Elucidating the CodY regulon in *Staphylococcus aureus* USA300 substrains

Ye Gao[1,2], Saugat Poudel[2], Yara Seif[2], Zeyang Shen[2], Bernhard O. Palsson[2,3,4,5,*]

1 Department of Biological Sciences, University of California San Diego, La Jolla, CA, 92093, USA
2 Department of Bioengineering, University of California San Diego, La Jolla, CA, 92093, USA
3 Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093, USA
4 Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA 92093, USA.
5 Novo Nordisk Foundation Center for Biosustainability, 2800, Kongens Lyngby, Denmark

*Correspondence should be addressed to: Palsson O. Bernhard, Email: palsson@ucsd.edu.

Abstract

CodY is a conserved broad acting transcription factor that regulates the expression of genes related to amino acid metabolism and virulence in methicillin-resistant *Staphylococcus aureus* (MRSA). CodY target genes have been studied by using *in vitro* DNA affinity purification and deep sequencing (IDAP-Seq). Here we performed the first *in vivo* determination of CodY target genes using a novel CodY monoclonal antibody in established ChIP-exo protocols. Our results showed, 1) the same 135 CodY promoter binding sites regulating 165 target genes identified in two closely related virulent *S. aureus* USA300 TCH1516 and LAC strains; 2) The differential binding intensity for the same target genes under the same conditions was due to sequence differences in the same CodY binding site in the two strains; 3) Based on transcriptomic data, a CodY regulon comprising 72 target genes that are differentially regulated relative to a CodY deletion strain, representing genes that are mainly involved in amino acid transport and metabolism, inorganic ion transport and metabolism, transcription and translation, and virulence; and 4) CodY systematically regulated central metabolic flux to generate branched-chain amino acids (BCAAs) by mapping the CodY regulon onto a genome-scale metabolic model of *S. aureus*. Our study performed the first system-level analysis of CodY in two closely related USA300 TCH1516 and LAC strains giving new insights into the similarities and differences of CodY regulatory roles between the closely related strains.

Importance

With the increasing availability of whole genome sequences for many strains within the same pathogenic species, a comparative analysis of key regulators is needed to understand how the different strains uniquely coordinate metabolism and expression of virulence. To successfully infect the human host, *Staphylococcus aureus* USA300 relies on the transcription factor CodY to reorganize metabolism and express virulence factors. While CodY is a known key transcription factor, its target genes are not characterized on a genome-wide basis. We performed a comparative analysis to describe the transcriptional regulation of CodY between two dominant USA300 strains. This study motivates the characterization of common pathogenic strains and an evaluation of the possibility of developing specialized treatments for major strains circulating in the population.

Key words

Staphylococcus aureus, CodY, Transcription factor binding sites, Metabolism

Introduction

Staphylococcus aureus is a ubiquitous, Gram-positive pathogen that causes a diverse range of bacterial infections, from skin and soft tissue infections to potentially fatal infections, such as pneumonia, endocarditis, osteomyelitis, sepsis, and toxic shock syndrome (1). Coupled with the growing prevalence of methicillin-resistant *S. aureus* (MRSA) strains, the worldwide threat posed by this pathogen is obvious (2). Current research provides insights into important features of these strains, including antibiotic resistance and extensive virulence factors. So far, a number of known virulence factors consist of surface-associated proteins, such as microbial surface components, and secreted proteins, like hemolysins, immunomodulators, and many exoenzymes (3). To combat the worldwide spread of *S. aureus*, significant effort is being focused on the investigation of the transcription factors that control virulence factors during infection (4, 5). With different strains circulating in the population, it is important to understand the fundamental differences between them.

CodY is an important broad acting transcription factor in *S. aureus* (6, 7). While the primary known role of CodY is to regulate the metabolic genes in response to cellular branched chain amino acid and GTP concentrations, it also controls the expression of several virulence factors, acting as a bridge between metabolism and virulence (6, 8–11). The CodY regulon, estimated to consist of 150 to more than 200 genes, can vary in size between *S. aureus* strains, and CodY can even have opposite effects on the expression of the same gene in different strains (10, 12, 13). At present, direct determination of CodY

87 DNA-binding sites using ChIP-exo and the identification of the corresponding target
88 genes has not been achieved.
89
90 To address this challenge, we developed a reliable ChIP-grade monoclonal antibody and
91 conducted in *vivo* genome-wide experiments (ChIP-exo) to identify 165 identical CodY
92 target genes in two common *S. aureus* USA300 isolates (TCH1516 and LAC). To
93 reconstruct the CodY regulon, we compared RNA-seq profiles of the wild-type strain and
94 *codY* mutant. To examine the network level effects of CodY, we used a genome-scale
95 metabolic model to simulate the flux state of central carbon metabolism, demonstrating
96 the regulatory activities of CodY to generate BCAA. Our study used a comprehensive
97 pipeline to characterize the CodY regulon between closely related USA300 substrains
98 that included genetic parameters and network level computational models.
99
100
101

Results

Comparative genomic analysis of *S. aureus* USA300 reveals a high level of identity between two closely related strains, TCH1516 and LAC

The community-associated methicillin-resistant *S. aureus* (CA-MRSA) clones belonging to the USA300 lineage have become the dominant sources of MRSA in the United States (14). They are distinct from other *S. aureus* clones such as USA200 (UAMS-1), USA400 (MW2), and ST8 MSSA (Newman) (14). Two dominant CA-MRSA USA300 strains, TCH1516 and LAC, isolated in Los Angeles and Houston, respectively, were chosen to study CA-MRSA clones as they represent well characterized strains in the USA300 lineage (15). These strains have become the epidemic clones spreading in the community (16).

First, using whole-genome alignment, we characterized the genomes of TCH1516 and LAC, with lengths of ~2.872 Mb and ~2.878 Mb, respectively (**Table 1**). We found a high level of identity at the nucleotide level between these two closely related strains, and on average, they displayed 99.5% identity for all coding genes (**Figure 1A**). Further, MUMmer was used to check for the synteny between the strains. We produced a MUMmer dot plot resulting from the alignment of their chromosomal sequences, demonstrating that they have high similarity between the genomes and large scale inversions located at positions ~ 2.18 Mb (TCH1516) and ~ 2.20 Mb (LAC) (**Figure 1B**).

Overall, these data revealed that they have a high degree of sequence similarity throughout their genome sequences.

Genome-wide identification of CodY-binding sites in *S. aureus* USA300 TCH1516

Previously, 57 CodY-binding sites in the *S. aureus* clinical strain UAMS-1 were identified *in vitro* deploying an affinity purification experiment using purified *S. aureus* His-tagged CodY and a related mutant strain (8). However, there are no direct *in vivo* measurements of the interaction between CodY and DNA in the recent USA300 isolates. Thus, using a novel monoclonal antibody, we performed Chromatin immunoprecipitation followed by exonuclease digestion (ChIP-exo) to identify the CodY-binding sites with single nucleotide resolution in *S. aureus* USA300 TCH1516 under RPMI with 10% LB medium (17).

Using a peak calling algorithm (MACE), a total of 165 CodY target genes (135 binding peaks) were identified from TCH1516 (**Figure 2A, upper panel**). As mentioned before, there were 57 CodY target genes identified in *S. aureus* UAMS-1 using IDAP-Seq, 57% (36/57) of which were also detected in TCH1516. Compared to the IDAP-Seq approach, this study detected the interaction between TF-DNA *in vivo* to enrich the DNA bound by CodY in its natural state, but also showed the location of each peak at single nucleotide resolution.

Our results showed that 88% (144/165) of CodY binding sites were located within the intergenic regions, and the remaining 12% of binding sites were found in coding regions

(**Figure 2B, upper panel**). Most of the binding sites located in intergenic regions were present upstream of assigned genes, indicating that CodY may play critical regulatory roles in the expression of these genes. A total of 128 novel CodY target genes were identified (**Figure 2B, bottom panel**). These findings expanded the list of CodY target genes in TCH1516 and enabled a better understanding of the global regulatory role of CodY in the USA300 lineage.

Identification of the CodY-binding motif in *S. aureus* TCH1516

To identify the DNA sequence motif of CodY-binding sites, we used the MEME motif-searching algorithm with the genomic sequences of binding sites, and then identified the conserved 20-bp CodY binding motif, which was consistent with the previously characterized CodY DNA-binding consensus sequence (AATTTTCWGAAAATT) in *S. aureus* UAMS-1 (**Figure 2C**). Furthermore, the *S. aureus* TCH1516 CodY binding motif is similar with the CodY binding motif (AATTTTCWGAATATTCWGAAAATT) reported in *Listeria monocytogenes* and *Bacillus subtilis* (18, 19). These results suggest that CodY likely has a conserved DNA-binding domain in Gram-positive bacteria (**Supplementary Figure 1**).

Comparison of *in vivo* CodY-binding sites in *S. aureus* TCH1516 and LAC strains

As two community-associated methicillin-resistant strains, *S. aureus* TCH1516 and LAC have an identical CodY sequence (**Supplementary Figure 2**). To investigate the direct

gene targets of CodY, we employed the same monoclonal antibody to perform the ChIP-exo assay in *S. aureus* LAC under RPMI with 10% LB medium (17).

S. aureus LAC has 165 CodY target genes identified at the genome that were identical to those from the TCH1516 strain, though their positions at each chromosome are different due to the inversions mentioned earlier (**Figure 2A, bottom panel**). The alignment of the binding motifs revealed that the two strains have 18 nucleotides overlap (*p*-value= 9.35e-09) (**Supplementary Figure 3**). Among the 165 target genes, there are ten genes directly related to virulence in *S. aureus* TCH1516 and LAC strains (**Table 2**), consistent with the report that CodY links metabolism with virulence gene expression (6). Taken together, these data demonstrated that the global regulator CodY controls the expression of metabolism and virulence genes in *S. aureus* (9).

Strain-specific binding intensity is associated with DNA sequence variations in the binding site CodY

Though the genome-wide CodY binding sites identified were the same in the two strains, we found that strain-specific binding peaks had differing intensities (**Supplementary Figure 4**). The data showed that some of the binding peaks in TCH1516 had a higher intensity for CodY than those in LAC, and vice versa. 45 of 135 binding sites had identical DNA sequences (**Supplementary Figure 5**).

