AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

Check for updates

| Computational Biology | Research Article

Elucidation of independently modulated genes in *Streptococcus pyogenes* reveals carbon sources that control its expression of hemolytic toxins

Yujiro Hirose,[1,2] Saugat Poudel,[3] Anand V. Sastry,[3] Kevin Rychel,[3] Cameron R. Lamoureux,[3] Richard Szubin,[3] Daniel C. Zielinski,[3] Hyun Gyu Lim,[3,4] Nitasha D. Menon,[2,5] Helena Bergsten,[2] Satoshi Uchiyama,[2] Tomoki Hanada,[1] Shigetada Kawabata,[1,6] Bernhard O. Palsson,[3] Victor Nizet[2,7]

AUTHOR AFFILIATIONS See affiliation list on p. 18.

ABSTRACT *Streptococcus pyogenes* can cause a wide variety of acute infections throughout the body of its human host. An underlying transcriptional regulatory network (TRN) is responsible for altering the physiological state of the bacterium to adapt to each unique host environment. Consequently, an in-depth understanding of the comprehensive dynamics of the *S. pyogenes* TRN could inform new therapeutic strategies. Here, we compiled 116 existing high-quality RNA sequencing data sets of invasive *S. pyogenes* serotype M1 and estimated the TRN structure in a top-down fashion by performing independent component analysis (ICA). The algorithm computed 42 independently modulated sets of genes (iModulons). Four iModulons contained the *nga-ifs-slo* virulence-related operon, which allowed us to identify carbon sources that control its expression. In particular, dextrin utilization upregulated the *nga-ifs-slo* operon by activation of two-component regulatory system CovRS-related iModulons, altering bacterial hemolytic activity compared to glucose or maltose utilization. Finally, we show that the iModulon-based TRN structure can be used to simplify the interpretation of noisy bacterial transcriptome data at the infection site.

IMPORTANCE *S. pyogenes* is a pre-eminent human bacterial pathogen that causes a wide variety of acute infections throughout the body of its host. Understanding the comprehensive dynamics of its TRN could inform new therapeutic strategies. Since at least 43 *S. pyogenes* transcriptional regulators are known, it is often difficult to interpret transcriptomic data from regulon annotations. This study shows the novel ICA-based framework to elucidate the underlying regulatory structure of *S. pyogenes* allows us to interpret the transcriptome profile using data-driven regulons (iModulons). Additionally, the observations of the iModulon architecture lead us to identify the multiple regulatory inputs governing the expression of a virulence-related operon. The iModulons identified in this study serve as a powerful guidepost to further our understanding of *S. pyogenes* TRN structure and dynamics.

KEYWORDS *Streptococcus pyogenes*, independent component analysis, modulon, transcription, regulatory network, maltose, dextrin

Streptococcus pyogenes is a significant human pathogen responsible for over 700 million infections and at least 517,000 deaths annually worldwide (1). This organism causes diverse diseases, ranging from superficial pharyngitis and impetigo to life-threatening invasive diseases, such as necrotizing fasciitis (NF) and streptococcal toxic shock syndrome (2). As an adaptation to many different host environments is reflected in its versatile pathogenicity, a comprehensive understanding of the dynamics of *S. pyogenes* gene regulation may be useful in guiding new therapeutic strategies.

Editor Xiao-Hua Zhang, Ocean University of China, Qingdao, China

Address correspondence to Yujiro Hirose, hirose.yujiro.dent@osaka-u.ac.jp.

The authors declare no conflict of interest.

See the funding table on p. 18.

Received 18 March 2023
Accepted 2 April 2023
Published 6 June 2023

An underlying transcriptional regulatory network (TRN) alters the physiological state of a bacterium to adapt to unique challenges presented by each host environment (3–7). *S. pyogenes* regulons have been determined based on direct molecular methods, including transcriptomics performed on single-gene knockout mutants (3, 5, 6, 8–21) (PubMed unique identifiers, PMIDs, are listed in Supplementary Data 3) and chromatin immunoprecipitation sequencing (22, 23). At least 13 *S. pyogenes* two-component regulatory systems and 30 transcriptional regulators are known (24). Despite these extensive and valuable efforts, it is often difficult to predict transcriptomic data from regulon annotations (25, 26). With many more potential interactions among each regulatory factor, comprehensive interpretation of the *S. pyogenes* global TRN has proven difficult.

We previously reported an independent component analysis (ICA)–based framework that decomposes a compendium of RNA-sequencing (RNA-seq) expression profiles to determine the underlying regulatory structure of a bacterial transcriptome (27–29). ICA computes independently modulated sets of genes (termed iModulons) and their corresponding activity levels under each experimental condition. The iModulons can be interpreted as data-driven regulons, since they rely on observed expression changes instead of predicted transcription factor (TF)–binding sites. Therefore, it is possible that some iModulons reveal sets of genes never known to move coordinately. Moreover, since the number of iModulons is substantially fewer than the number of genes, they provide a helpful framework to more easily interpret the bacterial transcriptome. Here, we apply ICA for the first time to the major human pathogen *S. pyogenes* to elucidate its iModulon architecture.

Among over 200 serotypes of *S. pyogenes*, serotype M1 is the most frequently identified from streptococcal pharyngitis (30) and invasive diseases worldwide (31). Todd-Hewitt medium supplemented with yeast extract (THY) is the most commonly used growth medium in which transcriptome profiling of *S. pyogenes* has been performed. To elucidate the TRN features of *S. pyogenes* serotype M1, we compiled 116 high-quality RNA-seq data sets of *S. pyogenes* serotype M1 cultured in THY and conducted ICA-based decomposition. ICA computed 42 iModulons that we characterized and analyzed their activities to formulate hypotheses. Users can search and browse all iModulons from this data set on iModulonDB.org and view or interrogate them using interactive dashboards (32).

Drilling down, we focused in this study on four iModulons that each includes the *nga-ifs-slo* operon encoding an NAD glycohydrolase (NADase), immunity factor, and the pore-forming cytolytic toxin streptolysin O that contribute to enhanced virulence of *S. pyogenes (*33, 34). Comparing iModulon activities across all 116 samples allowed us to formulate hypotheses and identify carbon sources that control *nga-ifs-slo* expression. Although maltose and dextrin both consist solely of glucose molecules, dextrin utilization upregulated the *nga-ifs-slo* operon and changed bacterial hemolytic activity in contrast to glucose or maltose utilization. Furthermore, dextrin utilization induced the activation of iModulons related to the two-component regulatory system CovRS, a global transcriptional control system of *S. pyogenes* virulence phenotypes (2). In the concluding part of this study, we show that the computed TRN structure from iModulon analyses can simplify the interpretation of noisy bacterial transcriptomes at the infection site using matrix projection. Our composite results suggest that *S. pyogenes*, in the inflammatory environment of the NF, senses and responds to both carbohydrate depletion and stresses affecting the CovRS system.

MATERIALS AND METHODS

Data acquisition and preprocessing for independent component analysis (ICA)

We downloaded RNA-seq data of *S. pyogenes* serotype M1 from National Center for Biotechnology Information Sequence Read Archive. To calculate transcripts per kilobase million (TPM) from RNA-seq data and conduct the subsequent quality

control (QC)/quality assurance (QA), the previously reported pipline was used with minor modifications (35). Briefly, the sequences were aligned to the *S. pyogenes* strain 5448 genome (Serotype M1, accession no. CP008776), using Bowtie2 (36). The aligned sequences were assigned to open reading frames using feature Counts (37). To reduce the effect of noise, genes with average counts per sample <10 were removed. The final count matrix with 1,723 genes was used to calculate TPM. To generate high-quality RNA-seq expression profiles, we removed RNA-seq samples that show low quality in FastQC (38) and low correlation between biological replicates (Pearson's R <0.92). Finally, QC-passed 116 RNA-seq data were used for ICA (8, 33, 39–42) (Supplementary Data 1).

Independent component analysis

ICA decomposes a transcriptomic matrix (X) into independent components (M) and their condition-specific activities (A) (Supplementary Data 2). The procedure for computing robust components with ICA has been described in detail previously (27). Log2(TPM + 1) values were centered on strain-specific reference conditions and used as ICA decomposition. Next, Scikit-learn (v0.20.3) implementation of the FastICA algorithm was used to calculate independent components with 100 iterations, convergence tolerance of 10^{-7}, log(cosh(x)) as contrast function and parallel search algorithm (27, 43). To determine the ideal number of components, we used our previously developed OptICA method (44). The resulting M matrix containing source components from the 100 iterations was clustered with Scikit-learn implementation of the DBSCAN algorithm with ε of 0.1 and a minimum cluster seed size of 50 samples (50% of the number of random restarts). If necessary, the component in each cluster was inverted such that the gene with the maximum absolute weighting of the component was positive. Centroids for each cluster were used to define the final weightings for M and the corresponding A matrix. The whole process was repeated 100 times to ensure that the final calculated components were robust. Finally, components with activity levels that deviated more than five times between samples in the same conditions were also filtered out.

Determining independently modulated sets of genes (iModulons)

ICA enriches components that maximize the non-Gaussianity of the data distribution. While most of the genes have weightings near 0 and fall under Gaussian distribution in each component, there exists a set of genes whose weightings in that component deviate from this significantly. To enrich these genes, we used Scikit-learn's implementation of the D'Agostino K2 test, which measures the skew and kurtosis of the sample distribution (D'Agostino, 1990). We first sorted the genes by the absolute value of their weightings and performed the K2 test after removing the gene with the highest weighting. This was done iteratively, removing one gene at a time, until the K2 statistic falls below a cutoff. We calculated this cutoff based on sensitivity analysis on the agreement between enriched iModulon genes and known regulons. Regulons predicted by RegPrecise (45) (*S. pyogenes* M1 GAS), shown in previous reports (PMIDs are listed in Supplementary Data 3), and prophage genes searched by PHASTER (46) are used for the information of known regulons (Supplementary Data 3). For a range of cutoffs, we ran the iterative D'Agostino K2 test on all components and checked for statistically significant overlap of iModulons with the known regulons using Fisher's exact test (Supplementary Data 4). For iModulons with significant overlap, we also calculated precision and recall. The cutoff of 150, which led to the highest harmonic average between precision and recall (F1 score), was chosen as the final cutoff. We also manually identified iModulons that consisted of genes with shared functions (e.g., cytolysins, transporter) or those that corresponded to other genomic features (Supplementary Information 1, Supplementary Data 4).

Bacterial strains and culture conditions

S. pyogenes M1T1 strain 5448 (accession no. CP008776) was isolated from a patient with toxic shock syndrome and NF and considered to be a genetically representative globally disseminated clone associated with the invasive infections (47). *S. pyogenes* were grown at 37°C in a screw-cap centrifuge tube (BD Biosciences, San Jose, CA, USA) filled with Todd-Hewitt broth supplemented with 0.2% yeast extract (THY) (Hardy Diagnostics, Santa Maria, CA, USA) in an ambient atmosphere and standing cultures. To obtain cultures for experiments, overnight cultures of *S. pyogenes* were back diluted 1:50 into fresh THY broth and grown at 37°C, with growth monitored by measuring optical density at 600 nm (OD_{600} = 0.35 for mid-exponential samples or OD_{600} = 0.75 for early-stationary samples). Colony-forming units (CFUs) were determined by plating diluted samples on THY agar.

Escherichia coli MC1061 strain was used as a host for derivatives of plasmids pHY304 (48). *E. coli* strains were cultured in Luria–Bertani medium (Hardy Diagnostics) at 37°C with agitation. For the selection and maintenance of strains, antibiotics were added to the medium at the following concentrations: erythromycin, 500 µg/mL for *E. coli* and 2 µg/mL for *S. pyogenes*; chloramphenicol, 2 µg/mL for *S. pyogenes*.

