Raman spectra identify vancomycin-resistant phenotypes and their transcriptomic features in *Staphylococcus aureus*

Ken-ichiro F. Kamei,[1*] Reiko Okura,[1] Koseki J. Kobayashi-Kirschvink,[2] Yuki Katayama,[3] Yuichi Wakamoto[1,4,5*]

[1]Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902, Japan

[2]Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

[3]KatayamaKikai Co., Ltd, Fukui 918-8014, Japan

[4]Research Center for Complex Systems Biology, The University of Tokyo, Tokyo 153-8902, Japan

[5]Universal Biology Institute, The University of Tokyo, Tokyo 153-8902, Japan

*To whom correspondence should be addressed; E-mail: kenichiro_kamei@cell.c.u-tokyo.ac.jp (K.F.K.), cwaka@mail.ecc.u-tokyo.ac.jp (Y.W.).

Staphylococcus aureus **is a pathogenic bacterium that has caused multiple epidemics linked with the emergence of new antibiotic resistance. Vancomycin is the first-line antibiotic to treat methicillin-resistant *S. aureus* (MRSA) infections. However, several types of vancomycin-non-susceptible MRSA strains have been isolated from patients to date. Rapid assessment of their resistance levels and underlying molecular profiles is crucial for preventing their spread and counteracting resistance; however, the broad resistance spectrum and the diversity of genetic changes have impeded this practice. Here, we demonstrate that the vancomycin resistance levels of various MRSA strains can be determined using dimension-reduced Raman spectra obtained from single cells. The transcriptome profiles of the different strains can also be predicted from their dimensionally reduced Raman spectra by simple linear regression. This Raman-transcriptome correspondence allows us to map the transcriptome components onto dimension-reduced Raman space and characterize groups of genes associated with different phenotypes. Furthermore, single-cell Raman spectra predicted a cell strain with greater phenotypic heterogeneity, which was confirmed by single-cell growth analysis. Overall, our results demonstrate the efficacy of Raman spectroscopy in identifying resistant phenotypes, their associated gene expression features, and intrapopulation phenotypic heterogeneity.**

Introduction

The emergence and spread of antibiotic-resistant bacteria is a huge threat to public health worldwide. *Staphylococcus aureus* is known for its ability to acquire antibiotic resistance and has caused several epidemic waves since the introduction of antibiotics (*1, 2*). Vancomycin was introduced in the 1980s to treat severe methicillin-resistant *S. aureus* (MRSA) infections (*3*). However, various vancomycin-nonsusceptible MRSA strains have been identified clinically (*3–9*), which are often associated with treatment failures (*10–13*).

Vancomycin-nonsusceptible *S. aureus* strains fall into three categories: vancomycin-resistant *S. aureus* (VRSA), vancomycin intermediately-resistant *S. aureus* (VISA), and heterogeneous vancomycin intermediately-resistant *S. aureus* (hVISA) (*3, 8, 9*). In particular, VISA and hVISA prevalence has been increasing recently (*8, 9*). The minimum inhibitory concentration (MIC) for VISA is 4–8 μg/mL, which is higher than that of vancomycin-susceptible MRSA (VSSA) strains ($\leq$2 μg/mL) but lower than that of VRSA ($\geq$16 μg/mL). hVISA strains are often judged as VSSA by the standard MIC test, but they contain small subpopulations with the VISA-level vancomycin resistance and could lead to treatment failures (*5, 14, 15*). hVISA strains are considered to be in intermediate evolutionary steps from VSSA to VISA (*15, 16*). As VISA and hVISA prevalence raises growing concern about longer hospitalization and higher risk of persistent infection (*8, 9*), it is critical to swiftly identify the resistant phenotypes of bacteria infecting patients and treat infections with appropriate drug regimens.

The methicillin resistance of MRSA is predominantly caused by the acquisition of the *mecA* or *mecC* gene on a mobile genetic element (*2, 3, 15, 17, 18*). However, the mechanisms behind vancomycin-nonsusceptibility are not yet fully understood and are believed to be diverse (*3, 9, 15, 16, 19–24*). Genetic analyses have suggested that multi-step mutations at various genetic loci can increase vancomycin resistance levels (*15, 16, 22, 23*). Consequently, it remains challenging

to identify the vancomycin resistance levels through targeted gene analysis.

Despite the diverse mutational profiles, it is suggested that physiological characteristic changes such as cell wall thickening and slow growth are prevalent amongst vancomycin-nonsusceptible MRSA strains (*19, 20, 24–26*). Therefore, analyzing their global molecular composition and gene expression profiles might allow us to identify their vancomycin-resistance levels. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) combined with a machine-learning classification can distinguish different pathogenic bacteria and has also been applied to identify MRSA and different phenotypes of vancomycin-nonsusceptible MRSA strains (*27–29*). However, MALDI-TOF MS requires culturing cells to obtain microbial biomass (for bacteria, 10^4 to 10^6 colony forming unit) for analysis (*27*). Furthermore, identifying the comprehensive molecular signatures of targeted cells based on the MALDI-TOF MS spectra remains intractable. In addition, due to the population-based approach of the analysis, it is intractable to unravel single-cell-level phenotypic heterogeneity within a bacterial population.

Raman microscopy also provides spectral information reflecting the comprehensive molecular composition of single cells (*30–35*) because a Raman spectrum from a cell is, in principle, the linear superposition of the spectra of its constituent biomolecules. The complexity of the molecular composition of cells and severe spectral overlap make it nearly impossible to comprehensively decompose a cellular Raman spectrum and determine the abundances of different cellular components (*30–32*). However, it is important to note that most changes in gene expression and metabolic profiles are inherently global and modular (*36–43*). Therefore, without finely and precisely decomposing the spectra and assigning peaks to molecular species, the correspondence between Raman spectra and comprehensive molecular profiles may well be unraveled. Indeed, recent studies have shown that global changes in the entire fingerprint region of cellular Raman spectra correlate with global gene expression profiles for environmentally

perturbed cells (*33, 44, 45*). Consequently, transcriptomic and proteomic profiles can be inferred from Raman spectra thanks to the intrinsic low-dimensional constraints on changes in global gene expression patterns (*33, 44, 45*). However, most Raman applications have focused on the correlations between the changes in specific Raman spectral peaks and the changes in the abundance of specific biomolecules in cells, or have employed machine learning approaches to distinguish microbial species or cell types without clarifying the underlying molecular profiles. As a result, the correspondence between Raman spectra and system-level molecular profiles of cells has remained largely unexplored in Raman spectroscopy applications.

This report first demonstrates that dimension-reduced Raman spectra of VSSA, hVISA and VISA strains under a common culture condition can predict vancomycin-resistance phenotypes. We then reveal that global transcriptomic profiles are also inferable from the low-dimensional Raman spectra, demonstrating that the Raman-transcriptome correspondence holds even for genetically distinct cells. Using the Raman-transcriptome correspondence, we propose a method for representing transcriptome components with Raman axes that correlate with different phenotypic features. This method identifies several gene sets associated with vancomycin resistance and virulence. Finally, we show that the variability of single-cell Raman spectra can be used to interrogate the phenotypic heterogeneity of different strains, a clinically important phenotypic feature of the hVISA strains.

Results

Raman spectra distinguish cellular states of MRSA strains with distinct vancomycin susceptibility. In this study, we analyzed nine MRSA strains. Of these, three strains (N315, Mu3, and Mu50) are clinical isolates with distinct susceptibility to vancomycin (*5*), while the other six (ΔIP, ΔIP1–ΔIP5) are strains created by introducing specific mutations or by selection under rifampicin exposure (*16*). Vancomycin susceptibility tests performed in (*16*) classified N315

and ΔIP as VSSA, Mu3, ΔIP1, and ΔIP2 as hVISA, and Mu50 and ΔIP3–ΔIP5 as VISA on the basis of population measurements such as MIC tests (Fig. 1A).

We cultured these strains in brain heart infusion (BHI) medium and sampled cells in exponential phase for Raman measurements. Raman spectra of 48 single cells per strain were acquired using a Raman microscope (Fig. 1B, 1C and S1). To eliminate systematic errors in weak spontaneous Raman signals, the spectra were acquired in three biological replicates on different dates. We focused on the fingerprint region (700 to 1,800 cm^{-1}) of the Raman spectra in the analysis, where the signals from biomolecules such as proteins and metabolites are mainly observed.

We applied principal component-linear discriminant analysis (PC-LDA) to the Raman spectra to determine whether they could distinguish differences in cellular states among the strains (*33, 45, 46*). The Raman spectra were classified in a low-dimensional (eight-dimensional) space (Fig. 1D). In this LDA space, each point represents a Raman spectrum from an individual cell. The results showed that the points for the cells in each strain were closely positioned and clustered in this LDA space (Fig. 1D). Visualizing the eight-dimensional Raman spectra by embedding them in a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) confirmed the cluster separation (Fig. 1E). Therefore, the Raman spectra can distinguish the distinct cellular states of the nine MRSA strains.

