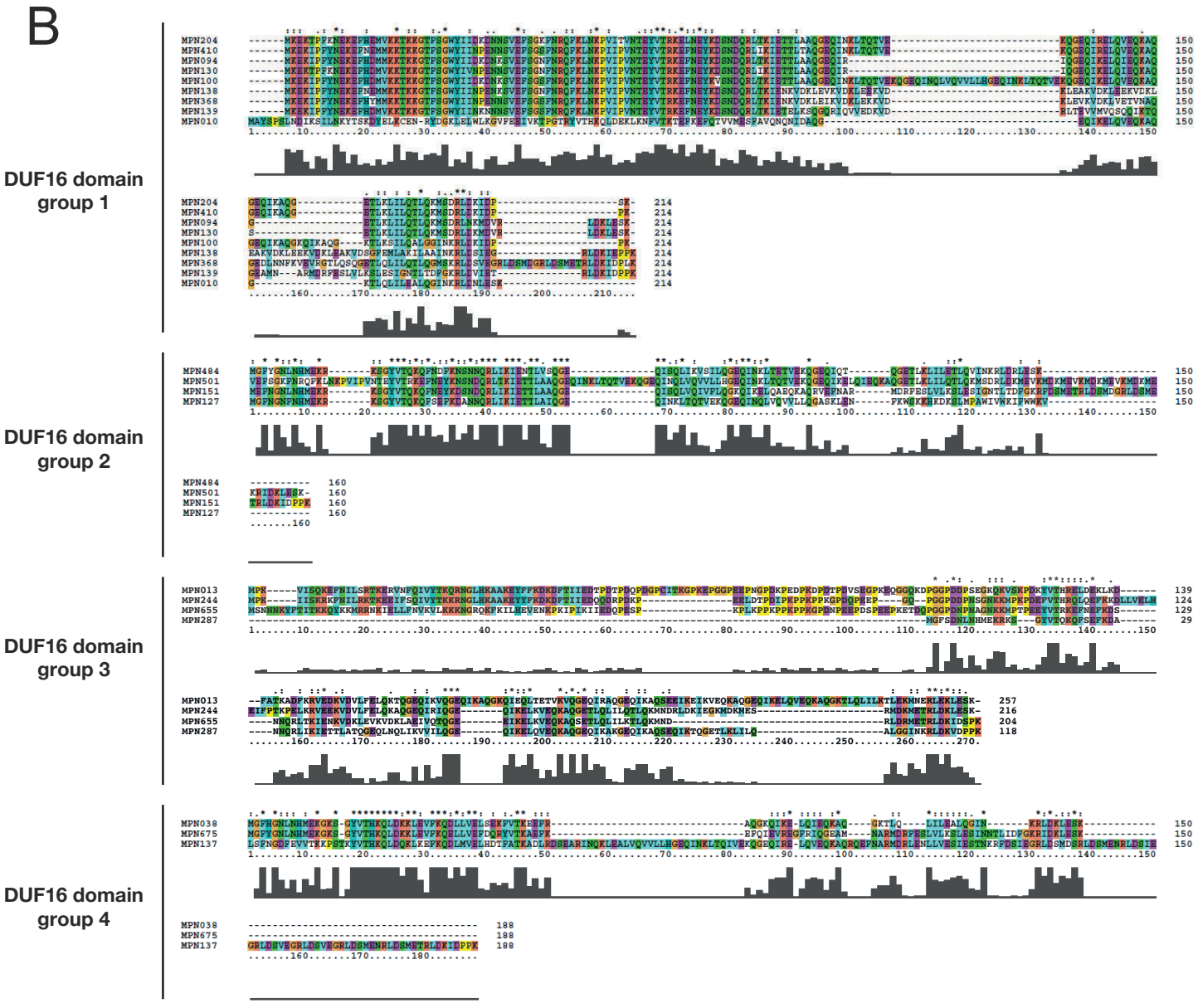
Table S8. Transcriptional regulation mechanisms (supporting Table for Figure 5 and Table 1)

Features regarding the promoter sequence (Start position, Pribnow motif, extended Pribnow, and AT% before the Pribnow box), the UTR (length, AT%), the transcript (TSS position relative to the Pribnow, -1/+1 nucleotides, +1/+2 nucleotides, and RNA decay

