

# Elucidating the CodY regulon in *Staphylococcus aureus* USA300 substrains TCH1516 and LAC

Ye Gao,<sup>1,2</sup> Saugat Poudel,<sup>2</sup> Yara Seif,<sup>2</sup> Zeyang Shen,<sup>2</sup> Bernhard O. Palsson<sup>2,3,4,5</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 15.

**ABSTRACT** CodY is a conserved broad-acting transcription factor that regulates the expression of genes related to amino acid metabolism and virulence in Gram-positive bacteria. Here, we performed the first *in vivo* determination of CodY target genes using a novel CodY monoclonal antibody in methicillin-resistant *Staphylococcus aureus* (MRSA) USA300. Our results showed (i) the same 135 CodY promoter binding sites regulating the 165 target genes identified in two closely related virulent *S. aureus* USA300 TCH1516 and LAC strains; (ii) 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; (iii) 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, all based on transcriptomic data; and (iv) 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, revealing 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.

**KEYWORDS** *Staphylococcus aureus*, CodY, Transcription factor binding sites, Metabolism

*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 ones, 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 remains critical (2, 3). 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 (4). To combat the worldwide spread of *S. aureus*, significant

**Editor** Neha Sachdeva, Colorado State University, Fort Collins, Colorado, USA

Address correspondence to Bernhard O. Palsson, palsson@ucsd.edu.

The authors declare no conflict of interest.

See the funding table on p. 15.

**Received** 30 March 2023

**Accepted** 25 April 2023

**Published** 13 June 2023

Copyright © 2023 Gao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

effort is being focused on the investigation of the transcription factors that control virulence factors during infection (5, 6). With different strains circulating in the population and the serial rise and fall of specific strain types, it is important to understand the fundamental differences between them (7).

CodY is an important broad-acting transcription factor in *S. aureus* (8, 9). While the primary known role of CodY is to regulate the metabolic genes in response to the changes in 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 (8, 10–13). 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). At present, the direct determination of CodY DNA-binding sites using ChIP-exo and the identification of the corresponding target genes have not been achieved.

To address this challenge, we developed a reliable ChIP-grade monoclonal antibody and conducted *in vivo* genome-wide experiments (ChIP-exo) to identify the 165 identical CodY target genes in two common *S. aureus* USA300 isolates (TCH1516 and LAC). To reconstruct the CodY regulon, we compared RNA-seq profiles of the wild-type strain and *codY* mutant. To examine the network-level effects of CodY, we used a genome-scale metabolic model to simulate the flux state of central carbon metabolism, demonstrating the regulatory activities of CodY to generate branched-chain amino acids (BCAAs). Our study used a comprehensive pipeline to characterize the CodY regulon between closely related USA300 substrains that included genetic parameters and network-level computational models.

## 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 Houston and Los Angeles, 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).

To understand the similarity between two strains, using whole-genome alignment, we characterized the genomes of TCH1516 and LAC, having lengths of ~2.872 Mb and ~2.878 Mb, respectively (Table 1). According to NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (17), we found a high level of identity at the nucleotide level between

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

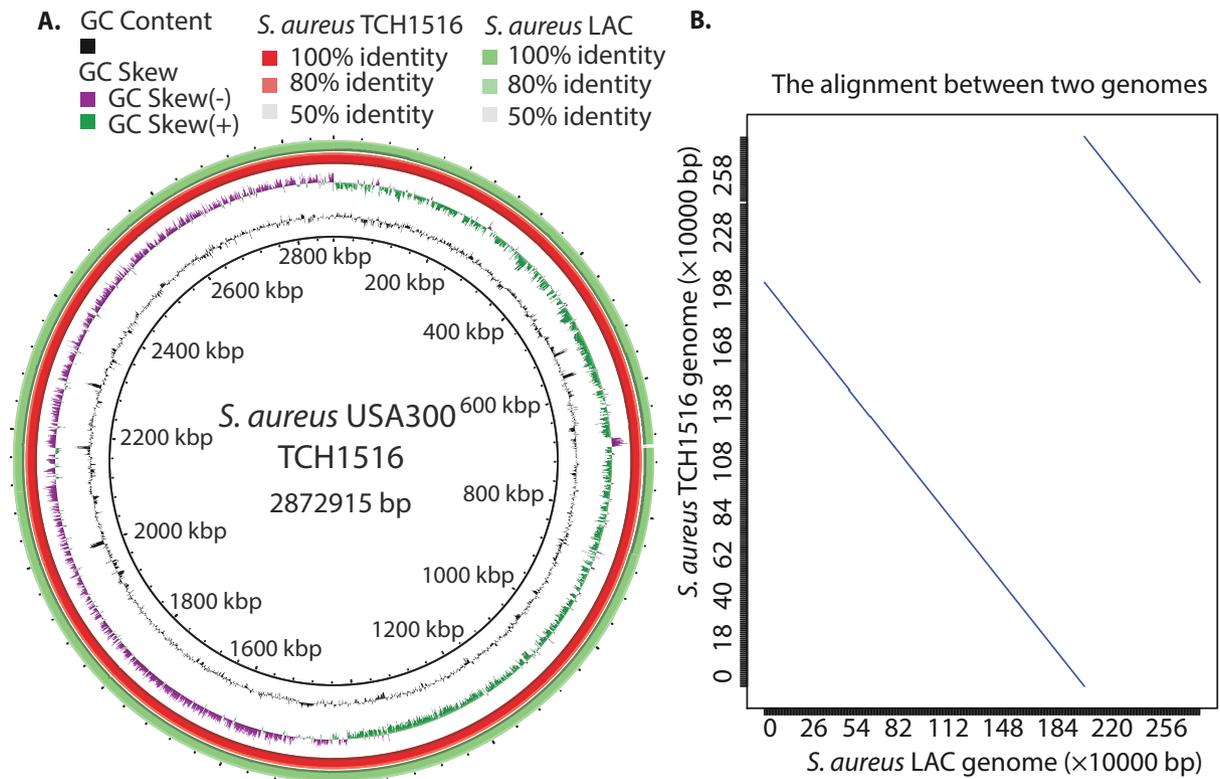
	<i>S. aureus</i> TCH1516	<i>S. aureus</i> LAC
Total sequence length (bp)	2,872,915	2,878,171
Total chromosome and plasmids	3, NC_010079.1 (chromosome) NC_012417.1 (plasmid 1) NC_010063.1 (plasmid 2)	2, CP035369.1 (chromosome) CP035370.1 (plasmid)
Gene (total)	2920	2892
CDS (total)	2841	2802
Genes/CDS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

these two closely related strains, and on average, they displayed 99.5% identity for all coding genes (Fig. 1A). Furthermore, MUMmer, a system for rapidly aligning large DNA sequences to one another, was used to check for the synteny between the strains (18). 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) (Fig. 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, 59 enriched CodY-binding regions that corresponded to 49 transcriptional units in the *S. aureus* NCTC 8325 were identified *in vitro* deploying an affinity purification experiment using purified *S. aureus* His-tagged CodY and a related mutant strain (10). 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, as it is more closely mimicking the nutritional environment of the human host (19).

