Cell Reports, Volume *37*

Supplemental information

Spatiotemporal localization of proteins in mycobacteria

Junhao Zhu, Ian D. Wolf, Charles L. Dulberger, Harim I. Won, Jemila C. Kester, Julius A. Judd, Samantha E. Wirth, Ryan R. Clark, Yawei Li, Yuan Luo, Todd A. Gray, Joseph T. Wade, Keith M. Derbyshire, Sarah M. Fortune, and Eric J. Rubin

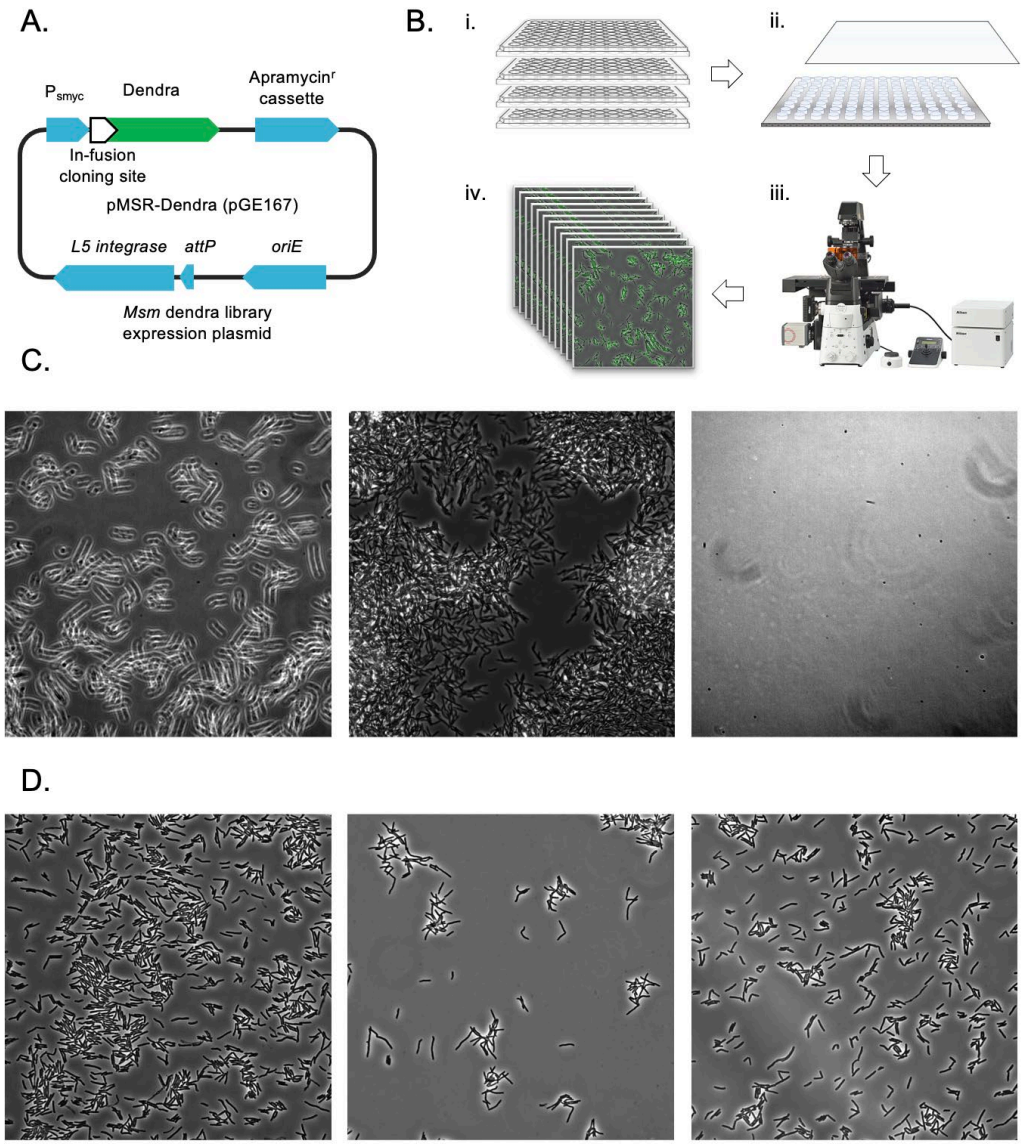
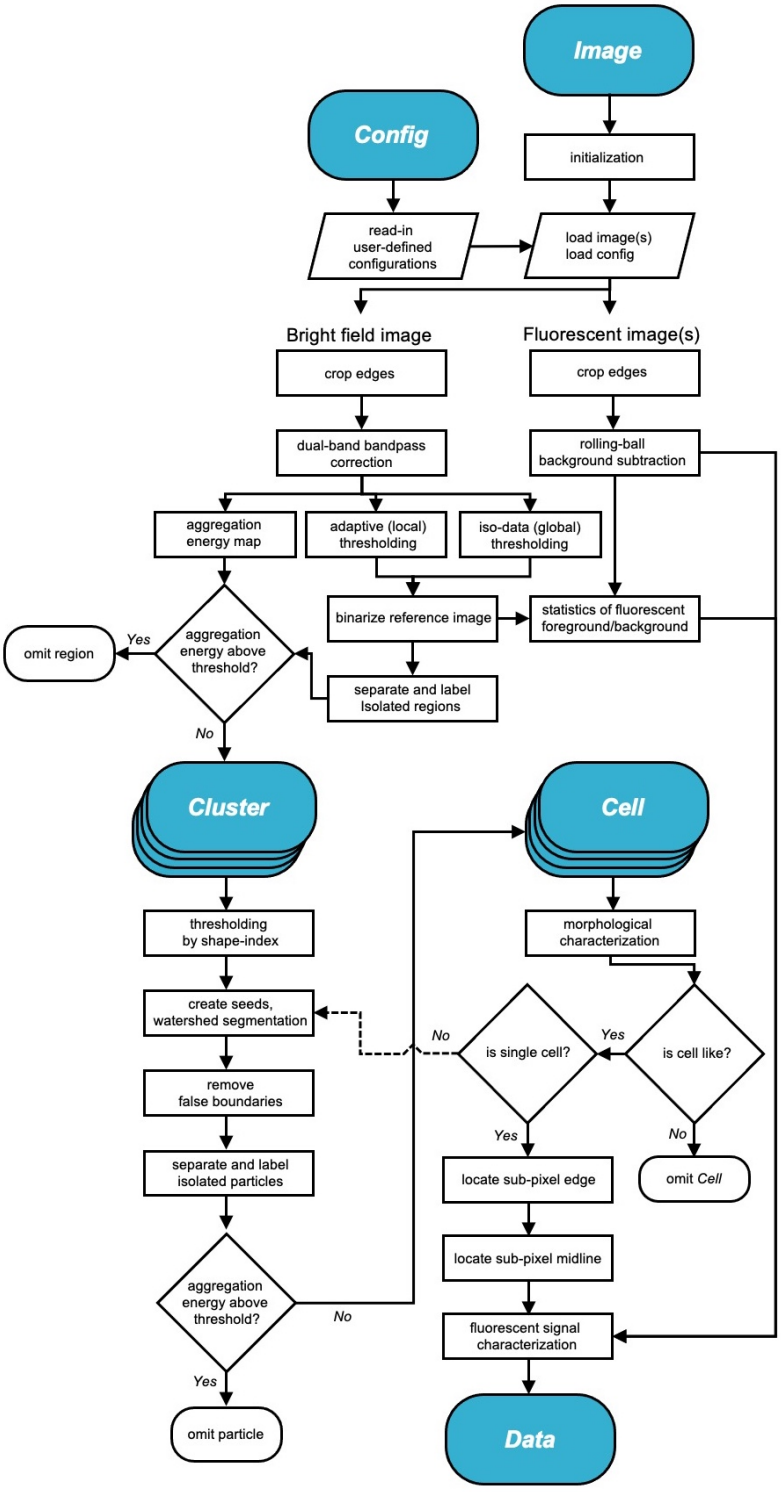


Fig. S1. Overview of the MSR-Dendra library, related to Figure 1 and STAR Methods. A. Schematic of the pMSR-Dendra plasmid backbone. B. Overview of the image acquisition method used to generate the MSR-Dendra dataset. After automated acquisition, raw images are scored manually: off- focus images (C, left panel) or the ones of extremely high or low cell density (C, middle and right panels) are rejected whereas the qualified remnants (D) are subjected to image segmentation and other downstream analysis.