Construction of *malR2* mutant strain

An in-frame *malR2* TF deletion mutant (Δ*malR2*) with a background of strain 5448 (wild type [WT]) was constructed using the pHY304 temperature-sensitive shuttle vector, as previously reported (49). Briefly, a pHY304-*malR2*KO plasmid harboring the DNA fragment, in which upstream of the *malR2* gene, a chloramphenicol acetyltransferase gene (*cat*), and downstream regions of the *malR2* gene were linked by overlapping PCR, was electroporated into WT strain and grown in the presence of erythromycin. The plasmid was then integrated into the chromosome via the first allelic replacement at 37°C, after which it was cultured at 30°C without antibiotics to induce the second allelic replacement. The deletion of *malR2* was confirmed by site-specific PCR using purified genomic DNA. Primers are listed in Table 1.

Red blood cell hemolysis assay

Human blood was collected by using BD Vacutainer EDTA (Cat. 366643m; Becton, Dickinson, Mountain View, CA, USA) tubes. Red blood cells (RBCs) were isolated by centrifuging and washing with phosphate-buffered saline (PBS). Bacterial cultures in THY broth at OD_{600} = 0.3–0.4 were centrifuged by using Eppen tubes and resuspended in an equal amount of chemically defined medium (CDM) (50). CDM was supplemented with 4.5 g/L of D-glucose (Cat. G7021; Sigma-Aldrich, St. Louis, MO, USA), D-maltose (Cat. M5885; Sigma-Aldrich), or dextrin (Cat. 31410; Sigma-Aldrich). At 2-hour postincubation at 37°C, 150 µL of whole culture or supernatants were collected for serial dilutions in PBS. All wells were added 50 µL of 2% RBC in PBS and incubated for 4 hours. Supernatants were collected from assay wells after centrifugation at 3,000× *g* for 15 minutes, and hemolysis was determined by absorbance with a SpectraMax M3 plate reader at 541 nm using SoftMax Pro software. Each titer was recorded as the point that the hemolysis

TABLE 1 Primers used in this study

Primer	Sequence (5′–3′)	Purpose
MalR2KOupF	CCCAAGCTTTCCTTTTGGAAAATGGTTTATGA	Construction of pHY304-*malR2*KO
MalR2KOupR	CCAGTGATTTTTTTCTCCATCATTTTGTTATATCCGAATTC	Construction of pHY304-*malR2*KO
MalR2KOdownF	GAGTGGCAGGGCGGGGCGTAACGCCTGAGAATTAACCTATC	Construction of pHY304-*malR2*KO
MalR2KOdownR	ATAGTTTAGCGGCCGCTCCCTTGGAACGGATTGTAA	Construction of pHY304-*malR2*KO
MalR2KOcatF	ATGGAGAAAAAAATCACTGGATATACC	Construction of pHY304-*malR2*KO
MalR2KOcatR	TTACGCCCCGCCCTGCCACTCATCGCA	Construction of pHY304-*malR2*KO
MalR2KOconfirmF	GCGAAGGAACTGCTTACGTC	Confirmation of *malR2* deletion
MalR2KOconfirmR	GCCGAGCTCAGTTTACTTGG	Confirmation of *malR2* deletion

reached half of the 100% RBC lysis (H_2O) control. No bacteria controls were used as the 0% RBC lysis.

RNA extraction and library preparation

The details of the new 16 RNA samples analyzed in this study are shown in Supplementary Data 5. All samples were collected in biological duplicates. Overnight cultures originating from different colonies were back diluted 1:50 into fresh THY broth and grown at 37°C, with growth monitored by measuring OD at 600 nm. In regards to RNA samples from *S. pyogenes* cultured in CDM, bacterial cultures in THY broth at the mid-exponential phase were centrifuged and resuspended in an equal amount of CDM for 2 hours. Three milliliter of samples were added to tubes containing 6-mL RNAprotect Bacteria Reagent (Qiagen, Hilden, Germany) and vortexed. After 5 minutes of incubation at room temperature, they were centrifuged to remove the supernatant. RNA was extracted from the pelleted cells using a Quick RNA Fungal/Bacterial Microprep kit (Zymo Research, Irvine, CA, USA). The cells were mechanically lysed with Mini-Beadbeater-96 (Biospec Products, Bartlesville, OK, USA) for 2 minutes, and DNA was removed with DNase I (New England Biolabs, Beverly, MA, USA) during the RNA purification. RNA quality was checked with an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) instrument. rRNA was removed using RiboRid (51) with designed probes for *S. pyogenes* M1T1 strain 5448 (Supplementary Data 6). The library was prepared with KAPA Hyper Prep kit (KAPA Biosystems, Wilmington, MA, USA) following the manufacturer's protocol.

RNA-seq and data analysis

Libraries were sequenced using Illumina NextSeq or NovaSeq systems, with 40 bp (sequenced at the University of California, Davis DNA sequencing facility) or 150 bp (sequenced at the University of California San Diego IGM Genomics Center) paired-end reads. Raw reads were trimmed using the fastp (52), mapped to the *S. pyogenes* strain 5448 genome, and used to calculate the TPM with the commercially available Geneious Prime 2019.2 software package (Biomatters, Auckland, New Zealand). Differential expression and global analyses of RNA-seq expression data were performed using iDEP (ge-lab.org). EdgeR log transformation was used for clustering and principal component analysis (iDEP). Hierarchical clustering was visualized using the average linkage method with correlation distance (iDEP). Functional annotations of *S. pyogenes* strain 5448 genome are obtained by using EggNOG (eggnog5.embl.de) or PATRIC (www.patricbrc.org).

RESULTS

42 biologically meaningful sets of genes are revealed from 116 high-quality gene expression profiles by ICA

To extract regulatory signals from transcriptomic data, we used our established protocol (53) and downloaded RNA-seq data from the public Sequence Read Archive (54), where 229 RNA-seq samples were available for the M1 serotype of *S. pyogenes*. After QC, which checks for read quality, alignment, and high replicate correlations, we selected 116 RNA-seq samples for ICA (Supplementary Data 1). These samples derived from *S. pyogenes* propagated in THY under various culture conditions and exhibiting diverse expression states (Fig. S1). Applying an extended ICA algorithm (27), we decomposed the expression compendium into 42 iModulons and their activities (Fig. 1A) (Supplementary Data 2).

Regulons are sets of co-regulated genes classified informally through scientific consensus on a variety of experimental results in the literature, while iModulons are derived solely from the measured transcriptome through an unbiased method (27). Nevertheless, the known regulon structure of the *S. pyogenes* TRN described in RegPrecise (45), PHASTER (46), or previous reports (see Supplementary Data 3 or Materials and

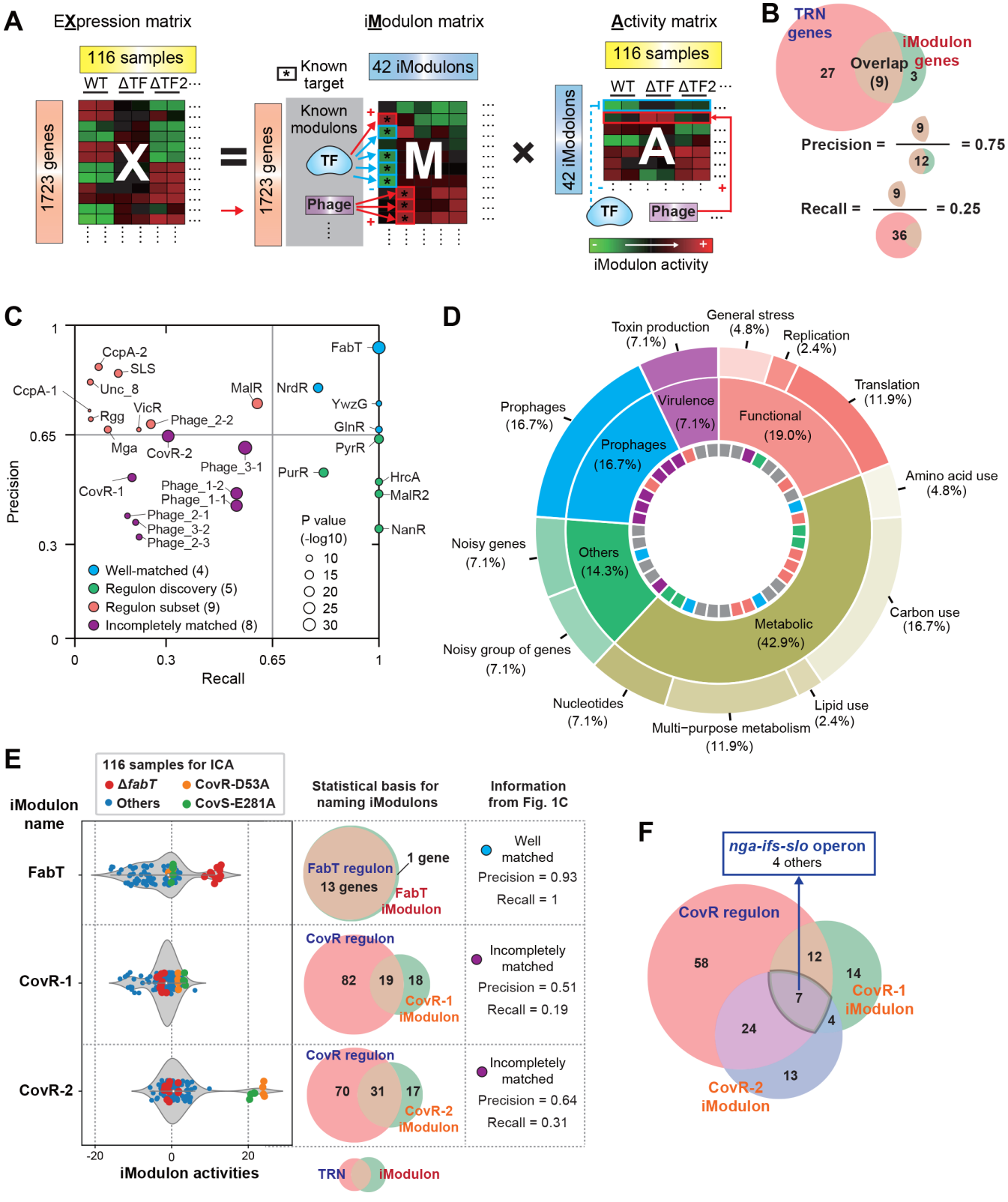

FIG 1 Independent component analysis (ICA) decomposes the compendium of transcriptomic data of *Streptococcus pyogenes* serotype M1 to biologically meaningful signals. (A) Schematic illustration of ICA applied to a gene expression compendium. ICA decomposes a transcriptomic matrix (X) into independent components (M) and their condition-specific activities (A). F: transcription factor. (B) The definitions of precision and recall with example numbers. (C) Scatter plot of precision and recall of the enrichments for the 26 (out of 42) iModulons that were matched to a regulon. The size of the circle indicates the negative log of

(Continued on next page)

FIG 1 (Continued)

P-values to the base 10. (D) Donut chart of iModulon functions. The outermost ring lists specific functions and the center ring lists broad functions. The coloring of innermost ring has reflected the colors of 26 regulatory iModulons as indicated in Fig. 1C, and gray indicates iModulons which do not match a regulon. (E) Examples to demonstrate the reliability of the annotated Modulons. Violin plot shows that FabT iModulon and CovRS-related iModulon activities (CovR-1 and CovR-2 iModulons) of all 116 samples. Transcriptomes from *fabT* deletion mutant strains (red) are for the positive control of FabT iModulon. Transcriptomes from strains with CovR-D53A (orange) and CovS-E281A (green) are for the positive control of CovR-1 and CovR-2 iModulons. The overlap of iModulons with the known regulons is shown as Venn diagram. iModulons are detailed in Supplementary Information 1. The list of the TRN is shown in Supplementary Data 3. The list of genes in each iModulons is shown in Supplementary Data 4. The imodulondb.org, where users can search and browse all iModulons from this data set and view them with interactive dashboards. (F) Overlapping genes among CovR regulon, CovR-1 iModulon, and CovR-2 iModulon.