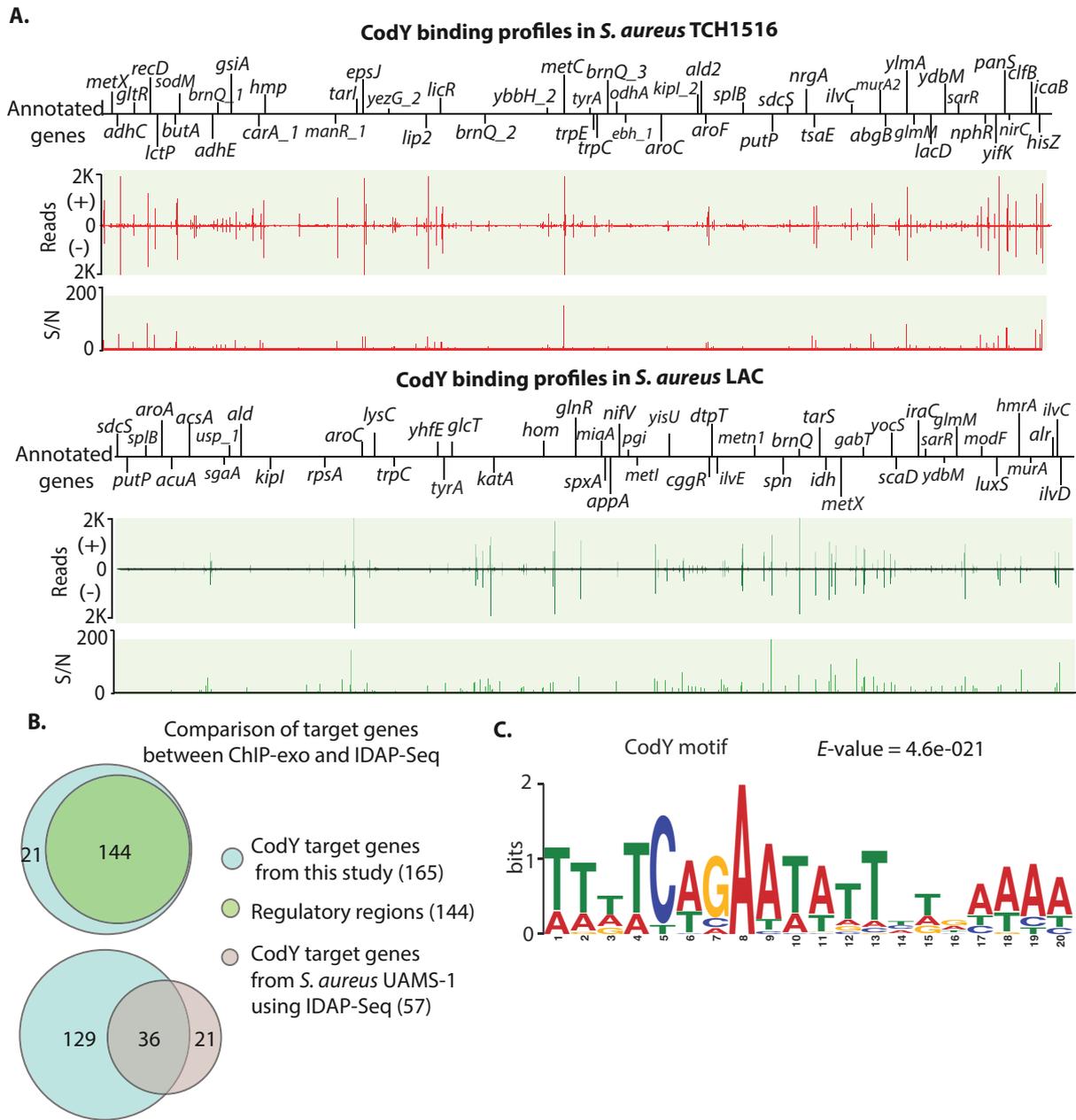
Using a peak calling algorithm (MACE), a total of 165 CodY target genes were identified from TCH1516 (Fig. 2A, full data set in Table S3). Of the 49 transcriptional



**FIG 1** The comparison of two dominant *Staphylococcus aureus* USA300 isolates (TCH1516 and LAC). (A) Circular representation of whole-genome comparison of *S. aureus* TCH1516 (red ring, reference strain) and LAC strains (green ring). Each ring of the circle represents a specific complete genome that corresponds to different colors in the legend above. 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.

units previously identified as strong candidates for direct regulation by CodY in *S. aureus* UAMS-1 using IDAP-Seq, 73% (36/49) 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 (Fig. 2B, upper panel). Most of the binding sites located in intergenic regions were



**FIG 2** Comparison of CodY-binding sites between TCH1516 and LAC strains. (A) Identification of CodY DNA-binding sites in *S. aureus* substr. USA300 TCH1516 and LAC at the genome. *S/N* denotes signal-to-noise ratio. (+) and (–) indicate forward and reverse reads, respectively. 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. 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.

present upstream of assigned genes, indicating that CodY may play critical regulatory roles in the expression of these genes. A total of 129 novel CodY target genes were identified (Fig. 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 we 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 (Fig. 2C). Furthermore, the *S. aureus* TCH1516 CodY-binding motif is similar with the CodY-binding motif (AATTTTCWGAATATTCWGAAAATT) reported in *Listeria monocytogenes* and *Bacillus subtilis* (20, 21). These results suggest that CodY likely has a conserved DNA-binding domain in Gram-positive bacteria (Fig. S1).

### 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 (Fig. S2). 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 (19).

*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 (Fig. 2A, bottom panel, Table S3). The alignment of the binding motifs revealed that the two strains have 18 nucleotides that overlap ( $P$ -value =  $9.35e-09$ ) by using TOMTOM. Among the 165 target genes, there are 10 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 (8). Taken together, these data demonstrated that the global regulator CodY controls the expression of metabolism and virulence genes in *S. aureus* (11).

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

As chromatin immunoprecipitation with high-throughput sequencing is widely used to identify genome-wide binding sites, many studies have initiated the investigation of the variation in the binding intensities of the same transcription factors between different conditions or homologous transcription factors across different strains. Though the genome-wide CodY-binding sites identified were the same in the two strains, we found that strain-specific binding peaks had differing binding intensities (Fig. S3), which

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

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

provided the case study to explore the causes of the variations of the binding intensities between two closely related strains.

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 on a merged set of CodY-binding peaks from the strains using MEME (22) (Fig. S4). We then utilized the computational method MAGGIE to measure differences in the DNA sequence in paired binding sites in the two strains (23). MAGGIE used two-sided Wilcoxon signed-rank tests to associate epigenomic changes at homologous sequences with motif mutations to identify motifs that likely contribute to the strain-specific difference. By performing the analysis on a consensus motif of both strains, we could eliminate the computation bias toward any one type of strain.

Among 135 pairs of peaks between the strains, 45 of 135 (33.3%) binding sites with identical DNA-binding sequences had a range of motif score difference between  $-1$  and  $1$  (Fig. S5). The pairs of peak heights for the 135 shared binding sites in LAC and TCH1516 were plotted (Fig. 3A). When we highlighted the 45 binding sites with near identical DNA sequences, we observed that they represent the dots closest to the 45-degree line. This indicates that these 45 sites have similar peak heights in the two strains. The binding sites where the DNA sequence differs are off the 45-degree line, showing that these differences lead to differential peak heights for the same binding site in the two strains.