11 **Fig. S2. Schematic of MOMIA pipeline structure, related to STAR Methods.**

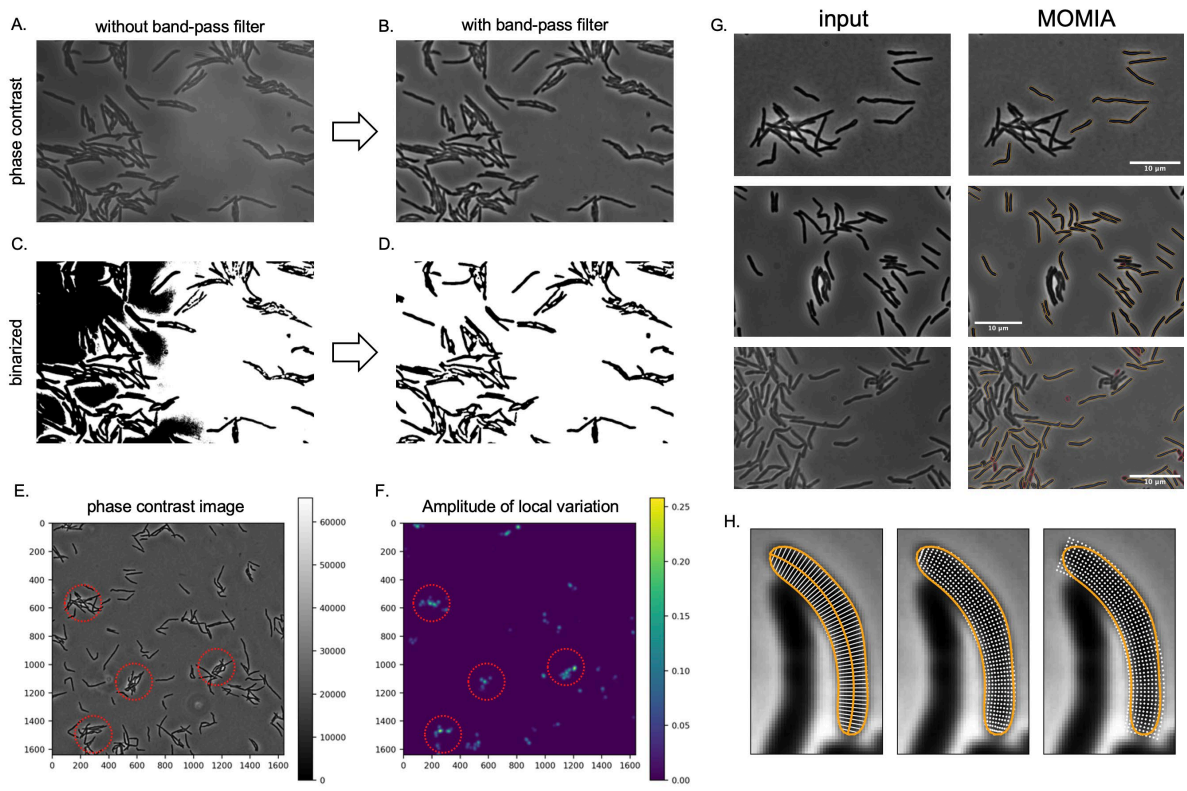


Fig. S3. Image processing with MOMIA, related to STAR methods. A-D. Dual-bandpass filter enhances segmentation performance. A. Example phase contrast image with uneven illumination. B. Dual-bandpass filtered output of A. C. Binarized mask of A. using iso-data thresholding method, separation of image background (white) and cells (black) is severely affected by uneven illumination. D. Binarized mask of B using the same threshold method. E-F. Amplitude of local variation as an indicator for cellular aggregation. E. Example phase contrast image with scattered cell aggregates. F. Computed amplitude of local variation (STAR Methods). Regions with cell aggregates are circled in red. G. Segmentation of different challenging imaging fields by MOMIA. Top row: recognition and omission of cellular clumps; middle row: segmentation of cells in dense clusters; bottom row: segmentation of cells in an image with aberrant illumination. Scale bars: 10 μm. Yellow lines: subpixel contours of segmented cells that are qualified for downstream analysis. Red lines: contours of omitted objects. H. Subcellular signal profiling. Left: cell contour and midline (orange) overlayed with orthogonal profile lines (white); middle: width-adapted profile mesh (white dots); right: regular profile mesh (white dots).