Notably, the positions along the first LDA axis (LDA1) correlated with their vancomycin resistance levels (Pearson correlation coefficient $r = 0.81 \pm 0.15$, Fig. 1F and 1G). Therefore, PC-LDA detects the Raman spectral changes correlated with the vancomycin resistance levels as the dominant differences. Additionally, the second LDA axis (LDA2) recognizes the differences between the clinically isolated vancomycin-nonsusceptible strains (Mu3 and Mu50) and the constructed vancomycin-nonsusceptible strains (ΔIP1–ΔIP5) (Fig. 1F). Therefore, the result shows that, even among the MRSA strains with equivalent vancomycin resistance levels,

there exist differences in their molecular profiles detectable by Raman spectra.

Raman spectral differences among the MRSA strains are explainable by their transcriptomes. We next investigated whether the differences in Raman spectra of nine MRSA strains can be explained by their gene expression profiles. To test this, we gathered transcriptome data of these strains cultured in the BHI medium from public databases (*16, 21, 26*) and examined their correspondences with the Raman data. We hypothesized a linear correspondence

$$\mathbf{r}'_s = A\mathbf{t}_s, \tag{1}$$

where $\mathbf{r}'_s$ is the population average of the eight-dimensional Raman points of strain *s* (Fig. 1F), and $\mathbf{t}_s$ is the 2,628-dimensional transcriptome vector of the same strain containing the expression levels of 2,628 genes. A is the strain-independent matrix that maps the transcriptomes onto the Raman LDA space.

We conducted leave-one-out cross-validation (LOOCV) to verify the linear correspondence between Raman and transcriptome data (*33*). First, we estimated matrix A by using the Raman and transcriptome data of eight strains through partial least squares regression (PLS-R) (*47*). We then applied this estimated matrix to the transcriptome data of the excluded strain (test strain) to predict an eight-dimensional Raman point, represented as $\hat{\mathbf{r}}'_s$, which was then compared to the actual population average point, represented as $\mathbf{r}'_s$ (Fig. 2A). We repeated this process by excluding one strain at a time and evaluated the overall estimation error using predicted residual error sum of squares (PRESS), calculated as $\mathrm{PRESS}_{\mathbf{r}'} = \sum_s \|\hat{\mathbf{r}}'_s - \mathbf{r}_s\|^2$. A small PRESS value validates the assumption of linear correspondence between the Raman and transcriptome data.

We quantitatively evaluated the resulting $\mathrm{PRESS}_{\mathbf{r}'}$ value through a permutation test (*33*). We created 1,000 datasets in which we randomly permuted the correspondences between the Raman and transcriptome data, calculated the $\mathrm{PRESS}_{\mathbf{r}'}$ values for these false datasets, and compared them with the original $\mathrm{PRESS}_{\mathbf{r}'}$ value. The result confirmed a significantly small

overall estimation error ($\mathrm{PRESS}_{\mathbf{r}'} = 4.46$, $p = 9.99 \times 10^{-4}$, permutation test, Fig. 2B). This result therefore suggests that the observed Raman spectral differences in the LDA space can be explained by the differences in the gene expression profiles of these MRSA strains.

Global transcriptome profiles can be inferred from low-dimensional Raman data. Next, we asked whether transcriptome profiles of different strains could be inferred from their low-dimensional Raman spectra in LDA space. To address this question, we examined another linear correspondence

$$\mathbf{t}_s = B \begin{pmatrix} 1 \\ \mathbf{r}'_s \end{pmatrix}, \tag{2}$$

where B is the strain-independent coefficient matrix that converts the Raman spectral points of each strain in the LDA space into 2,628-dimensional transcriptome profile data (*45*). We again conducted LOOCV to verify this linear correspondence. We excluded the Raman and transcriptome data of one strain and estimated the coefficient matrix B using ordinary linear regression with the data of the other eight strains (see Materials and Methods). Then, we predicted the transcriptome profile of the excluded strain from its Raman spectrum $\mathbf{r}'_s$ as $\hat{\mathbf{t}}_s$ from Eq. 2.

Noticing that transcriptome profiles of different strains are also strongly correlated (Fig. S2), we calculated the PRESS for the transcriptome data as $\mathrm{PRESS}_{\mathbf{t}} = \sum_s \|\hat{\mathbf{t}}_s - \mathbf{t}_s\|^2$ to quantify the overall estimation error. A permutation test with randomly permuted 1,000 datasets found a significantly small overall estimation error ($\mathrm{PRESS}_{\mathbf{t}} = 1.09 \times 10^4$, $p = 9.99 \times 10^{-4}$, Fig. 3C). We also quantified the overall estimation error using other distance metrics and consistently found significantly small estimation errors (Table S1). Therefore, differences in the Raman spectra of the MRSA strains contain the information that allows the prediction of global gene expression differences of these strains. Note that the transcriptome inference was performed using the low-dimensional LDA Raman data, which correlated with vancomycin MIC levels; the high precision of the Raman-transcriptome correspondence suggests a constraint on the

global transcriptomic changes associated with the vancomycin resistance of the genotypically distinct MRSA strains.

Raman-based representations of transcriptome components can identify gene groups linked to vancomycin resistance and virulence The linear correspondence between transcriptome profiles and Raman spectra in the LDA space (Eq. 2 and Fig. 3) allows us to analyze the dependence of each gene on different LDA axes (*45*). Importantly, the major LDA axes are strongly correlated with different phenotypic features, such as vancomycin resistance levels and strain origin (clinically isolated or laboratory-constructed, Fig. 1D, 1F and 1G). Thus, the dependence of each gene on the LDA axes may help us to identify groups of genes associated with different phenotypes.

From Eq. 2, the expression level of gene i in strain s is expressed as

$$t_{s,i} = b_{i0} + \sum_{j=1}^{N-1} b_{ij} r'_{s,j}, \tag{3}$$

where $t_{s,i}$ is the i-th element of $\mathbf{t}_s$, $\begin{pmatrix} b_{i0} & b_{i1} & \cdots & b_{iN-1} \end{pmatrix}$ is the i-th row of B, $r'_{s,j}$ is the j-th element of the Raman LDA vector $\mathbf{r}'_s$, and $N = 9$ is the number of strains in the analysis. Since b_{i0} represents the basal expression level of each gene, relative change in the expression level of gene i in strain s is

$$\frac{t_{s,i} - b_{i0}}{b_{i0}} = \sum_{j=1}^{N-1} \left(\frac{b_{ij}}{b_{i0}} \right) r'_{s,j}. \tag{4}$$

Therefore, the quantity $\beta_i^{\mathrm{LDA}j} := b_{ij}/b_{i0}$ determines the dependence of gene i on the j-th LDA axis.

Figure 4 shows the plot of the first and second normalized coefficients $\left(\beta_i^{\mathrm{LDA1}}, \beta_i^{\mathrm{LDA2}}\right)$ for all genes in the transcriptome data. Most genes are distributed near the center, but we can find a number of genes deviating significantly from the center. Since the LDA1 axis is linked to the vancomycin resistance, we first examined the expression patterns of the genes with high β_i^{LDA1}

values (Table 1) and found that all of these genes are highly expressed in the VISA strains (Fig. 4). Conversely, genes with low β_i^{LDA1} values are highly expressed in the vancomycin-susceptible strains but suppressed in the VISA strains (Fig. 4). Among the genes with high β_i^{LDA1} values, those with low β_i^{LDA2} values are exclusively expressed in the clinically isolated strain Mu50, while those with high β_i^{LDA2} values are exclusively expressed in the laboratory-constructed strains ΔIP4 and ΔIP5 (Fig. 4). Genes with the β_i^{LDA2} values close to zero are commonly expressed in both clinically isolated and laboratory-constructed strains (Fig. 4). These expression patterns correspond to the Raman spectral patterns in the LDA space (Fig. 1D and 1F). Thus, representing the transcriptome components using the low-dimensional Raman LDA axes allows us to distinguish genes linked with different phenotypes by corresponding Raman positions.

Some of the genes identified as having high β_i^{LDA1} values were previously suggested to contribute to increased vancomycin resistance and virulence in *S. aureus*. Four genes—*vraD*, *vraE*, SAS091, and SA2097—were listed as highly expressed in both clinically isolated and laboratory-constructed VISA strains (Table 1). *vraD*, *vraE*, and SAS091 are ABC transporter genes that are known to be upregulated in VISA strains (*48, 49*). SA2097 encodes a peptidoglycan hydrolase component (*50, 51*). *S. aureus* possesses at least 18 different peptidoglycan hydrolases that facilitate growth and division (*50*). Since vancomycin targets the final steps of peptidoglycan and cell wall biosynthesis, high expression of SA2097 may enable the *S. aureus* cell to circumvent the action of the drug and survive. While the functions of these genes have not been fully characterized, their co-expression may be considered as a strong indicator of VISA-level vancomycin resistance.