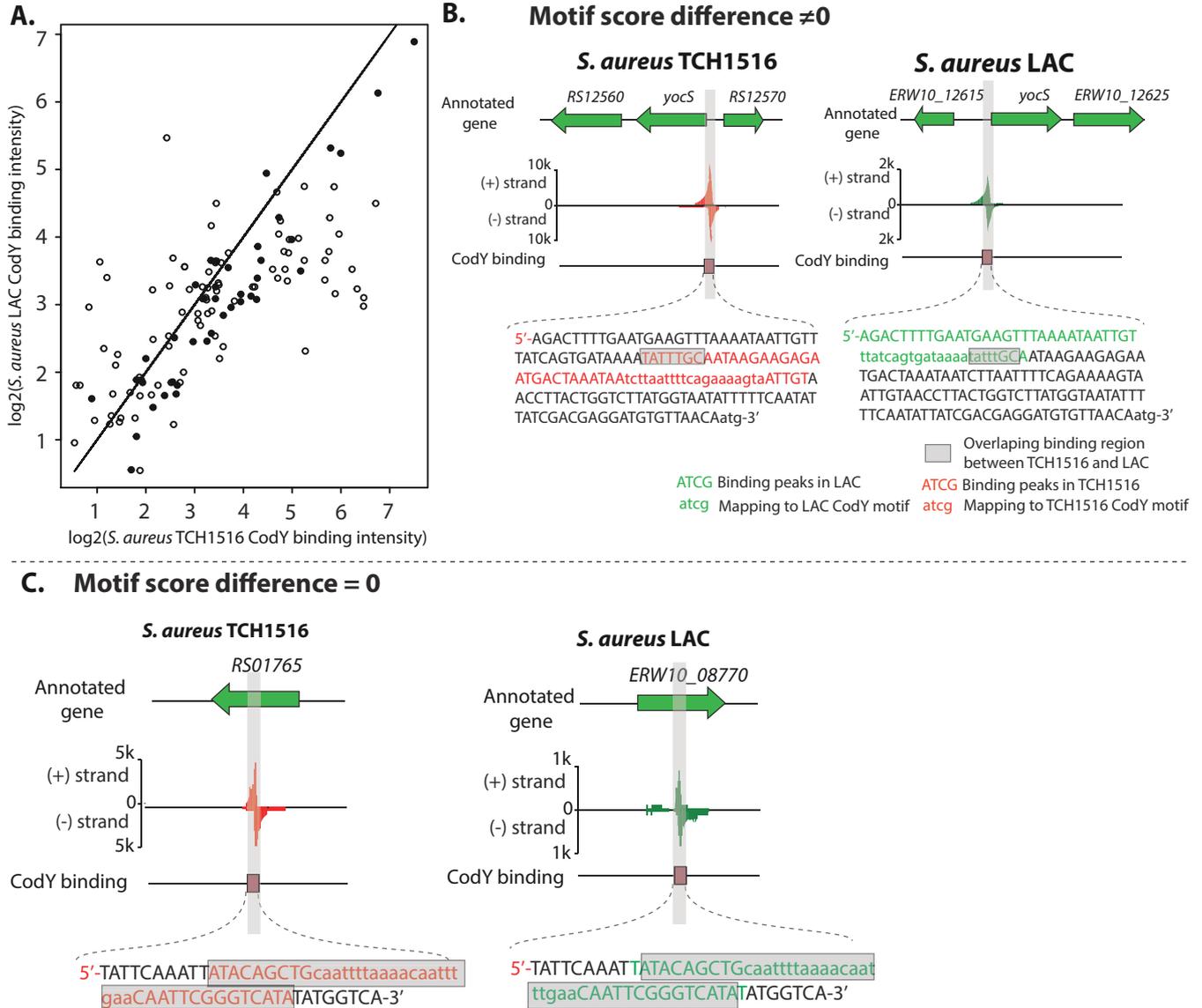
Furthermore, to visualize the differences between a pair of binding peaks, we mapped CodY-binding peaks to the reference genome. For example, the sequence of the peak located at the upstream of the gene *yocS* in TCH1516 slightly (7 bp) overlaps the corresponding peak sequence in LAC (Fig. 3B). Another example is from a pair of peaks from unknown genes USA300HOU\_RS01765 and ERW10\_08770. They had zero motif score difference, and thus we observed that both peak sequences nearly overlap each other (Fig. 3C).

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

The ChIP-exo data sets from this study expanded the number of target genes in the *S. aureus* USA300 lineage to 165, which included 129 novel CodY target genes. Of these, 37% (47 of 129) were metabolic genes, which were non-essential genes in *S. aureus* USA300 (24).

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 we found that there were 809 genes differentially expressed in the *codY* mutant [at least twofold change ( $P < 0.05$ ) in RNA-seq expression] (Fig. 4A). 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* (25). Furthermore, we used functional enrichment analysis by Clusters of Orthologous Groups (COGs) classification of 809 differentially expressed genes (DEGs). Of 809 DEGs, 193 are of the unknown function category; of those with known function, we found the top six differential enrichment pathways: 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 (Fig. 4B).

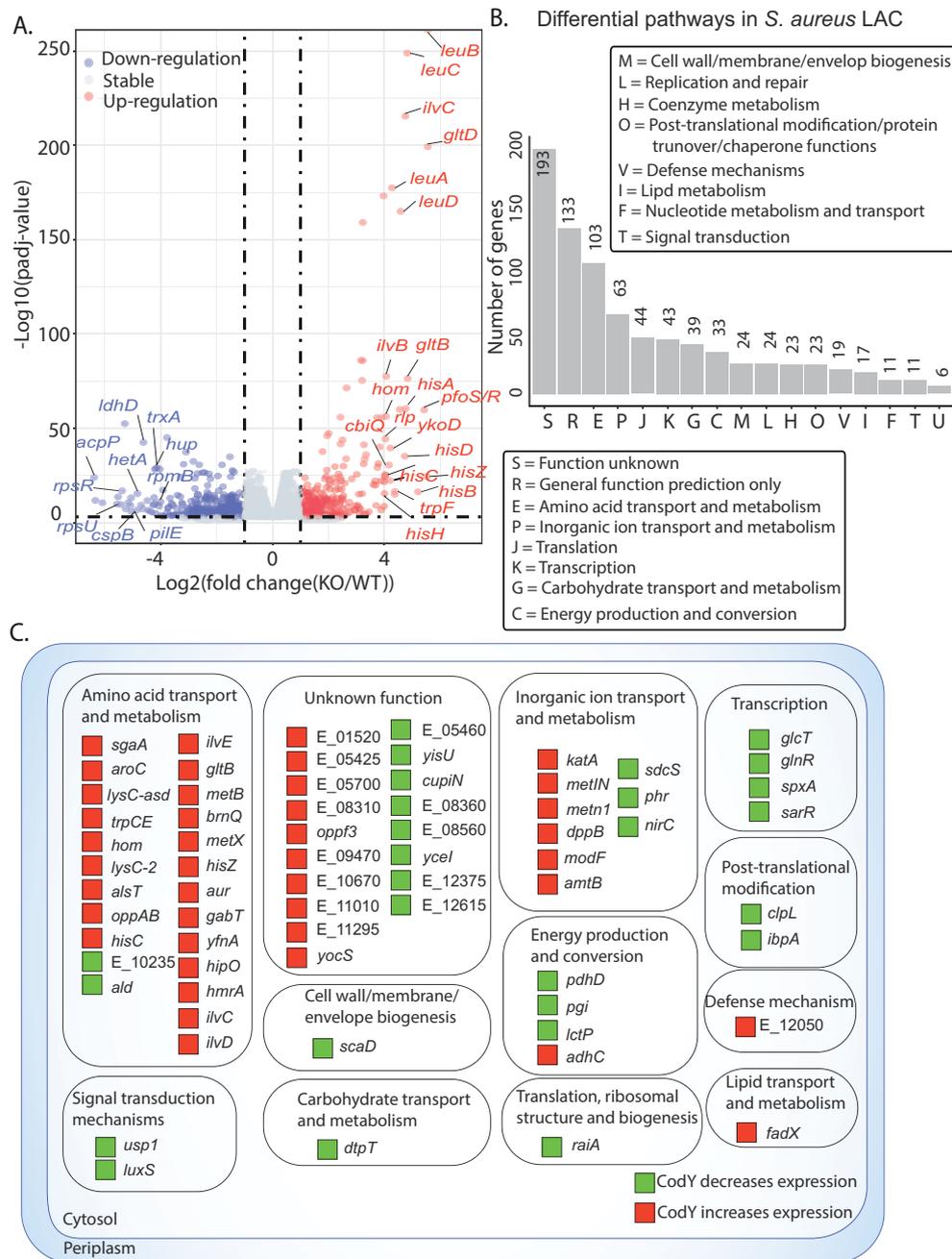
These 809 DEGs comprise the CodY regulon—a set of genes directly or indirectly regulated by CodY in the USA300 LAC strain. Prior studies integrated a global transcription factor DNA binding with expression profiles of cells lacking that same TF compared to wild-type cells to reconstruct the regulons of the transcription factor (26, 27). In this study, we integrated CodY-binding sites with transcriptomics to reconstruct the direct regulons of CodY—those genes that showed a change in expression when CodY was deleted and also had a corresponding binding peak upstream. Accordingly, we identified 72 direct regulons of CodY in the LAC strain (Fig. 4C) (28). Over half of the direct regulons (51%, 37 of 72) were related to the metabolic pathways (amino acid transport and metabolism, inorganic ion transport and metabolism, and energy production/