Methods section for more details) is largely captured by the iModulons. Overall, 26 of the 42 iModulons were successfully mapped to a known regulator (Supplementary Data 4), and we named iModulons based on known regulons that share the highest overlap with the given iModulon or by shared functionality of genes (e.g., cytolysins, transporter). The details for 42 iModulons are shown in Supplementary Information 1 and Supplementary Data 4. They are also freely available to browse, search, or download on iModulonDB.org (32).

Precision and recall between iModulons and regulons (Fig. 1B) were used to classify iModulons into four groups, namely well-matched, regulon subset, regulon discovery, and incompletely matched groups (Fig. 1C). (i) The well-matched group (n = 4) has precision and recall greater than 0.65. (ii) The regulon subset (n = 5) exhibits high precision and low recall. They contain only part of their enriched regulon, perhaps because the regulon is very large and only the genes with the most pronounced transcriptional changes are captured. (iii) A third group, regulon discovery (n = 5), has low precision but high recall. These iModulons contain some genes that are known to be co-regulated, along with other genes that are co-stimulated by the conditions in the data set. (iv) The remaining enriched iModulons are termed incompletely matched (n = 8) because neither their precision nor their recall met the cutoff; however, this grouping still had statistically significant enrichment levels and appropriate activity profiles. The difference in gene membership between these iModulons and their regulons provides excellent targets for discovery.

iModulons identified with no enrichments suggest that their composite gene sets may be regulated by unexplored transcriptional mechanisms or that there remains some noise due to large variance within the data. Functional categorization of iModulons provides a systems-level perspective on the transcriptome (Fig. 1D). Metabolism-related iModulons account for over 40% of the iModulons, while comparatively fewer iModulons deal with toxin production, stress response, replication, translation, and mobile genetic elements like prophages.

iModulon activities of mutant strains demonstrate the reliability of the annotation of iModulons

All iModulons have a computed activity in every sample (Supplementary Data 2), allowing for easy comparisons of iModulon activities across samples (Supplementary Information 1). To benchmark the reliability of iModulon annotation, we show the FabT and CovRS-related iModulon activities (CovR-1 and CovR-2 iModulons) of 116 all samples (Fig. 1E).

The FabT iModulon overlaps nearly perfectly with the FabT regulon. This iModulon contains all genes known to be regulated by FabT, plus a hypothetical protein (XK27_06930) that may be co-regulated with the known regulon. Additionally, transcriptomes from *fabT* deletion mutant strains (Fig. 1E, Red) (39) show high FabT iModulon activities, consistent with the role of FabT as a repressor. Thus, this iModulon captures a known regulon and its expected behavior using transcriptome data alone. Several other iModulons are similarly easily explainable, particularly those in the well-matched group.

The activity of the CovRS TF can be altered with exact point mutations. CovR-D53A or CovS-E281A point mutations induce considerable differential expression of genes

(DEGs) in *S. pyogenes* transcriptome (171 or 139, respectively) (40). Although the CovR-1 and CovR-2 iModulons incompletely overlap with the CovR regulon (6) (Supplementary Information 1), strains with CovR-D53A (Fig. 1E, orange) and CovS-E281A (Fig. 1E, green) show higher CovR-1 and CovR-2 iModulon activities compared to other samples. These correlations suggest that ICA results and iModulon annotations reflect previous findings properly and are useful for interpreting RNA-seq results.

Interestingly, the CovR regulon, CovR-1 iModulon, and CovR-2 iModulon overlap in seven genes that contain *nga-ifs-slo* operon (Fig. 1F). Increased expression of the *nga-ifs-slo* operon is linked to enhanced virulence of *S. pyogenes* (33, 34).

nga-ifs-slo operon is in SLO iModulon, MalR2 iModulon, CovR-1 iModulon, and CovR-2 iModulon

To decipher *S. pyogenes* pathogenicity, it is important to resolve the regulatory dynamics underlying two important virulence operons encoding potent cytolytic toxins: the abovementioned *nga-ifs-slo*, which encodes the gene for streptolysin O (SLO), and *sagA-I*, which encodes the biosynthetic machinery to produce streptolysin S (SLS), the toxin responsible for the hallmark β-hemolytic phenotype of the bacterium grown on blood agar. Both SLS and SLO are important for the establishment of *S. pyogenes* skin infection and the formation of necrotizing skin lesions in invasive disease (55–57).

Because iModulons are based on matrix decomposition, some genes may be present in multiple iModulons. The effects of all iModulons add together to reconstruct individual expression levels, which may be due to multiple TFs influencing the target genes. This is the case for the *nga-ifs-slo* operon, which is part of the designated SLO iModulon (Fig. 2A), MalR2 (a member of the LacI/GalR family of repressors) iModulon (Fig. 2B), CovR-1 iModulon, and CovR-2 iModulon (Fig. 1F). Indeed, the *nga-ifs-slo* operon are the only genes overlapping among these four iModulons (Fig. 2C). These findings indicate there are the multiple regulatory inputs governing the expression of the *nga-ifs-slo* operon. In the subsequent sections, we speculate and validate environment cues under which the *nga-ifs-slo operon* is regulated based on iModulon information and corresponding activities.

SLS and SLO iModulons suggests the antagonistic regulation between *sag A-I* and *nga-ifs-slo* operons in the stationary phase

The SLS iModulon contains the *sag A-I* operon encoding SLS (Fig. 2D). Among 42 iModulons, the SLS iModulon is the one containing all the *sagA-I* genes. Interestingly, there is a transcriptional tradeoff between the two major toxin-encoding virulence systems: the SLS iModulon activities are negatively correlated with SLO iModulon activities (Fig. 2E). SLS and SLO iModulon activities tend to be antagonized in samples obtained from stationary phase growth and a CcpA deletion mutant at mid-exponential phase (Fig. 2E). *S. pyogenes* stationary phase cultures in THY show evidence of glucose depletion (58). Carbon catabolite repressor CcpA is the major transcriptional regulator in carbon catabolite repression (CCR) and senses carbohydrate availability (11, 59, 60). These observations lead us to hypothesize that the antagonistic regulation between *sag A-I* and *nga-ifs-slo* operons depends on a change in carbon source or TFs sensing the availability of carbon sources (Fig. S2). SLO iModulon activities are also strongly deactivated in CcpA deletion mutants (Fig. 2E), suggesting CcpA deletion downregulates the *nga-ifs-slo* operon. However, other studies have indicated that the *ccpA* deletion in *S. pyogenes* serotype M1 induces the upregulation of the *slo* gene (6) or does not change the expression level of the *slo* gene (60). Therefore, the SLO iModulon alone is not enough to explain the regulation of the *nga-ifs-slo* operon.

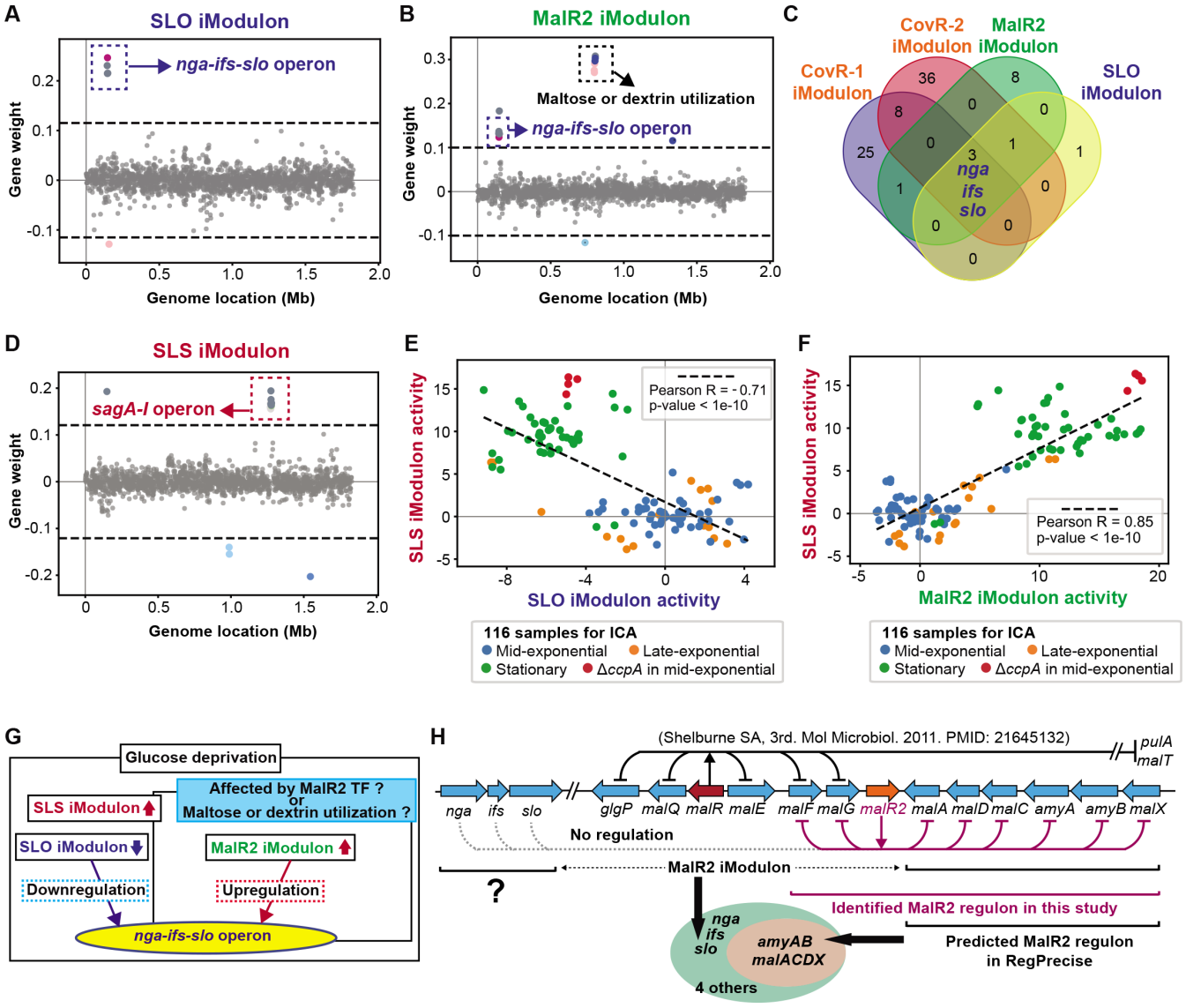

FIG 2 The *nga-ifs-slo operon* is in four iModulons, suggesting this operon is under multiple regulations. (A) Gene weights in SLO iModulon. Colored circles indicate clusters of orthologous groups of proteins (COG) categories, which are detailed in Supplementary Information 1. The list of genes in each iModulon is shown in Supplementary Data 4. (B) Gene weights in MalR2 iModulon. (C) Overlapping genes among CovR-1 iModulon, CovR-2 iModulon, MalR2 iModulon, and SLO iModulon. (D) Gene weights in SLS iModulon. (E) Comparison between SLS and SLO iModulon activities across 116 samples. (F) Comparison between SLS and MalR2 iModulon activities across 116 samples. (G) Schematic diagram of the hypothesis that *nga-ifs-slo* operon is antagonistically regulated in a specific condition. (H) Genomic organization of genes involved in maltose/dextrin transport, metabolism, and metabolic regulation. To identify the actual MalR2 regulon, transcriptomes of wild-type and *malR2* deletion mutant strains were compared at both the mid-exponential and stationary growth phases in THY broth (Fig. S3 and Supplementary Data 5).

MalR2 iModulon harboring *nga-ifs-slo* operon is activated in the stationary phase, contrary to SLO iModulon

The MalR2 iModulon contains not only the *nga-ifs-slo* operon but also an operon for maltose or dextrin utilization (*malACDX*, *amyAB*) (Fig. 2B). Maltose is a disaccharide of D-glucose, whereas dextrin is a polysaccharide of D-glucose. Since maltose and dextrin are the products of the action of salivary amylases on dietary starch, it is possible *S. pyogenes* utilizes these carbohydrates during throat colonization or the development of pharyngitis.