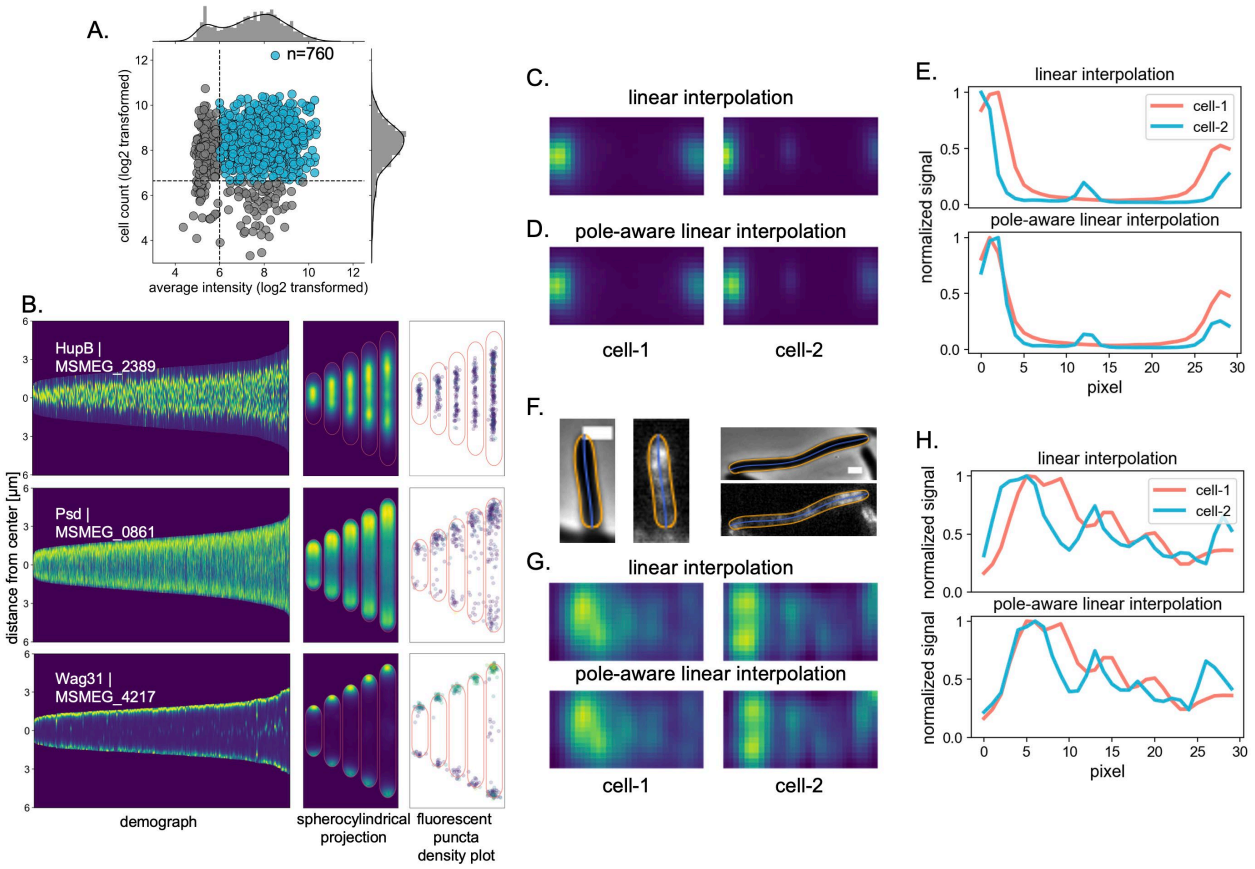


Fig. S4. Post-segmentation processing of MSR-Dendra dataset, related to Figure 1 and STAR methods. A. MSR-Dendra entries are scored by their corresponding cell counts and averaged fluorescent intensities. The present study included 760 entries which contain at least 150 cells and an average intensity over 64 (arbitrary cutoff). B. Populational localization profiles of Dendra tagged HupB, Psd, and Wag31(Top-bottom panels) depicted as: left-fluorescent demograph; midle-Spherocylindrical projection; right-foci density plot. C-E**.** Cell pole-aware, bimodal interpolation preserves polar topology. C. Linear interpolation renders varied polar signal distributions of Wag31-Dendra, exemplified by the two representative cells in Fig. 1H. D. Pole-aware, bimodal interpolation enables consistent representation of the morphologically inert cell poles. E. Interpolated data from C (upper panel) or D (bottom panel) are averaged along the longitudinal axes and aligned to demonstrate the polar signal representation by the two methods. F-H. Comparisons the two interpolation methods with example cells expressing subpolar-associated Gtf1-Dendra (MSMEG_0389). Scale bars: 1 μm.

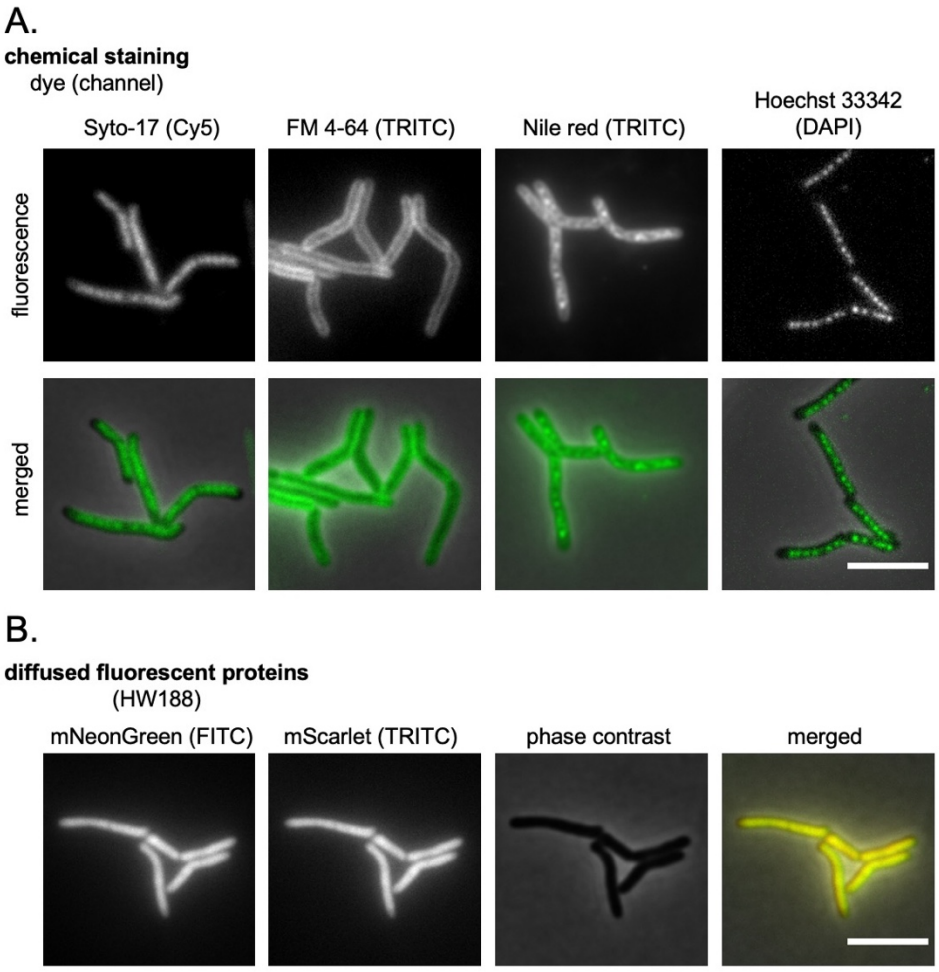


Fig. S5. Independently generated validation dataset 1, related to Figure 3. A. Example images of chemical fluorescent dye stained *Msm*. B. Examples images of untagged cytosolic fluorescent proteins mNeonGreen and mScarlet. Scale bars: 5 μm.