**FIG 3** Differential CodY-binding peaks area plot for TCH1516 and LAC highlighting binding sites with identical binding motifs. (A) Binding peak areas of  $\log_2(\text{TCH1516\_binding\_intensity})$  (x-axis) and  $\log_2(\text{LAC\_binding\_intensity})$  (y-axis) are shown. The diagonal line represents identical peak areas. The 45 binding sites with near zero sequence difference (Fig. S5) are shown with the solid dots, while those that are significantly 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. Gray 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.

conversion). In addition, 57% (41 of 72) of regulons were negatively regulated by CodY in the wild type. The remaining DEGs (737 of 809) were indirect regulons—those genes that show a change in expression when CodY is deleted but do not have a corresponding binding peak upstream.

Recent studies introduced an independent component analysis (ICA)-based framework in the bacteria that decomposes a compendium of RNA-sequencing (RNA-seq) expression profiles to determine the underlying regulatory structure and defined the term iModulon to describe the condition-invariant sets of genes regulated by the transcription factor. Each iModulon is calculated based on an ICA of RNA-seq data, which allows for a top-down determination of gene expression patterns (29). CodY iModulons were calculated based on RNA-seq data from *S. aureus* USA300 strains in the prior study



**FIG 4** Role of CodY regulon in *S. aureus* USA300. (A) A volcano plot representation of differential expression analysis performed between wild type and *codY* mutant. (B) Functional enrichment analysis by Clusters of Orthologous Groups (COGs) classification of 809 differentially expressed genes in the *S. aureus 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 regulons in the *S. aureus* USA300.

(30). Importantly, we compared CodY direct regulons from the experimental data with the reported CodY iModulon from ICA-based machine learning and found that 90% (65/72) of direct CodY regulons significantly overlap with CodY iModulons, which confirmed the regulatory roles of CodY (Fig. S6). The CodY iModulon contains a set of genes having significant overlap with previously predicted CodY regulons.

Furthermore, 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 BCAAs

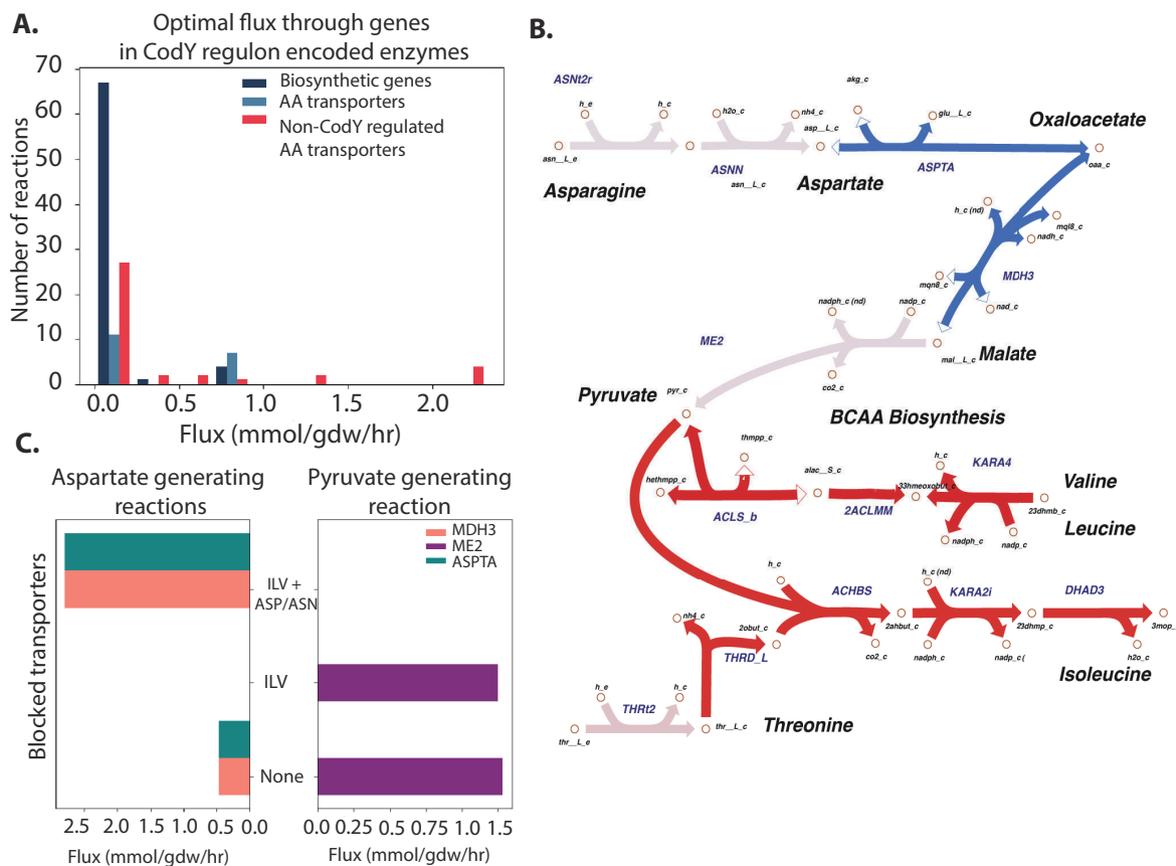
CodY is reported to regulate the expression of metabolic genes in response to changes in the pools of specific metabolites, i.e., the branched-chain amino acids (BCAAs); isoleucine, leucine, and valine (ILV); and nucleoside triphosphate (GTP), to regulate genes involved in the biosynthesis of these amino acids (31). Despite responding primarily to ILV levels, CodY also regulates the expression of genes involved in biosynthesis of other amino acids, such as threonine, histidine, and aspartate. Therefore, we sought to understand how *S. aureus* manages flux through these CodY regulated biosynthetic enzymes during ILV starvation.