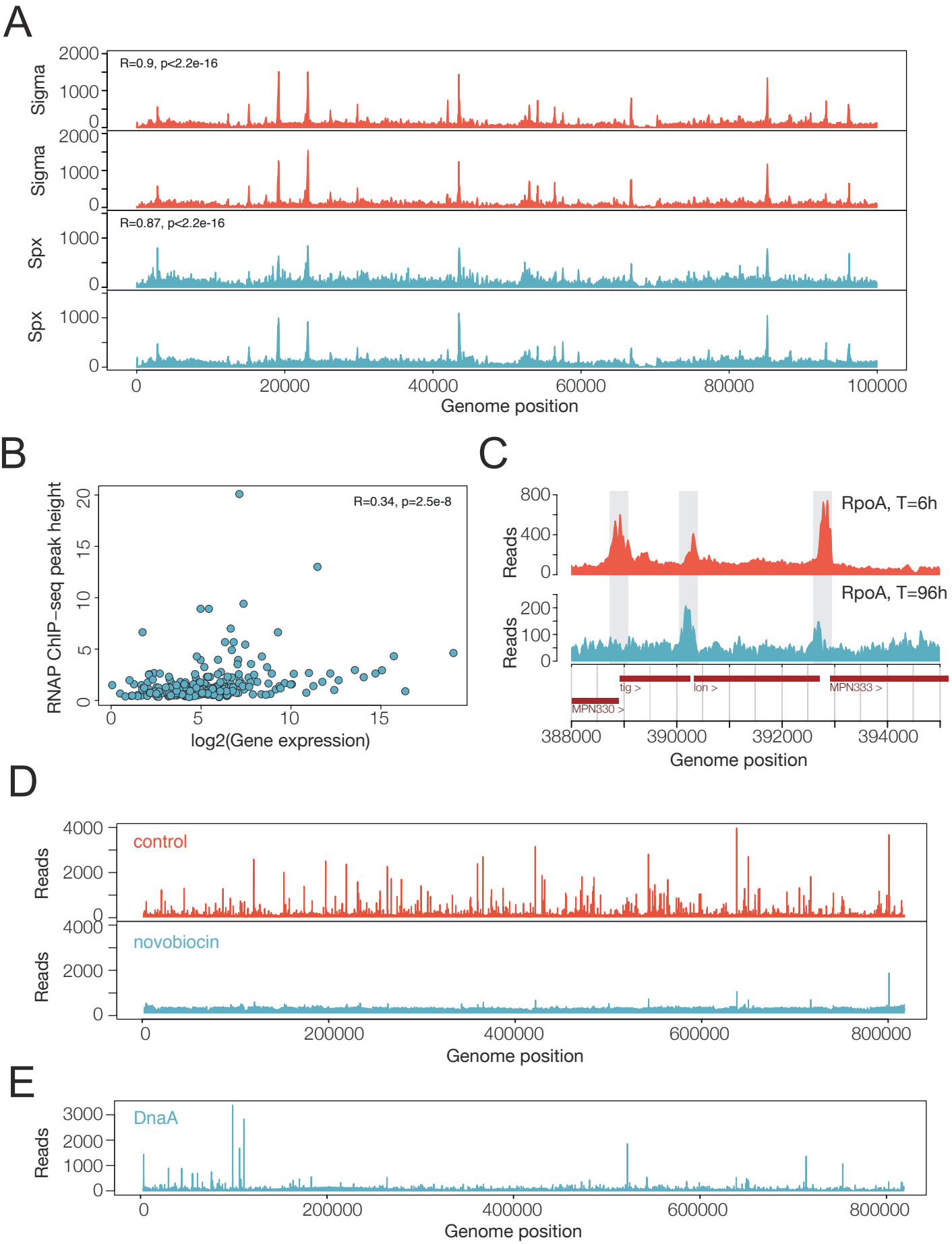
rate), the gene location in the chromosome (CID) and others (riboswitches, hairpins, transcriptional read-through, and behavior in supercoiling) are summarized in this table, together with the targets of the TFs and regulators. These features were considered in the Random Forest fit of each perturbation experiment to determine how much variance can be explained by each factor. The rows without content are genes that we excluded from the analysis because they had borderline expression or were too noisy.