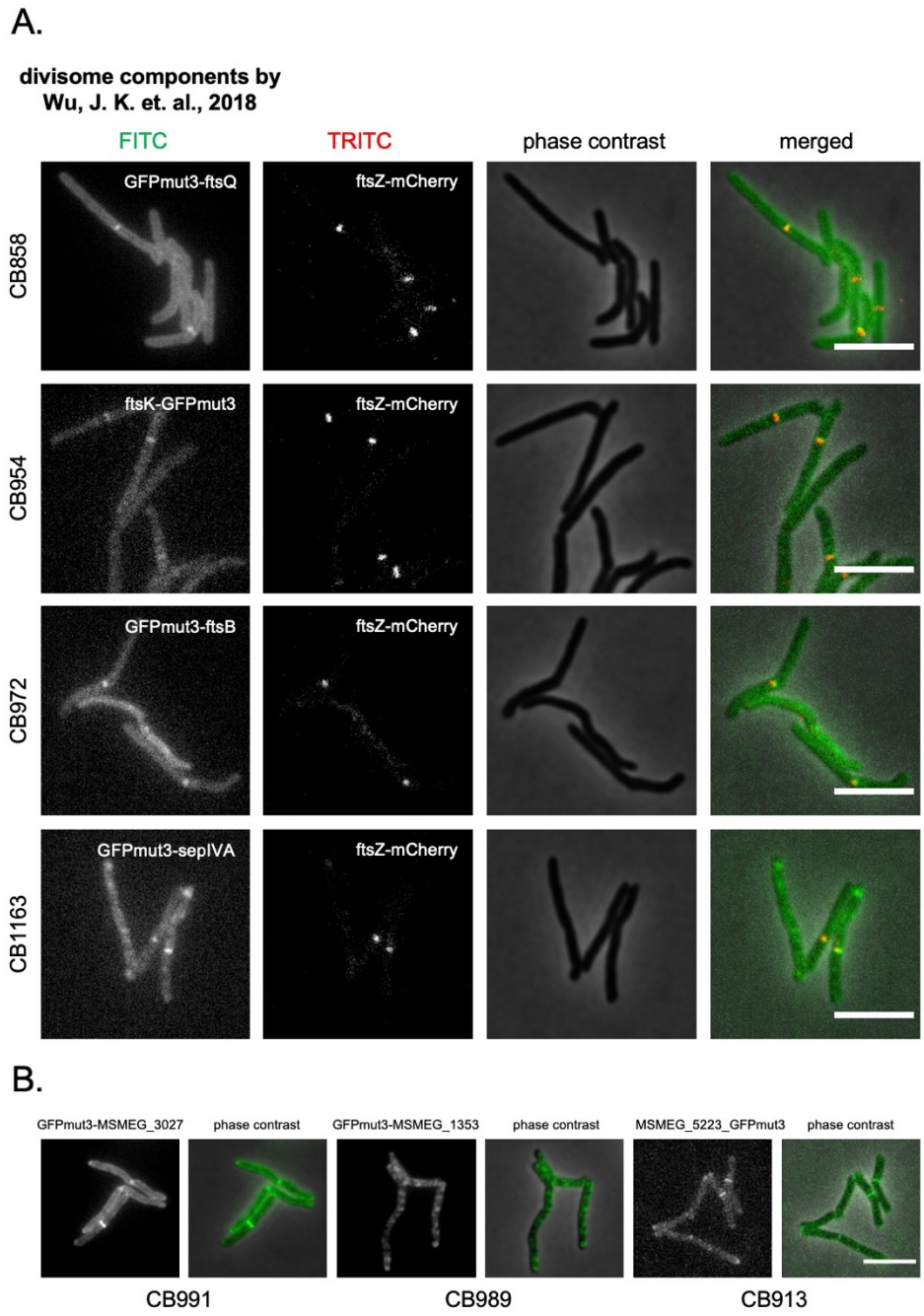


Fig. S6. Independently generated validation dataset 2, related to Figure 3. A. Example images of Msm expressing FtsZ-mCherry and one other mGFPmut3-tagged mycobacterial divisome components identified by Wu, J. K. et. al., 2018. B. Example images of *Msm* expressing postulated mycobacterial divisome components. Scale bars: 5 µm.

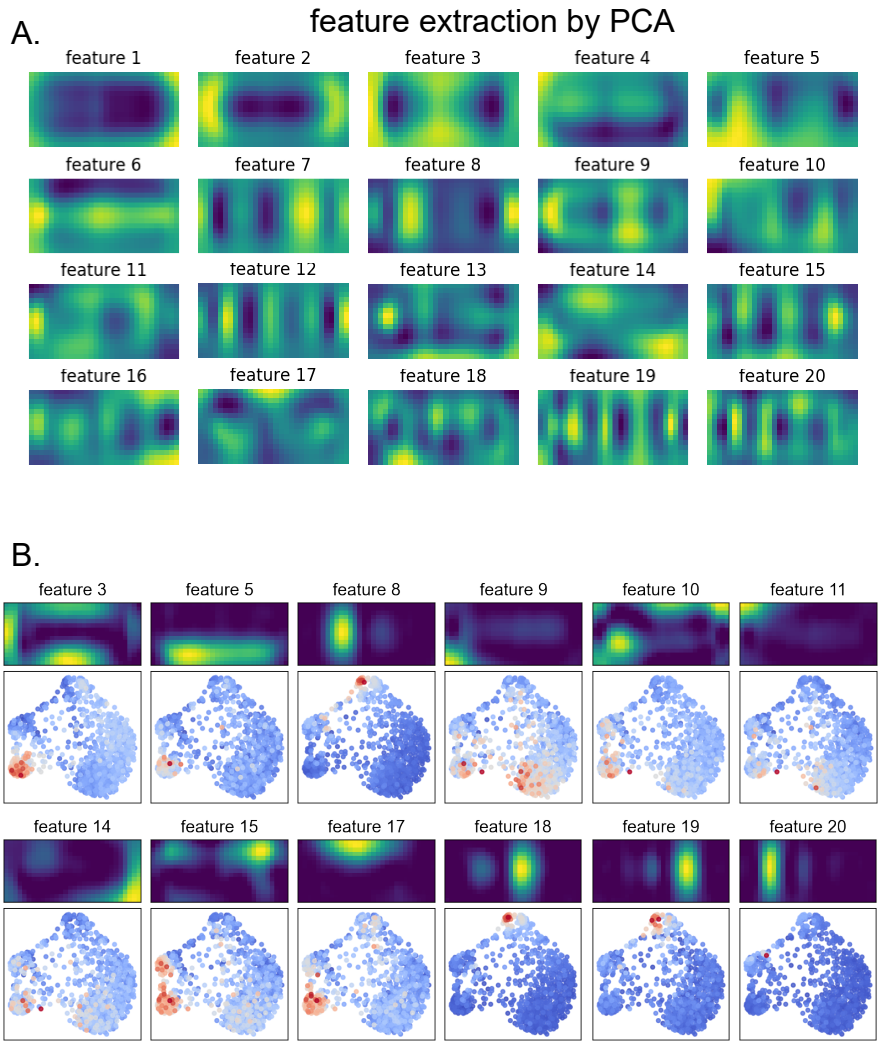


Fig. S7. MSR-Dendra data matrix decomposition by PCA and NMF, related to Figures 2 and 3. A. PCA decomposition infer holistic and complex features which are not visually intuitive**.** The matricized dataset in Fig. 3B is used to perform PCA decomposition (n=20). The extracted features are rescaled and depicted with pseudocolors. B. Illustration of NMF derived localization patterns (continuation of Figures 3A and 3B).

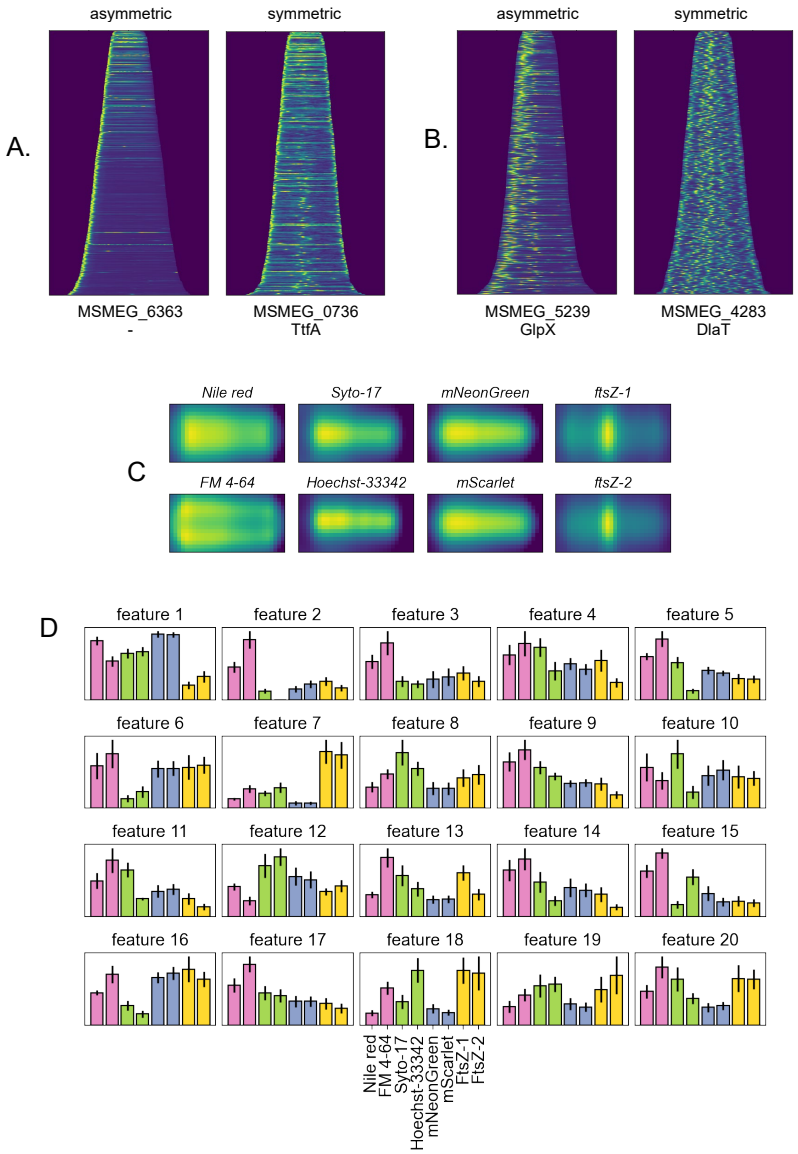


Fig. S8. GEMATRIA extracted features faithfully characterize various protein localization patterns, related to Figure 3. A and B. Demographs of representative strains discussed in Fig. 3C. C. Depiction of strain consensuses of the 8 selected validation datasets in Fig. 3D. Strain consensus is the average of the 10 frames of its length-binned fluorescence patterns. D. Complete feature profiles of the selected validation datasets. Error bars denote the standard deviation of corresponding feature weights over the 10 length bins.