To fully evaluate whether the strain-specific binding peaks are due to the changes in the sequence-specific affinity to which CodY is bound, we first identified a 20-bp motif based

187 on a merged set of CodY binding peaks from the strains using MEME (20)
188 (**Supplementary figure 6**). We then utilized the computational method MAGGIE to
189 measure differences in the DNA sequence in paired binding sites in the two strains (21).
190 Among 135 pairs of peaks between the strains, 33.3% (45/135) had a range of motif score
191 difference between -1 and 1. We plotted the pairs of peak heights for the 135 shared
192 binding sites in LAC and TCH1516 (**Figure 3A**). When we highlighted the 45 binding sites
193 with near identical DNA sequences, we observed that they represent the dots closest to
194 the 45 degree line. This shows that these 45 sites have similar peak heights in the two
195 strains. The binding sites where the DNA sequence differs are off the 45 degree line,
196 showing that these differences lead to differential peak heights for the same binding site
197 in the two strains.
198
199 Further, to visualize the differences between a pair of binding peaks, we mapped CodY
200 binding peaks to the reference genome. For example, the sequence of the peak located
201 at the upstream of gene *yocS* in TCH1516 slightly (8 bp) overlaps the corresponding peak
202 sequence in LAC **(Figure 3B)**. Another example is from a pair of peaks from unknown
203 genes USA300HOU_RS01765 and ERW10_08770. They had zero motif score
204 difference, and thus we observed that both peak sequences nearly overlap each other
205 **(Figure 3C)**.

Genome-wide reconstruction of CodY regulons in the *S. aureus* USA300 lineage

The ChIP-exo datasets from this study expanded the size of target genes to 165 in *S. aureus* USA300 lineage, which included 128 novel CodY target genes. Of these, 37% (47 of 128) were metabolic genes. Nearly all of these 47 metabolic genes were non-essential genes in *S. aureus* USA300.

To further characterize the regulatory roles of CodY in the *S. aureus* USA300 lineage, we compared gene expression profiling of the *codY* mutant to that of the wild type under RPMI with 10% LB medium, and found there were 809 genes differentially expressed in the *codY* mutant (at least 2-fold change ($P<0.05$) in RNA-seq expression) (**Figure 4A, Supplementary Figure 7**). The majority of genes related to metabolism were up-regulated in the *codY* mutant, which was consistent with the repressor role that CodY plays in *S. aureus* (22). Further, we used functional enrichment analysis by Clusters of Orthologous Groups (COG) classification of 809 differentially expressed genes to discover the top six differential enrichment pathways: amino acid transporter and metabolism, inorganic ion transport and metabolism, translation, ribosomal structure and biogenesis, transcription, carbohydrate transport and metabolism, and energy production and conversion (**Figure 4B**).

Next, to reconstruct the CodY regulon, we compared the expression profiling of the *codY* mutant to the wild type strain. Combining genome-wide target genes with transcriptomics, we expanded the size of CodY regulons to 72 target genes that were directly regulated

by CodY (**Figure 4C**). Over half of the regulons (51%, 37 of 72) were related to the metabolic pathways (amino acid transport and metabolism, inorganic ion transport and metabolism, and energy production/conversion). In addition, 57% (41 of 72) of regulons were negatively regulated by CodY in the wild type. We also found that 90% (65/72) of CodY regulons had been identified in the CodY iModulon, which confirmed the regulatory roles of CodY (**Supplementary Figure 8**) (23). Further, CodY regulons were directly involved in signal transduction mechanisms, transcription, translation, post-translational modification, and defense mechanisms. These data indicated that CodY contributes to global regulatory roles beyond the metabolism of *S. aureus* USA300.

Rerouting of flux through central carbon metabolism to generate the branched-chain amino acids (BCAAs)

CodY is reported to regulate the expression of metabolic genes in response to changes in the pools of specific metabolites, i.e., the branched-chain amino acids (BCAAs; isoleucine, leucine, and valine (ILV) and nucleoside triphosphate GTP), to regulate genes involved in the biosynthesis these amino acids (24). In addition to providing building blocks for proteins, BCAAs are also incorporated into the membrane as a part of branched chain fatty acids (BCFAs) (25). Therefore, sufficient uptake or biosynthesis of BCAAs is a key component of cellular homeostasis. Analysis of the codY iModulon and ChIP-exo data indicated that codY directly regulates genes required for both the transport and biosynthesis of BCAAs. For many other amino acids (e.g. histidine, aspartate, threonine), our data indicates that codY only directly regulates their biosynthetic genes and not their transport. Therefore, we sought to understand how *S. aureus* deals with BCAA starvation.

In order to understand the partition of fluxes required for BCAA synthesis we utilized parsimonious Flux Balance Analysis (pFBA), using a previously published metabolic model of *S. aureus* USA300 (26, 27). pFBA determines metabolic fluxes that maximize growth while minimizing the sum of fluxes through the system (28). Simulated growth in RPMI supplemented with BCAA led to growth with zero flux through the BCAA biosynthetic pathway in favor of direct transporters of the necessary metabolites. However, restricting flux through all BCAA transporters (see Materials and Methods), led to a spike in flux through the biosynthetic enzymes (**Figure 5**). Interestingly, aspartate, a precursor to isoleucine biosynthesis, was not generated by the TCA cycle intermediate oxaloacetate. Instead, aspartate was derived from the breakdown of asparagine, while flux through the TCA cycle was redirected towards generating pyruvate via malate. This result indicated that in the presence of sufficient asparagine and aspartate in the media, *S. aureus* generates pyruvate to increase precursor pools available for isoleucine biosynthesis. Indeed, flux through the two enzymes linking the TCA cycle and aspartate, malate dehydrogenase, and aspartate transaminase, increased when aspartate and asparagine transporters were blocked in addition to BCAA biosynthesis.

Discussion

With the increasing number of whole genome sequences becoming available for multiple strains of a pathogenic species, the importance of the differences in their genomes and gene content is becoming more appreciated (29). Although many properties of pathogenic strains can now be predicted from sequence alone (30–32), detailed experimental

273 characterization of differences for multiple strains is also needed. In this study, we
274 combined genome-wide experiments and computational modeling to address the
275 differences of CodY in the dominant CA-MRSA USA300 clinical isolates TCH1516 and
276 LAC (15).
277
278 The study resulted in a series of significant findings. First, through the genome genome-
279 wide identification of binding sites, we found the same 165 CodY target genes in both
280 strains. Second, an examination of the differential binding intensity for the same target
281 genes under the same conditions revealed that the variance was due to DNA sequence
282 differences in the same CodY binding site in the two strains, while the *codY* protein was
283 identical. This finding gives insights into the system-level analysis of CodY target genes
284 and differential binding intensity across closely related strains. Third, the study identified
285 ten virulence genes that belong to different types of virulence factors, which demonstrated
286 that CodY connects metabolism genes with virulence genes in *S. aureus* (6). Considering
287 that different *S. aureus* lineages may have distinct virulence factors, we could expand this
288 study to identify many other virulence factor genes coordinated by CodY across different
289 *S. aureus* strains. Finally, a genome-scale model of the metabolic network can be
290 constrained by the regulatory action of CodY, and the results show the subsequent
291 systematic rerouting of metabolism and pathway use.
292
293 Taken together, this study demonstrates, for the first time, the differences in the function
294 of a conserved globally acting transcription factor (e.g., CodY) between closely related
295 pathogenic strains. These results enable a wider range of studies to further decipher both

subtle and major differences between the phenotypes of sequenced strains. This study motivates the characterization of common pathogenic strains and an evaluation of the possibility of developing specialized treatments for major strains circulating in the population.

Materials and Methods

Bacterial strains, plasmids, and culture conditions

The bacterial strains used in this study were described in Dataset 1. Methicillin-resistant *Staphylococcus aureus* (MRSA) strain substr. USA300 TCH1516 (also named USA300-HOU-MR) was originally isolated from an outbreak in Houston, Texas and caused severe invasive disease in adolescents (33). MRSA USA300 LAC was originally isolated from the Los Angeles county jail (15). *S. aureus* JE2, as a parental strain, was used for all sequence-defined Tn mutagenesis experiments. *CodY* mutant was from the Nebraska Transposon Mutant Library in which each of non-essential genes were disrupted via mariner Tn mutagenesis. All *S. aureus* strains were grown in tryptic soy broth (TSB, Sigma-Aldrich) or RPMI-1640 (Gibco) with 10% Lysogeny broth (LB, Sigma-Aldrich, MO) containing 10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl with the appropriate antibiotics for plasmid maintenance or selection (ampicillin, 100 ug/mL; chloramphenicol, 10 ug/mL) with shaking (250 rpm) at 37℃, maintaining a flask-to-medium volume ratio of 9:1, unless otherwise specified.

Monoclonal antibody

317 The entire *S. aureus* CodY coding sequence was amplified, and introduced into *E. coli*
318 BL21. The resulting glutathione-S-transferase (GST):: phoP fusion constructs were
319 verified by DNA sequencing. To obtain the GST fusion proteins, *E. coli* cells were grown
320 in 2 × YT medium at 30°C in an orbital shaker (180 rpm) to an OD_{600} of 0.6. Expression
321 was induced with IPTG (0.1 mM final concentration) for 5 h. Cells were harvested by
322 centrifugation, washed twice with PBS, lysed by sonication and then mixed with
323 Glutathione Sepharose-4B beads (Pharmacia Biotech). Proteins were eluted with 10 mM
324 reduced glutathione (in 50 mM Tris·HCl, pH 8.0) following the manufacturer's
325 recommendations and conserved in 40% glycerol at −80°C before use. Monoclonal anti-
326 CodY antibody was generated by injection of CodY into the BALB/C mouse (Genscript,
327 USA). Mouse anti-CodY IgG monoclonal antibody (IgG) was generated and purified
328 (Genscript, USA).

329

330 **ChIP-exo experiments**

331 ChIP-exo experimentation was performed following the procedures described previously
332 (34). To identify CodY binding maps for each strain *in vivo*, the DNA bound to CodY from
333 formaldehyde cross-linked cells were isolated by chromatin immunoprecipitation (ChIP)
334 with the specific antibodies that specifically recognize CodY (Genscript, USA), and
335 Dynabeads Pan Mouse IgG magnetic beads (Invitrogen) followed by stringent washings

as described previously. ChIP materials (chromatin-beads) were used to perform on-bead enzymatic reactions of the ChIP-exo method. Briefly, the sheared DNA of chromatin-beads was repaired by the NEBNext End Repair Module (New England Biolabs) followed by the addition of a single dA overhang and ligation of the first adaptor (5'-phosphorylated) using dA-Tailing Module (New England Biolabs) and NEBNext Quick Ligation Module (New England Biolabs), respectively. Nick repair was performed by using PreCR Repair Mix (New England Biolabs). Lambda exonuclease- and RecJf exonuclease-treated chromatin was eluted from the beads and overnight incubation at 65 ℃ reversed the protein-DNA cross-link. RNAs- and Proteins-removed DNA samples were used to perform primer extension and second adaptor ligation with following modifications. The DNA samples incubated for primer extension as described previously were treated with dA-Tailing Module (New England Biolabs) and NEBNext Quick Ligation Module (New England Biolabs) for second adaptor ligation. The DNA sample purified by GeneRead Size Selection Kit (Qiagen) was enriched by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA Polymerase (New England Biolabs). The amplified DNA samples were purified again by GeneRead Size Selection Kit (Qiagen) and quantified using Qubit dsDNA HS Assay Kit (Life Technologies). Quality of the DNA sample was checked by running Agilent High Sensitivity DNA Kit using Agilent 2100 Bioanalyzer (Agilent) before sequenced using HiSeq 2500 (Illumina) following the manufacturer's instructions. Each modified step was also performed following the manufacturer's instructions. ChIP-exo experiments were performed in biological duplicates.