Supplementary figure 1

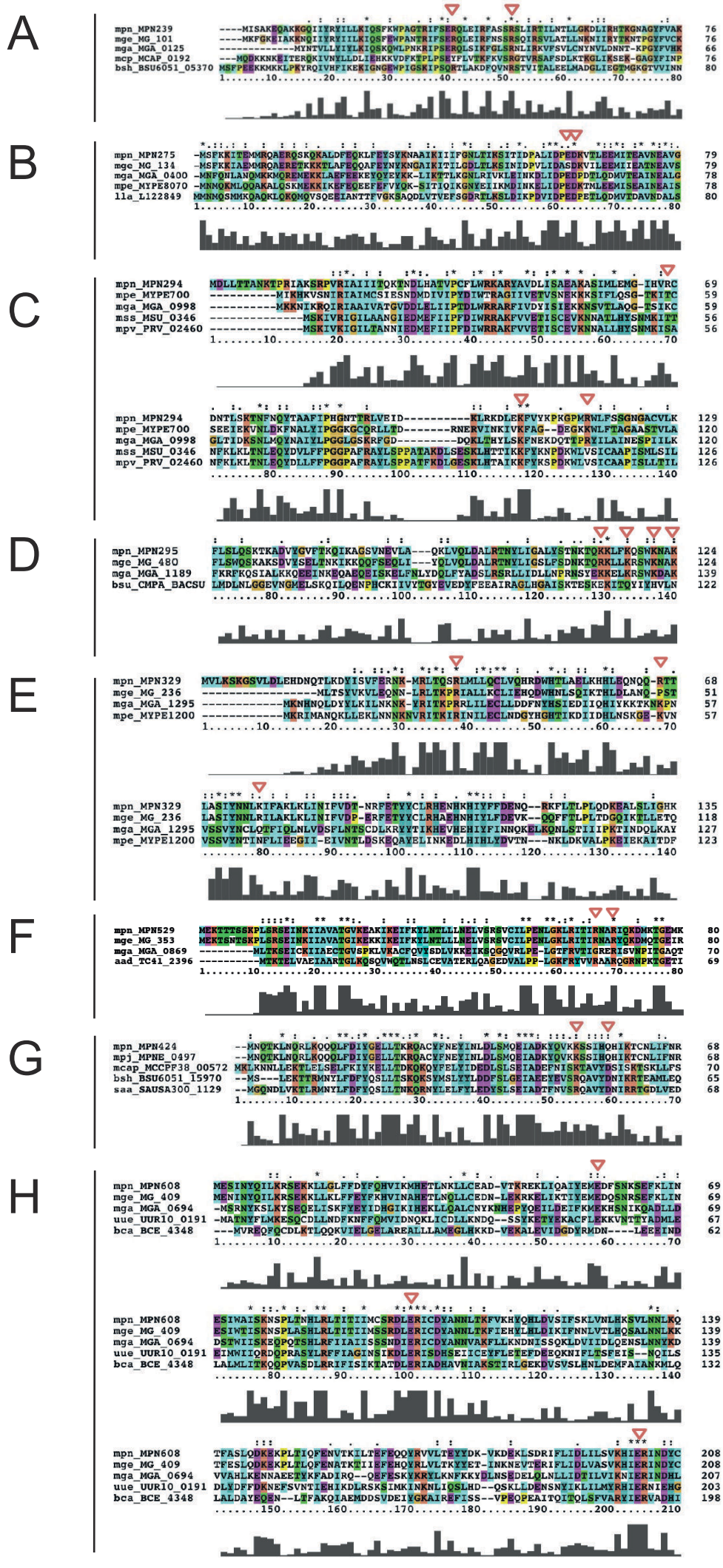




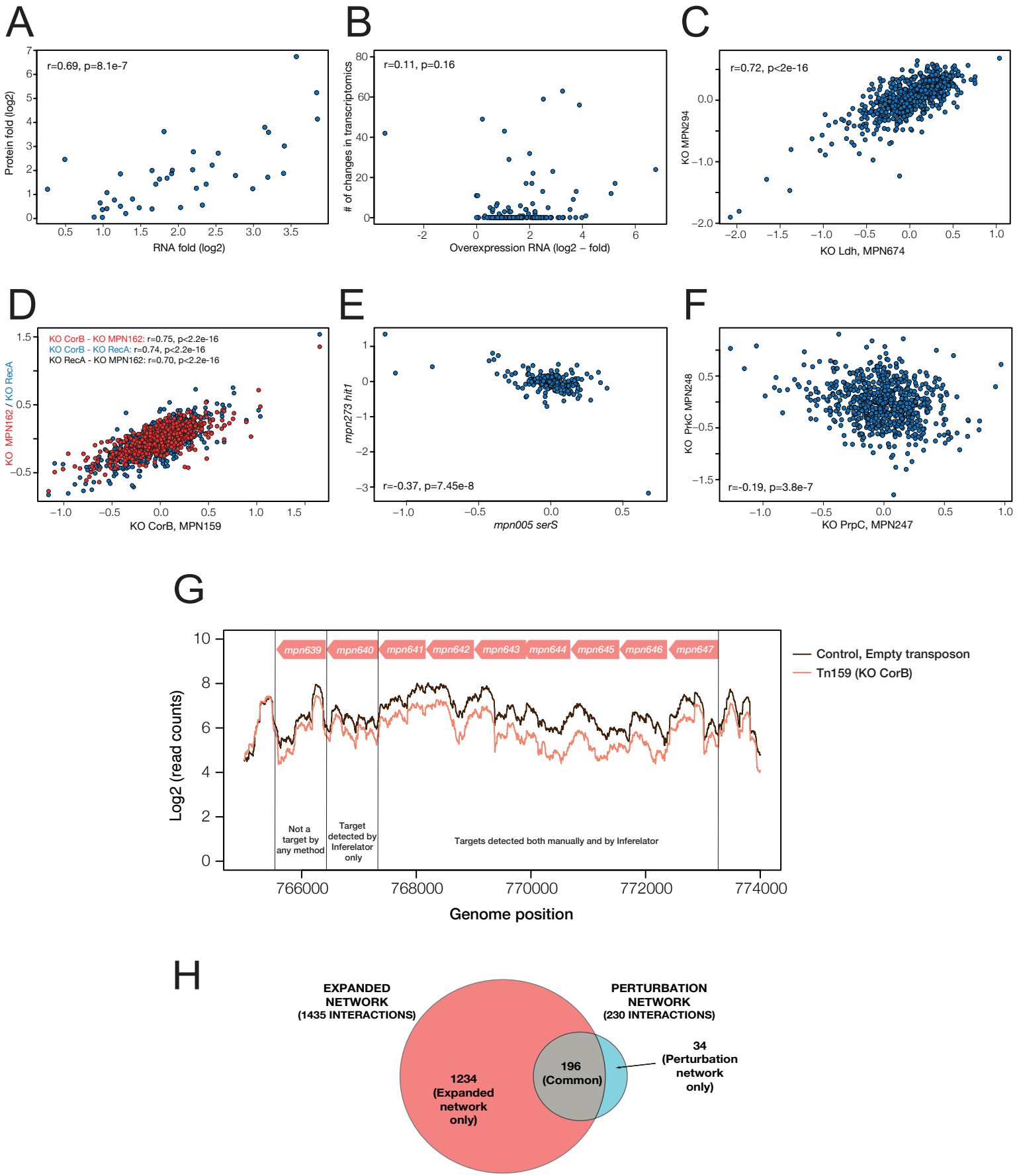
Supplementary figure 2



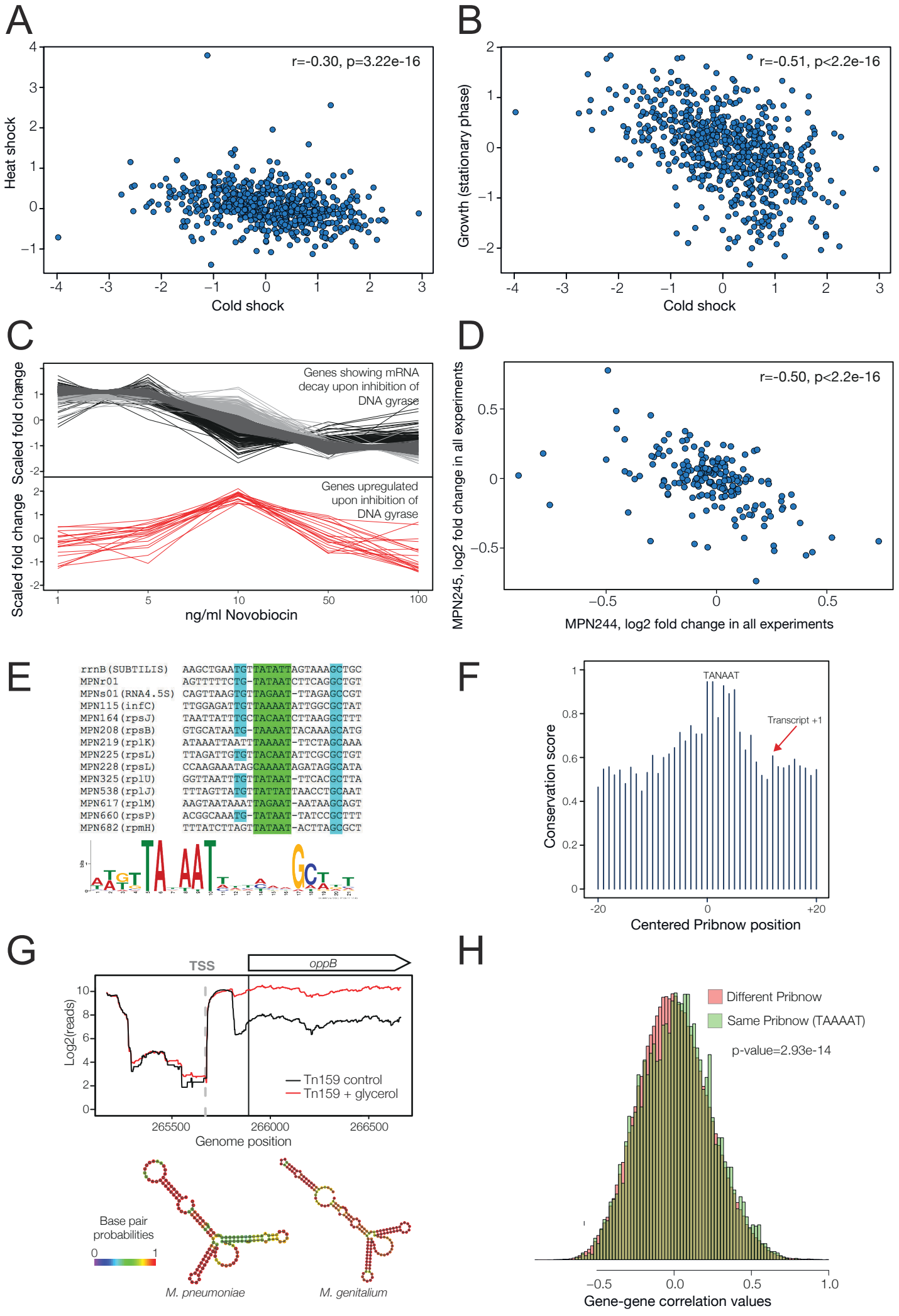
Supplementary figure 3



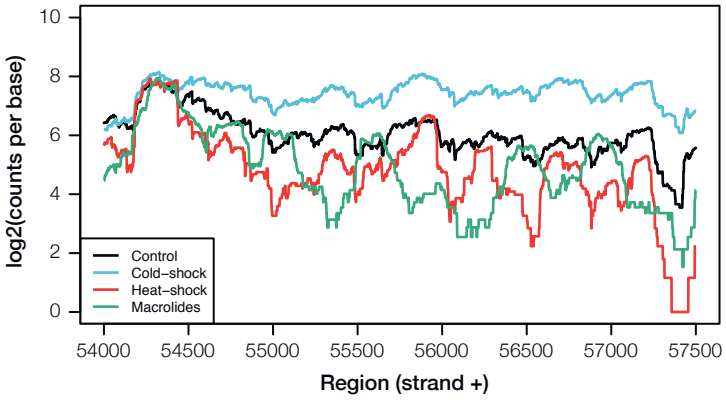
Supplementary figure 4

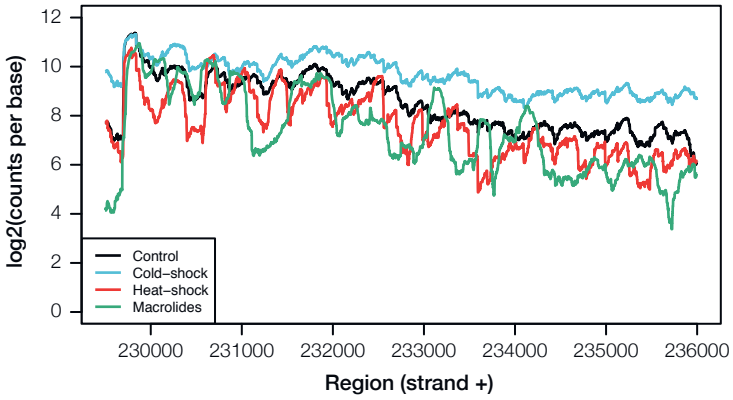


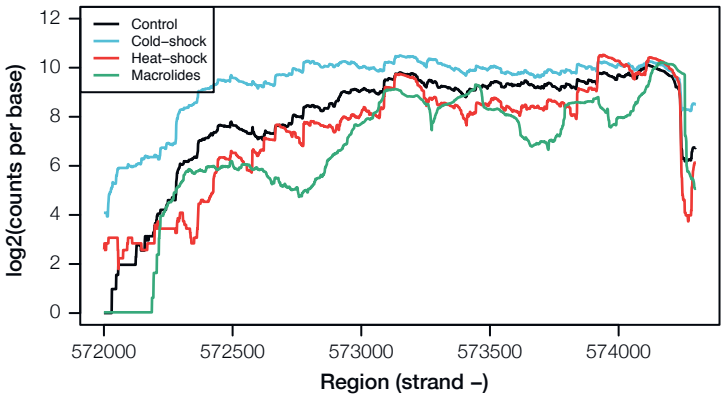
Supplementary figure 5