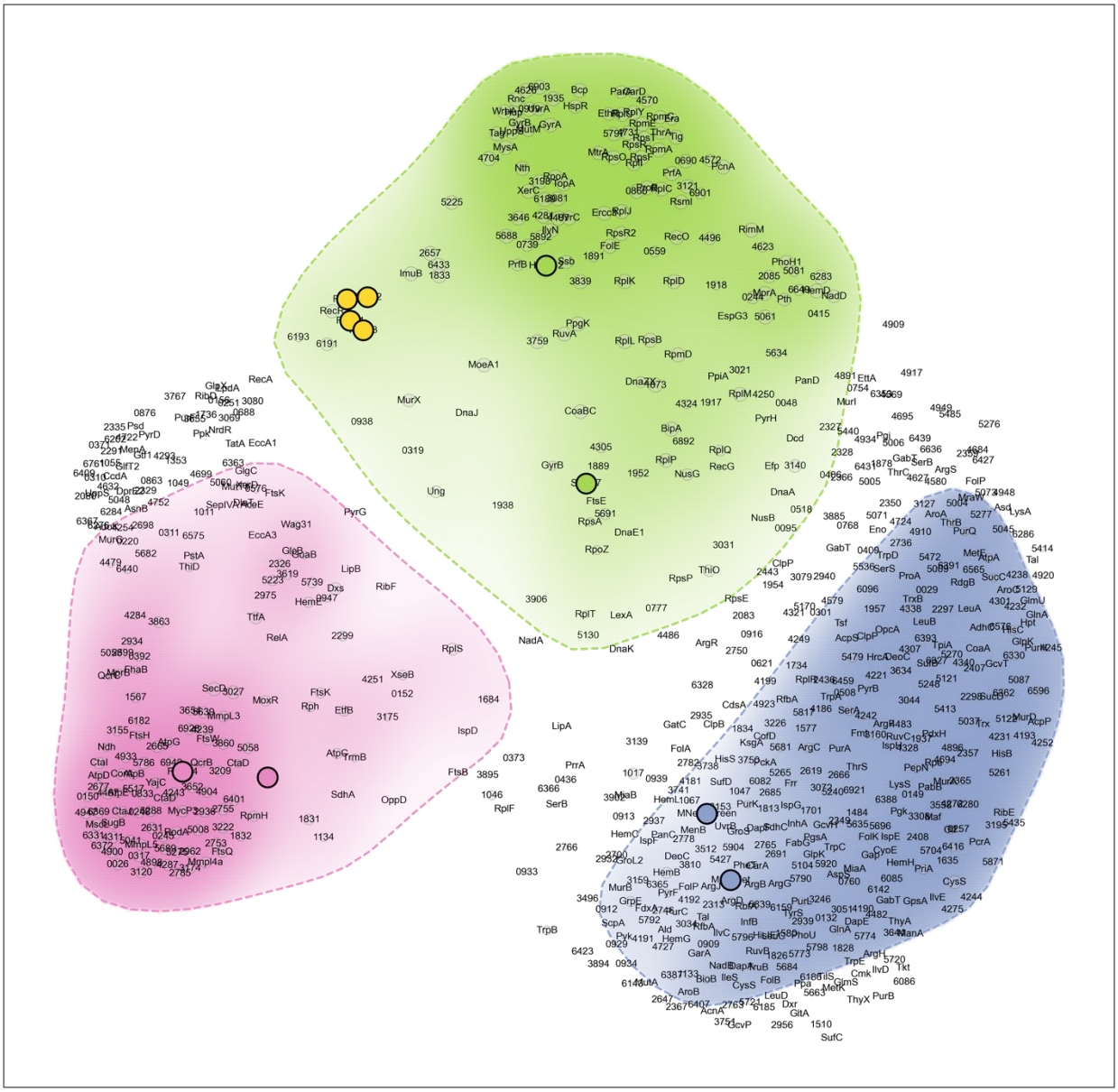


Fig. S9. SAFE defines three major domains in the GEMATRIA derived composite network (zoom-in view of Fig. 3E with labeled entry names or the abbreviated locus identifiers), related to Figure 3.

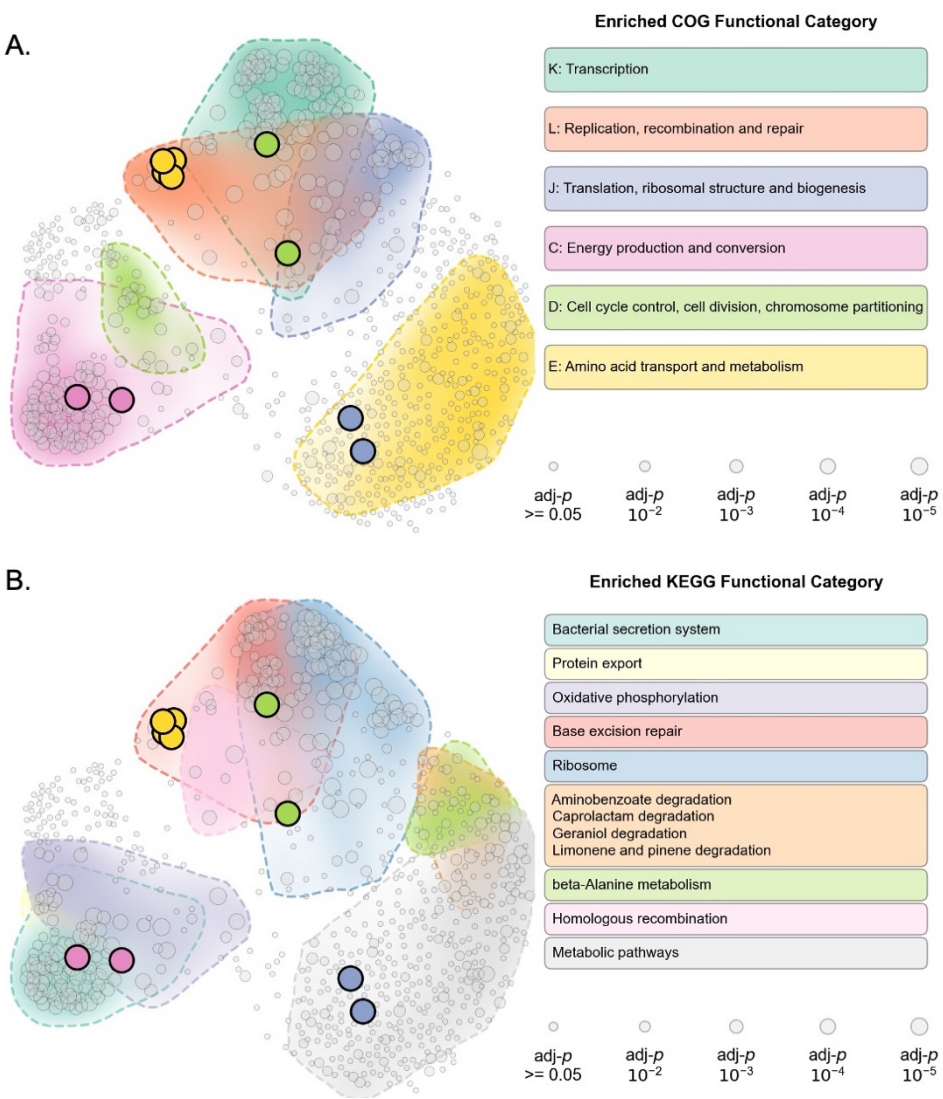


Fig. S10. Subdomains of various biological functions are found in the GEMATRIA derived composite network, related to Figure 3. Subdomains are identified by SAFE using a more stringent cutoff (<20% maximal distance) for merging overlapping network neighborhoods. A. Subdomains enriched of individual COG categories as revealed by SAFE. B. Subdomains enriched of KEGG (Kyoto Encyclopedia of Genes and Genomes) terms. The sizes of the nodes denote the FDR-corrected p values by hypergeometric test, as specified in the bottom right panels.