Contrary to the SLO iModulon, MalR2 iModulon activities are positively correlated with SLS iModulon activities and are very high in samples from the stationary phase or CcpA deletion mutants at the mid-exponential phase (Fig. 2F). These facts suggest that the transcription of the *nga-ifs-slo* operon is both downregulated (suggested by the SLO iModulon) and upregulated (suggested by the MalR2 iModulon) in stationary phase or under glucose-depleted conditions. These effects may counteract with one another, resulting in the upregulation of the *slo* gene in the MalR2 iModulon–activated state even under conditions that deactivate the SLO iModulon (Fig. 2G).

MalR2 TF does not contribute to the regulation of the *nga-ifs-slo* operon

The MalR2 iModulon is named due to a statistically significant overlap with the MalR2 regulon predicted by RegPrecise (45) (Fig. 2H). However, the predicted MalR2 regulon in RegPrecise does not contain the *nga-ifs-slo* operon. In the genome of serotype M1 *S. pyogenes*, two genes predicted to encode LacI family TFs, *malR* and *malR2*, are situated near the *malACDX* and *amyAB* operons (61, 62). The MalR regulon was experimentally defined using MalR-deletion mutant strains (5), but no such data have been generated for MalR2.

To investigate whether the *nga-ifs-slo* operon is directly regulated by MalR2 TF, we compared the transcriptomes of WT and *malR2* deletion mutant (Δ*malR2*) strains at both mid-exponential and stationary growth phases in THY broth. Deletion of *malR2* induced the upregulation of only the *malACD* operon in the mid-exponential phase (Fig. S3A, Supplementary Data 5), but in the stationary phase, it upregulated both the *malACDFGX* and *amyAB* operons (Fig. S3B, Supplementary Data 5). The *nga-ifs-slo* operon was not affected, which indicates that its inclusion with the MalR2 iModulon was due to co-stimulation, but not direct regulation. Therefore, we decided to explore the carbon sources that may stimulate the operon.

Maltose or dextrin utilization changes bacterial hemolytic activity as compared to the glucose utilization

Despite the lack of direct MalR2 regulation, the *nga-ifs-slo* operon may still have been included in the MalR2 iModulon due to co-stimulation by inducers such as maltose or dextrin. To investigate whether maltose or dextrin utilization changes bacterial hemolytic activity, we used a red blood cell hemolysis assay (Fig. 3A). We adjusted the CDM for *S. pyogenes* by following a previous report (50). CDM without carbohydrate sources could not support the growth of serotype M1 *S. pyogenes* strain 5448, but its viability (bacterial CFUs) was maintained. Supplementation of the CDM with glucose, maltose, or dextrin allowed *S. pyogenes* 5448 strain growth (Fig. 3B). In a hemolysis assay using WT *S. pyogenes* supernatant, the glucose (+) condition produced a high hemolytic titer, while the maltose (+) or dextrin (+) conditions yielded a low titer (Fig. 3C). To clarify whether this hemolysis was caused by SLS or SLO, we used SLS deletion mutant (Δ*sagA*) and SLO deletion mutant (Δ*slo*) strains (Fig. 3D). In the glucose-supplemented condition, SLO is completely responsible for the hemolytic activity. In contrast, in the maltose- or dextrin-supplemented conditions, SLS is completely responsible for the hemolytic activity. In a hemolysis assay by using a live culture of the WT *S. pyogenes* strain, the glucose and dextrin conditions produced high hemolytic titer, while the carbon (–) and maltose conditions resulted in low titer (Fig. 3E). Here again, we used Δ*sagA* and Δ*slo* strains (Fig. 3F). In the glucose-supplemented condition, SLO is principally responsible for the hemolytic activity, while SLS contributed only weakly. In other conditions, SLS is mainly responsible for the hemolytic activity. However, SLO-dependent hemolytic activity was confirmed in a dextrin-supplemented condition, while it was not present in the carbon (–) and maltose conditions. These results show how supplementation of CDM with different carbon sources changes the hemolytic activity and virulence factor expression in *S. pyogenes*. Therefore, we reasoned that the bacterial transcriptome obtained from

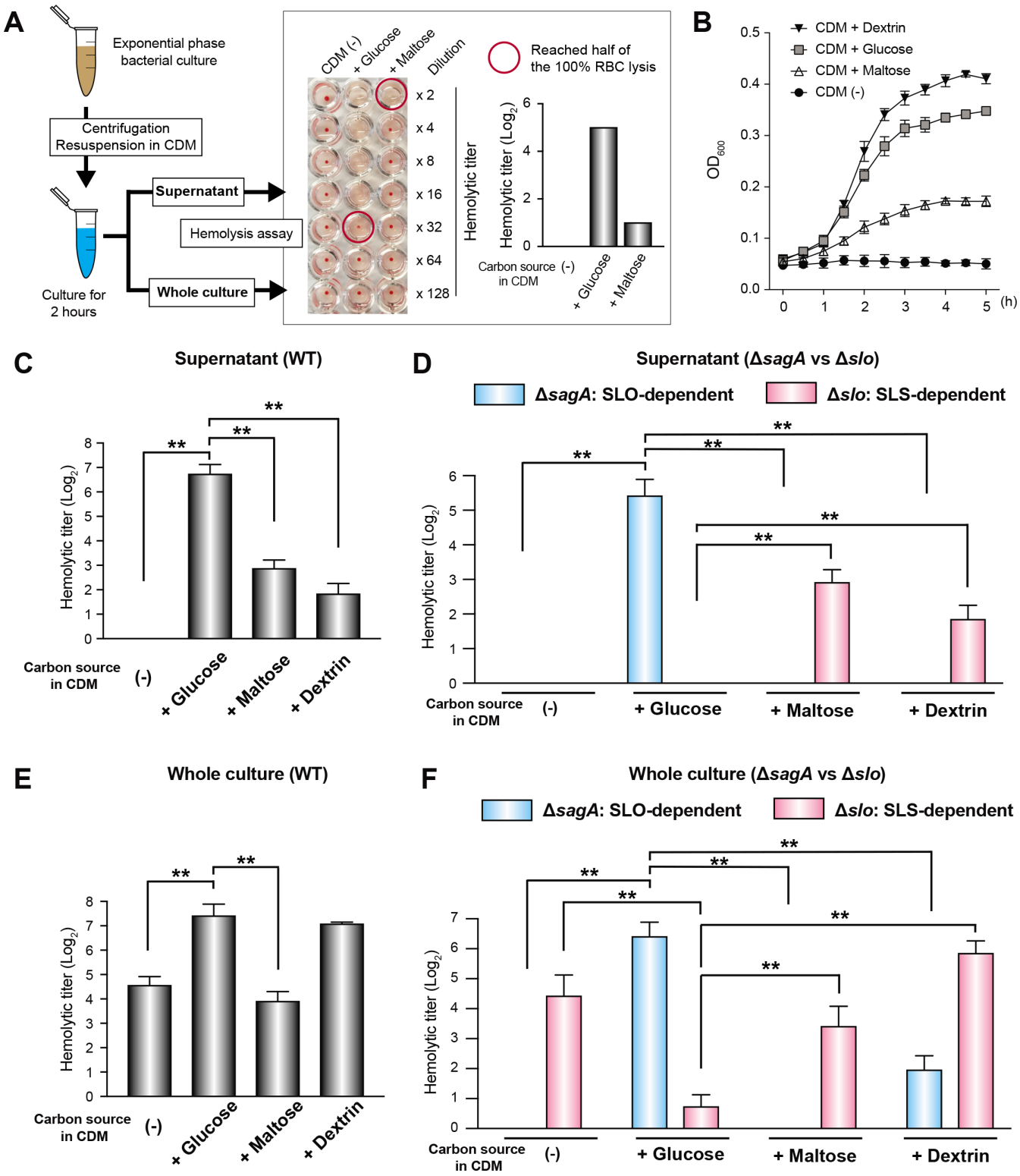

FIG 3 The switch between SLS and SLO activities is induced by the difference of supplemented carbon sources in CDM. (A) Experimental workflow for red blood cell hemolysis assay. Each titer was recorded as the point that the hemolysis reached half of the 100% RBC lysis (H_2O) control. No bacteria controls were used as the 0% RBC lysis. (B) Growth curves in CDM supplemented with indicated carbon sources. (C) Hemolytic activity of the supernatant from WT. Vertical lines represent the mean + SD. Values are presented as the mean of six wells from one of the three independent experiments. Differences between groups were analyzed using a Mann–Whitney *U* test. The significant differences as compared to CDM glucose (+) condition are only shown. Hemolytic titers were normalized

(Continued on next page)

FIG 3 (Continued)

using bacterial colony-forming units (CFUs. CFU from CDM glucose (+) condition was set to 1. $^{**}P < 0.01$. (D) Hemolytic activity of the supernatant from Δ*sagA* and Δ*slo* strains. (E) Hemolytic activity of the whole culture from WT. (F) Hemolytic activity of the whole culture from Δ*sagA* and Δ*slo* strains. Assay was conducted in biological duplicates and repeated six times. Values are presented as the mean of duplicates from a representative experiment.

similar experimental conditions might allow us to validate the hypothesis suggested in Fig. 2G.

Utilization of glucose, maltose, or dextrin significantly activates different iModulons, containing *nga-ifs-slo* operon

To examine the influence of different carbon sources on bacterial iModulon activities, we conducted RNA-seq analysis by using the RNA samples isolated from the *S. pyogenes* M1 strain after 2-hour incubation in CDM (detailed conditions and results are presented in Supplementary Data 5). Samples isolated from *S. pyogenes* cultured in CDM were evaluated together with those cultured in THY and shown earlier. In a hierarchical clustering (Fig. 4A), samples from CDM carbon (–) samples are located close to samples from stationary phase growth in THY. Therefore, the CDM carbon (–) condition seems to recapitulate the carbohydrate-depleted condition in THY. Although CDM glucose, maltose, and dextrin conditions supported *S. pyogenes* growth, samples from these conditions are located far from that mid-exponential phase in THY on the hierarchical clustering. Since THY broth is a nutrient-rich media containing many kinds of carbon sources, a CDM is better suited for assessing the influence of the supplementation of carbon sources.

In the CDM + maltose condition, a total of 648 S. *pyogenes* genes were altered in total compared to the CDM + glucose condition (Fig. 4B and Supplementary Data 5). Rather than analyzing so many DEGs individually, we propose that iModulon activities can be used to facilitate the interpretation of RNA-seq data. To that end, we calculated iModulon activities of samples in CDM + maltose and CDM + glucose conditions, which were centered on samples in CDM carbon (–) condition. After that, a Differential iModulon Activity plot (DiMA plot) allowed us to identify the significantly altered iModulons. In the CDM + maltose condition, antagonistic activation of the SLS and SLO iModulons was observed (Fig. 4B). The iModulons associated with the detection of carbohydrate depletion, including iModulons mapped to TFs of CcpA, MalR2, M protein transacting positive regulator Mga (9), and phase-dependent regulators Rgg (10), were likewise significantly activated in the CDM + maltose condition as compared to the CDM + glucose condition.

S. pyogenes grown in CDM + dextrin showed alterations in a total of 809 genes compared to the CDM + glucose condition (Fig. 4C and Supplementary Data 5). DiMA indicated that the significantly altered iModulons in the CDM + dextrin condition compared to CDM + glucose were similar to those significantly altered by maltose in Fig. 4B. However, SLO iModulon activity was not significantly altered. Notably, CovRS-related iModulons (CovR-1 and CovR-2) were significantly activated in the CDM + dextrin condition compared to CDM + glucose. Although the utilization of glucose, maltose, or dextrin each resulted in the upregulation of the *nga-ifs-slo* operon, the activated iModulons that contributed to operon upregulation differed by each carbon source (Fig. 4D).