Four of the five genes that are highly expressed exclusively in the clinically isolated VISA strain Mu50 are related to capsular polysaccharide synthesis (Table 1). It has been suggested that these genes enhance virulence by enabling *S. aureus* cells to withstand phagocytosis and

killing by polymorphonuclear phagocytes (*52–54*). Thus, while the upregulation of these genes may not directly contribute to vancomycin non-susceptibility, it may be indispensable for pathogenesis.

The genes that are highly expressed exclusively in laboratory-constructed VISA strains (*ilvH*, *leuA*, and *ilvC*) are all in the same operon and are involved in the biosynthesis of branched-chain amino acids (BCAAs) (*55*). Thus, the coordinated upregulation of these genes may alter BCAA levels in these strains. BCAAs act as signaling molecules that can affect a number of metabolic and virulence genes by binding to the global regulatory protein CodY (*56*). However, the relevance of these changes to vancomycin non-susceptibility and virulence in these VISA strains remains unclear.

The clinically isolated hVISA Mu3 strain is physiologically heterogeneous. Previous research has demonstrated that the hVISA Mu3 strain contains subpopulations that have the same level of resistance to vancomycin as VISA (*5*). This suggests that Mu3 is physiologically more diverse among clonal cells than those of N315 and Mu50.

In line with this picture, we noticed that the Raman spectra of the Mu3 strain were more heterogeneous among individual cells than those of the clinically isolated VSSA and VISA strains. We evaluated the standard deviations of normalized cellular Raman spectra at different wavenumbers among the clinically isolated MRSA strains (N315, Mu3, and Mu50), finding that the variations were consistently larger in Mu3 than those of N315 and Mu50 (Fig. 5A). This result suggests that the Mu3 cells are more heterogeneous in their molecular composition at the single-cell level.

To ask whether the heterogeneity recognized by their Raman spectra is physiologically relevant, we observed the growth and microcolony formation of the three clinically isolated strains on the BHI agar by time-lapse microscopy (Movies S1–S3). As expected, the growth curves of

Mu3 were more heterogeneous than those of N315 and Mu50 among the microcolonies grown from different individual cells (Fig. 5B–5D). The coefficient of variation (CV) of the growth rate of Mu3 was 36%, significantly higher than those of N315 and Mu50, which were both 18% (Fig. 5E–5G). The mean growth rate of Mu50 was smaller than that of N315 (0.47 h^{-1} for N315 and 0.26 h^{-1} for Mu50), which corroborates the slow-growing trait of the VISA strain. Although the median of the growth rate of Mu3 is close to that of N315, Mu3's growth rate distribution is skewed with a heavy left tail, indicating the presence of subpopulations in the Mu3 strain growing as slowly as the Mu50 strain (Fig. 5E–5G).

The fact that cell-to-cell heterogeneity of Mu3 is consistently larger in both Raman spectra and growth supports that this clinically isolated hVISA strain is physiologically more diverse among genetically identical cells. Furthermore, the results suggest that phenotypic heterogeneity in cellular population can be interrogated spectroscopically by evaluating their Raman spectral heterogeneity.

Discussion

Vancomycin resistance undermines the treatment efficacy of MRSA infection. Inappropriate administration of vancomycin could result in generating and enhancing the resistance of bacteria. Therefore, it is important to quickly detect vancomycin resistance to make the right choices in drug regimens and prevent the emergence and spread of vancomycin-nonsusceptible MRSA strains.

This study shows that the global transcriptome profiles and vancomycin resistance levels of the MRSA strains can be identified by their Raman spectral patterns (Fig. 1G, 3A, and 3B). Cellular Raman spectra can be obtained by exposing cells to light without the need for staining or tagging. Furthermore, unlike the MALDI-TOF MS-based method, it is not necessary to culture cells to target bacterial cells for Raman measurements, as long as they are recognizable

in specimens. As a result, Raman microscopy has significant potential for quickly uncovering antibiotic resistance of bacteria in patients.

Previous studies have shown that cellular Raman spectra obtained under different culture conditions can be used to infer transcriptomic and proteomic differences (*33, 44, 45*). The present study shows that transcriptomic differences among the MRSA strains with different genotypes under identical culture conditions can also be inferred from their Raman spectral patterns (Fig. 3A and 3B). Therefore, Raman-based inference of gene expression profiles applies not only to external environmental differences but also to internal genotypic differences.

In this study, we introduced a method for representing transcriptome components using low-dimensional Raman LDA axes based on Raman-transcriptome linear correspondence (Fig. 4). This method is particularly useful when the Raman LDA axes are closely associated with particular phenotypic characteristics. In our study, the major LDA axes (LDA1 and LDA2) were related to vancomycin susceptibility and virulence (Fig. 1D and 1F). It should be noted that the correspondence between the LDA axes and the phenotypes depends on the presence of systematic differences in their molecular profiles associated with phenotypes. In other words, proximity of different samples along a Raman LDA axis corresponding to some phenotype indicates the presence of shared molecular profile changes related to the phenotype across those samples.

This transcriptome representation facilitates identifying gene groups whose expression patterns correlate with different phenotypes. For instance, we identified several genes expressed in both clinically isolated and laboratory-constructed VISA strains (Table 1 and Fig. 4). Some of these genes have previously been shown to exhibit increased expression levels in VISA strains (*48,49*). However, their role in enhancing vancomycin resistance has not been characterized in detail. Furthermore, some of the identified genes are completely functionally unknown. Characterizing the roles of those genes is an important topic of future studies. Additionally, it

would be important to investigate whether only some of these genes contribute to resistance or if the co-expression of these multiple genes is indispensable.

The Mu3 strain was found to be phenotypically heterogeneous in both growth and Raman spectra. This result suggests that cellular Raman spectra can interrogate cell-to-cell heterogeneity within a cellular population. Such phenotypic heterogeneity could be linked to resistance of a subpopulation in the Mu3 strain under vancomycin exposures.

While our study focused on vancomycin resistance in MRSA, the broader applicability of this method to detect drug resistance in other bacterial and fungal pathogens represents an exciting avenue for future research. Exploring the adaptability of Raman spectroscopy to various pathogenic strains could significantly enhance our capacity to combat antibiotic resistance.

MATERIALS AND METHODS

Bacterial strains. Nine *S. aureus* strains were used in this study. Three strains, N315, Mu3, and Mu50, are clinical isolates (*5*). The other six strains, ΔIP, ΔIP1, ΔIP2, ΔIP3, ΔIP4, and ΔIP5, were constructed from N315 by introducing mutations or by selection under rifampin exposures (*16*). The vancomycin MIC values of the strains were determined by Etest in (*16*). The vancomycin resistant phenotypes of the strains were determined in (*16*) on the basis of MIC values and population curves on a vancomycin concentration-CFU plane: VSSA for N315 and ΔIP; hVISA for Mu3, ΔIP1, and ΔIP2; and VISA for Mu50, ΔIP3, ΔIP4, and ΔIP5.

Raman measurement and preprocessing of Raman signals. To prepare samples for Raman measurements, *S. aureus* cells were inoculated from a glycerol stock into $3\,\mathrm{mL}$ BHI medium (Eiken Chemical Co., Ltd.) and cultured with shaking (Taitec BR-32FP, $200\,\mathrm{r\,min^{-1}}$) at $37\,^{\circ}\mathrm{C}$ overnight. A $30\,\mu\mathrm{L}$ overnight culture was inoculated into $3\,\mathrm{mL}$ BHI medium and grown to an exponential phase (OD_{600} = 2.0–2.5) with shaking at $37\,^{\circ}\mathrm{C}$. $1\,\mathrm{mL}$ of cell suspension was taken in an eppentube, and the cells were washed with phosphate-buffered saline (PBS) twice. The cells were fixed with $4\,\%$ (w/w) paraformaldehyde for $15\,\mathrm{min}$ at room temperature. The fixed cells were washed twice with $0.1\,\mathrm{M}$ glycine. After substituting the liquid by $1\,\mathrm{mL}$ PBS, the cell suspensions were stored at $4\,^{\circ}\mathrm{C}$ until Raman measurement (max. one day).

Immediately before the Raman measurement, the liquid of cell suspension was substituted by $1\,\mathrm{mL}$ distilled water. $5\,\mu\mathrm{L}$ of the cell suspension was placed on a synthetic quartz slide glass (TOSHIN RIKO Co., LTD.) and dried.