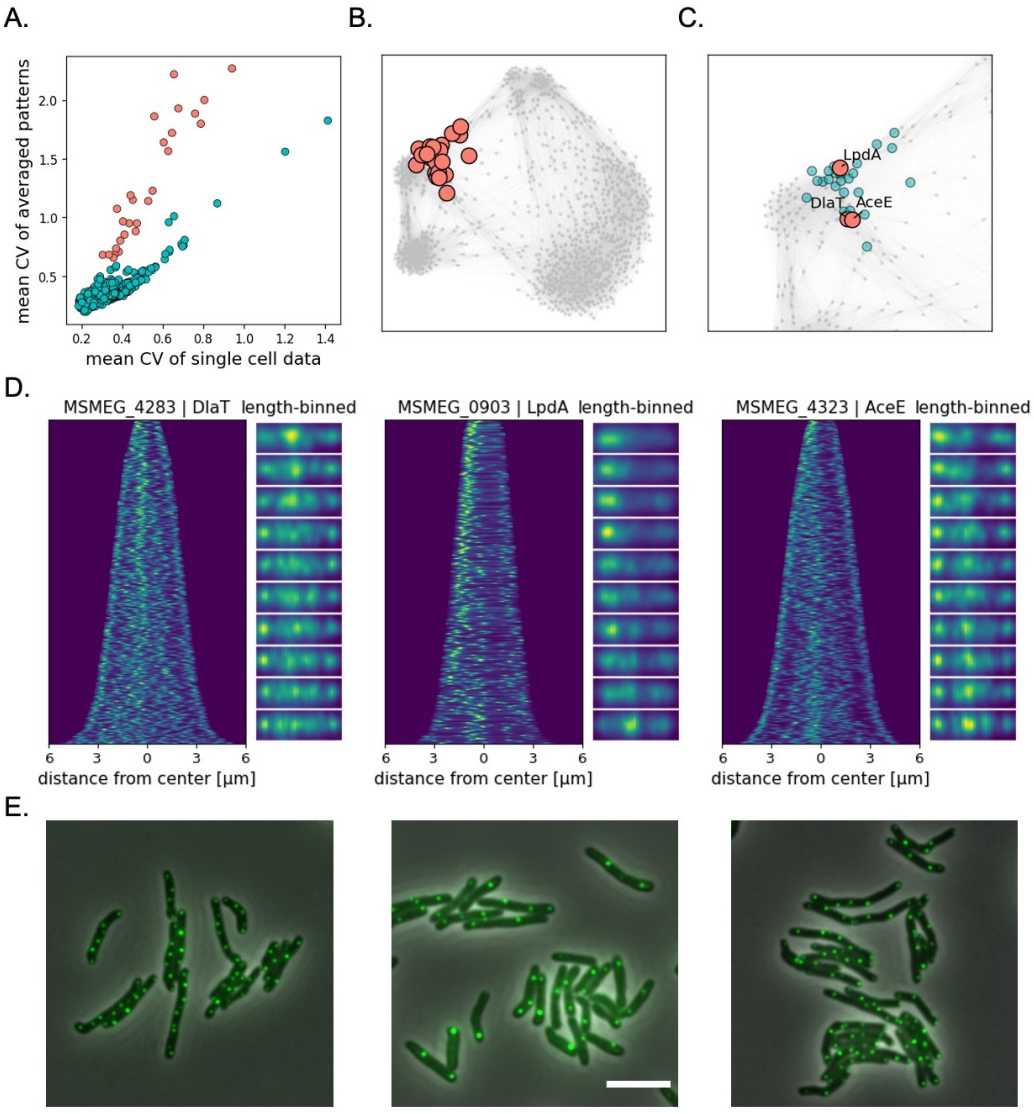


Fig. S11. Outliers with underrepresented signal variation are clustered together, related to Table S2. A. The comparison of mean signal coefficient of variation (CV) before (horizontal- axis) and after interpolation and data binning (vertical-axis). CV outliers are highlighted in red. B. Network locations of CV outliers. C. Zoom-in view of the outlier-enriched region. The three subunits of mycobacterial pyruvate dehydrogenase (AceE, DlaT, LpdA) are colored red wherein other CV outliers are colored blue. D. Fluorescent demographs (left panels) and length-binned patterns (right panels) of the pyruvate dehydrogenase subunits. E. Example microscopy images of the three pyruvate dehydrogenase subunits (DlaT, LpdA, and AceE). Scale bar: 5 μm.