The environment *S. pyogenes* were exposed in necrotizing fasciitis (NF) seemed both the carbohydrate depletion and stresses affecting the CovRS regulator

The iModulons presented in this study help us to interpret RNA-seq data from other studies. When we perform an analysis in this way, we can make certain observations about how data behave with respect to our TRN even if the data are of low quality. Here, we assess *S. pyogenes* transcriptome changes in the inflammatory environment

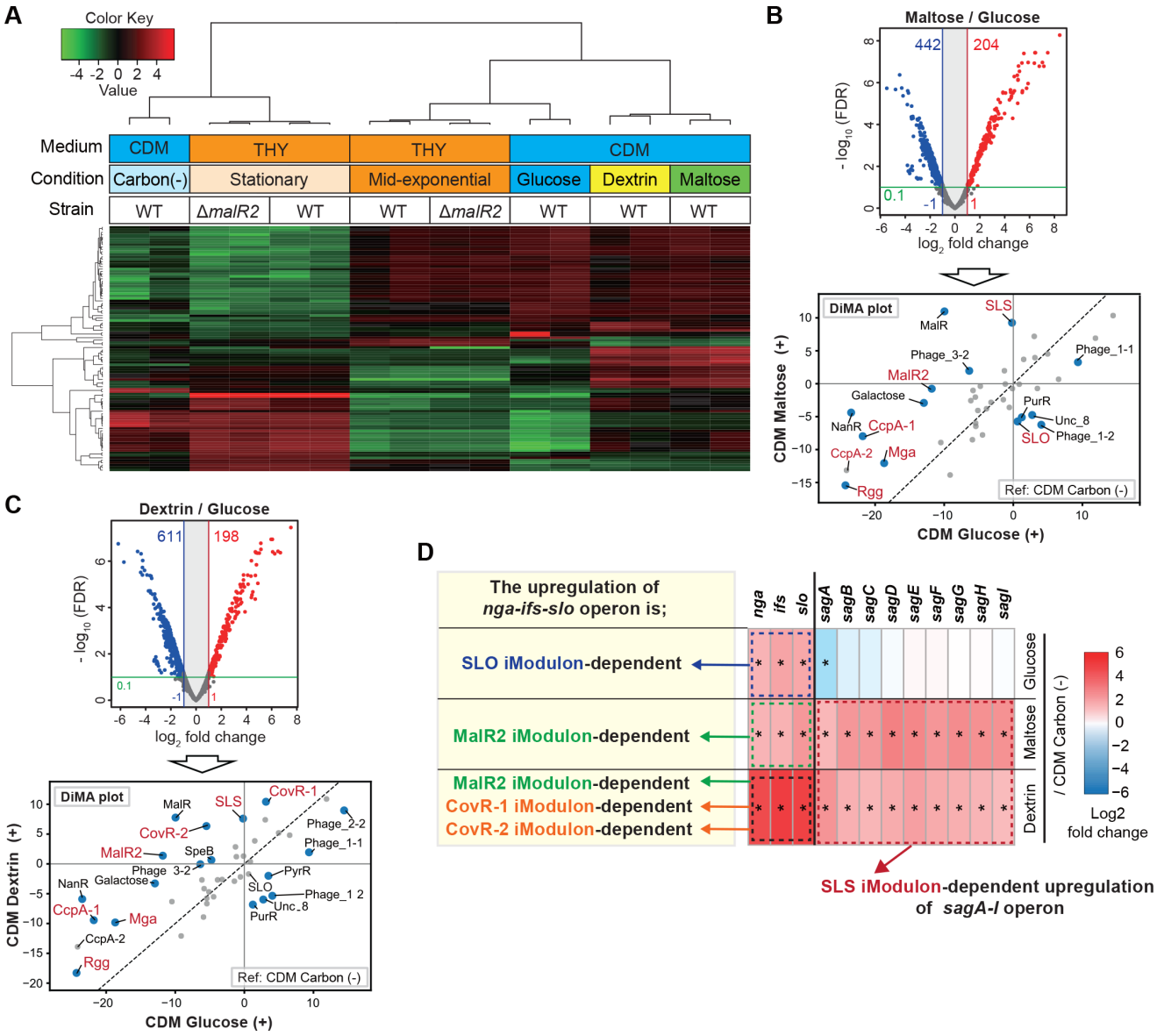

FIG 4 The activated iModulons that contribute to the upregulation of *nga-ifs-slo* operon differed by each carbon source. (A) Heat map of TPM data from 16 RNA-seq data set in this study. Heat map showing clustering of top 100 genes that contribute to the differences in each group. (B, C) Volcano plot and Differential iModulon Activity (DiMA) plot show transcriptome differences under the comparison conditions indicated in each figure. In volcano plots, colored circles indicate significantly upregulated (red) and downregulated (blue) genes (absolute $\log_2$ fold change >1; adjusted $P < 0.1$). In DiMA plots, blue circles indicate the significantly altered iModulons. (D) Expression levels of *nga-ifs-slo* and *sag A-I* operons in samples from CDM glucose (+), maltose (+), or dextrin (+) condition, as compared to those in samples from CDM carbon (–) condition. Based on the results of the DiMA plot, iModulons that contribute to the upregulation of *nga-ifs-slo* or *sag A-I* operons were determined.

of a murine model of NF, in which bacterial RNA samples were isolated from infected hind limbs obtained at 24-, 48-, and 96-hour postinfection (63). At first, we calculated iModulon activities of samples isolated from NF at 24, 48, and 96 hours, which were centered on samples isolated from THY culture at mid-exponential phase (THY-ME, control). After that, we identified the significantly altered iModulons at each time point. Then, a set of 18 were visualized (Fig. 5A). The iModulons associated with the detection of the carbohydrate depletion, such as CcpA-related, MalR, MalR2, Mga, and Rgg iModulons, were significantly activated in RNA samples isolated from NF *in vivo* compared with those obtained from THY-ME *in vitro*. In addition, CovRS-related

iModulons (CovR-1 and CovR-2) are significantly activated in RNA samples isolated from NF compared to THY-ME.

To compare data from NF with new data obtained in this study, we also re-calculated iModulon activities of samples in CDM carbon (–), glucose, maltose, and dextrin conditions, which were centered on samples from THY-ME of the WT strain. Of interest, the activities of CovRS-related iModulons in RNA samples isolated from NF are located very close to that from CDM dextrin condition (Fig. 5B). In addition, the activities of CcpA-related iModulons in the CDM carbon (–) condition are located relatively close to that those in NF (Fig. 5C). These results suggest *S. pyogenes* in NF sensed both carbohydrate depletion and stresses affecting the CovRS regulator. Taken with our other results about the regulation of the virulence factors, a coherent picture of carbon source depletion, CovRS, SLS, and SLO is presented. The most pronounced changes in the mouse model expression are in the same iModulons that we have been using to understand the regulation of virulence operons that contribute to NF pathology, which suggests this approach can be fruitful for studying *in vivo* disease models or clinical data.

To investigate whether or not we should take into account the global gene expression changes of bacterial growth, we computed the correlation of growth rate with the activities of iModulons on which we focused (Fig. 5D) (detailed conditions and results are presented in Supplementary Data 5). When comparing the iModulon activities between mid-exponential and stationary phases, no significant changes were seen in the activities of MalR2, SLO, FabT, and CovRS-related iModulons, while the activities of MalR2 and CovRS-related iModulons are consistently activated in RNA samples isolated from NF compared with THY-ME. These suggest that the significantly altered iModulon activities shown in Fig. 5A could capture a combination of both growth rate–independent and growth rate–dependent factors.

DISCUSSION

In this study, 116 existing, high-quality RNA-seq data sets of *S. pyogenes* serotype M1 were decomposed using ICA. This decomposition identified 42 iModulons and their iModulon activities; 26 of the iModulons correspond to specific biological functions or transcriptional regulators. Based on the gathered information on iModulons and their activities across samples, we could formulate hypotheses and identify carbon sources that modulate hemolysis activity in *S. pyogenes*. Calculation of iModulon activities in new RNA-seq data sets also allowed us to identify that *S. pyogenes* activated CovRS-related iModulons in a CDM + dextrin condition (Fig. 6). Furthermore, by computing iModulon activities of published bacterial transcriptomes, we estimate the stress to which *S. pyogenes* is exposed at the infection site of NF.

Sastry et al. (64) previously indicated that iModulons of *E. coli* directly capture groups of genes that vary relatedly, and independently, from other groups of genes, regardless of the magnitude of their variance. Lamoureux et al. (65) also demonstrated the overall similarity of the iModulon structure even after adding the 1,600 additional public samples from all sorts of different projects/laboratories/growth conditions. Therefore, an ICA-based statistical approach for the identification of iModulons can capture the general characterization of the gene regulatory network, although differences in data sets reflect some differences in gene sets of iModulons (64, 65). We cannot at present conduct the same analysis in *S. pyogenes* due to the small number of RNA-seq samples available in public databases. However, we could re-capture FabT and CovRS-related iModulons even if samples of *fabT* deletion mutants and CovRS mutants are erased from the RNA-seq data set (data not shown).

Some iModulons contained sets of genes that had never previously been reported as a regulon. This is likely because iModulons that are defined through an untargeted ICA-based statistical approach represent a set of genes that are co-expressed as a by-product of multiple regulatory influences. By focusing on *nga-ifs-slo* operon, we could identify here that dextrin utilization alters bacterial hemolytic activity compared to maltose or glucose utilization. The full data set is available for examination by researchers with