Peak calling for ChIP-exo dataset

359 Peak calling was performed as previously described (34). Sequence reads generated
360 from ChIP-exo were mapped onto the reference genome using bowtie (35) with default
361 options to generate SAM output files. MACE program was used to define peak candidates
362 from biological duplicates for each experimental condition with sequence depth
363 normalization (36). To reduce false-positive peaks, peaks with signal-to-noise (S/N) ratio
364 less than 1.5 were removed. The noise level was set to the top 5% of signals at genomic
365 positions because top 5% makes a background level in a plateau and top 5% intensities
366 from each ChIP-exo replicates across conditions correlate well with the total number of
367 reads (34, 37, 38). The calculation of S/N ratio resembles the way to calculate ChIP-chip
368 peak intensity where IP signal was divided by Mock signal. Genome-scale data were
369 visualized using MetaScope.
370 (https://sites.google.com/view/systemskimlab/software?authuser=0)
371

372 **Motif search from ChIP-exo peaks**

373 The sequence motif analysis for CodY binding sites was performed using the MEME
374 software suite (20). For each strain, sequences in binding regions were extracted from
375 the reference genome (*S. aureus* TCH1516: GenBank: NC_010079.1, NC_012417.1,
376 and NC_010063.1; *S. aureus* LAC: GenBank: CP035369.1 and CP035370.1). To achieve
377 a more accurate motif, the sequence of each binding site was extended by 10bp at each
378 end. The width parameter was fixed at 20bp and the minsites parameter was fixed at 90%
379 of the total number of the sequence. All other parameters followed the default setting.
380

381 **Clusters of Orthologous groups (COGs) enrichment**

CodY regulons were categorized according to their annotated COG database (39). Functional groups in core, accessory, and unique CodY-regulated genes were determined by COG categories.

Multiple genome comparison and alignment

MUMmer was used to run the complete nucleotide based alignments to check for synteny amongst the sequences (40). BLAST Ring Image Generator (BRIG) was used to show a genome wide visualization of coding sequences identity between *S. aureus* USA300 TCH1516 and LAC (41). Multiple genomes were analyzed by the M-GCAT, which is a tool for rapidly visualizing and aligning the most highly conserved regions in multiple prokaryote genomes. M-GCAT is based on a highly efficient approach to anchor-based multiple genome comparison using a compressed suffix graph (42).

Determining Core Genome with Bidirectional BLAST Hits

To combine the data from the two strains, core genome-containing conserved genes between the LAC (GenBank: CP035369.1 and CP035370.1) and TCH1516 (GenBank: NC_010079.1, NC_012417.1, and NC_010063.1) were first established using bidirectional BLAST hits (43). In this analysis, all protein sequences of CDS from both genomes are BLASTed against each other twice with each genome acting as reference once. Two genes were considered conserved (and therefore part of the core genome) if 1) the two genes have the highest alignment percent to each other than to any other genes in the genome, and 2) the coverage is at least 80%.

405 **RNA Extraction and Library Preparation**

406 *S. aureus* USA300 isolates JE2, and its derivatives of *codY* mutant were used for this
407 study. The growth conditions and RNA preparation methods for data acquired from Choe
408 et al. has been previously described(44). Detailed growth conditions, RNA extraction, and
409 library preparation methods for other samples have also been already described (45).
410 Briefly, an overnight culture of *S. aureus* was used to inoculate a preculture and were
411 grown to mid-exponential growth phase (OD600 = 0.4) in RPMI+10%LB medium. Once
412 in the mid-exponential phase, the preculture was used to inoculate the media containing
413 appropriate supplementation or perturbations. All samples were collected in biological
414 duplicates, originating from different overnight cultures. Samples for control conditions
415 were collected for each set to account for batch effect.

416

417 **RNA-Seq Data Processing**

418 The RNA-seq pipeline used to analyze and perform QC/QA has been described in detail
419 previously (45). Briefly, the sequences were aligned to respective genomes, JE2, LAC or
420 TCH1516, using Bowtie2 (46). The aligned sequences were assigned to ORFs using
421 HTSeq-counts (47). Differential expression analysis was performed using DESeq2 with a
422 P-value threshold of 0.05 and an absolute fold-change threshold of 2 (48). To create the
423 final counts matrix, counts from conserved genes in LAC samples were represented by
424 the corresponding ortholog in TCH1516. The counts for accessory genes were filled with
425 0s if the genes were not present in the strain (i.e., LAC-specific genes had counts of 0 in
426 TCH1516 samples and vice versa). Finally, to reduce the effect of noise, genes with

427 average counts per sample <10 were removed. The final counts matrix with 2,581 genes
428 was used to calculate transcripts per million (TPM).
429
430 **Metabolic modeling and assessment of significant differences in flux distribution**
431 We used BiGG model iYS854 and set the lower bound to the corresponding nutrient
432 exchange to -1 mmol/gDW/h (the negative sign is a modeling convention to allow for the
433 influx of nutrients) and -13 mmol/gDW/h for oxygen exchange (as measured
434 experimentally) (27). Next, we compared two conditions with: 1) amino acid rich medium
435 and 2) amino acid poor medium. In the first condition, assuming that in the presence of
436 amino acid, CodY mediates the repression of multiple target genes. Specifically, we
437 turned off the set of 65 reactions that CodY regulated related to biosynthetic pathways.
438 No regulatory constraints were added. FBA was implemented with the biomass formation
439 set as the functional network objective. Next, flux balance analysis was run in both
440 conditions and the fluxes were sampled 10,000 times. All reaction fluxes were normalized
441 by dividing by the growth rate to account for growth differences across the two media
442 types. The flux distribution for each metabolic process was compared across both
443 conditions using the Kolmogorov-Smirnov test, a non-parametric test which compares
444 two continuous probability distributions. The distribution across two reactions was
445 deemed to be significantly different when the Kolmogorov-Smirnoff statistic was larger
446 than 0.99 with an adjusted p-value < 0.01.
447
448 **Motif mutation analysis**

To gain insights into the CodY motifs from TCH1516 and LAC strains, MAGGIE was used for motif mutation analysis (21). For CodY, given a pair of CodY binding sequences of the same target gene from TCH1516 and LAC strains, the peak sequence with a higher binding intensity was considered as a positive sequence; the other one with a lower binding intensity was considered as a negative sequence. Here, CodY has 135 pairs of positive and negative sequences from TCH1516 and LAC. MAGGIE computes differences of representative motif scores (i.e., motif mutations) within each sequence pair by subtracting the maximal scores of negative from positive sequences and then statistically tests for the association between motif mutations and the differences in CodY binding intensity. Positive significant p-values from MAGGIE indicate that higher-affinity motif is associated with stronger CodY binding.

Data availability

The ChIP-exo and RNA-seq datasets are accessible through GEO under accession number GSE159856 (review token: clgbaokcxhujhmt) and GSE163312 (review token: qfsvcukmnpwlvmb).

Acknowledgements

We thank Richard Szubin for help with ChIP-exo library sequencing, Charles J. Norsigian for the insights about flux balance analysis of CodY, and Marc Abrams for reviewing and editing the manuscript.

Funding

This work was funded by the Novo Nordisk Foundation Center for Biosustainability (Grant Number NNF10CC1016517) and the NIH NIAID (Grant number U01AI124316)

471

References

1. Bagnoli F, Rappuoli R, Grandi G. 2018. Staphylococcus aureus: Microbiology, Pathology, Immunology, Therapy and Prophylaxis. Springer.

2. Medveďová A, Györiová R. 2019. Prevalence of Staphylococcus aureus and antibiotic resistant Staphylococcus aureus in public transport in Bratislava, Slovakia. Acta Chimica Slovaca.

3. Chen S-Y, Liao C-H, Wang J-L, Chiang W-C, Lai M-S, Chie W-C, Chen W-J, Chang S-C, Hsueh P-R. 2012. Methicillin-resistant Staphylococcus aureus (MRSA) staphylococcal cassette chromosome mec genotype effects outcomes of patients with healthcare-associated MRSA bacteremia independently of vancomycin minimum inhibitory concentration. Clin Infect Dis 55:1329–1337.

4. Kane TL, Carothers KE, Lee SW. 2018. Virulence Factor Targeting of the Bacterial Pathogen Staphylococcus aureus for Vaccine and Therapeutics. Curr Drug Targets 19:111–127.

5. Chu AJ, Qiu Y, Harper R, Lin L, Ma C, Yang X. 2020. Nusbiarylins Inhibit Transcription and Target Virulence Factors in Bacterial Pathogen Staphylococcus aureus. International Journal of Molecular Sciences.

6. Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, Goerke C, Schrenzel J, Wolz C. 2009. CodY in Staphylococcus aureus: a regulatory link between metabolism and virulence gene expression. J Bacteriol 191:2953–2963.

7. Ibberson CB, Jones CL, Singh S, Wise MC, Hart ME, Zurawski DV, Horswill AR. 2014. Staphylococcus aureus hyaluronidase is a CodY-regulated virulence factor. Infect Immun 82:4253–4264.

8. Majerczyk CD, Dunman PM, Luong TT, Lee CY, Sadykov MR, Somerville GA, Bodi K, Sonenshein AL. 2010. Direct targets of CodY in Staphylococcus aureus. J Bacteriol 192:2861–2877.

9. Waters NR, Samuels DJ, Behera RK, Livny J, Rhee KY, Sadykov MR, Brinsmade SR. 2016. A spectrum of CodY activities drives metabolic reorganization and virulence gene expression in *Staphylococcus aureus*. Mol Microbiol 101:495–514.

10. Basu A, Shields KE, Eickhoff CS, Hoft DF, Yap M-NF. 2018. Thermal and nutritional regulation of ribosome hibernation in Staphylococcus aureus. J Bacteriol https://doi.org/10.1128/JB.00426-18.