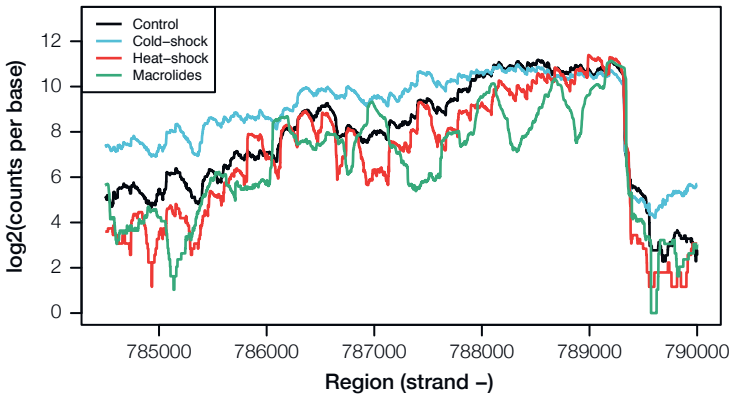


Supplementary figure 6

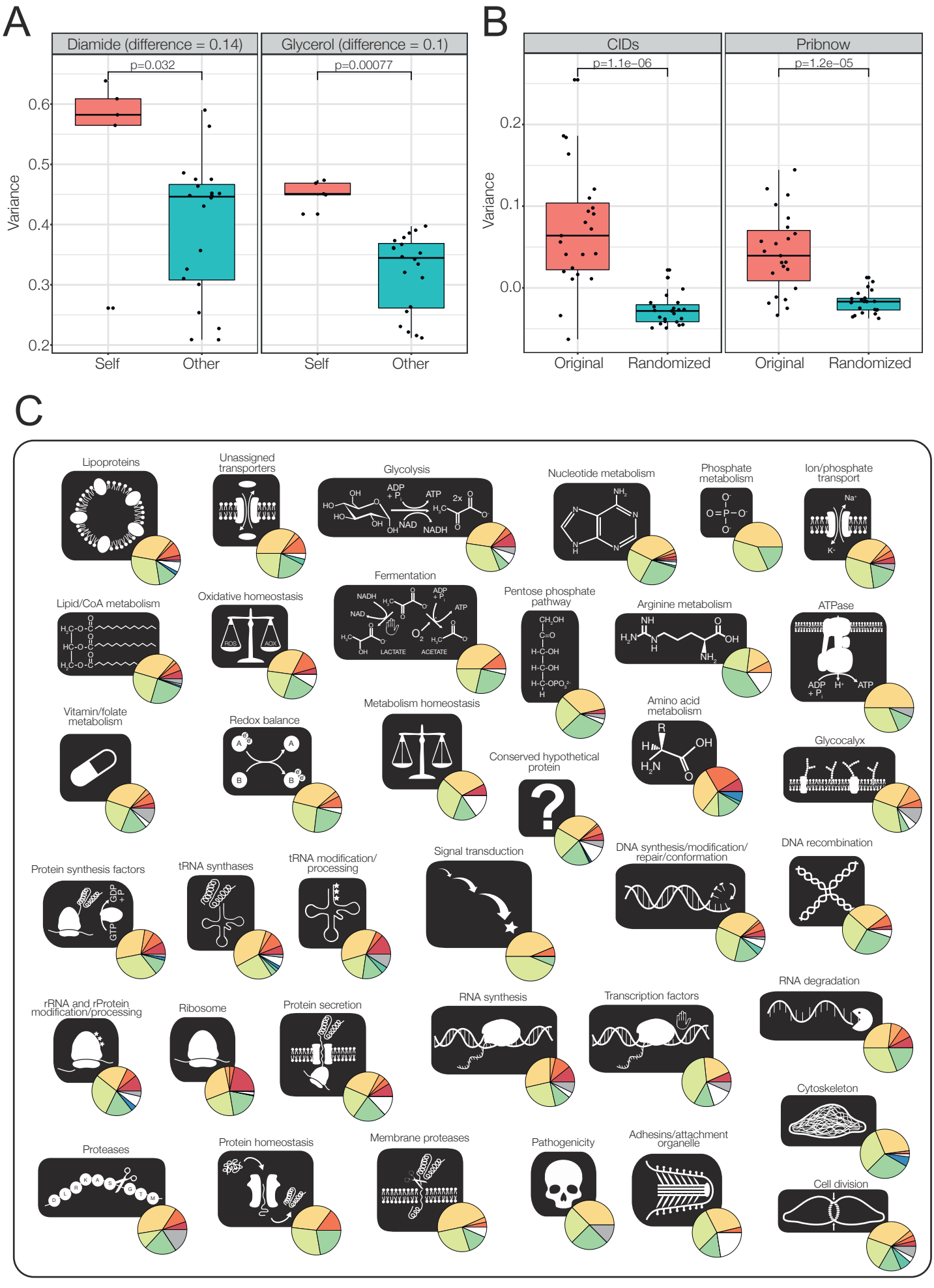








Supplementary figure 7



Supplementary figure 8

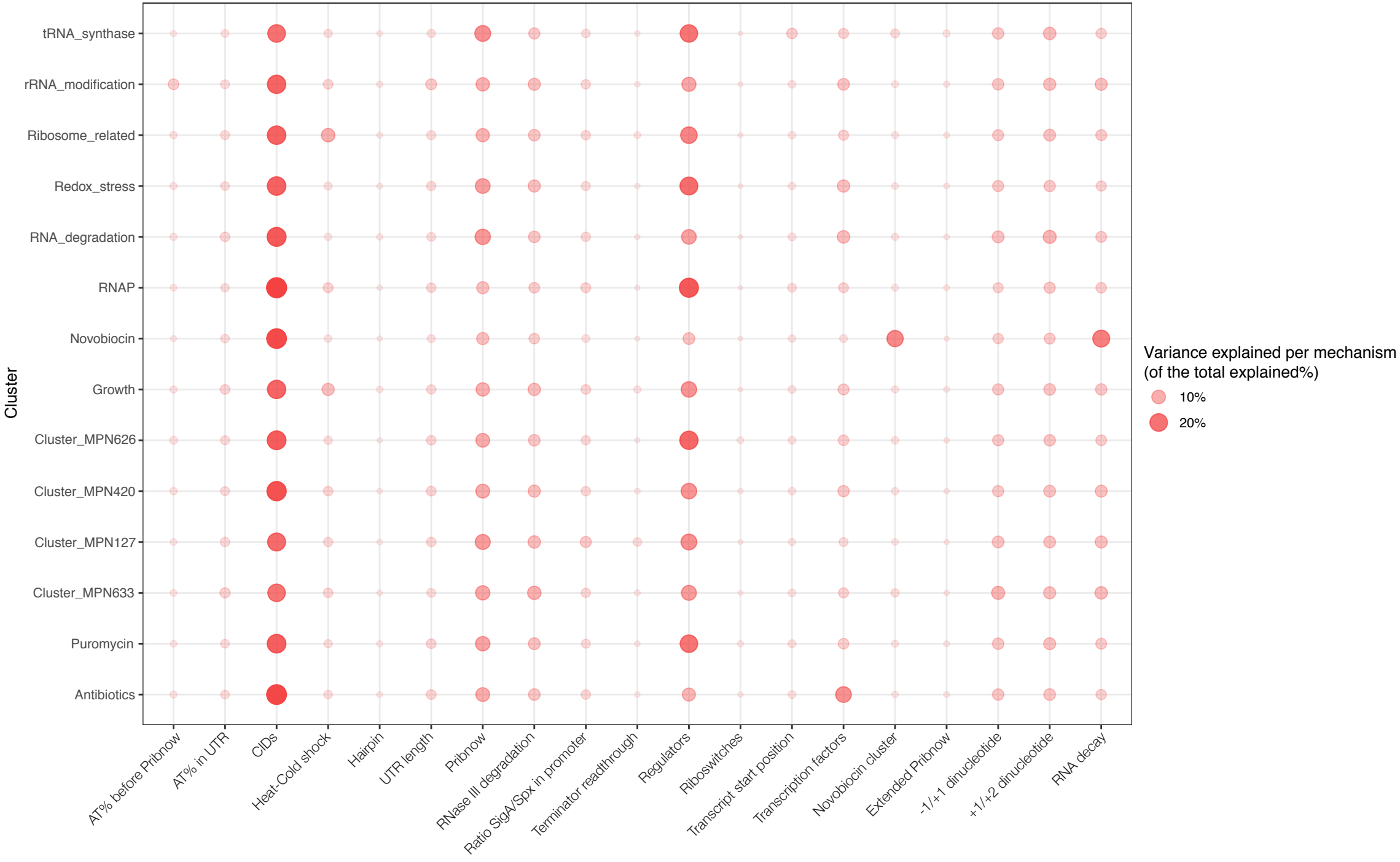


Table S7. Phenotypic analysis (supporting Table for Figure 3)

Strain (Gene overexpressed, mutated or KO)	Growth phenotype	Acidification phenotype	Changes in transcriptomics of this strain (<10 considered not significant)