Raman spectra of the cells were measured using a custom-built Raman microscope (STR-Raman, AIRIX corp.) with $532\,\mathrm{nm}$ excitation wavelength (Gem 532, Laser Quantum). Laser power at the sample plane was approximately $17.5\,\mathrm{mW}$. Raman spectra were obtained from single cells with exposure time of $5\,\mathrm{s}$. For each strain, 16 cells were measured per replicate.

Three biological replicates were measured on different days. Consequently, we obtained Raman spectra from 48 cells for each strain. Background Raman spectra were also measured in regions adjacent to target cells. Raman signals were detected by an sCMOS camera (ORCA-Flash4.0 V2, Hamamatsu Photonics) water-cooled at $15\,^{\circ}\mathrm{C}$. The optical setup was the same as that used for the Raman measurement of *Escherichia coli* in (*33*).

An sCMOS-specific noise reduction filter developed in (*33*) was applied to the obtained images. Then, the region where the spectra were recorded was cropped, and pixel counts were summed along the direction perpendicular to the wavenumber. Next, background Raman spectra were subtracted from single-cell Raman spectra, and the region of 632 to 1862 cm^{-1} was cropped. Subsequently, each spectrum was normalized by subtracting the average and dividing it by the standard deviation.

Dimension reduction of cellular Raman spectra. Principal component analysis (PCA) was first applied to the normalized spectra to reduce noise. Principal components that explained $98\,\%$ of the variance were used. After PCA, linear discriminant analysis (LDA) was applied. LDA is a supervised method, which calculates the most discriminatory bases by maximizing the ratio of the between-group (in the case of this research, between-strain) variances to the sum of within-group (within-strain) variances. The dimension was reduced to $n - 1$ by LDA, where $n\ (= 9)$ was the number of groups (strains).

Linearly estimating dimension-reduced Raman spectra from transcriptome profiles. Since the number of genes is much larger than that of the strains, and in general, expression patterns of many genes are highly correlated, a linear regression method called partial least squares regression (PLS-R) was employed. The transcriptomes of the nine *S. aureus* strains were obtained from the public database (NCBI GEO, accession number GSE43643, GSE42218, and GSE50826). The dimensions of the transcriptomes were reduced to $n-3$ in PLS-R. To verify the

linearity between the LDA Raman spectra and transcriptomes, leave-one-out cross-validation (LOOCV) was performed. We calculated the predicted residual error sum of squares (PRESS) as estimation errors. Statistical significance of the estimation was tested with permutation test. See (*33*) for the detail of LOOCV and permutation test.

Linearly inferring global transcriptome profiles from dimension-reduced Raman sepctra. To investigate whether global transcriptome profiles can be inferred from the low-dimensional LDA Raman spectra, ordinary linear regression was conducted. We verified the linear correspondence by LOOCV again. In the case of LOOCV, the regression is underdetermined, and the minimum norm solution among all least-squares solutions was adopted. Overall estimation errors were calculated with some distance metrics including the predicted residual error sum of squares (PRESS). Statistical significance of the inference was examined with permutation test.

Mapping transcriptome profiles on dimension-reduced Raman space. Linear transformation between the transcriptome profiles and the dimension-reduced LDA Raman profiles is

$$[\mathbf{t}_1 \cdots \mathbf{t}_n] = B \left[\begin{pmatrix} 1 \\ \mathbf{r}'_1 \end{pmatrix} \cdots \begin{pmatrix} 1 \\ \mathbf{r}'_n \end{pmatrix} \right]. \tag{5}$$

As $\left[\begin{pmatrix} 1 \\ \mathbf{r}'_1 \end{pmatrix} \cdots \begin{pmatrix} 1 \\ \mathbf{r}'_n \end{pmatrix} \right]$ is full rank, the strain-independent coefficient matrix B can be obtained by

$$B = [\mathbf{t}_1 \cdots \mathbf{t}_n] \left[\begin{pmatrix} 1 \\ \mathbf{r}'_1 \end{pmatrix} \cdots \begin{pmatrix} 1 \\ \mathbf{r}'_n \end{pmatrix} \right]^{-1}. \tag{6}$$

Writing the j-th column of B as $\mathbf{b}_j$, $B = [\mathbf{b}_0 \cdots \mathbf{b}_{n-1}]$. From the viewpoint of linear regression, $\mathbf{b}_0$ is the constant terms, and $\mathbf{b}_j$ $(1 \leq j \leq n-1)$ is the coefficients for the j-th LDA dimension. The normalized coefficients discussed in the main text is the columns of $\mathrm{diag}\,(\mathbf{b}_0)^{-1} B$, where $\mathrm{diag}\,(\mathbf{x}) = (\delta_{ij} x_i)$ for a vector $\mathbf{x} = (x_i)$. For instance, the first and second axes in Fig. 4 are the second and third axes of $\mathrm{diag}\,(\mathbf{b}_0)^{-1} B$, namely, $\mathrm{diag}\,(\mathbf{b}_0)^{-1} \mathbf{b}_1$ and $\mathrm{diag}\,(\mathbf{b}_0)^{-1} \mathbf{b}_2$. See (*45*)

for the detailed mathematical explanation of the phenotype-expression pattern correspondence between the LDA Raman space and the normalized coefficient gene space.

Heterogeneity analysis of Raman spectra of the clinically isolated strains. To make the intensities of the average spectra approximately equal among the three strains, three outlier single-cell Raman spectra were removed for each strain. Then, the standard deviation of the remaining 45 normalized spectra was calculated at each wavenumber for each strain.

Time-lapse microscopy. To prepare samples for time-lapse microscopy, *S. aureus* cells were inoculated from a glycerol stock into $2\,\mathrm{mL}$ BHI medium and cultured with shaking ($200\,\mathrm{r\,min^{-1}}$) at $37\,^{\circ}\mathrm{C}$ overnight. A $15\,\mu\mathrm{L}$ overnight culture was inoculated into $3\,\mathrm{mL}$ BHI medium and grown to exponential phase (OD_{600} = 2.4–2.5) with shaking at $37\,^{\circ}\mathrm{C}$. $1\,\mu\mathrm{L}$ of cell suspension was taken on a coverslip (Neo No.1, $24\,\mathrm{mm}\times60\,\mathrm{mm}$, Matsunami) and covered by a BHI agarose pad ($20\,\mathrm{mm}\times25\,\mathrm{mm}\times7\,\mathrm{mm}$ (thickness)) solidified with $1.5\,\%$ low-melt agarose (GeneMate LowMelt Agarose, BioExpress). The assembled sample was placed on a microscope XY stage and covered by a transparent plastic dish to prevent the agarose from drying.

We used Nikon Eclipse Ti-E inverted microscope equipped with digital sCMOS camera (ORCA-Flash4.0 V3, Hamamatsu Photonics), phase-contrast objective (Plan Apo λ, 100×, NA1.45). We obtained phase-contrast images from 50 XY positions with $1\,\mathrm{min}$ interval in each experiment.

The acquired time-lapse images were analyzed using ImageJ (*57*). We applied Mexican Hat Filter to find outlines of microclonies of *S. aureus* and measured the areas to quantify the growth.

Heterogeneity analysis of microcolony growth of the clinically isolated strains. Growth rates were calculated by taking logarithm to 2 of relative colony sizes at $t = 3\,\mathrm{h}$ to those at $t =$

$0\,\mathrm{h}$ and dividing them by $3\,\mathrm{h}$.

Data analysis. All the data analyses in this study were conducted using MATLAB (R2019a, R2023b and R2024b).

Declaration of interests

YK is the president and representative director of KatayamaKikai Co., Ltd. YK holds 50% stock ownership of KatayamaKikai Co., Ltd. YW and KJKK are listed as inventors of patents (JP6993682 and US10,379,052 B2) filed by The University of Tokyo. The other authors declare no competing interests.

Acknowledgments

We thank Tetsuo Yamaguchi for reading the manuscript and providing comments. This work was supported by JST CREST Grant Number JPMJCR1927 (Y.W.); JST ERATO Grant Number JPMJER1902 (Y.W.); JSPS KAKENHI Grant Numbers 19J22448 (K.F.K.); and Takeda Science Foundation Award (Y.K. and Y.W.).