11. Roux A, Todd DA, Velázquez JV, Cech NB, Sonenshein AL. 2014. CodY-mediated regulation of the Staphylococcus aureus Agr system integrates nutritional and population density signals. J Bacteriol 196:1184–1196.

12. Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, Goerke C, Schrenzel J, Wolz C. 2009. CodY in Staphylococcus aureus: a regulatory link between metabolism and virulence gene expression. J Bacteriol 191:2953–2963.

13. Majerczyk CD, Dunman PM, Luong TT, Lee CY, Sadykov MR, Somerville GA, Bodi K, Sonenshein AL. 2010. Direct targets of CodY in Staphylococcus aureus. J Bacteriol 192:2861–2877.

513 14. Thurlow LR, Joshi GS, Richardson AR. 2012. Virulence strategies of the dominant
514 USA300 lineage of community-associated methicillin-resistant Staphylococcus
515 aureus (CA-MRSA). FEMS Immunol Med Microbiol 65:5–22.

516 15. Vitko NP, Richardson AR. 2013. Laboratory maintenance of methicillin-resistant
517 Staphylococcus aureus (MRSA). Curr Protoc Microbiol Chapter 9:Unit 9C.2.

518 16. Kennedy AD, Porcella SF, Martens C, Whitney AR, Braughton KR, Chen L, Craig
519 CT, Tenover FC, Kreiswirth BN, Musser JM, DeLeo FR. 2010. Complete nucleotide
520 sequence analysis of plasmids in strains of Staphylococcus aureus clone USA300
521 reveals a high level of identity among isolates with closely related core genome
522 sequences. J Clin Microbiol 48:4504–4511.

523 17. Wijesinghe G, Dilhari A, Gayani B, Kottegoda N, Samaranayake L, Weerasekera
524 M. 2019. Influence of Laboratory Culture Media on in vitro Growth, Adhesion, and
525 Biofilm Formation of Pseudomonas aeruginosa and Staphylococcus aureus. Med
526 Princ Pract 28:28–35.

527 18. Biswas R, Sonenshein AL, Belitsky BR. 2020. Genome-wide identification of
528 Listeria monocytogenes CodY-binding sites. Mol Microbiol 113:841–858.

529 19. Belitsky BR, Sonenshein AL. 2013. Genome-wide identification of Bacillus subtilis
530 CodY-binding sites at single-nucleotide resolution. Proc Natl Acad Sci U S A
531 110:7026–7031.

532 20. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW,
533 Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic

534 Acids Res 37:W202–8.

535 21. Shen Z, Hoeksema MA, Ouyang Z, Benner C, Glass CK. 2020. MAGGIE:
536 leveraging genetic variation to identify DNA sequence motifs mediating
537 transcription factor binding and function. Bioinformatics 36:i84–i92.

538 22. Majerczyk CD, Sadykov MR, Luong TT, Lee C, Somerville GA, Sonenshein AL.
539 2008. Staphylococcus aureus CodY negatively regulates virulence gene
540 expression. J Bacteriol 190:2257–2265.

541 23. Poudel S, Tsunemoto H, Seif Y, Sastry AV, Szubin R, Xu S, Machado H, Olson CA,
542 Anand A, Pogliano J, Nizet V, Palsson BO. 2020. Revealing 29 sets of
543 independently modulated genes in Staphylococcus aureus, their regulators, and
544 role in key physiological response. Proc Natl Acad Sci U S A 117:17228–17239.

545 24. Brinsmade SR, Kleijn RJ, Sauer U, Sonenshein AL. 2010. Regulation of CodY
546 activity through modulation of intracellular branched-chain amino acid pools. J
547 Bacteriol 192:6357–6368.

548 25. Sen S, Sirobhushanam S, Johnson SR, Song Y, Tefft R, Gatto C, Wilkinson BJ.
549 2016. Growth-Environment Dependent Modulation of Staphylococcus aureus
550 Branched-Chain to Straight-Chain Fatty Acid Ratio and Incorporation of
551 Unsaturated Fatty Acids. PLoS One 11:e0165300.

552 26. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins
553 JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith
554 RD, Palsson BØ. 2010. Omic data from evolved E. coli are consistent with

555 computed optimal growth from genome-scale models. Molecular Systems Biology.

556 27. Seif Y, Monk JM, Mih N, Tsunemoto H, Poudel S, Zuniga C, Broddrick J, Zengler K,
557 Palsson BO. 2019. A computational knowledge-base elucidates the response of
558 Staphylococcus aureus to different media types. PLoS Comput Biol 15:e1006644.

559 28. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins
560 JN, Schramm G, Purvine SO, Lopez-Ferrer D, Others. 2010. Omic data from
561 evolved E. coli are consistent with computed optimal growth from genome-scale
562 models. Mol Syst Biol 6:390.

563 29. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, Karpinets T, Lund O,
564 Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of
565 bacterial genome sequencing. Funct Integr Genomics 15:141–161.

566 30. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM,
567 Palsson BØ. 2013. Genome-scale metabolic reconstructions of multiple
568 Escherichia coli strains highlight strain-specific adaptations to nutritional
569 environments. Proc Natl Acad Sci U S A 110:20338–20343.

570 31. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W,
571 Zhang Z, Mori H, Feist AM, Palsson BO. 2017. iML1515, a knowledgebase that
572 computes Escherichia coli traits. Nat Biotechnol 35:904–908.

573 32. Norsigian CJ, Danhof HA, Brand CK, Oezguen N, Midani FS, Palsson BO, Savidge
574 TC, Britton RA, Spinler JK, Monk JM. 2020. Systems biology analysis of the
575 Clostridioides difficile core-genome contextualizes microenvironmental evolutionary

576 pressures leading to genotypic and phenotypic divergence. NPJ Syst Biol Appl
577 6:31.

578 33. Hultén KG, Kaplan SL, Gonzalez BE, Hammerman WA, Lamberth LB, Versalovic J,
579 Mason EO Jr. 2006. Three-year surveillance of community onset health care-
580 associated staphylococcus aureus infections in children. Pediatr Infect Dis J
581 25:349–353.

582 34. Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Palsson BO. 2014. Deciphering Fur
583 transcriptional regulatory network highlights its complex role beyond iron
584 metabolism in Escherichia coli. Nat Commun 5:4910.

585 35. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
586 alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

587 36. Wang L, Chen J, Wang C, Uusküla-Reimand L, Chen K, Medina-Rivera A, Young
588 EJ, Zimmermann MT, Yan H, Sun Z, Zhang Y, Wu ST, Huang H, Wilson MD,
589 Kocher J-PA, Li W. 2014. MACE: model based analysis of ChIP-exo. Nucleic Acids
590 Res 42:e156.

591 37. Seo SW, Kim D, Szubin R, Palsson BO. 2015. Genome-wide Reconstruction of
592 OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in
593 Escherichia coli K-12 MG1655. Cell Rep 12:1289–1299.

594 38. Ogasawara H, Ohe S, Ishihama A. 2015. Role of transcription factor NimR (YeaM)
595 in sensitivity control of Escherichia coli to 2-nitroimidazole. FEMS Microbiol Lett
596 362:1–8.

597 39. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool
598 for genome-scale analysis of protein functions and evolution. Nucleic Acids Res
599 28:33–36.

600 40. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. 1999.
601 Alignment of whole genomes. Nucleic Acids Res 27:2369–2376.

602 41. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image
603 Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics.

604 42. Treangen TJ, Messeguer X. 2006. M-GCAT: interactively and efficiently
605 constructing large-scale multiple genome comparison frameworks in closely related
606 species. BMC Bioinformatics 7:433.

607 43. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. 1999. The use of gene
608 clusters to infer functional coupling. Proc Natl Acad Sci U S A 96:2896–2901.

609 44. Choe D, Szubin R, Dahesh S, Cho S, Nizet V, Palsson B, Cho B-K. 2018. Genome-
610 scale analysis of Methicillin-resistant Staphylococcus aureus USA300 reveals a
611 tradeoff between pathogenesis and drug resistance. Sci Rep 8:2215.

612 45. Poudel S, Tsunemoto H, Meehan M, Szubin R, Olson CA, Lamsa A, Seif Y, Dillon
613 N, Vrbanac A, Sugie J, Dahesh S, Monk JM, Dorrestein PC, Pogliano J, Knight R,
614 Nizet V, Palsson BO, Feist AM. 2019. Characterization of CA-MRSA TCH1516
615 exposed to nafcillin in bacteriological and physiological media. Sci Data 6:43.

616 46. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2.

617 Nature Methods.

618 47. Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-
619 throughput sequencing data. Bioinformatics 31:166–169.

620 48. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
621 dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

622

FIGURE LEGENDS:

Figure 1 | The comparison of two dominant *Staphylococcus aureus* USA300 isolates (TCH1516 and LAC).
(A) Circular representation of whole genome comparison of *S. aureus* TCH1516 (internal ring) and LAC strains. Each ring of the circle represents a specific complete genome that corresponds to different colors in the legend on the right. The similarity between strains is represented by the intensity of the color. Darker colors represent higher similarities than lighter ones. Deleted regions are represented by blank spaces inside the circles. The whole genome comparisons were generated by BRIG. Alignment identity cutoffs of 0.8 (upper) and 0.5 (lower) were used to determine missing regions in the query genome (*S. aureus* LAC) compared to the *S. aureus* TCH1516 reference. Since *S. aureus* TCH1516 was the first genome to be annotated in the USA300 lineage, it was used as a reference genome for this study. (B) Dot plot of a nucleotide-based alignment of the genomes between *S. aureus* USA300 TCH1516 and LAC.

Figure 2 | Comparison of CodY-binding sites between TCH1516 and LAC strains.
(A) Direct comparison of CodY target genes between *S. aureus* substr. USA300 TCH1516 and LAC at the genome. Upper panel: an overview of CodY-binding profiles across the *S. aureus* TCH1516 genome at mid-exponential growth phase in RPMI 1640 + 10% LB medium. S/N denotes signal-to-noise ratio. (+) and (-) indicate forward and reverse reads, respectively. Bottom panel: an overview of CodY-binding profiles across the *S. aureus* LAC strain genome at mid-exponential growth phase under RPMI 1640 + 10% LB medium. (B) Distribution of *in vivo* CodY genome-wide binding sites at the genome of the TCH1516 strain (upper panel). Comparison of CodY-binding sites obtained from this study (ChIP-exo) with CodY target genes from *S. aureus* USA200 UAMS-1 using IDAP-Seq (bottom panel). (C) The consensus DNA sequence for *S. aureus* TCH1516 CodY binding motif. (D) The consensus DNA sequences for *S. aureus* LAC CodY binding motif.