MPN002	Slow growth	-	0
MPN004	Slow growth	-	0
MPN015	Fast growth	-	7
MPN020	-	Slow acidification	21
MPN024	-	-	0
MPN024_Mut	-	-	0
MPN027	-	-	54
MPN027_KO	-	-	4
MPN030	-	-	0
MPN032	-	-	0
MPN038	-	-	0
MPN051	-	Slow acidification	0
MPN051_KO	Slow growth	Slow acidification	4
MPN053	-	-	3
MPN053_Mut	-	-	7
MPN055	-	-	0
MPN063	-	-	17
MPN064	Fast growth	-	0
MPN067	-	-	0
MPN069	-	-	0
MPN076_KO	-	Slow acidification	0
MPN077_KO	-	-	0
MPN081	-	-	0
MPN082	Fast growth	Slow acidification	0
MPN106	Slow growth	-	0
MPN114	-	Slow acidification	0
MPN114_KO	-	Fast acidification	3
MPN119	-	-	0
MPN122	-	Slow acidification	0
MPN124_MUT	-	Fast acidification	24
MPN127	-	-	0
MPN133_KO	-	-	8
MPN140	-	-	5
MPN148	Fast growth	Fast acidification	0
MPN154	-	-	0
MPN159	-	-	0
MPN159_KO	-	-	48
MPN162_KO	-	Fast acidification	34
MPN164	-	Fast acidification	0
MPN165	-	-	0
MPN166	-	-	0
MPN168	-	-	0
MPN173	-	-	0
MPN178	-	-	0
MPN191	Fast growth	Fast acidification	51
MPN192	-	-	0
MPN194	-	-	0
MPN197	-	-	0
MPN208	-	-	0
MPN222	-	-	0
MPN223	Fast growth	-	0
MPN223_KO	Slow growth	-	0
MPN229	-	-	0
MPN239	Slow growth	Fast acidification	43