In order to understand the partition of fluxes required for ILV synthesis, we performed parsimonious Flux Balance Analysis (pFBA) using the metabolic model iYS854 of *S. aureus* USA300 (32, 33). pFBA determines metabolic fluxes that maximize growth while minimizing the sum of fluxes through the system (32). As expected, simulated growth in RPMI supplemented with ILV led to growth with zero flux through the CodY regulated biosynthetic pathway in favor of direct transport of the necessary metabolites (Fig. 5A). Restricting flux through all ILV transporters (see Materials and Methods) led to a spike in flux through the ILV biosynthetic enzymes. Interestingly, in ILV starvation conditions, 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 toward generating pyruvate (another precursor for ILV biosynthesis) via malate (Fig. 5B). This result indicated that in a BCAA-deficient environment with sufficient asparagine and aspartate, *S. aureus* generates pyruvate to increase precursor pools available for isoleucine biosynthesis. Indeed, flux through the two reactions linking the TCA cycle and aspartate, malate dehydrogenase (MDH3), and aspartate transaminase (ASPT) increased when aspartate and asparagine transporters were blocked in addition to BCAA transporters (Fig. 5C). Flux through malic enzyme (ME2) was only detectable in the presence of aspartate and asparagine, i.e., when the flux through those enzyme transporters were not blocked. Taken together, our pFBA analysis demonstrates how fluxes through CodY-regulated biosynthetic enzymes are rerouted in ILV starvation conditions to generate required precursors. It suggests that the unified regulation of genes encoding enzymes involved in various amino acid biosynthesis pathways by CodY is due to the need for *S. aureus* to coordinate the fluxes through these pathways to efficiently generate ILV.

## 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 (34). Although many properties of pathogenic strains can now be predicted from sequence alone (35–37), detailed experimental characterization of differences in multiple strains is also needed. In this study, we combined genome-wide experiments and computational modeling to address the differences of CodY in the dominant CA-MRSA USA300 clinical isolates TCH1516 and LAC (15). The study resulted in a series of significant findings.

First, through the genome-wide identification of *in vivo* binding sites at single-nucleotide level, we found the same 165 CodY target genes in two clinical USA300 strains. Although the prior study by Majerczyk et al. (10) identified 59 regions bound by CodY in *S. aureus* NCTC8325 through *in vitro* DNA affinity technology, those results could not represent the scope of CodY target genes in USA300 strains due to genome sequence differences and pathogenicity between USA300 strains and NCTC8325. In addition, the ChIP-exo method reveals a snapshot of the binding sites obtained in a particular strain



**FIG 5** Central metabolic flux rerouting to generate branched-chain amino acids. (A) Optimal flux through CodY-regulated biosynthetic enzymes, transporters, and non-CodY-regulated transporters in RPMI medium. Y-axis represents the number of reactions with a given flux. A histogram showing fluxes through different categories of reactions (i.e., CodY-regulated biosynthetic enzymes, CodY-regulated AA transporters, and non-CodY-regulated transporters). When amino acids are present, *S. aureus* imports them via transporters (light blue and red bars). (B) pFBA solution of *S. aureus* grown in RPMI without ILV predicts rerouting of several central carbon metabolic fluxes to generate ILV precursors. Red arrows represent reactions with increased flux during ILV starvation, and blue arrows represent those with decreased flux relative to ILV rich conditions. Malate dehydrogenase and aspartate synthase (blue arrows) lower flux when BCAA transporters are blocked. Note: Full ILV biosynthesis pathways are not shown. (C) A histogram showing the number of reactions under different nutrient conditions. X-axis shows different conditions (different amino acids conditions, i.e., none, ILV, ILV + ASP/ASN). Double y-axis: left y-axis shows the flux; right y-axis shows the number of reactions. Flux through pyruvate generating malate enzyme (ME2) and aspartate generating malate dehydrogenase (MDH3) and aspartate transaminase (ASPTA) are dependent on the presence of aspartate and asparagine in the media.

under tested conditions. As a global regulator, CodY participated in the regulation of many cellular processes, including amino acid biosynthesis, transport of nutrient molecules, and virulence. Some regulations may require the cofactors to activate the binding affinity between CodY and DNA. Thus, ChIP-exo is appropriate to provide a snapshot of target genes that reflects the *in vivo* status of CodY under the conditions used. Furthermore, it provides a near base pair (bp)-level resolution of protein-DNA interaction, which greatly improves the accuracy of identifying CodY-binding sites at the genome and the sequences of the CodY motif.

Second, an examination of the differential binding intensity of CodY in the promoter regions of the same target genes in the two different strains under the same conditions revealed that the variance was due to DNA sequence differences in the same CodY-binding site in the two strains, while the CodY protein was identical in both strains. This novel finding reveals insights into the system-level analysis of CodY target genes and differential binding intensity across closely related strains.

Third, the study identified 10 virulence genes that belong to different types of virulence factors, which demonstrated that CodY connects metabolism genes with virulence genes in *S. aureus* (8). In addition, our combined ChIP-exo and RNA-sequencing

analysis also shed light on how CodY affects *agr* expression. As previously described and confirmed by our ChIP-exo data, CodY represses the expression of *agr* locus but does not bind strongly to the *agr* promoter (13). However, our data also show that CodY both binds to the *sarR* promoter and the expression level of *sarR* gene goes down in *codY* knock-out mutant. The product of this gene, SarR, is a repressor of the *agr* locus, which suggests a possible avenue by which CodY is able to repress *agr* expression (38). It is worth noting that the LAC strain used for gene expression profiling in this study expresses *agr* at high levels even in the exponential phase, and therefore it was not significantly affected by *codY* knockout as previously noted. Considering that different *S. aureus* lineages may have distinct virulence factors, we could expand this study to identify many other virulence factor genes coordinated by CodY across different *S. aureus* strains.

Finally, an existing genome-scale model of the metabolic network in *S. aureus* can be constrained by the regulatory action of CodY (33). This model computes the consequent systematic rerouting fluxes through metabolism of BCAAs as a result of CodY regulatory action.

Taken together, this study demonstrates, for the first time, the similarities in the function of a conserved globally acting transcription factor (i.e., CodY) between closely related pathogenic strains of *S. aureus* USA300. Understanding the CodY regulon will allow deciphering the network of CodY controlling the amino acid metabolism and virulence and motivate an evaluation of the possibility of developing appropriate treatments for major strains circulating in the population.

## MATERIALS AND METHODS

### Bacterial strains and culture conditions

The bacterial strains used in this study were described in Table S1. 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 (39). MRSA USA300 LAC was originally isolated from the Los Angeles county jail (15). MRSA USA300 LAC contains three small plasmids, one encoding resistance to tetracycline and another encoding erythromycin resistance. The third plasmid is cryptic. For ease of genetic manipulation and to avoid interference, all three plasmids were cured, yielding strain *S. aureus* USA300 JE2. Thus, *S. aureus* JE2, as a parental strain, was used for all sequence-defined Tn mutagenesis experiments. The *codY* mutant was from *S. aureus* JE2 in which each of non-essential genes was disrupted via mariner *bursa aurealis* Tn mutagenesis. *S. aureus* JE2 and a transposon mutant strain that integrated into *codY* are from Nebraska Transposon Mutant Library. All *S. aureus* strains were grown in tryptic soy broth (TSB, Sigma-Aldrich, St. Louis, MO) or RPMI-1640 (Gibco, Houston, TX) with 10% lysogeny broth (LB, Sigma-Aldrich, St. Louis, MO) containing 10 g/L peptone, 5 g/L yeast extract, and 10 g/L NaCl with shaking (250 rpm) at 37°C, maintaining a flask-to-medium volume ratio of 9:1, unless otherwise specified.