Figure 3 | Differential CodY binding peaks area plot for TCH1516 and LAC highlighting binding sites with identical binding motifs.
(A) Binding peak areas of log2(TCH1516_binding_intensity) (x-axis) and log2(Lac_binding_intensity) (y-axis) is shown. The diagonal line represents identical peak areas. The 45 binding sites with near zero sequence difference (Supplemental Figure 5) are shown with the solid dots, while those that are different are shown with open circles. (B) Case study I: the binding peak at the upstream of gene *yocS*, which has non-zero

motif score difference between TCH1516 and LAC. Annotations of color code nucleotides are shown in the legend. Nucleotides in red and green represent the CodY peak sequences in TCH1516 and LAC, respectively. Grey denotes the overlap between a pair of peak sequences. (C) Case study II: the binding peak at the gene USA300HOU_RS01765, which has zero motif score difference between TCH1516 and LAC.

Figure 4 | Role of CodY in regulation of *S. aureus* USA300 LAC genes differentially expressed in the *codY* mutant.

(A) The *S. aureus* USA300 LAC direct CodY regulon. LAC genes that had a CodY ChIP-exo binding and had at least 2-fold change (P<0.05) in RNA-seq expression between the codY mutant to the wild-type strain were assigned to the direct regulon. (B) Functional enrichment analysis by Clusters of Orthologous Groups (COG) classification of 809 differentially expressed genes in *S. aureus* LAC *codY* mutant compared to wild type. The number of genes are based on the annotated genome. The top six enriched pathways were amino acid transport and metabolism, inorganic ion transport and metabolism, translation, ribosomal structure and biogenesis, transcription, carbohydrate transport and metabolism, and energy production and conversion. The functional enrichment was analyzed by performing the hypergeometric test. The asterisk indicates hypergeometric P-value <0.05. (C) Reconstruction of 72 CodY regulon in *S. aureus* LAC strain.

Figure 5 | Central metabolic flux rerouting to generate BCAA.

(A) Optimal flux through CodY regulated biosynthetic enzymes, transporters and non-CodY regulated transporters in RPMI medium. When amino acids are present, *S. aureus* imports them via transporters (light blue and red bars). (B) pFBA solution of *S. aureus* grown in a chemically defined medium (CDMG) without BCAA predicts rerouting of several central carbon metabolic fluxes to generate BCAA precursors for BCAA. Red arrows represent reactions with increased flux during BCAA starvation and blue arrows represent those with decreased flux relative to starvation conditions. Malate dehydrogenase and aspartate synthase (blue arrows) lower flux when BCAA transporters were blocked. Note: Full ILV biosynthesis pathways are not shown. (C) Flux through pyruvate generating malate enzyme (ME2) and aspartate generating malate dehydrogenase (MDH3) and aspartate transaminase (ASPTA). The pathways in bold correspond to the biosynthesis pathways that contain CodY target genes.

TABLE CAPTIONS

Table 1 Comparison of annotation features in *S. aureus* USA300 lineage

Table 2 The CodY target genes considered to be virulence factors in the TCH1516 and LAC strains

SUPPLEMENTARY MATERIALS

Supplementary Figure 1 | Comparison of CodY sequences among *S. aureus* USA300, *L. monocytogenes*, and *B. subtilis*. Red rectangle denotes the CodY DNA-binding domain (Helix-Turn-Helix domain).

Supplementary Figure 2 | Comparison of CodY sequences between *S. aureus* USA300 TCH1516 and LAC strains.

Supplementary Figure 3 | The similarity between TCH1516 CodY binding motif and LAC CodY binding motif using TOMTOM. There are 18 bp nucleotides overlapping between them (*p*-value = 9.35e-09).

Supplementary Figure 4 | Comparison of CodY binding intensity between TCH1516 and LAC strains.

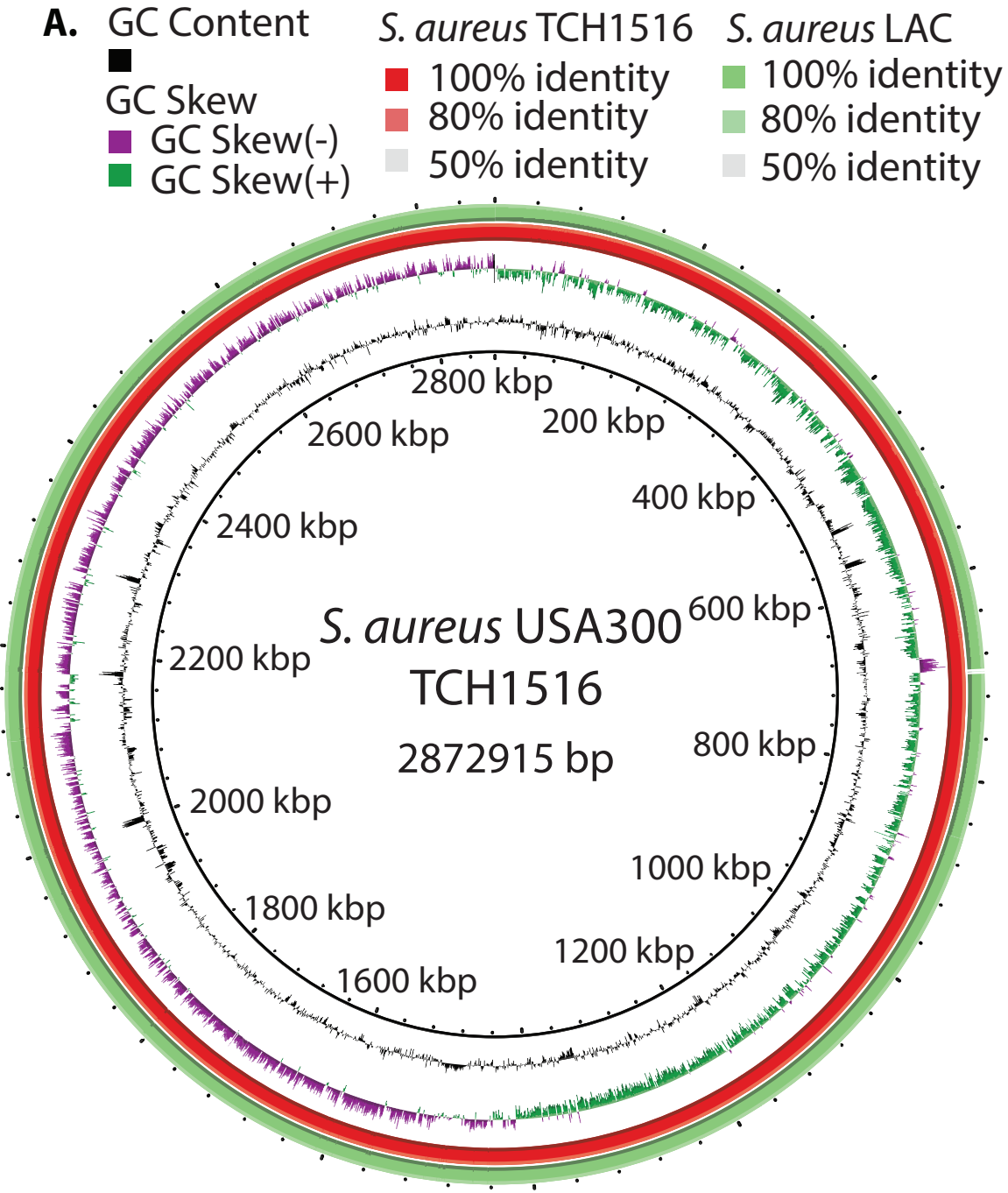
Supplementary Figure 5 | Distribution of CodY peaks based on the range of motif score difference. X-axis represents the range of motif score difference. Y-axis represents the number of peaks that fall into each range. Black-color bar showed 45 peaks with the motif score difference at the range of (-1, 1).

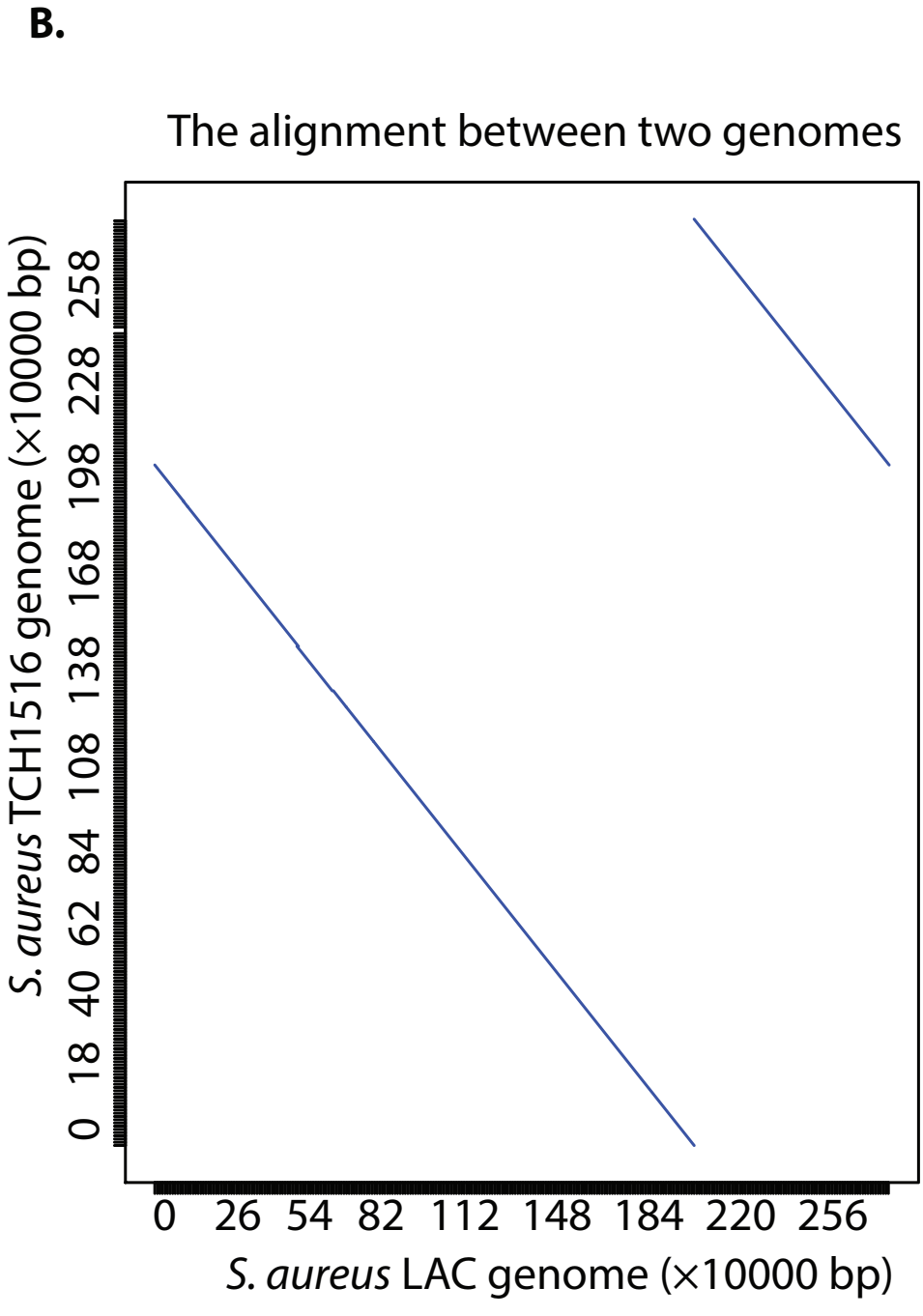
Supplementary Figure 6 | *S. aureus* USA300 CodY binding motif from a merged set of TCH1516 and LAC binding peaks.

