Supplementary data for Characterization of Bacterial Transcriptional Regulatory Networks in *Escherichia coli* through Genome-Wide *in vitro* Run-Off Transcription/RNA-SEq (ROSE)

**Table S1.** Technical comparison between the two genome-wide *in vitro* transcriptome sequencing techniques ROSE and RIViT-seq [1].

|  |  |  |
| --- | --- | --- |
|  | ROSE | RIViT-seq |
| Isolation of genomic DNA | * Three different isolation methods tested (Quick-DNA Universal Kit by Zymo, NucleoSpin Microbial DNA Kit by Macherey Nagel and Phenol-Chloroform Isoamyl alcohol DNA extraction)
 | * GenElute Bacterial Genomic DNA kit (MilliporeSigma)
 |
| Digestion of genomic DNA | * Random fragmentation (average size 6 kb) using gTubes (Covaris)
 | * Enzymatic digestion with *Eco*RI, *Hind*III, *Bam*HI and *Xho*I
 |
| RNA purification after *in vitro* transcription | * RNeasy MinElute Kit (Qiagen)
 | * RNeasy MinElute Kit (Qiagen)
 |
| Primary transcript library generation | * 100 ng RNA input
* Fragmented to an average size of 500 nt
* Digestion of transcripts with 5’ di- and monophosphates by terminator exonuclease (Epicentre)
* Ligation of RNA index adapters
* Treatment with RNA 5’-polyphosphatase (New England Biolabs)
* Ligation of RNA adapters
* Reverse-transcription to cDNA using a sequence-independent loop adapter
* Amplification of cDNA with barcoded primers
 | * 6 µg RNA input
* /
* /
* /
* Treatment with RNA 5’-polyphosphatase (New England Biolabs)
* Ligation of RNA adapters
* Reverse-transcription to cDNA using a random hexamer with an adaptor sequence
* Amplification of cDNA with barcoded primers
 |
| Whole-genome transcriptomics | * /
 | * Stranded cDNA library generation using TrueSeq Stranded RNA Library Prep Kit (Illumina) and low sample protocol
 |
| Sequencing and data processing | * High throughput single-end sequencing (1x75bp)
* Illumina MiSeq
* Trimming using Trimmomatic
* Mapping with Bowtie2
 | * High throughput paired-end sequencing (2x150bp)
* Illumina MiSeq
* Trimming using BBDuk
* Mapping with HISAT2
 |
| Sequence analysis  | * TSS detection and visualization with ReadXplorer, Improbizer, and MEME/Weblogo
 | * TSS detection manually with read counts of samtools, motif visualization with MEME
* Differential Expression analysis with DESeq2
 |

**Table S2.** Mapping statistics for all six 5’-end specific ROSE- Eσ70 libraries.Mapping statistics for ROSE- Eσ70 libraries with the isolation methods by Zymo Research (Z1; Z2), Macherey-Nagel (M1; M2) and with Phenol-Chloroform Isoamyl alcohol (P1; P2).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Z1 | Z2 | M1 | M2 | P1 | P2 |
| Mappings |  | 2,061,241 | 2,086,867 | 1,618,341 | 2,076,393 | 2,210,864 | 1,737,821 |
| Unique Mappings |  | 2,061,241 | 2,086,867 | 1,618,341 | 2,076,393 | 2,210,864 | 1,737,821 |
| Single Perfect Mappings |  | 1,572,385 | 1,553,410 | 1.240,224 | 1,533,963 | 1,680,056 | 1,305,272 |

**Table S3.** Mapping statistics for four 5’-end specific *in vivo* libraries.Mapping statistics for *in vivo* libraries of the wildtype strain (WT), the *Δfur* knockout strain (*Δfur*), the *Δfis* knockout strain (*Δfis*) and of the *Δhns* knockout strain (*Δhns*).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | WT | *Δfur* | *Δfis* | *Δhns* |
| Mappings | 1,605,710 | 928,706 | 493,254 | 597,949 |
| Unique Mappings | 1,605,710 | 928,706 | 493,254 | 597,949 |
| SinglePerfectMappings | 1,288,944 | 760,574 | 344,796 | 438,708 |

**Table S4.** Transcription start site detection parameters for ROSE-Eσ70 libraries in ReadXplorer [2,3].

|  |  |
| --- | --- |
| Parameter | Value |
| Minimum number of read starts: | 7 |
| Minimum percent of coverage increase: | 148 |
| Maximum low coverage read start count: | 0 |
| Minimum low coverage read starts: | 0 |
| Detect novel transcripts? | Yes |
| Minimum transcript extension coverage: | - |
| Maximum distance to feature of leaderless transcripts: | 300 |
| Associate nearby neighboring TSS? | Yes |
| Associate neighboring TSS in a bp window of: | 3 |
| Minimum mapping quality: | 0 |
| Single Perfect Match included: | Yes |
| Perfect Match included: | No |
| Single Best Match included: | Yes |
| Best Match included: | No |
| Common Match included: | No |
| Include multiple mapped reads: | Yes |
| Mapping strand selection: | Feature/analysis strand |