### Protein expression and monoclonal antibody production

The monoclonal antibody used was developed with GenScript (Piscataway, NJ). Briefly, the *S. aureus* USA300 *codY* gene (see Table S2) with 6X His tag was synthesized and cloned into a pET-30a(+) plasmid (EMD Millipore, Burlington, MA) using the NdeI and HindIII sites. The protein was expressed in *E. coli* BL21(DE3) in the Terrific Broth medium containing 50 µg/mL kanamycin. The culture was incubated at 37°C with shaking at 200 rpm. When the optical density at 600 nm (OD<sub>600</sub>) reached about 0.5, the cell culture was transferred to 15°C. Then, protein expression was induced with addition of 0.4 mM IPTG (final concentration) and shaking for an additional 12 h. Then, cells were harvested by centrifugation and lysed by sonication at 4°C, using a tip probe at cycles of 5 s ON and 25 s OFF at a power level at 6.0 until full lysis (Misonix Sonicator 3000

Ultrasonic Cell Disruptor, Cole-Parmer, Vernon Hills, IL). The His-tagged CodY protein was purified as previously described using Ni-NTA affinity columns (Qiagen, Germany) (40). The concentration was determined by the BCA protein assay (Thermo Fisher Scientific, Waltham, MA) with bovine serum albumin as the standard. Using purified His-tagged CodY, monoclonal anti-CodY antibody was generated by injection of CodY into the BALB/C mouse. The custom-made monoclonal antibody (IgG) that binds the CodY protein was produced and purified by GenScript (Piscataway, NJ).

### ChIP-exo experiments

ChIP-exo experimentation was performed following the procedures described previously using a custom-made antibody that binds the CodY protein (27). To identify CodY-binding maps for each strain *in vivo*, the culture with an OD<sub>600</sub> of ~0.5 was used to generate ChIP-DNA. Formaldehyde was added at a final concentration of 1% and incubated at room temperature for 20 min in a platform rocker. Cross-linking was quenched by adding 250 mM of glycine and incubating for 5 min. Cells were washed three times with ice-cold 1 X PBS (Thermo Fisher Scientific, Waltham, MA). The resulting cells were lysed with Ready-lyse lysozyme solution (Epicentre, Waltham, MA). Lysate was sheared by sonication (Misonix Sonicator 3000) with the settings (the output power level 9.0 for 30 min at 4 degree) to generate 300–500 bp randomly sheared chromosomal DNA fragments. Then, immunoprecipitation was carried out at 4°C with overnight incubation by using a custom-made antibody that binds the CodY protein (GenScript, Piscataway, NJ). In contrast, mouse IgG (MilliporeSigma™ Upstate) was used as internal control. The CodY protein, together with its cross-linked DNA and covalently bound mouse antibody, was captured with 50-μL Dynabeads Pan mouse IgG (Invitrogen) and washed with buffer I [50 mM Tris-HCl (pH 7.5), 140 mM NaCl, 1 mM EDTA, and 1% Triton X-100]. ChIP materials (chromatin beads) were used to perform on-bead enzymatic reactions of the ChIP-exo method (41). 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 a first adaptor (5'-phosphorylated) using the dA-Tailing Module (New England Biolabs, Ipswich, MA) and the NEBNext Quick Ligation Module (New England Biolabs), respectively. The first adaptor was designed to have different indices to distinguish different DNA samples after the sequencing. After ligation, multiple ChIP materials could be pooled together. 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 incubated overnight at 65°C to reverse 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 (27) 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, Germany) and quantified using Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA). Quality of the DNA sample was checked by running Agilent High Sensitivity DNA Kit using Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) before sequenced using HiSeq 2500 (Illumina) following the manufacturer's instructions.

### Peak calling for ChIP-exo data set

Peak calling was performed as previously described (27). Sequence reads generated from ChIP-exo were mapped onto the reference genome using bowtie (42) with default options to generate SAM output files. MACE program was used to define peak candidates from biological duplicates for each experimental condition with sequence depth normalization (43). To reduce false-positive peaks, peaks with signal-to-noise (S/N) ratio less than 1.5 were removed. The noise level was set to the top 5% of signals at

genomic positions because the top 5% makes a background level in a plateau and top 5% intensities from each ChIP-exo replicates across conditions correlate well with the total number of reads (27, 44, 45). The calculation of S/N ratio resembles the way to calculate ChIP-chip peak intensity where the IP signal was divided by Mock signal. Genome-scale data were visualized using MetaScope (<https://sites.google.com/view/systemskimlab/software?authuser=0>).

### Motif search from ChIP-exo peaks

The sequence motif analysis for CodY-binding sites was performed using the MEME software suite (22). For each strain, sequences in binding regions were extracted from the reference genome (*S. aureus* TCH1516: GenBank: NC\_010079.1, NC\_012417.1, and NC\_010063.1; *S. aureus* LAC: GenBank: CP035369.1 and CP035370.1). To achieve a more accurate motif, the sequence of each binding site was extended by 10 bp at each end. The width parameter was fixed at 20 bp, and the minsites parameter was fixed at 90% of the total number of the sequence. All other parameters followed the default setting.

### COG enrichment

CodY regulons were categorized according to their annotated COG database (46). 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 among the sequences (18). BLAST Ring Image Generator (BRIG) was used to show a genome-wide visualization of coding sequences identity between *S. aureus* USA300 TCH1516 and LAC (47). 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 (48).

### Determining the orthologs between different *S. aureus* strains with bidirectional BLAST hits

To determine the orthologs between different strains of *S. aureus*, core genome-containing conserved genes between the TCH1516 and LAC were first established using bidirectional BLAST hits (49). 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 (i) the two genes have the highest alignment percent to each other than to any other genes in the genome and (ii) the coverage is at least 80%.

### RNA extraction and library preparation

The culture of *S. aureus* JE2 and its derivatives of *codY* mutant were incubated in RPMI+10% LB medium at 37 degree overnight with agitation, and then they were used to inoculate the fresh media (1/200 dilution). The volume of the fresh media was 50 mL for each biological replicate. The fresh culture was incubated in RPMI+10% LB medium at 37 degree with agitation to the mid-log phase ( $OD_{600} \approx 0.5$ ). RNA-seq was performed using two biological replicates. Once the cell culture reached the mid-exponential phase ( $OD_{600} = 0.5$ ), the cell culture was collected for RNA extraction. Transcripts were stabilized by mixing 3 mL of cell cultures at the mid-log phase with 6 mL of RNAprotect Bacteria Reagent (Qiagen). Samples were immediately vortexed for 5 s, incubated for 5 min at room temperature, and then centrifuged at 5000 g for 10 min. The supernatant was decanted and any residual supernatant was removed by inverting the tube once onto a paper towel. Total RNA samples were then isolated using the RNeasy Plus Mini kit

(Qiagen) following the manufacturer's instruction. Samples were then quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific), and quality of the isolated RNA was checked by running RNA 6000 Pico Kit using Agilent 2100 Bioanalyzer (Agilent). Paired-end, strand-specific RNA-seq library was prepared using KAPA RNA Hyper Prep kit (KAPA Biosystems, Wilmington, MA), following the instructions (50). Resulting libraries were analyzed on an Agilent Bioanalyzer DNA 1000 chip (Agilent). Sequencing was performed on a HiSeq 2500 sequencer at the Genomics Core facility of University of California, San Diego, CA.

### RNA-Seq data processing

The RNA-seq pipeline used to analyze and perform QC/QA has been described in detail previously (51). Briefly, the sequences were aligned to the reference genomes using Bowtie2 (52). The aligned sequences were assigned to ORFs using HTSeq-counts (53). Differential expression analysis was performed using DESeq2 with a P-value threshold of 0.05 and an absolute fold-change threshold of 2 (54). To create the final counts matrix, counts from conserved genes in LAC samples were represented by the corresponding ortholog in TCH1516. The counts for accessory genes were filled with 0s if the genes were not present in the strain (i.e., LAC-specific genes had counts of 0 in TCH1516 samples and vice versa). Finally, to reduce the effect of noise, genes with average counts per sample <10 were removed. The final counts matrix with 2,581 genes was used to calculate transcripts per million (TPM).