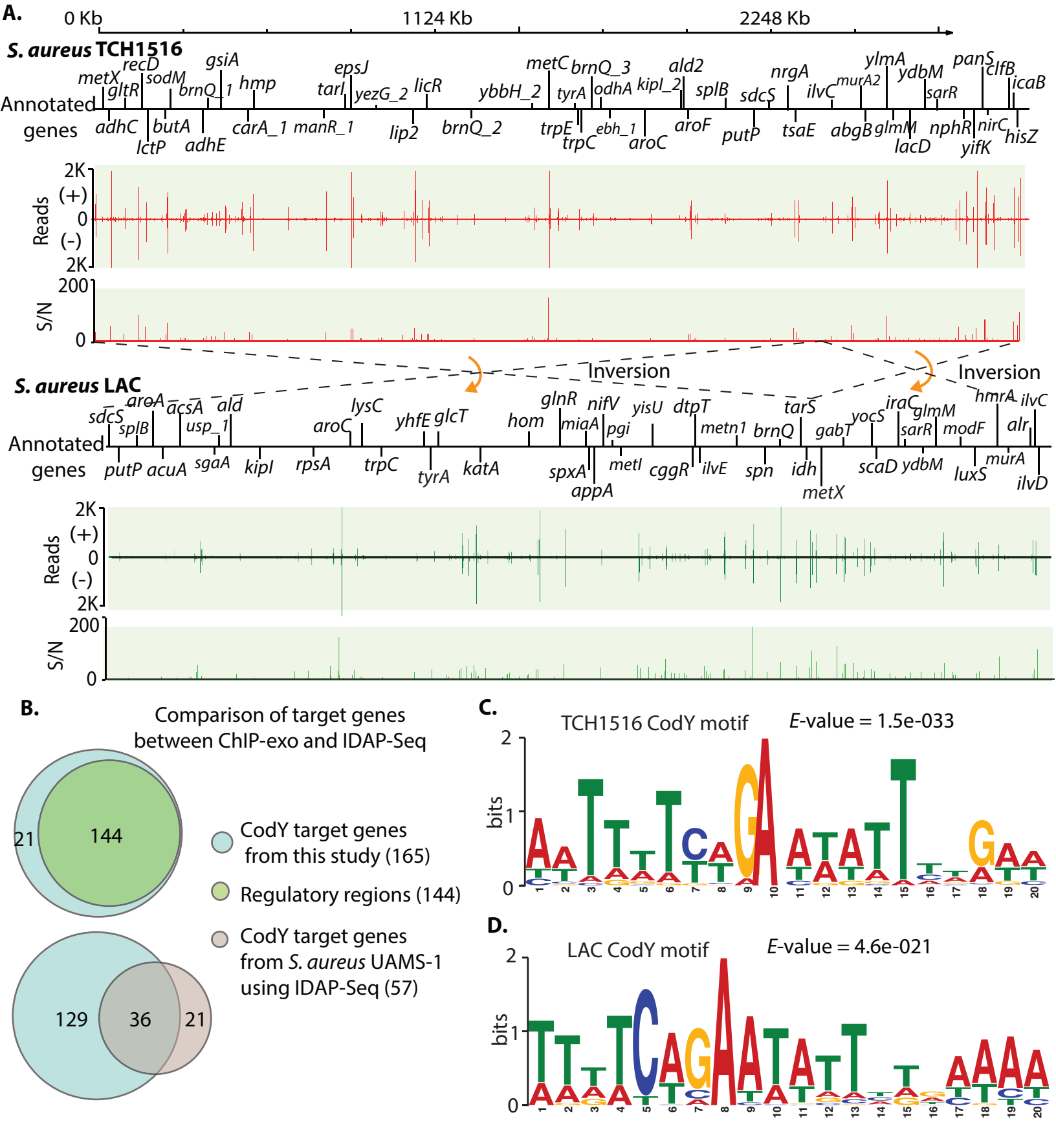
Supplementary figure 7 | A heatmap for wild type and *codY* mutant samples for expression profiling data from *S. aureus* USA300 LAC.

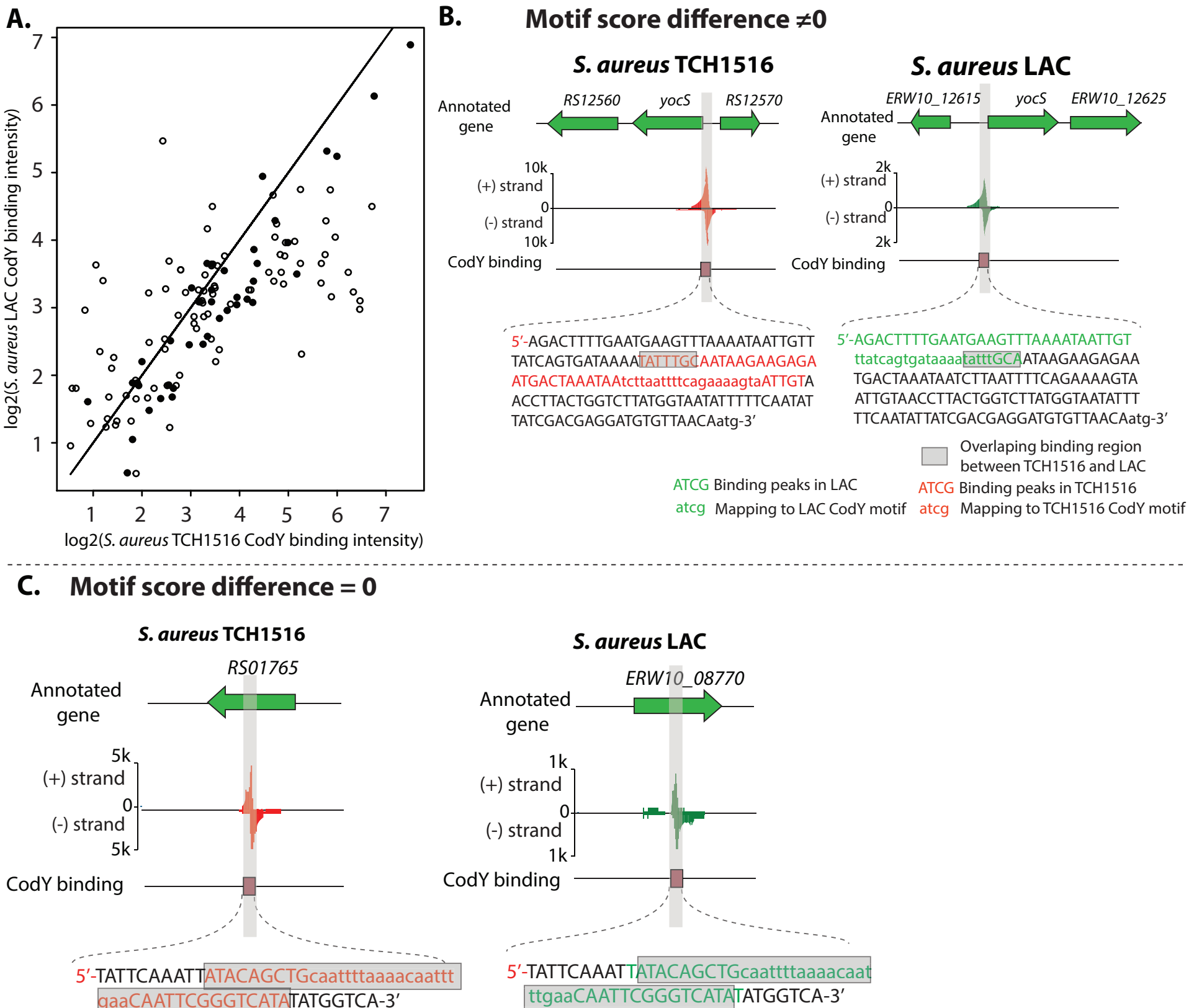
Supplementary Figure 8 | Comparison of genes in the CodY regulon (red) and CodY iModulon (green). The number in the circle represents the amount of genes in the category. CodY regulon is defined by a binding peak in the genes promoter region and detection of differential gene expression of the wild type and the *codY* mutant during growth in RPMI 1640 and 10% LB medium.