References

1. H. F. Chambers, F. R. Deleo, *Nature Reviews Microbiology* **7**, 629 (2009).

2. H. Grundmann, M. A. de Sousa, J. Boyce, E. Tiemersma, *The Lancet* **368**, 874 (2006).

3. Y. Cong, S. Yang, X. Rao, *Journal of Advanced Research* **21**, 169 (2020).

4. K. Hiramatsu, *et al.*, *Journal of Antimicrobial Chemotherapy* **40**, 135 (1997).

5. K. Hiramatsu, *et al.*, *The Lancet* **350**, 1670 (1997).

6. S. Chang, *et al.*, *New England Journal of Medicine* **348**, 1342 (2003).

7. L. M. Weigel, *et al.*, *Science* **302**, 1569 (2003).

8. S. Zhang, X. Sun, W. Chang, Y. Dai, X. Ma, *PLOS ONE* **10**, e0136082 (2015).

9. A. Shariati, *et al.*, *Scientific Reports* **10**, 1 (2020).

10. P. G. Charles, P. B. Ward, P. D. Johnson, B. P. Howden, M. L. Grayson, *Clinical Infectious Diseases* **38**, 448 (2004).

11. H. M. Neoh, *et al.*, *Annals of Clinical Microbiology and Antimicrobials* **6**, 1 (2007).

12. Y. Maor, *et al.*, *Journal of Clinical Microbiology* **45**, 1511 (2007).

13. R. K. Fong, J. Low, T. H. Koh, A. Kurup, *European Journal of Clinical Microbiology and Infectious Diseases* **28**, 983 (2009).

14. Y. Katayama, H. Murakami-Kuroda, L. Cui, K. Hiramatsu, *Antimicrobial Agents and Chemotherapy* **53**, 3190 (2009).

15. K. Hiramatsu, *et al.*, *Journal of Infection and Chemotherapy* **20**, 593 (2014).

16. Y. Katayama, M. Sekine, T. Hishinuma, Y. Aiba, K. Hiramatsu, *Antimicrobial Agents and Chemotherapy* **60**, 3730 (2016).

17. T. Ito, *et al.*, *Antimicrobial Agents and Chemotherapy* **45**, 1323 (2001).

18. K. Hiramatsu, L. Cui, M. Kuroda, T. Ito, *Trends in Microbiology* **9**, 486 (2001).

19. H. Hanaki, *et al.*, *Journal of Antimicrobial Chemotherapy* **42**, 199 (1998).

20. K. Sieradzki, A. Tomasz, *Journal of Bacteriology* **185**, 7103 (2003).

21. Y. Aiba, *et al.*, *Antimicrobial Agents and Chemotherapy* **57**, 4861 (2013).

22. S. Gardete, A. Tomasz, *The Journal of Clinical Investigation* **124**, 2836 (2014).

23. Q. Hu, H. Peng, X. Rao, *Frontiers in Microbiology* **7**, 221405 (2016).

24. L. Cui, H. Murakami, K. Kuwahara-Arai, H. Hanaki, K. Hiramatsu, *Antimicrobial Agents and Chemotherapy* **44**, 2276 (2000).

25. L. Cui, *et al.*, *Journal of Clinical Microbiology* **41**, 5 (2003).

26. M. Saito, *et al.*, *Antimicrobial agents and chemotherapy* **58**, 5024 (2014).

27. A. Wieser, L. Schneider, J. Jung, S. Schubert, *Applied Microbiology and Biotechnology* **93**, 965 (2012).

28. C. A. Mather, B. J. Werth, S. Sivagnanam, D. J. SenGupta, S. M. Butler-Wu, *Journal of Clinical Microbiology* **54**, 883 (2016).

29. K. Asakura, *et al.*, *PLoS ONE* **13**, 1 (2018).

30. I. W. Schie, T. Huser, *Applied Spectroscopy* **67**, 813 (2013).

31. H. J. Butler, *et al.*, *Nature Protocols* **11**, 664 (2016).

32. R. Smith, K. L. Wright, L. Ashton, *Analyst* **141**, 3590 (2016).

33. K. J. Kobayashi-Kirschvink, *et al.*, *Cell systems* **7**, 104 (2018).

34. A. Germond, *et al.*, *Communications Biology* **1**, 85 (2018).

35. K. S. Lee, *et al.*, *Nature Reviews Methods Primers* **1**, 80 (2021).

36. E. Segal, *et al.*, *Nature Genetics* **34**, 166 (2003).

37. L. Keren, *et al.*, *Molecular systems biology* **9**, 701 (2013).

38. K. Kaneko, C. Furusawa, T. Yomo, *Physical Review X* **5**, 011014 (2015).

39. S. Hui, *et al.*, *Molecular Systems Biology* **11**, 784 (2015).

40. G. Heimberg, R. Bhatnagar, H. El-Samad, M. Thomson, *Cell Systems* **2**, 239 (2016).

41. S. Biswas, *et al.*, *Nature Communications* **8**, 15309 (2017).

42. K. Husain, A. Murugan, *Molecular Biology and Evolution* **37**, 2865 (2020).

43. T. U. Sato, K. Kaneko, *Physical Review Research* **2**, 013197 (2020).

44. K. J. Kobayashi-Kirschvink, *et al.*, *Nature Biotechnology* **42**, 1726 (2024).

45. K. F. Kamei, *et al.*, *eLife* **14**, RP101485 (2025).

46. R. Goodacre, *et al.*, *Microbiology* **144**, 1157 (1998).

47. P. S. Gromski, *et al.*, *Analytica Chimica Acta* **879**, 10 (2015).

48. H. Neoh, *et al.*, *Antimicrobial Agents and Chemotherapy* **52**, 45 (2008).

49. K. Hiramatsu, *et al.*, *Journal of Global Antimicrobial Resistance* **2**, 213 (2014).

50. M. Wang, G. Buist, J. M. van Dijl, *FEMS Microbiology Reviews* **46**, fuac025 (2022).

51. M. Nega, P. M. Tribelli, K. Hipp, M. Stahl, F. Götz, *Communications Biology* **3**, 695 (2020).

52. I. M. Nilsson, J. C. Lee, T. Bremell, C. Rydén, A. Tarkowski, *Infection and Immunity* **65**, 4216 (1997).

53. B. P. Howden, J. K. Davies, P. D. R. Johnson, T. P. Stinear, M. L. Grayson, *Clinical Microbiology Reviews* **23**, 99 (2010).

54. M. Rausch, *et al.*, *Nature Communications* **10**, 1404 (2019).

55. T. Lei, J. Yang, Y. Ji, *Virulence* **6**, 75 (2015).

56. J. C. Kaiser, *et al.*, *PLoS Genetics* **14** (2018).

57. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nature Methods* **9**, 671 (2012).

Figure legends

Figure 1 Single-cell Raman spectra of MRSA strains and their correlation with vancomycin resistance. **(A)** Nine MRSA strains analyzed in this study. Three strains are clinical isolates, while the other six are strains constructed in a laboratory (*16*). Vancomycin susceptibility of these nine strains are classified into VSSA, hVISA, and VISA. Groups to which the strains used in this study belong were judged by (*16*). Colors and markers in this figure are used in some of the other figures throughout this paper. **(B)** Schematic figure representing our Raman measurements. Raman spectra were obtained from single cells placed on a synthetic quartz slide glass by focusing laser light on each targeted cell. **(C)** Representative single-cell Raman spectra. One is from ΔIP (top), and the other is from N315 (bottom). **(D)** Cellular Raman spectra in LDA space. LDA reduces the dimensionality of the spectra to $8\ (= 9 - 1)$. Each point represents a spectrum from a single cell. Each ellipse shows the 95% concentration ellipse. Projection to the plane spanned by the first two axes (LDA1 and LDA2) is shown. **(E)** t-SNE visualization of the eight-dimensional LDA space. **(F)** The population average of the single-cell Raman spectra of each strain in the LDA space. The arrows indicate the changes of the averages along the strain construction and selection steps in (*16*). **(G)** Correlation between the positions of cellular Raman spectra along the LDA1 axis and vancomycin resistance levels measured by (*16*), in which MIC of N315 was not reported. Pearson correlation coefficient, $r = 0.81 \pm 0.15$.

Figure 2 Transcriptome profiles of different MRSA strains can explain their Raman spectral differences. **(A)** Estimation of Raman spectra in the LDA space from transcriptomes. Thick colored points represent Raman spectra estimated from the transcriptomes assuming the linear correspondence (Eq. 1). **(B)** Histogram of $\mathrm{PRESS}_{\mathbf{r}'}$ values of 1,000 randomly permuted data. The $\mathrm{PRESS}_{\mathbf{r}'}$ of the original dataset is 4.46. The p-value was 9.99×10^{-4} (no occurrence

of smaller $\mathrm{PRESS}_{\mathbf{r}'}$ was observed in the 1,000 permuted datasets).