### Metabolic modeling and assessment of significant differences in flux distribution

We used BiGG model iYS854 and set the lower bound to the corresponding nutrient exchange to  $-1$  mmol/gDW/h (the negative sign is a modeling convention to allow for the influx of nutrients) and  $-13$  mmol/gDW/h for oxygen exchange (as measured experimentally) (33). Next, we compared two conditions with (i) amino acid-rich medium and (ii) amino acid-poor medium. In the first condition, assuming that in the presence of amino acid, CodY mediates the repression of multiple target genes. Specifically, we turned off various combinations of amino acid transporters as reported in Table S4 by setting both the minimum and maximum bounds through their corresponding exchange reactions to 0. No regulatory constraints were added. Flux-balance analysis (FBA) was implemented with the biomass formation set as the functional network objective. Next, flux balance analysis was run in both conditions and the fluxes were sampled 10,000 times. All reaction fluxes were normalized by dividing by the growth rate to account for growth differences across the two media types. The flux distribution for each metabolic process was compared across both conditions using the Kolmogorov-Smirnov test, a non-parametric test which compares two continuous probability distributions. The distribution across two reactions was deemed to be significantly different when the Kolmogorov-Smirnov statistic was larger than 0.99 with an adjusted *P*-value < 0.01.

### Motif mutation analysis

To gain insights into the CodY motifs from TCH1516 and LAC strains, MAGGIE was used for motif mutation analysis (23). 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.

## ACKNOWLEDGMENTS

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

This work was funded by the NIH NIAID (Grant number U01AI124316).

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Biological Sciences, University of California San Diego, La Jolla, California, USA

<sup>2</sup>Department of Bioengineering, University of California San Diego, La Jolla, California, USA

<sup>3</sup>Department of Pediatrics, University of California San Diego, La Jolla, California, USA

<sup>4</sup>Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, California, USA

<sup>5</sup>Novo Nordisk Foundation Center for Biosustainability, Kongens Lyngby, Denmark

## AUTHOR ORCID<sub>s</sub>

Ye Gao  <http://orcid.org/orcid.org/0000-0001-8519-1392>

## FUNDING

Funder	Grant(s)	Author(s)
HHS   National Institutes of Health (NIH)	U01AI124316	Bernhard O. Palsson

## DATA AVAILABILITY STATEMENT

The ChIP-exo and RNA-seq data sets are accessible through GEO under accession numbers [GSE159856](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159856) and [GSE163312](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163312).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Fig S1 (mSystems00279-23-S001.eps).** 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).

**Fig S2 (mSystems00279-23-S002.eps).** Comparison of CodY sequences between *S. aureus* USA300 TCH1516 and LAC strains. The sequence alignment shows identical CodY amino acid residues in two distinct but related *S. aureus* strains.

**Fig S3 (mSystems00279-23-S003.eps).** Comparison of CodY-binding intensity between TCH1516 and LAC strains. Though the genome-wide CodY-binding sites identified were the same in the two strains, strain-specific binding peaks had differing intensities.

**Fig S4 (mSystems00279-23-S004.eps).** *S. aureus* USA300 CodY-binding motif from a merged set of TCH1516 and LAC binding peaks. There is a 20-bp CodY motif based on a merged set of CodY-binding peaks from the strains using MEME (E-value = 1.5e-082).

**Fig S5 (mSystems00279-23-S005.eps).** Distribution of CodY peaks based on the range of motif score differences. X-axis represents the range of motif score differences. Y-axis represents the number of peaks that fall into each range. Black-color bar shows 45 peaks with the motif score difference at the range of (-1, 1).

**Fig S6 (mSystems00279-23-S006.eps).** Comparison of genes in the CodY regulon (red) and CodY iModulon (blue). 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.

**Table S1 (mSystems00279-23-S007.docx).** The strains used in this study.

**Table S2 (mSystems00279-23-S008.docx).** The primers and plasmids used in this study.

**Table S3 (mSystems00279-23-S009.docx).** The binding sites of CodY in *S. aureus* USA300 LAC.

**Table S4 (mSystems00279-23-S010.docx).** The combinations of amino acids and transporters in the model.

## REFERENCES

- Bagnoli F, Rappuoli R, Grandi G. 2017. *Staphylococcus aureus*: microbiology, pathology, immunology, therapy and prophylaxis. Springer. <https://doi.org/10.1007/978-3-319-72063-0>
- Lee AS, de Lencastre H, Garau J, Kluytmans J, Malhotra-Kumar S, Peschel A, Harbarth S. 2018. Methicillin-resistant *Staphylococcus aureus*. Nat Rev Dis Primers 4:18033. <https://doi.org/10.1038/nrdp.2018.33>
- World Health Organization. 2017. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis
- Foster TJ, Geoghegan JA, Ganesh VK, Höök M. 2014. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nat Rev Microbiol 12:49–62. <https://doi.org/10.1038/nrmicro3161>
- 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. <https://doi.org/10.2174/1389450117666161128123536>
- 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*. Int J Mol Sci 21:5772. <https://doi.org/10.3390/ijms21165772>
- Turner NA, Sharma-Kuinkel BK, Maskarinec SA, Eichenberger EM, Shah PP, Carugati M, Holland TL, Fowler VG. 2019. Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. Nat Rev Microbiol 17:203–218. <https://doi.org/10.1038/s41579-018-0147-4>
- 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. <https://doi.org/10.1128/JB.01492-08>
- 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. <https://doi.org/10.1128/IAI.01710-14>
- 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. <https://doi.org/10.1128/JB.00220-10>
- 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. <https://doi.org/10.1111/mmi.13404>
- Basu A, Shields KE, Eickhoff CS, Hoft DF, Yap MN. 2018. Thermal and nutritional regulation of ribosome hibernation in *Staphylococcus aureus*. J Bacteriol 200:e00426-18. <https://doi.org/10.1128/JB.00426-18>
- 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. <https://doi.org/10.1128/JB.00128-13>
- Thurlow LR, Joshi GS, Richardson AR. 2012. Virulence strategies of the dominant USA300 lineage of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA). FEMS Immunol Med Microbiol 65:5–22. <https://doi.org/10.1111/j.1574-695X.2012.00937.x>
- Vitko NP, Richardson AR. 2013. Laboratory maintenance of methicillin-resistant *Staphylococcus aureus* (MRSA). Curr Protoc Microbiol Chapter 9:Unit 9C.2. <https://doi.org/10.1002/9780471729259.mc09c02s28>
- Kennedy AD, Porcella SF, Martens C, Whitney AR, Braughton KR, Chen L, Craig CT, Tenover FC, Kreiswirth BN, Musser JM, DeLeo FR. 2010. Complete nucleotide sequence analysis of plasmids in strains of *Staphylococcus aureus* clone USA300 reveals a high level of identity among isolates with closely related core genome sequences. J Clin Microbiol 48:4504–4511. <https://doi.org/10.1128/JCM.01050-10>
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. 1999. Alignment of whole genomes. Nucleic Acids Res 27:2369–2376. <https://doi.org/10.1093/nar/27.11.2369>
- Machado H, Weng LL, Dillon N, Seif Y, Holland M, Pekar JE, Monk JM, Nizet V, Pálsson BO, Feist AM. 2019. Strain-specific metabolic requirements revealed by a defined minimal medium for systems analyses of *Staphylococcus aureus*. Appl Environ Microbiol 85:e01773-19. <https://doi.org/10.1128/AEM.01773-19>
- Biswas R, Sonenshein AL, Belitsky BR. 2020. Genome-wide identification of *Listeria monocytogenes* cody-binding sites. Mol Microbiol 113:841–858. <https://doi.org/10.1111/mmi.14449>
- Belitsky BR, Sonenshein AL. 2013. Genome-wide identification of *Bacillus subtilis* cody-binding sites at single-nucleotide resolution. Proc Natl Acad Sci U S A 110:7026–7031. <https://doi.org/10.1073/pnas.1300428110>
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Shen Z, Hoeksema MA, Ouyang Z, Benner C, Glass CK. 2020. MAGGIE: leveraging genetic variation to identify DNA sequence motifs mediating transcription factor binding and function. Bioinformatics 36:i84–i92. <https://doi.org/10.1093/bioinformatics/btaa476>
- Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, Bayles KW. 2013. A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. mBio 4:e00537–12. <https://doi.org/10.1128/mBio.00537-12>
- Majerczyk CD, Sadykov MR, Luong TT, Lee C, Somerville GA, Sonenshein AL. 2008. *Staphylococcus aureus* CodY negatively regulates virulence gene expression. J Bacteriol 190:2257–2265. <https://doi.org/10.1128/JB.01545-07>
- Seo SW, Kim D, O'Brien EJ, Szubin R, Pálsson BO. 2015. Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in *Escherichia coli*. Nat Commun 6:7970. <https://doi.org/10.1038/ncomms8970>
- Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Pálsson BO. 2014. Deciphering fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. Nat Commun 5:4910. <https://doi.org/10.1038/ncomms5910>
- Myers KS, Park DM, Beauchene NA, Kiley PJ. 2015. Defining bacterial regulons using ChIP-Seq. Methods 86:80–88. <https://doi.org/10.1016/j.ymeth.2015.05.022>
- Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA, Pálsson BO. 2019. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. Nat Commun 10:5536. <https://doi.org/10.1038/s41467-019-13483-w>