MPN241	Slow growth	Slow acidification	40
MPN243	-	-	0
MPN244	-	-	3
MPN246	-	-	0
MPN247	-	-	0
MPN247_KO	-	-	47
MPN248_KO	-	-	120
MPN250	-	-	0
MPN252	Fast growth	-	0
MPN255	Slow growth	Slow acidification	0
MPN263	-	-	4
MPN265	-	-	0
MPN266	-	-	0
MPN266_Mut1	Slow growth	Slow acidification	0
MPN266_Mut2	-	-	0
MPN266_Mut3	-	-	0
MPN269	-	-	1
MPN273	-	-	0
MPN275	-	-	0
MPN275_DN	-	-	0
MPN275_KO	-	-	6
MPN280	-	-	0
MPN284_KO	-	-	0
MPN287	-	-	0
MPN294	-	-	0
MPN294_DN	-	-	19
MPN294_KO	-	-	47
MPN295	-	-	2
MPN300	Slow growth	Slow acidification	2
MPN301	-	-	0
MPN303	Fast growth	Slow acidification	0
MPN314	-	-	5
MPN314_Mut	-	-	4
MPN315	-	-	0
MPN316	-	-	0
MPN329	-	-	4
MPN330	-	-	0
MPN332	-	-	4
MPN348	-	-	0
MPN349	-	-	4
MPN352	-	-	0
MPN368	-	-	0
MPN372_KO	-	-	1
MPN397	Fast growth	Fast acidification	2
MPN397_KO	-	-	27
MPN397_Mut1	-	Slow acidification	0
MPN397_Mut2	-	-	0
MPN397_Mut3	-	-	75
MPN400	-	-	0
MPN401	-	-	0
MPN420	Fast growth	-	17
MPN420_KO	Fast growth	Slow acidification	151
MPN421_KO	Slow growth	Slow acidification	3
MPN424	-	-	0
MPN424_DNKO	-	-	5
MPN426	-	Slow acidification	0
MPN428	Fast growth	Fast acidification	6
MPN430	-	-	0
MPN440	-	-	0
MPN443	-	-	0
MPN446	-	-	0

MPN473	Slow growth	Slow acidification	9
MPN475	-	Slow acidification	0
MPN478	-	-	0
MPN481	-	-	0
MPN482	-	-	0
MPN484	-	-	0
MPN485	-	-	0
MPN487	-	-	0
MPN490	-	Fast acidification	57
MPN490_KO	Fast growth	Fast acidification	26
MPN499	-	-	0
MPN506_KO	-	-	29
MPN507	-	Slow acidification	0
MPN516	-	-	11
MPN518	-	-	0
MPN525	-	-	0
MPN526	-	-	1
MPN529	-	-	0
MPN545	-	-	43
MPN545_KO	-	Fast acidification	587
MPN547	-	-	0
MPN549	-	-	0
MPN554	-	-	0
MPN555	-	-	0
MPN559	-	-	0
MPN563	-	-	0
MPN566_KO	Fast growth	-	2
MPN568	-	-	0
MPN569	-	-	0
MPN572	-	-	6
MPN574	-	-	0
MPN576	-	-	3
MPN590	-	-	0
MPN606	-	-	0
MPN608	Slow growth	Slow acidification	0
MPN615	-	Slow acidification	0
MPN617	-	-	0
MPN621	-	-	0
MPN626	-	-	25
MPN627	-	Fast acidification	0
MPN633	-	-	0
MPN633.634	-	-	42
MPN634	-	-	22
MPN635	-	-	0
MPN638	-	-	3
MPN651	-	Fast acidification	7
MPN663	-	-	0
MPN667	-	Slow acidification	4
MPN673	-	-	0
MPN674	-	-	3
MPN674_KO	-	Slow acidification	51
MPN677	-	-	0
MPN683	-	-	0
MPN686	-	Slow acidification	17
FP Venus (negative contro	-	-	0