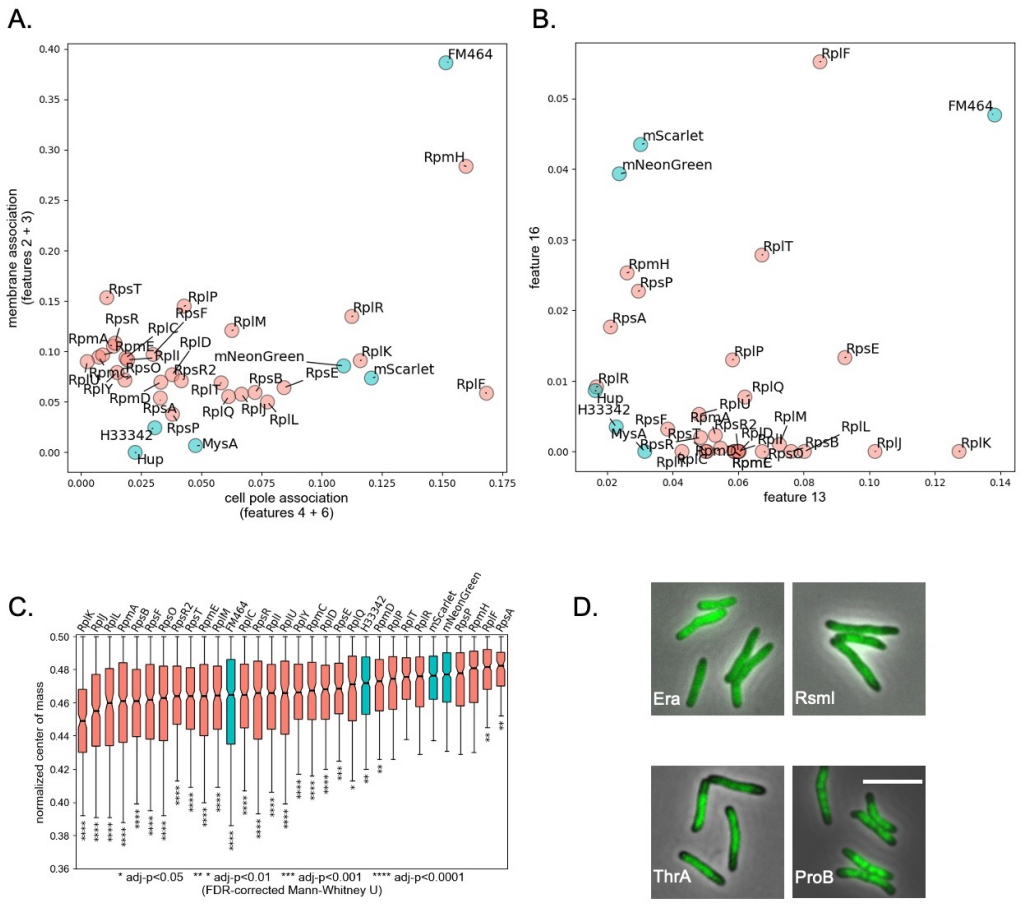


Fig. S12. Ribosomal proteins are asymmetrically distributed in mycobacteria cells, related to Figure 4. A. Ribosomal proteins exhibit cytosolic signals with diminished prevalence at the cell poles. The scatter plot denotes the summed means of the corresponding features for 29 ribosomal proteins (light red) and 6 non-ribosomal entries (light blue). B. Disproportional peripolar prevalence of mycobacterial ribosomal proteins as revealed by features 13 and 16. C. Mycobacterial ribosomal proteins exhibit significantly lower longitudinal symmetry compared to diffused cytosolic proteins. Normalized center-of-mass (NCM) is calculated as specified in STAR Methods and is used to evaluate the symmetry of longitudinal signals. Mann-Whitney U test is used to compare the NCM profiles of each entry to that of the cytosolic entry, mNeonGreen. Multiple testing is corrected using the Benjamini-Hochberg method. The symbolic representations of adjusted-p values are listed beneath the boxplot. D. Example images of MSR-Dendra entries that closely associate with the ribosome cluster. Scale bar: 5 μm.

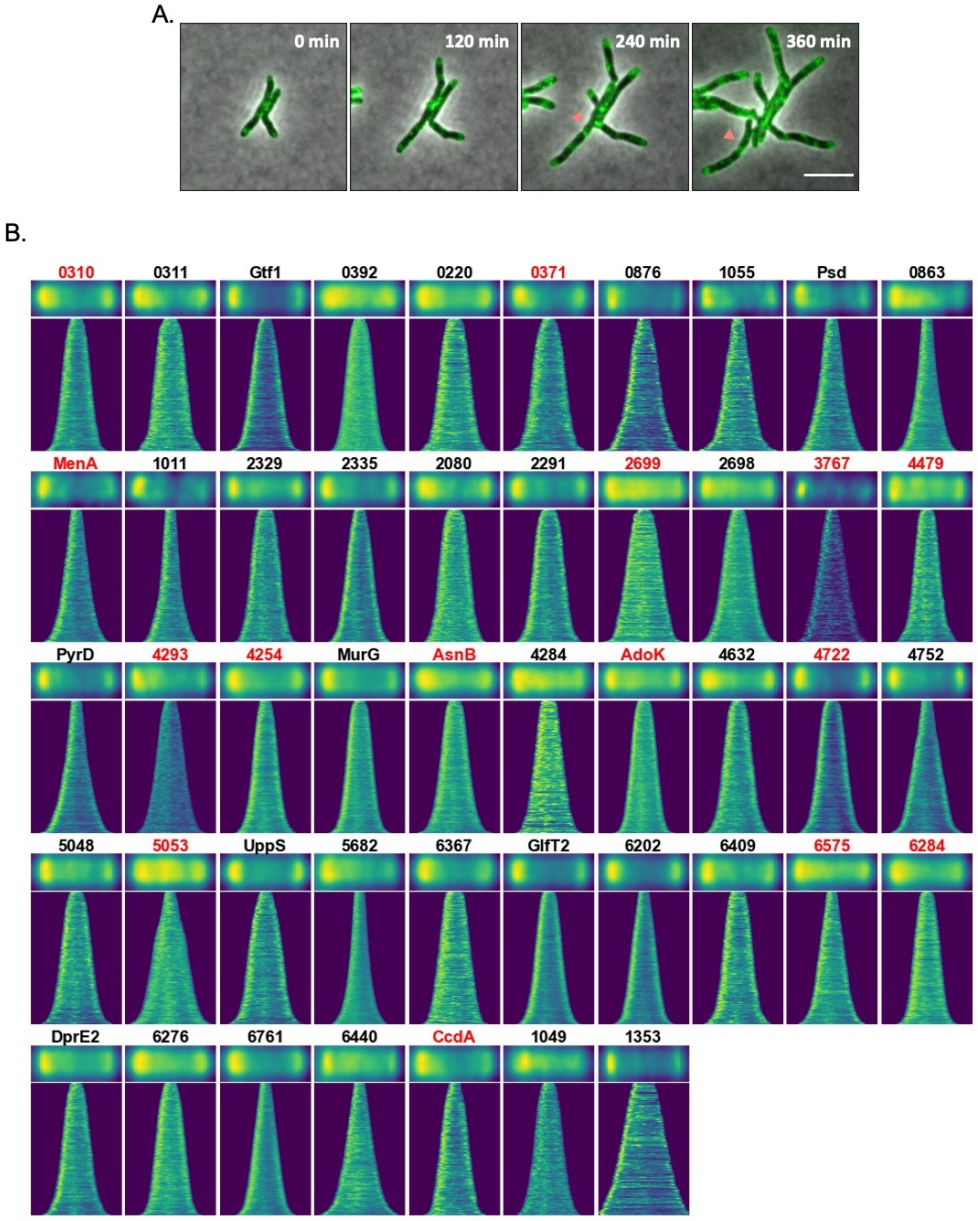


Fig. S13: GEMATRIA unveils subdomains enriched for known and novel IMD proteins, related to Figure 5. A. Representative frames of MSMEG_4479 time-lapse data (Video S3). Scale bar: 5 µm. B. In total 47 IMD-like entries are found using a SAFE based analysis and the biochemically characterized IMD proteins (Hayashi et al., 2016) as the reference. Protein localization patterns are demonstrated by the strain consensuses (top panels) and the *demographs* (bottom panels). Protein names or abbreviated locus identifiers of novel IMD proteins are colored in red.