**Table S5.** Transcription start site detection parameters for *in vivo* libraries in ReadXplorer [2,3].

|  |  |
| --- | --- |
| Parameter | Value |
| Minimum number of read starts: | 1 |
| Minimum percent of coverage increase: | 32 |
| Maximum low coverage read start count: | 0 |
| Minimum low coverage read starts: | 0 |
| Detect novel transcripts? | Yes |
| Minimum transcript extension coverage: | 20 |
| Maximum distance to feature of leaderless transcripts: | 300 |
| Associate nearby neighboring TSS? | Yes |
| Associate neighboring TSS in a bp window of: | 3 |
| Minimum mapping quality: | 0 |
| Single Perfect Match included: | Yes |
| Perfect Match included: | No |
| Single Best Match included: | Yes |
| Best Match included: | No |
| Common Match included: | No |
| Include multiple mapped reads: | Yes |
| Mapping strand selection: | Feature/analysis strand |



**Figure S1.** Mapped reads of the 5’-end-specific transcript library in exemplary genomic regions.The reads (black) are filtered and mapped to the reference genome U00096.3. The coverage of the reads associated to their respective location on the genome is visualized with ReadXplorer [2]. (A) Representation of the mapped read distribution in the genomic region 110,000-150,000 bp of the *E. coli* genome. The genomic DNA for the ROSE-Eσ70 experiment was isolated with Phenol-Chloroform Isoamyl alcohol (top), with the NucleoSpin Microbial DNA Kit from Macherey-Nagel (middle) and with the Quick-DNA Universal Kit from Zymo (bottom). (B) Exemplary read stack for the *rpsU-dnaG-rpoD* operon generated by ROSE-Eσ70. (C) Exemplary read stacks for the genes *nadB* on the sense strand and *sigE* on the antisense strand generated by ROSE-Eσ70.



**Figure S2.** Distribution of the identified transcription start sites (TSS) by ROSE and by Thomason *et al.* in relation to their distance to the published TSS in RegulonDB [4].



**Figure S3.** Coverage of mapped reads on the reference genome with an emphasis on gene *stpA.*Readcount in the promoter region of *stpA* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*hns* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].



**Figure S4.** Coverage of mapped reads on the reference genome with an emphasis on gene *ftnA.*Readcount in the promoter region of *ftnA* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*hns* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].



**Figure S5.** Coverage of mapped reads on the reference genome with an emphasis on gene *yjjZ*. Readcount in the promoter region of *yjjZ* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*fur* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].



**Figure S6.** Coverage of mapped reads on the reference genome with an emphasis on gene *fepA*. Readcount in the promoter region of *fepA* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*fur* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].



**Figure S7.** Coverage of mapped reads on the reference genome with an emphasis on gene *glcC*.Readcount in the promoter region of *glcC* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*fis* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].



**Figure S8.** Coverage of mapped reads on the reference genome with an emphasis on gene *aer.*Readcount in the promoter region of *aer* from the *E. coli* ROSE-Eσ70 (top), *E. coli in vivo* Wildtype strain (middle) and *E. coli in vivo* Δ*fis* knockout strain (bottom). The mapping took place on the respective reference genome (U00096.3) and is visualized with ReadXplorer [2].

References

1. Otani, H.; Mouncey, N.J. RIViT-seq enables systematic identification of regulons of transcriptional machineries. *Nat. Commun.* **2022**, *13*, 3502, doi:10.1038/s41467-022-31191-w.

2. Hilker, R.; Stadermann, K.B.; Doppmeier, D.; Kalinowski, J.; Stoye, J.; Straube, J.; Winnebald, J.; Goesmann, A. ReadXplorer--visualization and analysis of mapped sequences. *Bioinformatics* **2014**, *30*, 2247–2254, doi:10.1093/bioinformatics/btu205.

3. Hilker, R.; Stadermann, K.B.; Schwengers, O.; Anisiforov, E.; Jaenicke, S.; Weisshaar, B.; Zimmermann, T.; Goesmann, A. ReadXplorer 2-detailed read mapping analysis and visualization from one single source. *Bioinformatics* **2016**, *32*, 3702–3708, doi:10.1093/bioinformatics/btw541.

4. Santos-Zavaleta, A.; Salgado, H.; Gama-Castro, S.; Sánchez-Pérez, M.; Gómez-Romero, L.; Ledezma-Tejeida, D.; García-Sotelo, J.S.; Alquicira-Hernández, K.; Muñiz-Rascado, L.J.; Peña-Loredo, P.; et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* **2019**, *47*, D212-D220, doi:10.1093/nar/gky1077.