Dataset S1 The strains used in this study

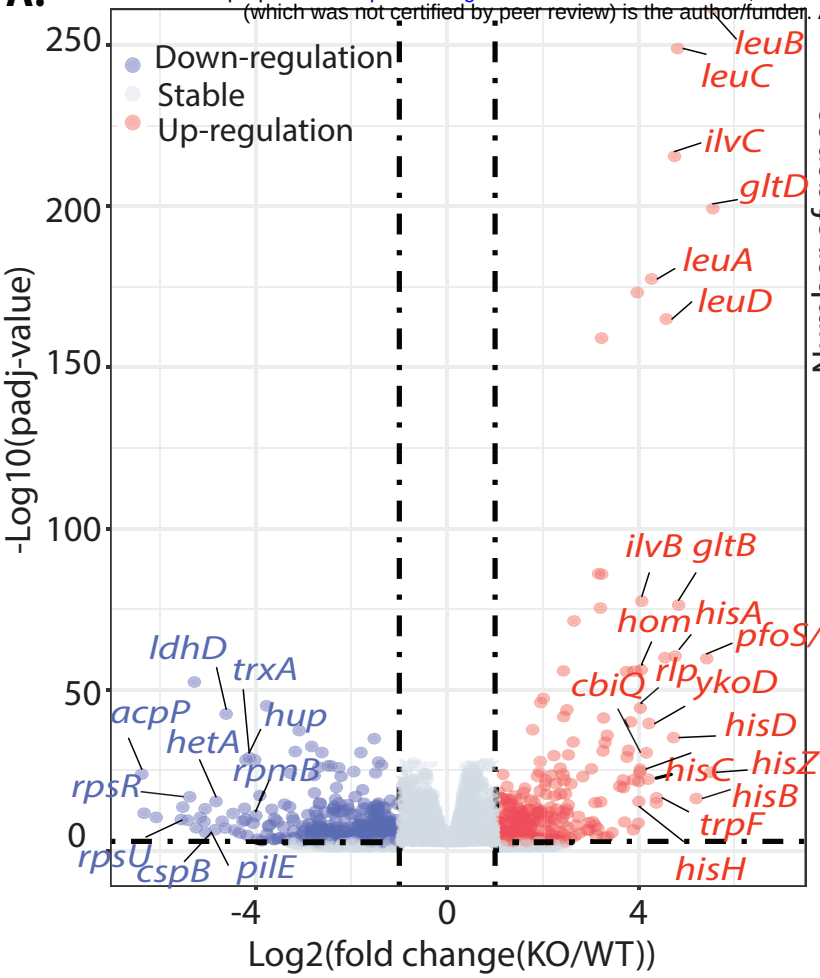

A.
GC Content
GC Skew
GC Skew(-)
GC Skew(+)
S. aureus TCH1516
100% identity
80% identity
50% identity
S. aureus LAC
100% identity
80% identity
50% identity
2800 kbp
200 kbp
2600 kbp
400 kbp
2400 kbp
600 kbp
S. aureus USA300
TCH1516
2872915 bp
2200 kbp
800 kbp
2000 kbp
1000 kbp
1800 kbp
1200 kbp
1600 kbp


B.
The alignment between two genomes
S. aureus TCH1516 genome (×10000 bp)
0 18 40 62 84 108 138 168 198 228 258
0 26 54 82 112 148 184 220 256
S. aureus LAC genome (×10000 bp)

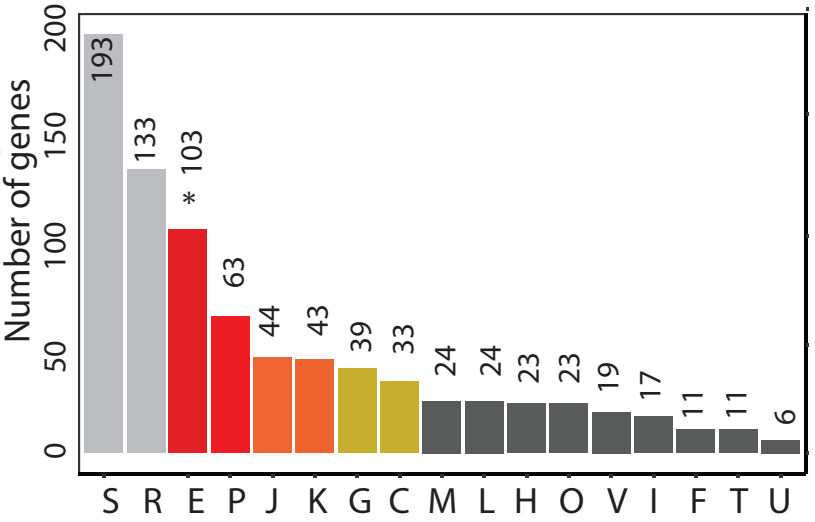



A.
log2(S. aureus LAC CodY binding intensity)
log2(S. aureus TCH1516 CodY binding intensity)
B.
Motif score difference ≠0
S. aureus TCH1516
S. aureus LAC
Annotated gene
RS12560
yocS
RS12570
ERW10_12615
yocS
ERW10_12625
10k
(+) strand
0
(-) strand
10k
2k
2k
CodY binding
5′-AGACTTTTGAATGAAGTTTAAAATAATTGTT
TATCAGTGATAAAATATTTGCAATAAGAAGAGA
ATGACTAAATAAtcttaattttcagaaaagtaATTGTA
ACCTTACTGGTCTTATGGTAATATTTTTCAATAT
TATCGACGAGGATGTGTTAACAatg-3′
5′-AGACTTTTGAATGAAGTTTAAAATAATTGT
ttatcagtgataaaatatttGCAATAAGAAGAGAA
TGACTAAATAATCTTAATTTTCAGAAAAGTA
ATTGTAACCTTACTGGTCTTATGGTAATATTT
TTCAATATTATCGACGAGGATGTGTTAACAatg-3′
Overlaping binding region between TCH1516 and LAC
ATCG Binding peaks in LAC
atcg Mapping to LAC CodY motif
ATCG Binding peaks in TCH1516
atcg Mapping to TCH1516 CodY motif
C.
Motif score difference = 0
S. aureus TCH1516
S. aureus LAC
RS01765
ERW10_08770
Annotated gene
5k
(+) strand
0
(-) strand
5k
1k
1k
CodY binding
5′-TATTCAAATTATACAGCTGcaattttaaaacaattt
gaaCAATTCGGGTCATATATGGTCA-3′
5′-TATTCAAATTATACAGCTGcaattttaaaacaat
ttgaaCAATTCGGGTCATATATGGTCA-3′

A.

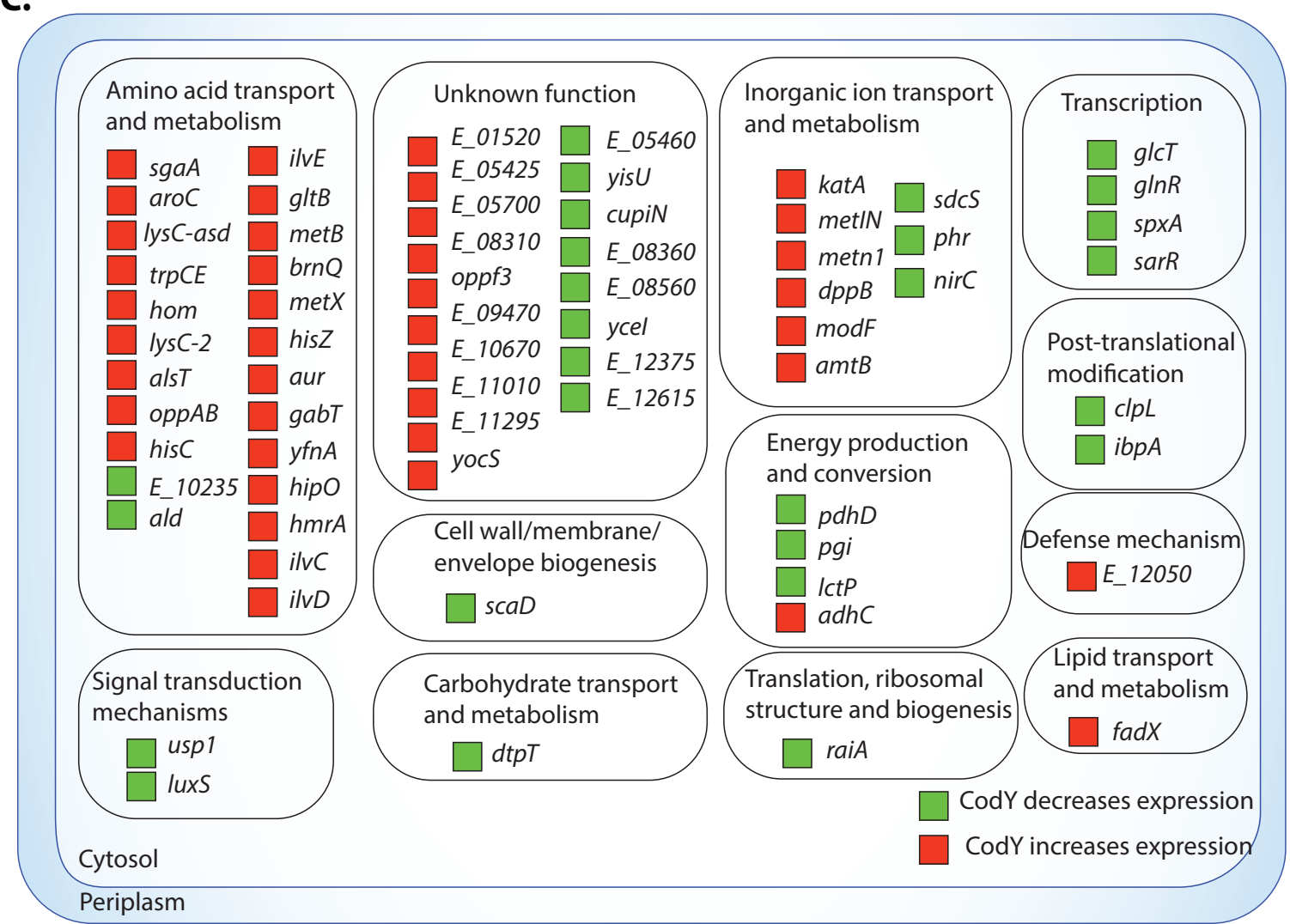


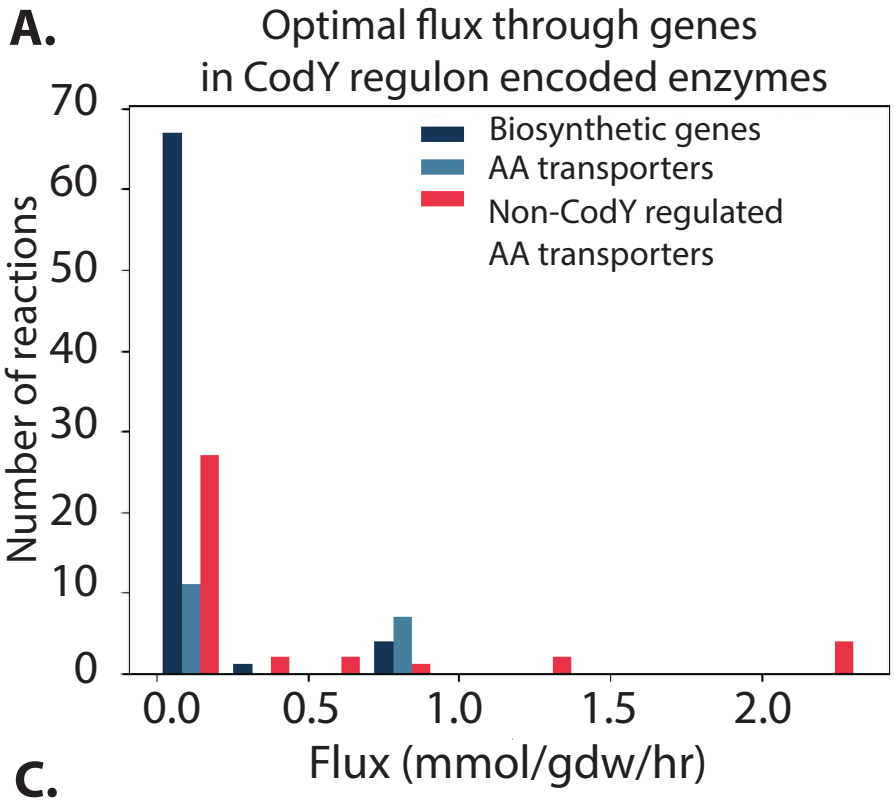
B. Differential pathways in *S. aureus* LAC