Figure 3 Inference of global transcriptome profiles from cellular Raman spectra. (A and B) Estimation of transcriptomes from Raman spectra in the LDA space. In (A), grey points represent the measured RNA abundance (average of the replicate measurements; error bar, SE), which are sorted along the horizontal axis. Green points represent the RNA abundance estimated from Raman spectra. (B) is scatterplots between the measured and the estimated RNA abundances. Straight lines represent $x = y$. **(C)** Histogram of $\mathrm{PRESS}_{\mathbf{t}}$ values of 1,000 randomly permuted data. The $\mathrm{PRESS}_{\mathbf{t}}$ of the original dataset is 1.07×10^4. The p-value was 9.99×10^{-4} (there was no occurrence of smaller $\mathrm{PRESS}_{\mathbf{t}}$ in the 1,000 permuted datasets).

Figure 4 Mapping transcriptome components on dimension-reduced Raman space identifies genes associated with distinct phenotypes. All genes in the transcriptome data are represented on a plane where the coordinates of each gene i are its first and second normalized coefficients $(\beta_i^{\mathrm{LDA1}}, \beta_i^{\mathrm{LDA2}})$. Each gray dot corresponds to a gene. All genes with $\beta_i^{\mathrm{LDA1}} > 1.5$ and several representative genes with $\beta_i^{\mathrm{LDA1}} < -2.5$ are highlighted by red circles. Expression patterns of these genes are also shown (average of the replicate measurements; error bar, SE). Since normalized coefficients $\beta_i^{\mathrm{LDA}j}$ determines the dependence of gene i on the j-th LDA axis, the first and second axes (β^{LDA1} and β^{LDA2}) in this figure correspond to LDA1 and LDA2 axes, respectively (Fig. 1D and 1F). For example, Raman spectra of VISA strains are placed on the right side of the LDA1-LDA2 plane in Fig. 1D and 1F. Correspondingly, genes highly expressed in the VISA strains are placed on the right side of β^{LDA1}-β^{LDA2} plane in this figure.

Figure 5 Heterogeneity of clinically isolated hVISA strain Mu3. (A) Comparison of standard deviations of normalized single-cell Raman spectra among the clinically isolated VSSA, hVISA, and VISA strains. **(B–D)** Microcolony growth of the three clinically isolated strains

on BHI agar. Single lines represent relative size growths of different single colonies. **(E–G)** Growth rate distributions of the microcolonies on BHI agar shown in Fig. 5B–5D.

Table 1 List of genes with a high b_{i1}/b_{i0} value. Genes with a b_{i1}/b_{i0} value above 1.5, which are located on the right side and highlighted by red circles in Fig. 4, are listed.

Fig. 1

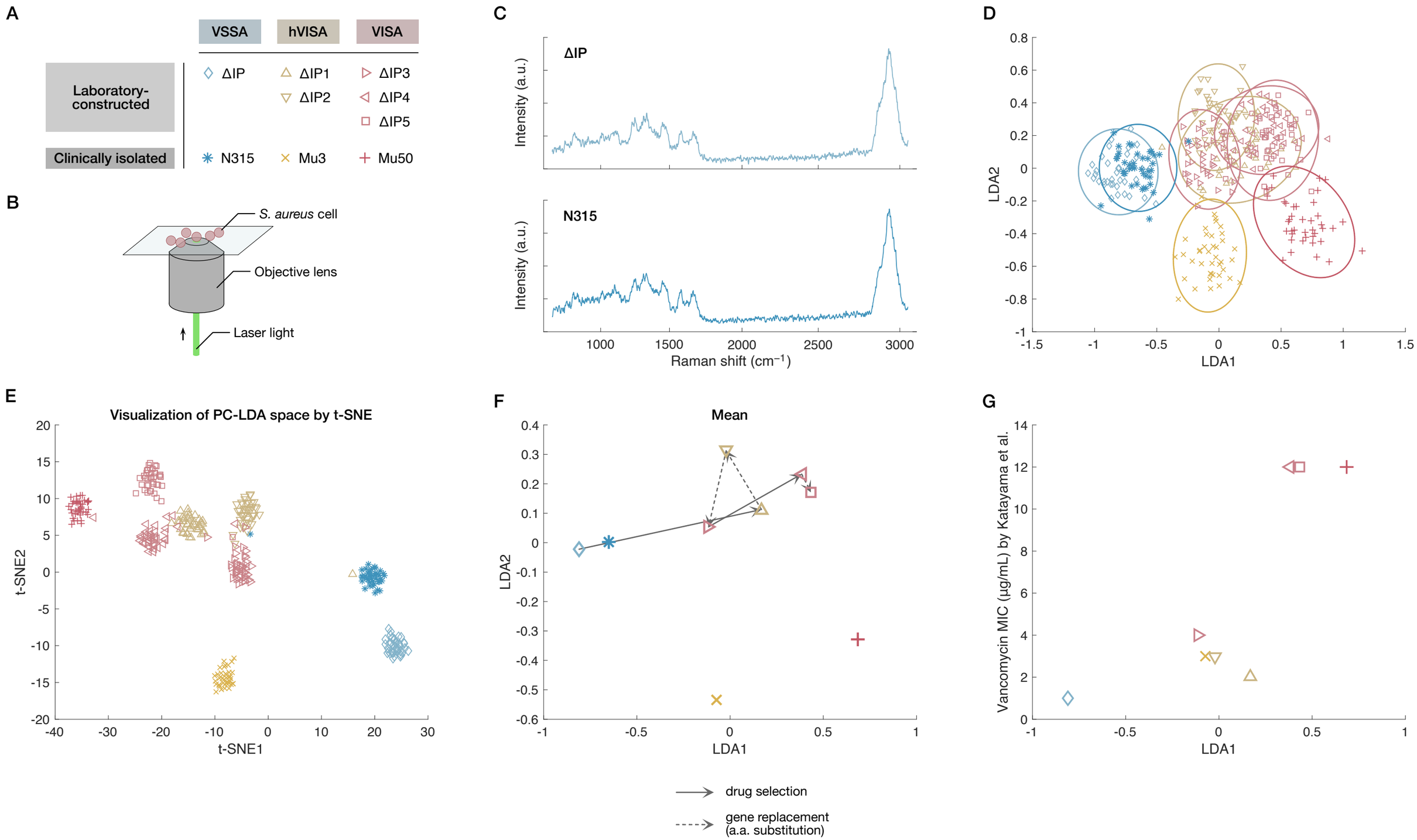
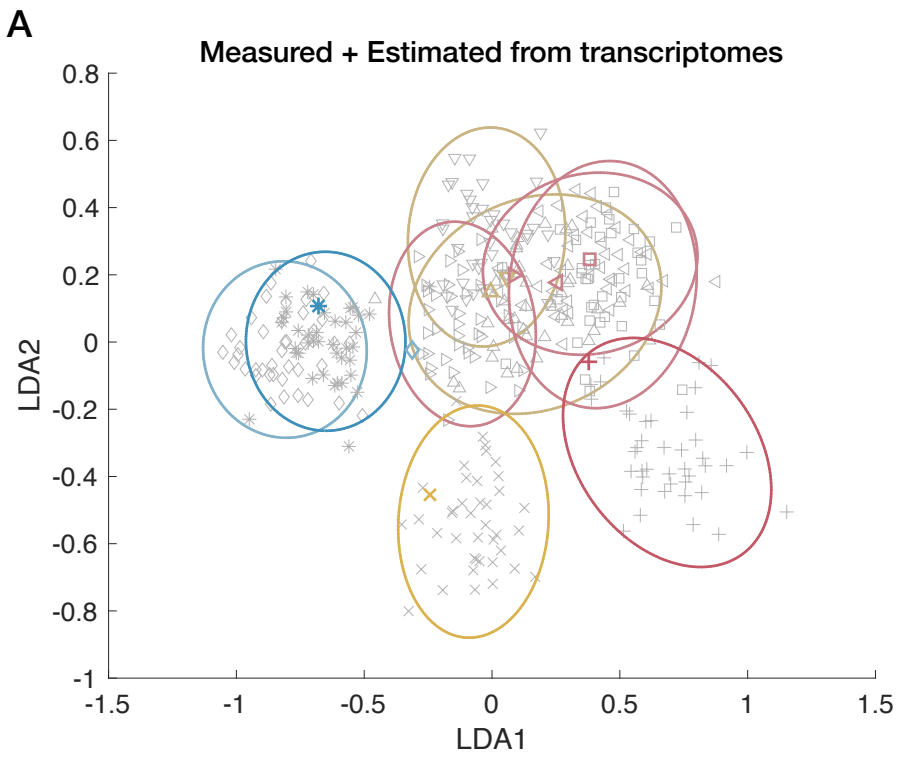
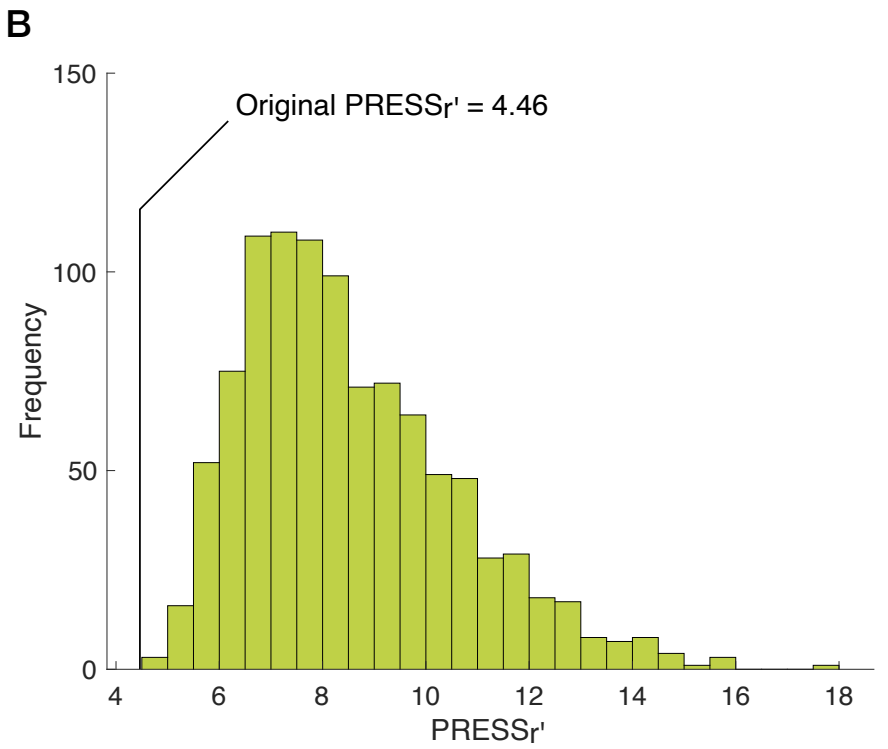