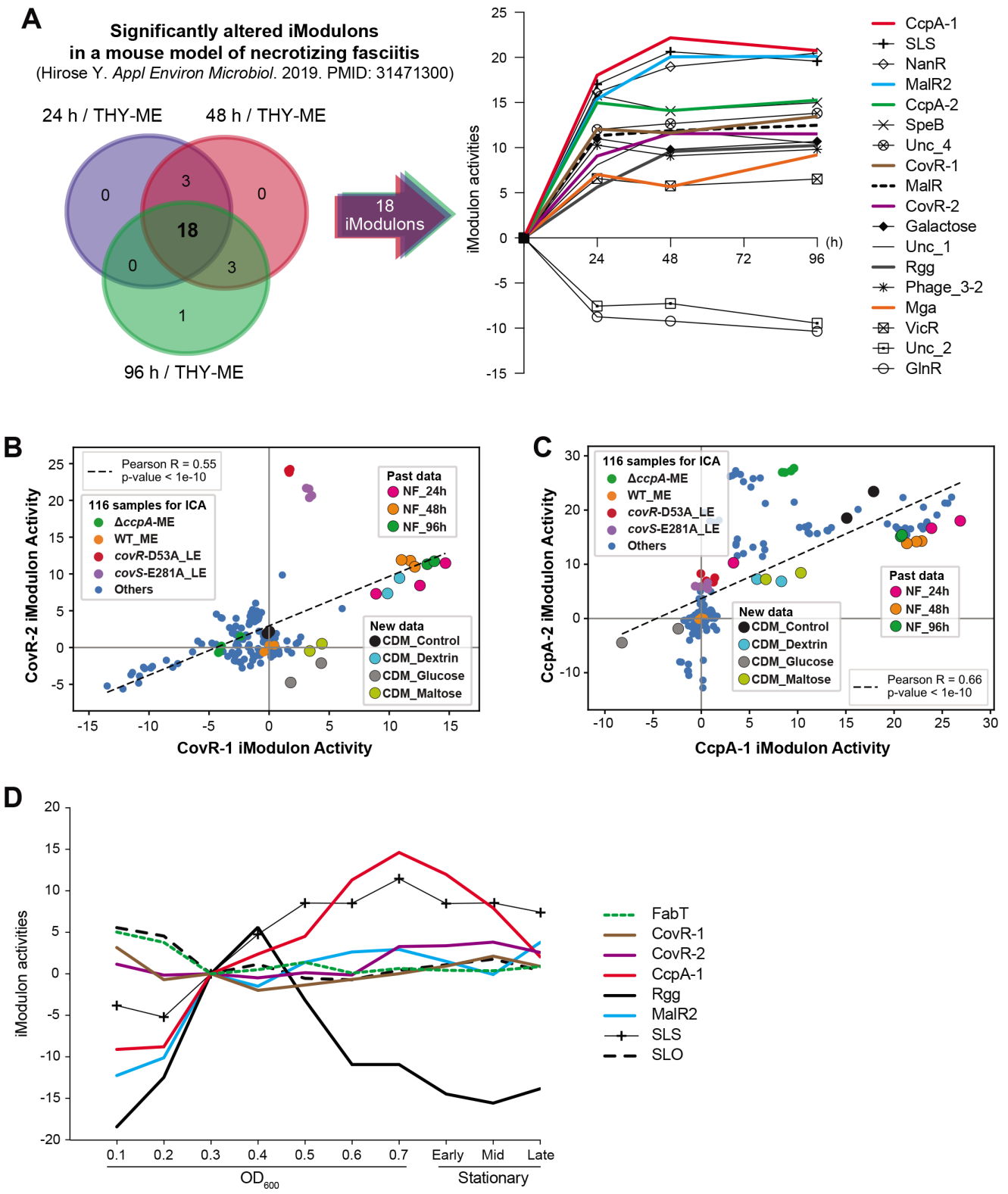

FIG 5 Bacterial iModulon activities in necrotizing fasciitis (NF) suggest that *Streptococcus pyogenes* sensed both the carbohydrate depletion and stresses affecting the CovRS regulator at the infected site. (A) Three-way Venn diagram illustrating the consistently altered iModulons during infection relative to samples isolated from THY culture at mid-exponential phase (THY-ME) (24 hours versus THY-ME, 48 hours versus THY-ME, and 96 hours versus THY-ME). (B) Comparison between CovR-1 and CovR-2 iModulon activities across 116 samples, new samples, and past samples. (C) Comparison between CcpA-1 and CcpA-2 iModulon activities across 116 samples, new samples, and past samples. (D) Relationship between the growth phase and the activities of iModulons. Log2(TPM + 1) values were centered on OD0.3 condition (mid-exponential phase).

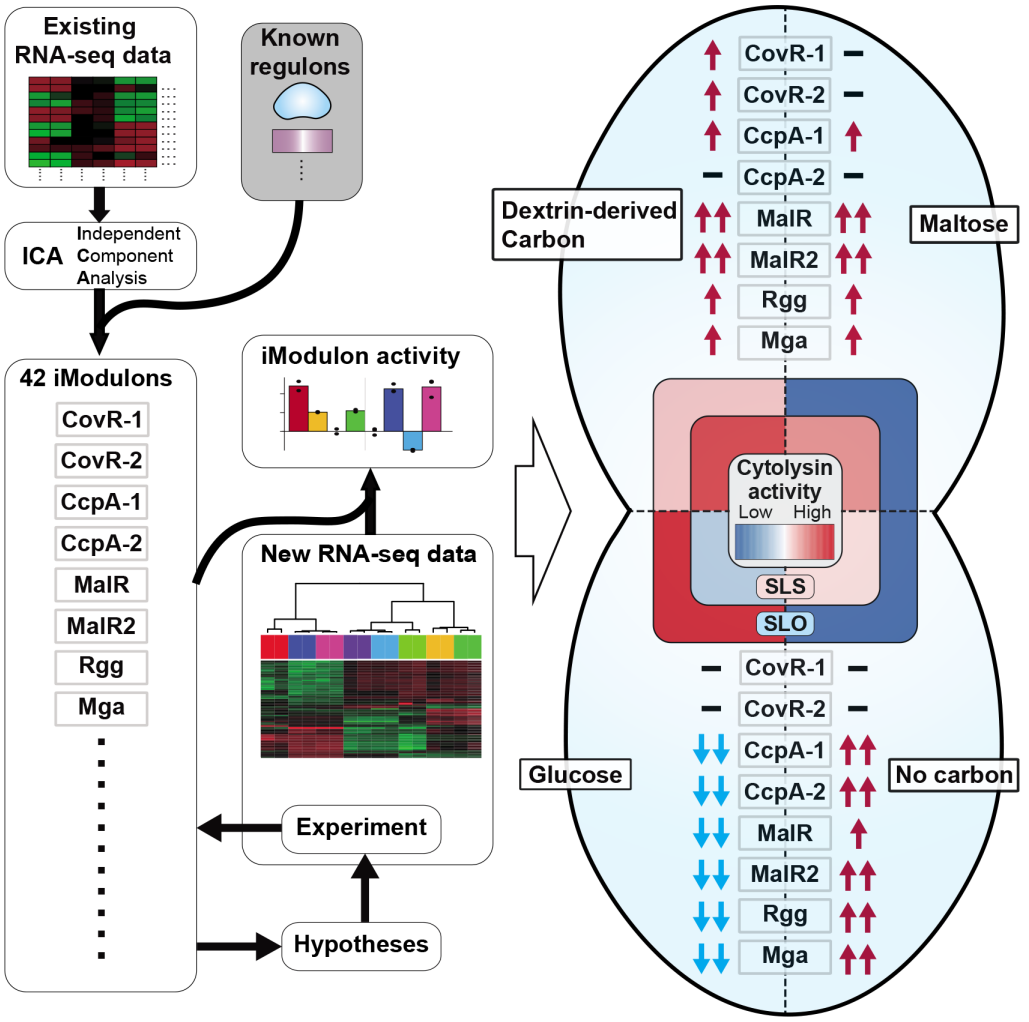

FIG 6 Independently modulated genes in *Streptococcus pyogenes* reveal carbon sources that change its hemolytic activity. In this study, we compiled existing high-quality 116 RNA-seq data sets of *S. pyogenes* serotype M1 and conducted ICA-based decomposition. This decomposition identified 42 iModulons and these activities, which allowed us to formulate hypotheses and identify carbon sources that change bacterial hemolysis activity. In particular, dextrin utilization changes bacterial hemolytic activity as compared to maltose or glucose utilization. Furthermore, visualizing the iModulon activities in the transcriptome help us to interpret RNA-seq data. In particular, we identified that *S. pyogenes* activated CovRS-related iModulons in the CDM dextrin (+) condition.

interests in the biology and pathogenesis of *S. pyogenes* infections, allowing them to explore the contents of other iModulons in detail and the relations between them. At iModulonDB.org, users can search and browse all iModulons from this data set and view them with interactive dashboards (32).

Previous reports have identified regulons using single-gene knockout mutants to conclude that the *nga-ifs-slo* operon is regulated by many TFs, including CovRS (6), Mga (9), Rgg (10), and TrxR (3). In our unbiased global study of available transcriptomes, only the MalR2 iModulon, CovRS-related iModulon, and SLO iModulon contained the *nga-ifs-slo* operon. This distinction suggests that *nga-ifs-slo* operon is under multiple regulations at the same time. By comprehensively comparing iModulon activities across all samples, we speculate three environmental cues increase the expression of *nga-ifs-slo* operon: (i) glucose-rich conditions (through activation of the SLO iModulon); (ii) glucose-depleted conditions (through activation of the MalR2 iModulon); and (iii) environmental stress sensed by CovRS (through activation of CovRS-related iModulon). The existence of multiple mechanisms for the upregulation of *nga-ifs-slo* operon may

link this operon with bacterial pathogenicity in various experimental conditions (33, 34, 66, 67).

Despite the structural similarities of glucose, maltose, and dextrin, *S. pyogenes* changed its hemolytic activity and transcriptome according to the carbon source. *S. pyogenes* in CDM + maltose or dextrin conditions activated iModulons associated with CcpA, MalR, and MalR2 TFs compared with those in CDM + glucose conditions. These TFs are repressors of the LacI/GalR family associated with CCR (5, 62). Glucose is the primary carbohydrate source of energy for many bacteria, including *S. pyogenes*, which generates ATP from its conversion to pyruvate through glycolysis. However, utilization of maltose or dextrin requires energy input to first convert them to glucose to be metabolized. Therefore, it was believed that *S. pyogenes* suppressed genes involved in alternative carbohydrate utilization when glucose was available, leading to CCR (61, 68). In the physiological niche of *S. pyogenes*, the human body, carbon sources are available in different combinations and levels depending on the type of tissue and state of starvation and could be amenable to further studies with these tools.

We identified the MalR2 regulon, *malACDFGX* and *amyAB* operons. Gene mutations of the *malR2* gene in *S. pyogenes* M1 strain significantly increase fitness in the non-human primate model of necrotizing myositis (50). In another point of note, the *malR2*, *malACDX*, and *amyAB* genes exist in the M1, M2, M4, and M28 serotypes but are absent from M3, M5, M6, M12, M18, and M49 serotypes (62). The MalR2 iModulon may underpin differences in metabolic and virulence properties among serotypes of *S. pyogenes*. Maltose and dextrin are the products of the action of salivary amylases on dietary starch. AmyA, α-amylase, increased murine mortality following mucosal challenge (62). It is possible that the products of salivary amylase may encourage upregulation of the *nga-ifs-slo* virulence-associated operon. It is presently unclear which particular carbon sources, concentrations, and enzymes are most utilized by the pathogen at infected sites.

CovRS-related iModulons were activated by dextrin utilization. Mutations of the CovRS virulence regulator derepress several different kinds of virulence genes, leading to a hypervirulent phenotype of *S. pyogenes* (4, 69). Ikebe et al. (70) reported that CovRS mutations in *S. pyogenes* were more common in clinical isolates from severe invasive infections compared with isolates from non-invasive infections. CovRS mediates a general stress response in *S. pyogenes* whereby specific environmental stimuli, such as increased temperature, acidic pH, and high salt concentrations (2). It had not been previously reported that dextrin utilization releases the expression of CovRS-repressed genes. However, further study is needed to clarify that dextrin utilization affects CovRS directly or indirectly.

Growth rate is one of the several basic characteristics of a bacterial culture that may introduce variance in gene expression. In fact, the activities of Rgg, SLS, and CcpA-related iModulons were quite variant with growth rate. Therefore, it may be necessary to control the growth rate difference if making comparisons between the activity of these iModulons across samples. In particular, this artifact should be taken into account when interpreting the noisy *in vivo* data such as the NF data from the murine model analyzed herein. This point represents a current limitation of interpretation in this study.

In this study, 116 all RNA-seq data sets for ICA come from *S. pyogenes* serotype M1 cultured in THY broth. Therefore, identified iModulons are specific for this most commonly studied media growth condition. Future studies can append other conditions to this data set and observe the changes to the set of iModulons that ensue. Although iModulon activities can often be explained by prior knowledge, they can also present surprising relationships that lead to the generation of hypotheses. We describe new information about iModulons and their activities enabled us to query metabolic and regulatory cross-talk, discover new potential relationships, find coordination between metabolism and virulence, and facilitate the interpretation of the *S. pyogenes* response *in vivo*. These versatilities may allow iModulons identified in this study or in near future to serve as a powerful guidepost to further our understanding of *S. pyogenes* TRN structure and dynamics.

ACKNOWLEDGMENTS

We acknowledge the University of California, Davis DNA sequencing facility for their support with RNA-seq. This study was supported in part by the Ministry of Health, Labour and Welfare of Japan and the Japan Agency for Medical Research and Development (AMED) (JP20fk0108130, JP23wm0325066); Japanese Society for the Promotion of Science (JSPS) KAKENHI (20K18474, 20KK02100, and 22K09924, 22H03262); JSPS Overseas Research Fellowships; and the Nippon Foundation (The Nippon Foundation—Osaka University Infectious Disease Response Project). Hyun Gyu Lim appreciates the Inha University Research Grant. The funders had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript.