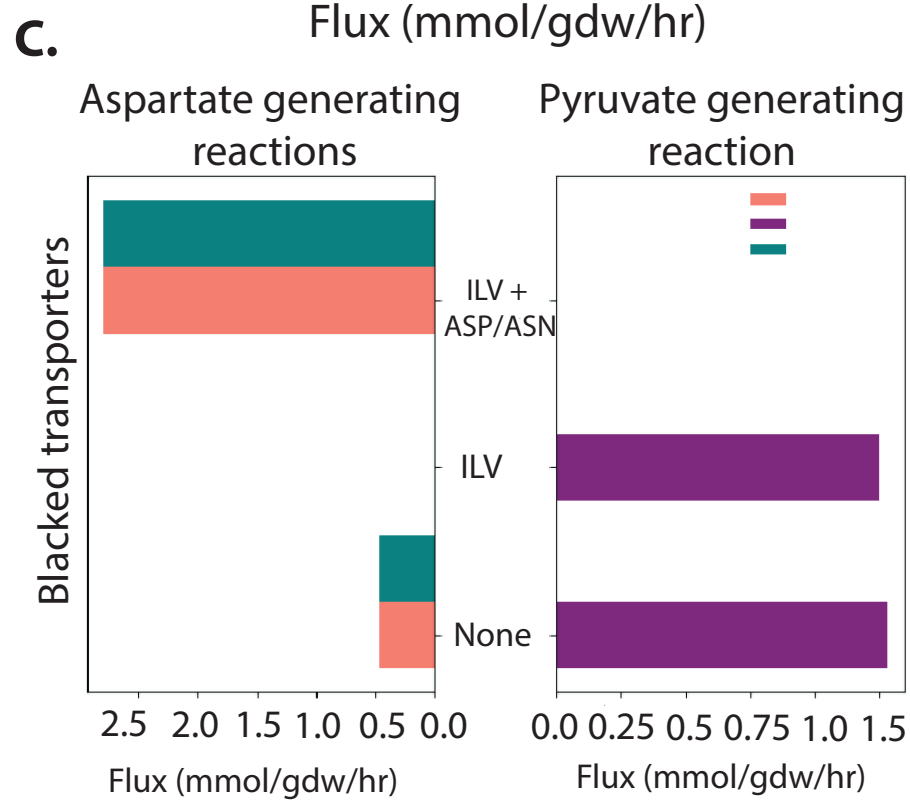


S = Function unknown
R = General function prediction only
E = Amino acid transport and metabolism
P = Inorganic ion transport and metabolism
J = Translation
K = Transcription
G = Carbohydrate transport and metabolism
C = Energy production and conversion

C.




A.
Optimal flux through genes
in CodY regulon encoded enzymes
Biosynthetic genes
AA transporters
Non-CodY regulated
AA transporters
Number of reactions
70
60
50
40
30
20
10
0
0.0
0.5
1.0
1.5
2.0
Flux (mmol/gdw/hr)


C.
Aspartate generating
reactions
Pyruvate generating
reaction
Blacked transporters
ILV +
ASP/ASN
ILV
None
2.5 2.0 1.5 1.0 0.5 0.0
Flux (mmol/gdw/hr)
0.0 0.25 0.5 0.75 1.0 1.5
Flux (mmol/gdw/hr)

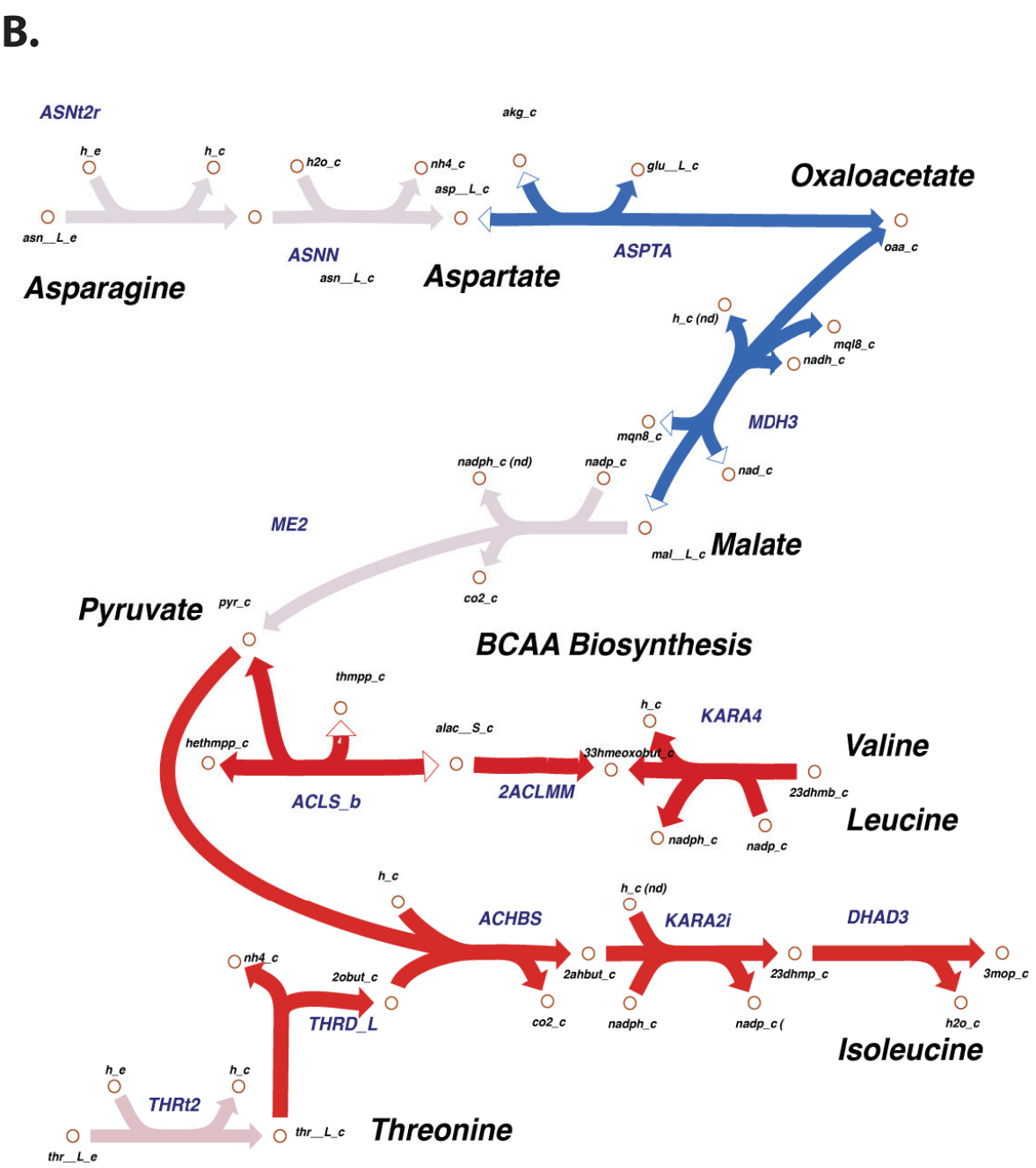

B.
ASNt2r
h_e
h_c
h2o_c
nh4_c
akg_c
glu__L_c
asp__L_c
Oxaloacetate
asn__L_e
ASNN
asn__L_c
Asparagine
Aspartate
ASPTA
oaa_c
h_c (nd)
mql8_c
nadh_c
MDH3
mqn8_c
nad_c
nadph_c (nd)
nadp_c
ME2
mal__L_c
Malate
co2_c
Pyruvate
pyr_c
BCAA Biosynthesis
thmpp_c
hethmpp_c
alac__S_c
h_c
KARA4
33hmeoxobut_c
Valine
ACLS_b
2ACLMM
23dhmb_c
Leucine
nadph_c
nadp_c
h_c
h_c (nd)
ACHBS
KARA2i
DHAD3
nh4_c
2obut_c
2ahbut_c
23dhmp_c
3mop_c
THRD_L
co2_c
nadph_c
nadp_c (
h2o_c
Isoleucine
h_e
h_c
THRt2
thr__L_c
Threonine
thr__L_e

Table 1 Comparison of annotation features in *S. aureus* USA300 lineage

	***S. aureus* TCH1516**	***S. aureus* LAC**
Total sequence length (bp)	2,872,915	2,878,171
Total chromosome and plasmids	3, NC_010079.1 (chromosome) NC_012417.1 (plasmid 1) NC_010063.1 (plasmid 2)	2, CP035369.1 (chromosome) CP035370.1 (plasmid)
Gene (total)	2920	2892
CDS (total)	2841	2802
Genes/CDS (coding)	2763	2733
RNA	79	90
rRNAs (5S, 16S, 23S)	6, 5, 5	7, 6, 6
tRNAs	59	67
ncRNAs	4	4
Pseudo gene (total)	78	69

Table 2 The CodY target genes considered to be virulence factors in the TCH1516 and LAC strains

Virulence factors	**Gene**	***S. aureus* TCH1516**	***S. aureus* LAC**
Cell wall associated fibronectin binding protein	*ebh*	Yes	Yes
Clumping factor B	*clfB*	Yes	Yes
Extracellular adherence protein/MHC analogous protein	*eap/map*	Yes	Yes
Intercellular adhesin	*icaB*	Yes	Yes
Hyaluronate lyase	*hysA*	Yes	Yes
Lipase	*geh*	Yes	Yes
Serine protease	*splB*	Yes	Yes
Sbi	*sbi*	Yes	Yes
Type VII secretion system	*esaG*	Yes	Yes
Enterotoxin-like K	*selk*	Yes	Yes

N/A denotes the gene name unavailable in the reference genome.