A
VSSA
hVISA
VISA
Laboratory-constructed
ΔIP
ΔIP1
ΔIP2
ΔIP3
ΔIP4
ΔIP5
Clinically isolated
N315
Mu3
Mu50
B
S. aureus cell
Objective lens
Laser light
C
ΔIP
N315
Intensity (a.u.)
Raman shift (cm⁻¹)
D
LDA1
LDA2
E
Visualization of PC-LDA space by t-SNE
t-SNE1
t-SNE2
F
Mean
LDA1
LDA2
G
Vancomycin MIC (μg/mL) by Katayama et al.
LDA1
drug selection
gene replacement
(a.a. substitution)

Fig. 2

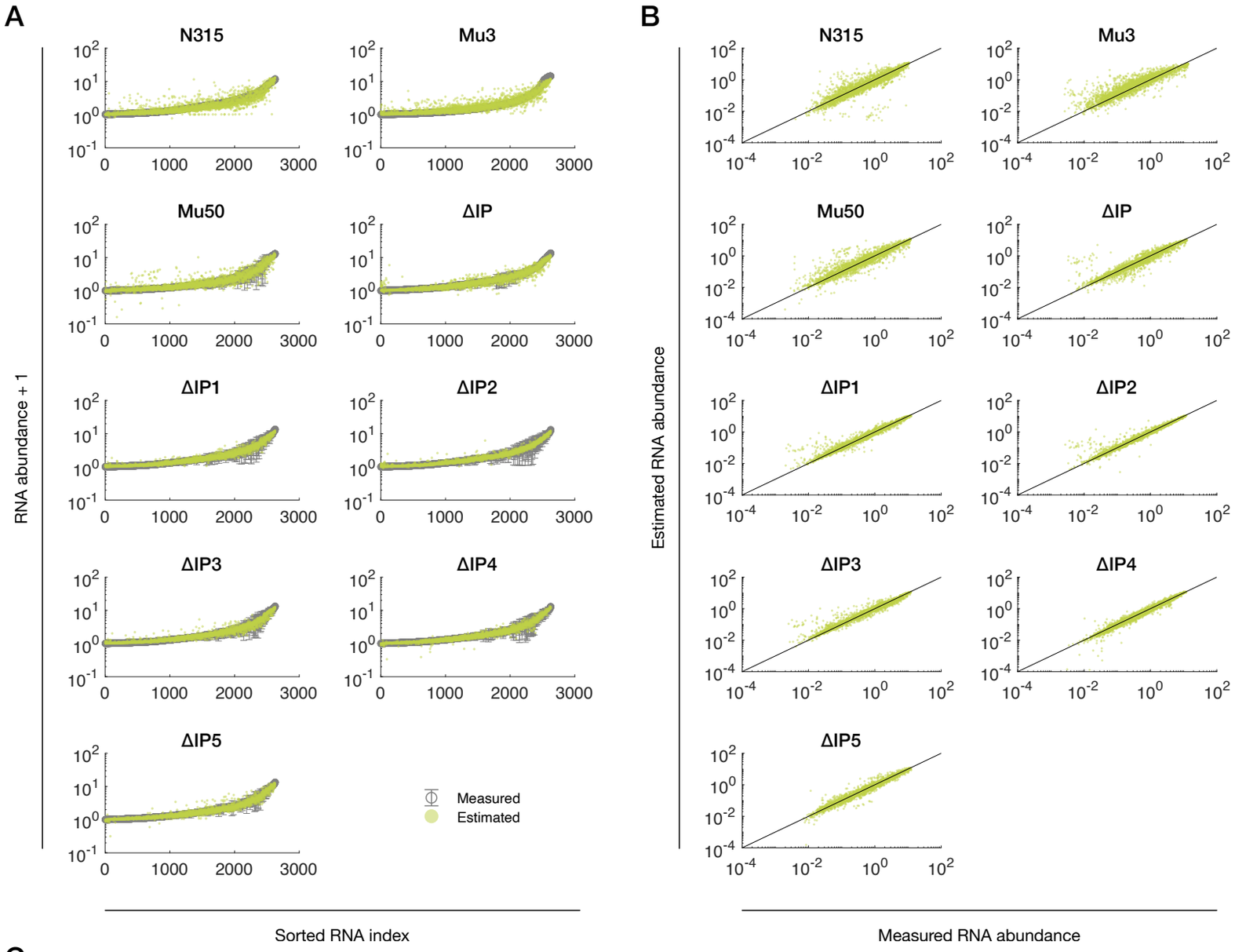




Gray points measured (same as Fig. 1C)

Colored points estimated from transcriptomes

Fig. 3


A
N315
Mu3
Mu50
ΔIP
ΔIP1
ΔIP2
ΔIP3
ΔIP4
ΔIP5
RNA abundance + 1
Sorted RNA index
Measured
Estimated
B
Estimated RNA abundance
Measured RNA abundance