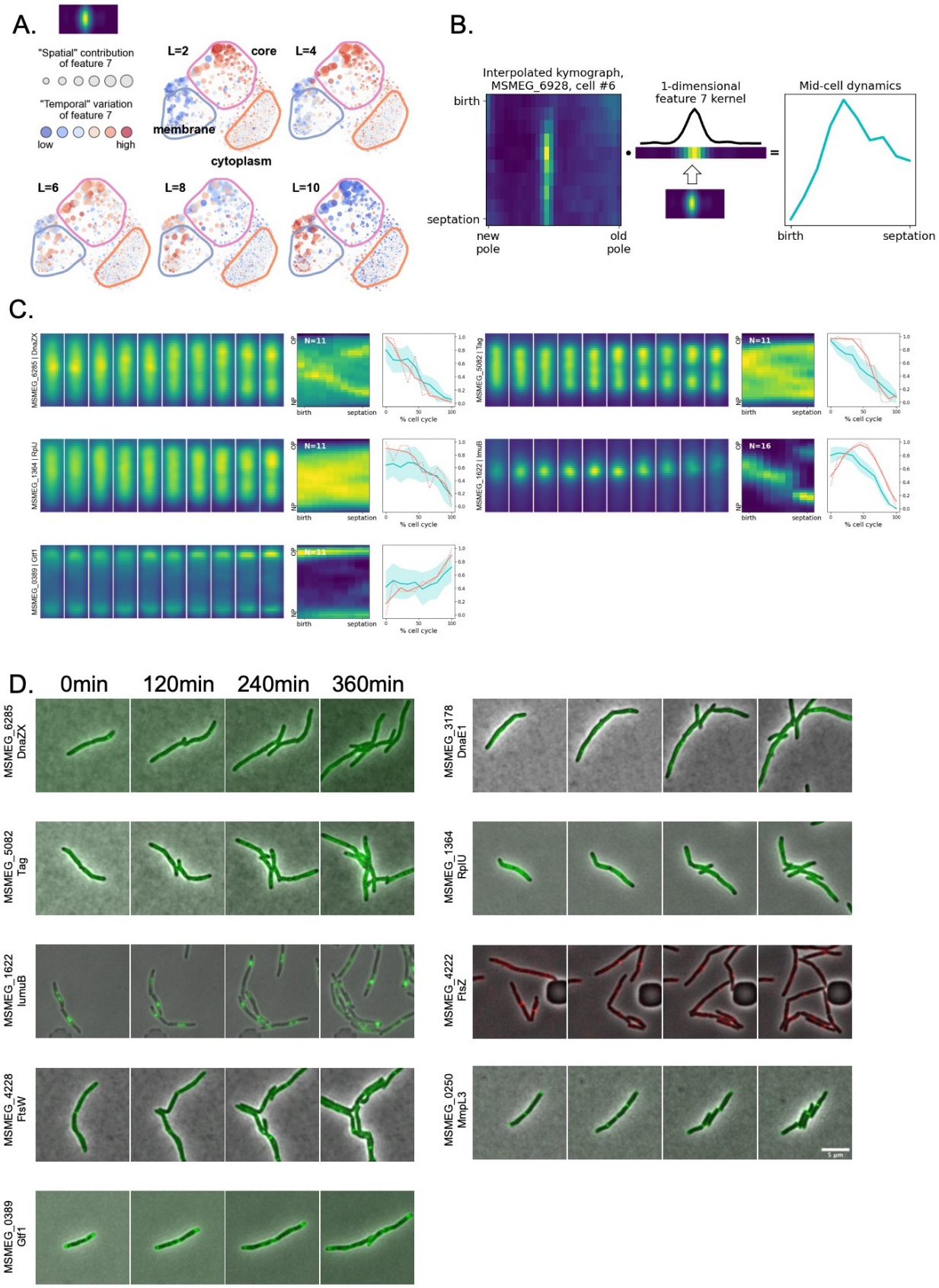

A.
"Spatial" contribution of feature 7
"Temporal" variation of feature 7
low
high
L=2
core
L=4
membrane
cytoplasm
L=6
L=8
L=10
B.
Interpolated kymograph, MSMEG_6928, cell #6
birth
septation
new pole
old pole
1-dimensional feature 7 kernel
Mid-cell dynamics
birth
septation
C.
MSMEG_6285 | DnaZX
MSMEG_5082 | Tag
MSMEG_1364 | RplJ
MSMEG_1622 | lmuB
MSMEG_0389 | Gtf1
N=11
N=16
birth
septation
% cell cycle
D.
0min
120min
240min
360min
MSMEG_6285 DnaZX
MSMEG_3178 DnaE1
MSMEG_5082 Tag
MSMEG_1364 RplU
MSMEG_1622 lumuB
MSMEG_4222 FtsZ
MSMEG_4228 FtsW
MSMEG_0250 MmpL3
MSMEG_0389 Gtf1
5 µm

Fig. S14: Approximation of cell-cycle-dependent mid-cell dynamics using GEMATRIA derived feature 7, related to Figure 6. A. Mid-cell associated feature 7 displays coordinated reallocation from the *core* domain to the *membrane* domain on a global scale. Here five sliced views (L=2, 4, 6, 8, and 10) of feature 7 dynamics are demonstrated, the full-length representations are animated in the Video S4. Node sizes represent the average feature 7 weights of MSR-Dendra entries, which remain constant over "time". The color-warmth (blue to red) of each node defines the normalized feature 7 intensity as a function of cell length, or "time". B. Schematic of extracting mid-cell dynamics from time- lapse kymographs using feature-7 *basis* matrix as a prior. Axial signal profiles of a single cell were oriented from time of birth to time of division and from new pole to old pole, normalized, then interpolated into a 10 × 30 matrix. A 1-dimensional feature-7 kernel (1 × 30) was derived by calculating the lateral mean of the feature-7 *basis* matrix. Mid-cell dynamics is represented by the dot product of the interpolated kymograph and the 1-dimensional feature-7 kernel. C. Length- binned patterns, fluorescence kymographs and mid-cell dynamics of test subjects (Fig. 6E) that were not listed in Fig. 6A. D. Representative time-lapse image slices of proteins presented in Fig. 6E. Fluorescence signals for Dendra-tagged proteins were pseudo-colored green where as mCherry-tagged FtsZ (reference strain CB954) signals were pseudo-colored red. Scale bar: 5 μm.

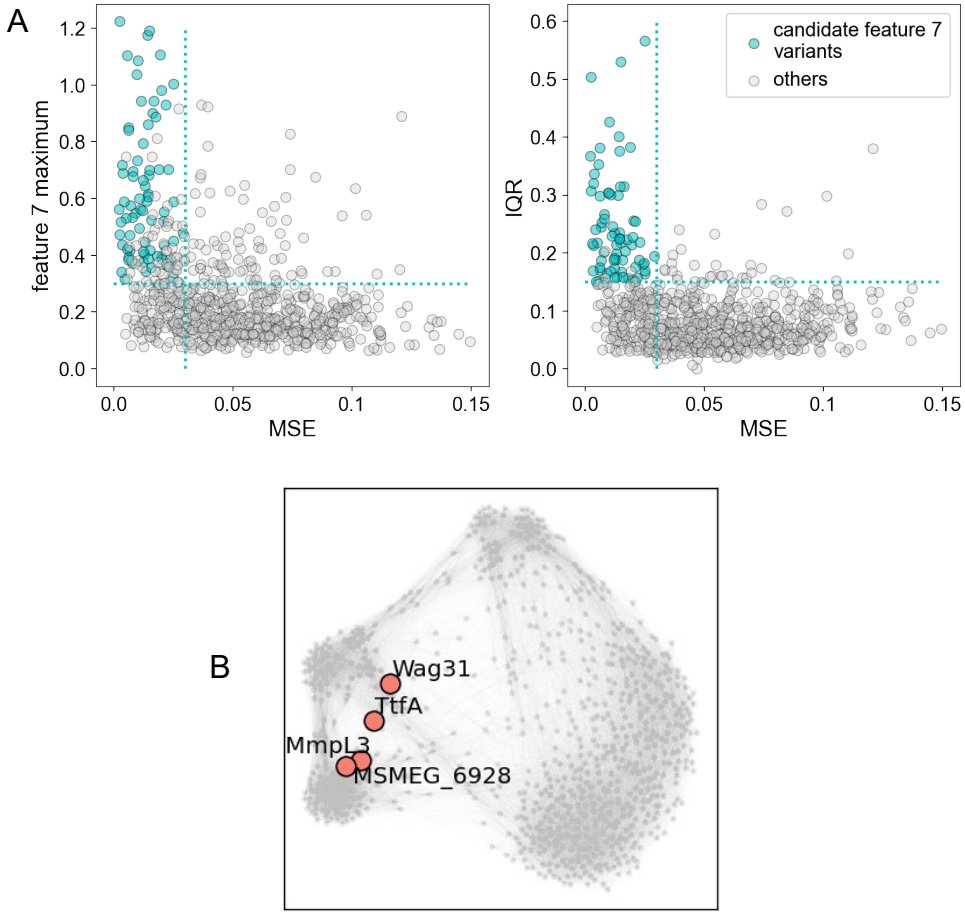


Fig. S15: Identification and case analysis of feature 7 variants, related to Figure 6 and STAR methods. A. Criteria defining candidate cell-cycle-dependent feature 7 variants. The maximum, the IQR, and the mean squared errors (MSE) of feature 7 profiles are calculated as specified in STAR Methods. Dashed lines denote the arbitrary cutoffs that separate candidate feature 7 variants from the remnants. B. MSMEG_6928 and the mycolic-acid transporter MmpL3 co-localize on the composite network.