30. Poudel S, Tsunemoto H, Seif Y, Sastry AV, Szubin R, Xu S, Machado H, Olson CA, Anand A, Pogliano J, Nizet V, Pálsson BO. 2020. Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. *Proc Natl Acad Sci U S A* 117:17228–17239. <https://doi.org/10.1073/pnas.2008413117>
31. Brinsmade SR, Kleijn RJ, Sauer U, Sonenshein AL. 2010. Regulation of CodY activity through modulation of intracellular branched-chain amino acid pools. *J Bacteriol* 192:6357–6368. <https://doi.org/10.1128/JB.00937-10>
32. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez - Ferrer D, Weitz KK, Eils R, König R, Smith RD, Pálsson BØ. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome - scale models . *Mol Syst Biol* 6:390. <https://doi.org/10.1038/msb.2010.47>
33. Seif Y, Monk JM, Mih N, Tsunemoto H, Poudel S, Zuniga C, Broddrick J, Zengler K, Pálsson BO. 2019. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLoS Comput Biol* 15:e1006644. <https://doi.org/10.1371/journal.pcbi.1006644>
34. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, Karpinetz T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15:141–161. <https://doi.org/10.1007/s10142-015-0433-4>
35. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Pálsson BØ. 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A* 110:20338–20343. <https://doi.org/10.1073/pnas.1307797110>
36. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, Feist AM, Pálsson BO. 2017. IML1515, a knowledge-base that computes *Escherichia coli* traits. *Nat Biotechnol* 35:904–908. <https://doi.org/10.1038/nbt.3956>
37. Norsigian CJ, Danhof HA, Brand CK, Oezguen N, Midani FS, Pálsson BO, Savidge TC, Britton RA, Spinler JK, Monk JM. 2020. Systems biology analysis of the clostridioides difficile core-genome contextualizes microenvironmental evolutionary pressures leading to genotypic and phenotypic divergence. *NPJ Syst Biol Appl* 6:31. <https://doi.org/10.1038/s41540-020-00151-9>
38. Reyes D, Andrey DO, Monod A, Kelley WL, Zhang G, Cheung AL. 2011. Coordinated regulation by AgrA, SarA, and SarR to control agr expression in *Staphylococcus aureus*. *J Bacteriol* 193:6020–6031. <https://doi.org/10.1128/JB.05436-11>
39. HulténKGKaplan SL, Gonzalez BE, Hammerman WA, Lamberth LB, Versalovic J, Mason EO. 2006. Three-year surveillance of community onset health care-associated *Staphylococcus aureus* infections in children. *Pediatr Infect Dis J* 25:349–353. <https://doi.org/10.1097/O1.inf.0000207404.50143.1e>
40. Frenzel E, Doll V, Pauthner M, Lücking G, Scherer S, Ehling-Schulz M. 2012. CodY orchestrates the expression of virulence determinants in emetic *Bacillus cereus* by impacting key regulatory circuits. *Mol Microbiol* 85:67–88. <https://doi.org/10.1111/j.1365-2958.2012.08090.x>
41. Rhee HS, Pugh BF. 2012. Chip-Exo method for identifying Genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol* Chapter 21:Unit 21.24. <https://doi.org/10.1002/0471142727.mb2124s100>
42. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
43. Wang L, Chen J, Wang C, Uusküla-Reimand L, Chen K, Medina-Rivera A, Young EJ, Zimmermann MT, Yan H, Sun Z, Zhang Y, Wu ST, Huang H, Wilson MD, Kocher J-P, Li W. 2014. MACE: model based analysis of chip-exo. *Nucleic Acids Res* 42:e156. <https://doi.org/10.1093/nar/gku846>
44. Seo SW, Kim D, Szubin R, Pálsson BO. 2015. Genome-Wide reconstruction of oxyR and SoxRS transcriptional regulatory networks under oxidative stress in *Escherichia coli* K-12 MG1655. *Cell Rep* 12:1289–1299. <https://doi.org/10.1016/j.celrep.2015.07.043>
45. Ogasawara H, Ohe S, Ishihama A. 2015. Role of transcription factor NimR (YeaM) in sensitivity control of *Escherichia coli* to 2-nitroimidazole. *FEMS Microbiol Lett* 362:1–8. <https://doi.org/10.1093/femsle/fnu013>
46. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>
47. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <https://doi.org/10.1186/1471-2164-12-402>
48. Treangen TJ, Messeguer X. 2006. M-GCAT: interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species. *BMC Bioinformatics* 7:433. <https://doi.org/10.1186/1471-2105-7-433>
49. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* 96:2896–2901. <https://doi.org/10.1073/pnas.96.6.2896>
50. Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan JA, Swerdlow HP, Oyola SO. 2012. Optimal enzymes for amplifying sequencing libraries. *Nat Methods* 9:10–11. <https://doi.org/10.1038/nmeth.1814>
51. Poudel S, Tsunemoto H, Meehan M, Szubin R, Olson CA, Lamsa A, Seif Y, Dillon N, Vrbancac A, Sugie J, Dahesh S, Monk JM, Dorrestein PC, Pogliano J, Knight R, Nizet V, Pálsson BO, Feist AM. 2019. Characterization of CA-MRSA TCH1516 exposed to nafcillin in bacteriological and physiological media. *Sci Data* 6:43. <https://doi.org/10.1038/s41597-019-0051-4>
52. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
53. Anders S, Pyl PT, Huber W. 2015. HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>
54. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>