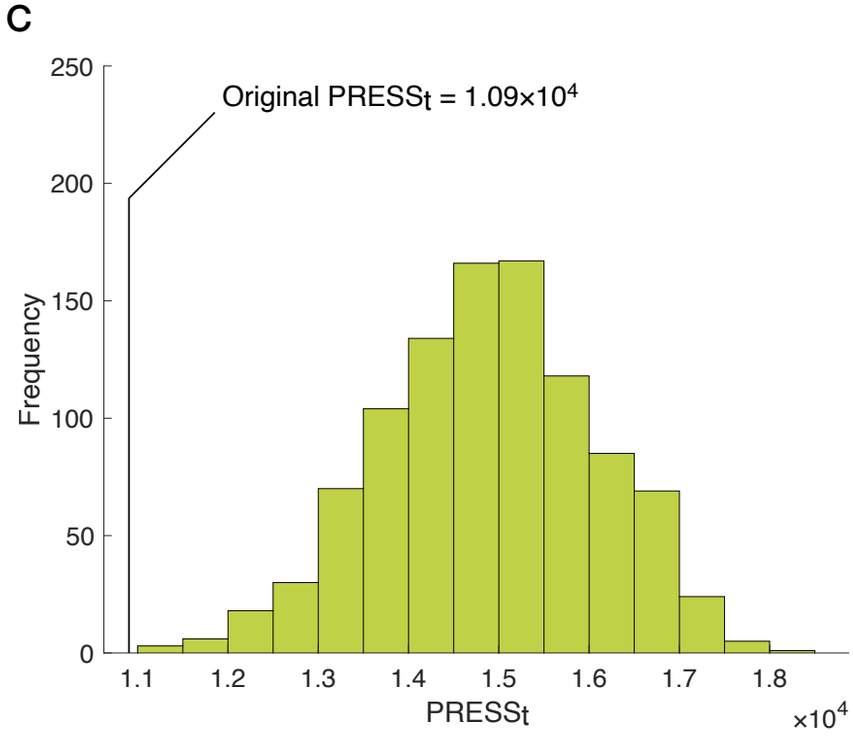

C
Original PRESSt = 1.09×10^4
Frequency
PRESSt
×10^4

Fig. 4

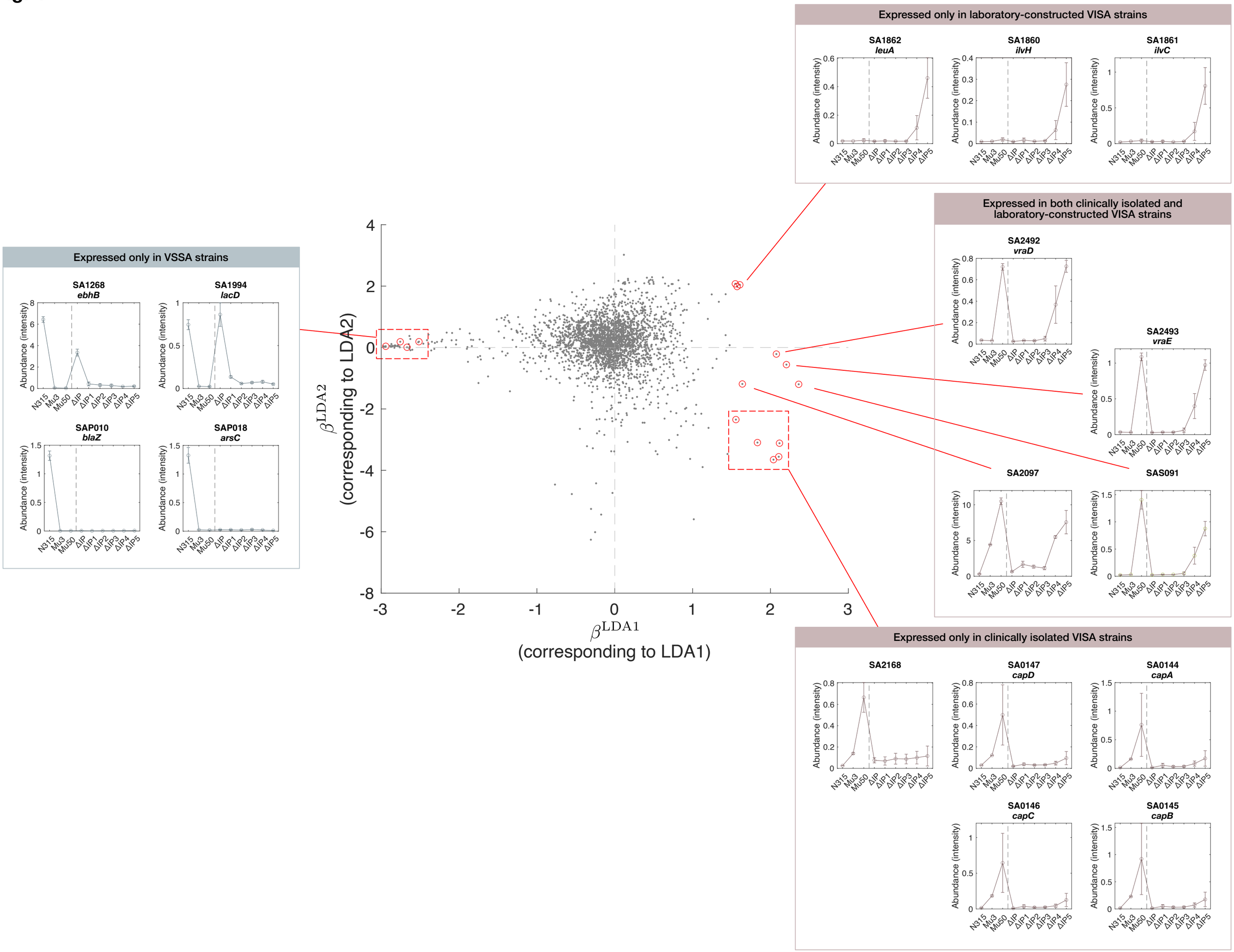
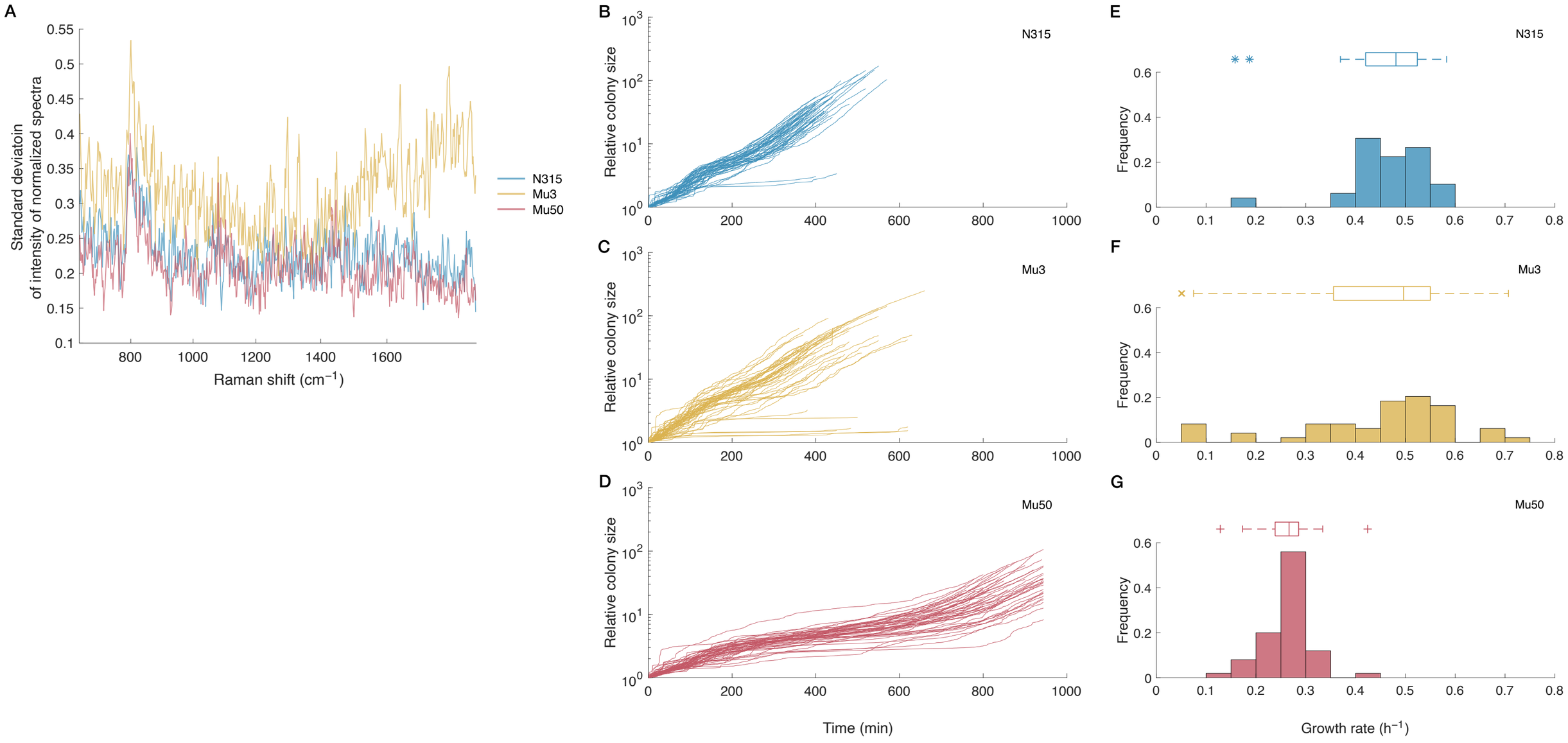

Expressed only in laboratory-constructed VISA strains
SA1862
leuA
SA1860
ilvH
SA1861
ilvC
Expressed only in VSSA strains
SA1268
ebhB
SA1994
lacD
SAP010
blaZ
SAP018
arsC
Expressed in both clinically isolated and laboratory-constructed VISA strains
SA2492
vraD
SA2493
vraE
SA2097
SAS091
Expressed only in clinically isolated VISA strains
SA2168
SA0147
capD
SA0144
capA
SA0146
capC
SA0145
capB
Abundance (intensity)
N315
Mu3
Mu50
ΔIP
ΔIP1
ΔIP2
ΔIP3
ΔIP4
ΔIP5
β^{LDA2}
(corresponding to LDA2)
β^{LDA1}
(corresponding to LDA1)

Fig. 5


A
Standard deviatoin of intensity of normalized spectra
Raman shift (cm^{-1})
N315
Mu3
Mu50
B
C
D
Relative colony size
Time (min)
E
F
G
Frequency
Growth rate (h^{-1})