AUTHOR AFFILIATIONS

[1]Department of Microbiology, Graduate School of Dentistry, Osaka University, Suita, Osaka, Japan
[2]Department of Pediatrics, University of California at San Diego School of Medicine, La Jolla, California, USA
[3]Department of Bioengineering, University of California San Diego, La Jolla, California, USA
[4]Department of Biological Engineering, Inha University, Michuhol-gu, Incheon, South Korea
[5]School of Biotechnology, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India
[6]Center for Infectious Diseases Education and Research, Osaka University, Suita, Osaka, Japan
[7]Skaggs School of Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA

AUTHOR ORCIDs

Yujiro Hirose http://orcid.org/0000-0001-6338-4767

FUNDING

Funder	Grant(s)	Author(s)
Japan Agency for Medical Research and Development (AMED)	JP20fk0108130	Shigetada Kawabata
MEXT | Japan Society for the Promotion of Science (JSPS)	20K18474	Yujiro Hirose
MEXT | Japan Society for the Promotion of Science (JSPS)	20KK02100	Yujiro Hirose
MEXT | Japan Society for the Promotion of Science (JSPS)	22K09924	Shigetada Kawabata
MEXT | Japan Society for the Promotion of Science (JSPS)	22H032620	Shigetada Kawabata

AUTHOR CONTRIBUTIONS

Yujiro Hirose, Conceptualization, Data curation, Formal analysis, Writing – original draft | Saugat Poudel, Data curation, Methodology, Software, Supervision | Anand V. Sastry, Data curation, Formal analysis, Methodology, Software, Supervision | Kevin Rychel, Visualization, Writing – original draft, Writing – review and editing | Cameron R. Lamoureux, Data curation, Visualization | Richard Szubin, Methodology | Daniel C. Zielinski, Methodology, Software | Hyun Gyu Lim, Methodology | Nitasha D. Menon, Methodology | Helena Bergsten, Formal analysis | Satoshi Uchiyama, Formal analysis | Tomoki Hanada, Formal analysis | Shigetada Kawabata, Funding acquisition | Bernhard O. Palsson, Conceptualization, Funding acquisition, Methodology, Software, Visualization

| Victor Nizet, Conceptualization, Funding acquisition, Project administration, Writing – review and editing

DATA AVAILABILITY STATEMENT

The data accession numbers can be found in Supplementary Data 1. The normalized log TPM (X) and the calculated M and A matrix of the model can be found in Supplementary Data 2. The code for QC/QA, ICA, and assessment of the regulator enrichment can be found on Github (https://github.com/avsastry/modulome-workflow). Python package for analyzing and visualizing iModulons (PyModulon) can be also found on Github (https://github.com/SBRG/pymodulon). All new 16 RNA-seq data obtained in this study have been deposited into DDBJ sequence read archive (DRA) under the accession number DRA014564. Interactive online dashboards for all iModulons and all data are available at https://imodulondb.org under the data set name "*S. pyogenes*."

ADDITIONAL FILES

The following material is available online.

Supplemental Material

FIG S1 (303210_0_supp_6693103_rrpr4p.pdf). Diversity of 116 RNA-seq data sets used for ICA decomposition. Loadings of the first two principal components (PC). The variation in locations across 116 RNA-seq samples demonstrates the diversity of the compendium.
FIG S2 (303210_0_supp_6693104_rrpr4p.pdf). Conditions that *S. pyogenes* may show the antagonistic expression of sag A-I and nga-ifs-slo operons.
FIG S3 (303210_0_supp_6693105_rrp94p.pdf). The actual MalR2 regulon. Differentially expressed genes (DEGs) from comparisons of the ΔmalR2 mutant and WT strains at (A) mid-exponential, and (B) stationary growth phases in THY broth. Colored circles indicate significantly upregulated (red) and downregulated (blue) genes (absolute log2 fold change, > 1; adjusted $P < 0.1$).
Supplementary information 1 (303210_0_supp_6693108_rrpp4p.pdf). Graphical representation of 42 iModulons.
Supplementary Data 1 (303210_0_supp_6684225_rrjryg.xlsx). Sample metadata and RNA-seq summary statistics.
Supplementary Data 2 (303210_0_supp_6693101_rrpm4p.xlsx). Transcriptomic matrix (X), independent components (M), and their condition-specific activities (A).
Supplementary Data 3 (303210_0_supp_6684228_rrjryg.xlsx). Transcriptional regulatory network (TRN) used to automatically annotate iModulons.
Supplementary Data 4 (303210_0_supp_6684227_rrjryg.xlsx). The iModulon structure and annotation statistics.
Supplementary Data 5 (303210_0_supp_6693102_rrpv4p.xlsx). Metadata of new 36 RNA-seq samples, and the part of the result of RNA-seq analyses.
Supplementary Data 6 (303210_0_supp_6684226_rrjryg.xlsx). Anti-rRNA oligonucleotide probes for *S. pyogenes* 5448 strain.

REFERENCES

1. Carapetis JR, Steer AC, Mulholland EK, Weber M. 2005. The global burden of group A streptococcal diseases. Lancet Infect Dis 5:685–694. https://doi.org/10.1016/S1473-3099(05)70267-X
2. Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V. 2014. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. Clin Microbiol Rev 27:264–301. https://doi.org/10.1128/CMR.00101-13
3. Baruch M, Belotserkovsky I, Hertzog BB, Ravins M, Dov E, McIver KS, Le Breton YS, Zhou Y, Cheng CY, Hanski E. 2014. An extracellular bacterial pathogen modulates host metabolism to regulate its own sensing and proliferation. Cell 156:97–108. https://doi.org/10.1016/j.cell.2013.12.007
4. Mayfield JA, Liang Z, Agrahari G, Lee SW, Donahue DL, Ploplis VA, Castellino FJ. 2014. Mutations in the control of virulence sensor gene from *Streptococcus pyogenes* after infection in mice lead to clonal bacterial variants with altered gene regulatory activity and virulence. PLoS One 9: e100698. https://doi.org/10.1371/journal.pone.0100698
5. Shelburne SA, Sahasrobhajane P, Suber B, Keith DB, Davenport MT, Horstmann N, Kumaraswami M, Olsen RJ, Brennan RG, Musser JM. 2011. Niche-specific contribution to streptococcal virulence of a MalR-regulated carbohydrate binding protein. Mol Microbiol 81:500–514. https://doi.org/10.1111/j.1365-2958.2011.07708.x
6. Shelburne SA, Olsen RJ, Suber B, Sahasrabhojane P, Sumby P, Brennan RG, Musser JM. 2010. A combination of independent transcriptional

regulators shapes bacterial virulence gene expression during infection. PLoS Pathog 6: e1000817. https://doi.org/10.1371/journal.ppat.1000817
7. Paluscio E, Watson ME, Caparon MG. 2018. CcpA coordinates growth/damage balance for *Streptococcus pyogenes* pathogenesis. Sci Rep 8:14254. https://doi.org/10.1038/s41598-018-32558-0
8. DebRoy S, Saldaña M, Travisany D, Montano A, Galloway-Peña J, Horstmann N, Yao H, González M, Maass A, Latorre M, Shelburne SA. 2016. A multi-serotype approach clarifies the catabolite control protein a regulon in the major human pathogen group A *Streptococcus*. Sci Rep 6:32442. https://doi.org/10.1038/srep32442
9. Valdes KM, Sundar GS, Belew AT, Islam E, El-Sayed NM, Le Breton Y, McIver KS. 2018. Glucose levels alter the Mga virulence regulon in the group A *streptococcus*. Sci Rep 8:4971. https://doi.org/10.1038/s41598-018-23366-7
10. Dmitriev AV, McDowell EJ, Kappeler KV, Chaussee MA, Rieck LD, Chaussee MS. 2006. The Rgg regulator of *Streptococcus pyogenes* influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon. J Bacteriol 188:7230–7241. https://doi.org/10.1128/JB.00877-06
11. Kietzman CC, Caparon MG. 2011. Distinct time-resolved roles for two catabolite-sensing pathways during *Streptococcus pyogenes* infection. Infect Immun 79:812–821. https://doi.org/10.1128/IAI.01026-10
12. Toukoki C, Gold KM, McIver KS, Eichenbaum Z. 2010. MtsR is a dual regulator that controls virulence genes and metabolic functions in addition to metal homeostasis in the group A *Streptococcus*. Mol Microbiol 76:971–989. https://doi.org/10.1111/j.1365-2958.2010.07157.x
13. Brenot A, Weston BF, Caparon MG. 2007. A PerR-regulated metal transporter (PmtA) is an interface between oxidative stress and metal homeostasis in *Streptococcus pyogenes*. Mol Microbiol 63:1185–1196. https://doi.org/10.1111/j.1365-2958.2006.05577.x
14. Treviño J, Liu Z, Cao TN, Ramirez-Peña E, Sumby P. 2013. RivR is a negative regulator of virulence factor expression in group A *Streptococcus*. Infect Immun 81:364–372. https://doi.org/10.1128/IAI.00703-12
15. Kreth J, Chen Z, Ferretti J, Malke H. 2011. Counteractive balancing of transcriptome expression involving CodY and CovRS in *Streptococcus pyogenes*. J Bacteriol 193:4153–4165. https://doi.org/10.1128/JB.00061-11
16. Chen Z, Itzek A, Malke H, Ferretti JJ, Kreth J. 2013. Multiple roles of RNase Y in *Streptococcus pyogenes* mRNA processing and degradation. J Bacteriol 195:2585–2594. https://doi.org/10.1128/JB.00097-13
17. Graham MR, Smoot LM, Migliaccio CAL, Virtaneva K, Sturdevant DE, Porcella SF, Federle MJ, Adams GJ, Scott JR, Musser JM. 2002. Virulence control in group A *Streptococcus* by a two-component gene regulatory system: global expression profiling and *in vivo* infection modeling. Proc Natl Acad Sci USA 99:13855–13860. https://doi.org/10.1073/pnas.202353699
18. Kreikemeyer B, Nakata M, Köller T, Hildisch H, Kourakos V, Standar K, Kawabata S, Glocker MO, Podbielski A. 2007. The *Streptococcus pyogenes* serotype M49 Nra-Ralp3 transcriptional regulatory network and its control of virulence factor expression from the novel eno ralp3 epf sagA pathogenicity region. Infect Immun 75:5698–5710. https://doi.org/10.1128/IAI.00175-07
19. Reid SD, Chaussee MS, Doern CD, Chaussee MA, Montgomery AG, Sturdevant DE, Musser JM. 2006. Inactivation of the group A *Streptococcus* regulator srv results in chromosome wide reduction of transcript levels, and changes in extracellular levels of Sic and SpeB. FEMS Immunol Med Microbiol 48:283–292. https://doi.org/10.1111/j.1574-695X.2006.00150.x
20. Voyich JM, Braughton KR, Sturdevant DE, Vuong C, Kobayashi SD, Porcella SF, Otto M, Musser JM, DeLeo FR. 2004. Engagement of the pathogen survival response used by group A *Streptococcus* to avert destruction by innate host defense. J Immunol 173:1194–1201. https://doi.org/10.4049/jimmunol.173.2.1194
21. Liu M, Hanks TS, Zhang J, McClure MJ, Siemsen DW, Elser JL, Quinn MT, Lei B. 2006. Defects in *ex vivo* and *in vivo* growth and sensitivity to osmotic stress of group A *Streptococcus* caused by interruption of response regulator gene vicR. Microbiology 152:967–978. https://doi.org/10.1099/mic.0.28706-0
22. Finn MB, Ramsey KM, Dove SL, Wessels MR, Kline KA. 2021. Identification of group A *Streptococcus* genes directly regulated by CsrRS and novel intermediate regulators. mBio 12:e0164221. https://doi.org/10.1128/mBio.01642-21
23. DebRoy S, Aliaga-Tobar V, Galvez G, Arora S, Liang X, Horstmann N, Maracaja-Coutinho V, Latorre M, Hook M, Flores AR, Shelburne SA. 2021. Genome-wide analysis of *in vivo* CcpA binding with and without its key co-factor HPr in the major human pathogen group A *streptococcus*. Mol Microbiol 115:1207–1228. https://doi.org/10.1111/mmi.14667
24. Kreikemeyer B, McIver KS, Podbielski A. 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. Trends Microbiol 11:224–232. https://doi.org/10.1016/s0966-842x(03)00098-2
25. Fang X, Sastry A, Mih N, Kim D, Tan J, Yurkovich JT, Lloyd CJ, Gao Y, Yang L, Palsson BO. 2017. Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. Proc Natl Acad Sci USA 114:10286–10291. https://doi.org/10.1073/pnas.1702581114
26. Larsen SJ, Röttger R, Schmidt H, Baumbach J. 2019. E. *Coli* gene regulatory networks are inconsistent with gene expression data. Nucleic Acids Res 47:85–92. https://doi.org/10.1093/nar/gky1176
27. Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA, Palsson 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
28. Poudel S, Tsunemoto H, Seif Y, Sastry AV, Szubin R, Xu S, Machado H, Olson CA, Anand A, Pogliano J, Nizet V, Palsson BO. 2020. Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. Proc Natl Acad Sci USA 117:17228–17239. https://doi.org/10.1073/pnas.2008413117
29. Rychel K, Sastry AV, Palsson BO. 2020. Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. Nat Commun 11:6338. https://doi.org/10.1038/s41467-020-20153-9
30. Shulman ST, Tanz RR, Kabat W, Kabat K, Cederlund E, Patel D, Li Z, Sakota V, Dale JB, Beall B, US Streptococcal Pharyngitis Surveillance Group. 2004. Group A streptococcal pharyngitis serotype surveillance in North America, 2000-2002. Clin Infect Dis 39:325–332. https://doi.org/10.1086/421949
31. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. 2009. Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis 9:611–616. https://doi.org/10.1016/S1473-3099(09)70178-1
32. Rychel K, Decker K, Sastry AV, Phaneuf PV, Poudel S, Palsson BO. 2021. iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. Nucleic Acids Res 49:D112–D120. https://doi.org/10.1093/nar/gkaa810
33. Zhu L, Olsen RJ, Nasser W, Beres SB, Vuopio J, Kristinsson KG, Gottfredsson M, Porter AR, DeLeo FR, Musser JM. 2015. A molecular trigger for intercontinental epidemics of group A *Streptococcus*. J Clin Invest 125:3545–3559. https://doi.org/10.1172/JCI82478
34. Zhu L, Olsen RJ, Nasser W, de la Riva Morales I, Musser JM. 2015. Trading capsule for increased cytotoxin production: contribution to virulence of a newly emerged clade of *emm*89 *Streptococcus pyogenes*. mBio 6: e01378-15. https://doi.org/10.1128/mBio.01378-15
35. Poudel S, Tsunemoto H, Meehan M, Szubin R, Olson CA, Lamsa A, Seif Y, Dillon N, Vrbanac A, Sugie J, Dahesh S, Monk JM, Dorrestein PC, Pogliano J, Knight R, Nizet V, Palsson 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
36. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
37. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656
38. Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
39. Eraso JM, Olsen RJ, Beres SB, Kachroo P, Porter AR, Nasser W, Bernard PE, DeLeo FR, Musser JM. 2016. Genomic landscape of intrahost variation in group A *Streptococcus*: repeated and abundant mutational inactivation of the fabT gene encoding a regulator of fatty acid synthesis. Infect Immun 84:3268–3281. https://doi.org/10.1128/IAI.00608-16

40. Horstmann N, Tran CN, Brumlow C, DebRoy S, Yao H, Nogueras Gonzalez G, Makthal N, Kumaraswami M, Shelburne SA. 2018. Phosphatase activity of the control of virulence sensor kinase CovS is critical for the pathogenesis of group A *Streptococcus*. PLoS Pathog 14: e1007354. https://doi.org/10.1371/journal.ppat.1007354
41. Kachroo P, Eraso JM, Olsen RJ, Zhu L, Kubiak SL, Pruitt L, Yerramilli P, Cantu CC, Ojeda Saavedra M, Pensar J, Corander J, Jenkins L, Kao L, Granillo A, Porter AR, DeLeo FR, Musser JM, Rappuoli R. 2020. New pathogenesis mechanisms and translational leads identified by multidimensional analysis of necrotizing myositis in primates. mBio 11: e03363-19. https://doi.org/10.1128/mBio.03363-19
42. Freiberg JA, Le Breton Y, Tran BQ, Scott AJ, Harro JM, Ernst RK, Goo YA, Mongodin EF, Goodlett DR, McIver KS, Shirtliff ME, Dorrestein PC. 2016. Global analysis and comparison of the transcriptomes and proteomes of group A *Streptococcus* biofilms . mSystems 1. https://doi.org/10.1128/mSystems.00149-16
43. Pedregosa Fet al. 2011. Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830.
44. McConn JL, Lamoureux CR, Poudel S, Palsson BO, Sastry AV. 2021. Optimal dimensionality selection for independent component analysis of transcriptomic data. BMC Bioinformatics 22: 584. https://doi.org/10.1186/s12859-021-04497-7
45. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA. 2013. RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics 14: 745. https://doi.org/10.1186/1471-2164-14-745
46. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387
47. Kansal RG, McGeer A, Low DE, Norrby-Teglund A, Kotb M. 2000. Inverse relation between disease severity and expression of the streptococcal cysteine protease, SpeB, among clonal M1T1 isolates recovered from invasive group A streptococcal infection cases. Infect Immun 68:6362–69. https://doi.org/10.1128/IAI.68.11.6362-6369.2000
48. Pritzlaff CA, Chang JC, Kuo SP, Tamura GS, Rubens CE, Nizet V. 2001. Genetic basis for the *β*-haemolytic/cytolytic activity of group B *Streptococcus*. Mol Microbiol 39:236–247. https://doi.org/10.1046/j.1365-2958.2001.02211.x
49. Lauth X, von Köckritz-Blickwede M, McNamara CW, Myskowski S, Zinkernagel AS, Beall B, Ghosh P, Gallo RL, Nizet V. 2009. M1 protein allows group A streptococcal survival in phagocyte extracellular traps through cathelicidin inhibition. J Innate Immun 1:202–214. https://doi.org/10.1159/000203645
50. Zhu L, Olsen RJ, Beres SB, Eraso JM, Saavedra MO, Kubiak SL, Cantu CC, Jenkins L, Charbonneau ARL, Waller AS, Musser JM. 2019. Gene fitness landscape of group A *streptococcus* during necrotizing myositis. J Clin Invest 129:887–901. https://doi.org/10.1172/JCI124994
51. Choe D, Szubin R, Poudel S, Sastry A, Song Y, Lee Y, Cho S, Palsson B, Cho B-K. 2021. RiboRid: a low cost, advanced, and ultra-efficient method to remove ribosomal RNA for bacterial transcriptomics. PLoS Genet 17: e1009821. https://doi.org/10.1371/journal.pgen.1009821
52. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560
53. Sastry AV, Poudel S, Rychel K, Yoo R, Lamoureux CR, Chauhan S, Haiman ZB, Al Bulushi T, Seif Y, Palsson BO. 2021. Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. Bioinformatics. https://doi.org/10.1101/2021.07.01.450581
54. Leinonen R, Sugawara H, Shumway M. 2011. International nucleotide sequence database C. The sequence read archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019
55. Fontaine MC, Lee JJ, Kehoe MA. 2003. Combined contributions of streptolysin O and streptolysin S to virulence of serotype M5 *Streptococcus pyogenes* strain Manfredo. Infect Immun 71:3857–3865. https://doi.org/10.1128/IAI.71.7.3857-3865.2003
56. Nizet V, Beall B, Bast DJ, Datta V, Kilburn L, Low DE, De Azavedo JCS. 2000. Genetic locus for streptolysin S production by group A *Streptococcus*. Infect Immun 68:4245–4254. https://doi.org/10.1128/IAI.68.7.4245-4254.2000
57. Datta V, Myskowski SM, Kwinn LA, Chiem DN, Varki N, Kansal RG, Kotb M, Nizet V. 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol Microbiol 56:681–695. https://doi.org/10.1111/j.1365-2958.2005.04583.x
58. Hirose Y, Yamaguchi M, Sumitomo T, Nakata M, Hanada T, Okuzaki D, Motooka D, Mori Y, Kawasaki H, Coady A, Uchiyama S, Hiraoka M, Zurich RH, Amagai M, Nizet V, Kawabata S. 2021. *Streptococcus pyogenes* upregulates arginine catabolism to exert its pathogenesis on the skin surface. Cell Rep 34:108924. https://doi.org/10.1016/j.celrep.2021.108924
59. Shelburne SA, Keith D, Horstmann N, Sumby P, Davenport MT, Graviss EA, Brennan RG, Musser JM. 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*. Proc Natl Acad Sci USA 105:1698–1703. https://doi.org/10.1073/pnas.0711767105
60. Kinkel TL, McIver KS. 2008. CcpA-mediated repression of streptolysin S expression and virulence in the group A *Streptococcus*. Infect Immun 76:3451–3463. https://doi.org/10.1128/IAI.00343-08
61. Pancholi V, Caparon M. 2016. *Streptococcus pyogenes* Metabolism, . In Ferretti JJ, DL Stevens, VA Fischetti (ed), Streptococcus pyogenes: basic biology to clinical manifestations
62. Shelburne Iii SA, Keith DB, Davenport MT, Beres SB, Carroll RK, Musser JM. 2009. Contribution of AmyA, an extracellular alpha-glucan degrading enzyme, to group A streptococcal host-pathogen interaction. Mol Microbiol 74:159–174. https://doi.org/10.1111/j.1365-2958.2009.06858.x
63. Hirose Y, Yamaguchi M, Okuzaki D, Motooka D, Hamamoto H, Hanada T, Sumitomo T, Nakata M, Kawabata S. 2019. *Streptococcus pyogenes* transcriptome changes in the inflammatory environment of necrotizing fasciitis. Appl Environ Microbiol 85: e01428-19. https://doi.org/10.1128/AEM.01428-19
64. Sastry AV, Hu A, Heckmann D, Poudel S, Kavvas E, Palsson BO. 2021. Independent component analysis recovers consistent regulatory signals from disparate datasets. PLoS Comput Biol 17: e1008647. https://doi.org/10.1371/journal.pcbi.1008647
65. Lamoureux CR, Decker KT, Sastry AV, Rychel K, Gao Y, McConn JL, Zielinski DC, Palsson BO. 2021. A multi-scale transcriptional regulatory network knowledge base for *Escherichia coli*. Bioinformatics. https://doi.org/10.1101/2021.04.08.439047
66. Limbago B, Penumalli V, Weinrick B, Scott JR. 2000. Role of streptolysin o in a mouse model of invasive group A streptococcal disease. Infect Immun 68:6384–6390. https://doi.org/10.1128/IAI.68.11.6384-6390.2000
67. Timmer AM, Timmer JC, Pence MA, Hsu L-C, Ghochani M, Frey TG, Karin M, Salvesen GS, Nizet V. 2009. Streptolysin O promotes group A *Streptococcus* immune evasion by accelerated macrophage apoptosis. J Biol Chem 284:862–871. https://doi.org/10.1074/jbc.M804632200
68. Deutscher J, Francke C, Postma PW. 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol Mol Biol Rev 70:939–1031. https://doi.org/10.1128/MMBR.00024-06
69. Kansal RG, Datta V, Aziz RK, Abdeltawab NF, Rowe S, Kotb M. 2010. Dissection of the molecular basis for hypervirulence of an *in vivo*-selected phenotype of the widely disseminated M1T1 strain of group A *Streptococcus* bacteria. J Infect Dis 201:855–865. https://doi.org/10.1086/651019
70. Ikebe T, Ato M, Matsumura T, Hasegawa H, Sata T, Kobayashi K, Watanabe H. 2010. Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. PLoS Pathog 6: e1000832. https://doi.org/10.1371/journal.ppat.1000832