In situ single particle classification reveals distinct 60S maturation intermediates in cells

Bronwyn A. Lucas[1,2,*], Kexin Zhang[1,2], Sarah Loerch[2,‡], and Nikolaus Grigorieff[1,2,*]

1. RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA
2. Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA, USA

*Correspondence: bronwyn.lucas@umassmed.edu, niko@grigorieff.org

‡Current address: Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA, USA

Electron cryo-microscopy (cryo-EM) can generate high-resolution views of cells with faithful preservation of molecular structure. *In situ* cryo-EM, therefore, has enormous potential to reveal the atomic details of biological processes in their native context. However, in practice, the utility of *in situ* cryo-EM is limited by the difficulty of reliably locating and confidently identifying molecular targets (particles) and their conformational states in the crowded cellular environment. We recently showed that 2DTM, a fine-grained template-based search applied to cryo-EM micrographs, can localize particles in two-dimensional views of cells with high precision. Here we demonstrate that the signal-to-noise ratio (SNR) observed with 2DTM can be used to differentiate related complexes in focused ion beam (FIB)-milled cell sections. We apply this method in two contexts to locate and classify related intermediate states of 60S ribosome biogenesis in the *Saccharomyces cerevisiae* cell nucleus. In the first, we separate the nuclear pre-60S population from the cytoplasmic mature 60S population, using the subcellular localization to validate assignment. In the second, we show that relative 2DTM SNRs can be used to separate mixed populations of nuclear pre-60S that are not visually separable. We use a maximum likelihood approach to define the probability of each particle belonging to each class, thereby establishing a statistic to describe the confidence of our classification. Without the need to generate 3D reconstructions, 2DTM can be applied even when only a few target particles exist in a cell.

Introduction:

Locating and characterizing molecules in cells is an important goal of molecular, structural, and cell biology. Cryogenic electron microscopy (cryo-EM) enables simultaneous visualization of all cellular molecules in their native cellular environment while preserving high-resolution molecular architecture. Therefore, cryo-EM holds the promise of delivering an atomistic view of the cell. However, realizing this promise is limited by the high density of molecules in a cell, making it difficult to identify molecules of interest (Lučić et al., 2013). As one way to address this, electron cryo-tomography (cryo-ET) can be used to build 3D maps of cellular structures in their native context (*in situ*) by constructing tomograms from a series of tilted 2D images (Kürner et al., 2005; Lučić et al., 2013; Mahamid et al., 2016). In a tomogram, molecules overlapping in any given view can be separated and large molecular assemblies (particles) with distinctive shapes can be identified. Once identified, subtomogram averaging can yield *in situ* molecular structures at <4 Å resolution (Himes and Zhang, 2018; Tegunov et al., 2021). However, since the effective resolution of a raw tomogram is below 15-20 Å (Vilas et al., 2020), identification of specific targets in tomograms is limited to abundant particles that are sufficiently distinct at this resolution to be identified.

Many potential cell biological applications require accurate categorization of individual molecule identity at a specific subcellular localization. Examples are the characterization of the spatial organization of a biosynthetic process such as ribosome biogenesis, and the assignment of molecular identities in small volumes such as synapses and vesicles. 3D classification of subtomograms can differentiate between structural states (Himes and Zhang, 2018; Xue et al., 2021). However, the assignment of states is unreliable for similar structures that can only be distinguished using high-resolution detail, and statistical approaches to quantitatively assess classification results are lacking. Machine learning has been employed for particle classification in tomograms, but currently only performs as well as a human operator (Moebel et al., 2021). While machine learning algorithms performed better than 3D template matching at molecule localization in tomograms, classification remained challenging for all algorithms (Gubins et al., 2020). *In situ* molecule classification, therefore, remains a major challenge.

We recently described an alternate method to locate particles that may improve structural classification in cells. By using 2D cryo-EM images, rather than tomograms, and fine-grained, high-resolution template matching (2DTM), specific particles can be located in cells with high

precision using their atomic structures (Lucas et al., 2021; Rickgauer et al., 2020, 2017). 2DTM uses molecular models, from *in vitro* structure determination or *in silico* structure prediction (e.g., Alphafold2 (Jumper et al., 2021)) to generate a 3D density. This 3D density (hereafter referred to as the template) is then projected in 2D along millions of orientations. A pixel-wise cross-correlation of the 2D projections with a high-resolution 2D cryo-EM image is performed, yielding a 2DTM signal-to-noise ratio (SNR) at every pixel location (Rickgauer et al., 2017). The 2DTM SNR values are subjected to a significance test, which identifies peaks with a desired level of confidence (Lucas et al., 2021; Rickgauer et al., 2017). In the following, we refer to targets passing this test as significant targets (Lucas et al., 2021; Rickgauer et al., 2017).

The 2DTM SNR is proportional to template mass and negatively affected by non-matching elements between template and target (Lucas et al., 2021; Rickgauer et al., 2020, 2017). We have shown that a template generated from a *Bacillus subtilis* 50S large ribosomal subunit was able to detect 50S in 2D cryo-EM images of *Mycoplasma pneumoniae* cells, but with a lower average 2DTM SNR compared to a *M. pneumoniae* 50S template (Lucas et al., 2021). This demonstrated that (1) 2DTM using partially matching templates can be sufficiently sensitive to yield significant targets and (2) the mean 2DTM SNR of detected targets provides a read-out of the relative similarity between different templates and populations of particle species.

In this study, we investigate whether the ratio of 2DTM SNRs obtained using different templates can be used to identify the template that more closely resembles the cellular target, and thereby classify particles in cells. As a model system, we chose to examine the late stages of 60S ribosomal subunit biogenesis in the yeast *Saccharomyces cerevisiae* because (1) intermediates are of a similar size and share significant structure with one another, making them difficult to separate at low resolution, (2) molecular models spanning multiple late intermediate states have recently been described, and (3) the maturation events that occur before and after nuclear export have been characterized. Subcellular localization can thereby validate the assignment of intermediate and mature states.

We show that 2DTM can locate and distinguish nuclear intermediates of 60S maturation in 2D cryo-EM images of FIB-milled yeast cells. We confirm that 2DTM can distinguish predefined 60S populations separated by subcellular localization and identify compositional differences between them. We then applied a maximum likelihood-based approach to identify two sub-populations of nuclear intermediates that were not otherwise separable and provide a

confidence of single particle classification. We show that using this approach, we can observe a shift in the nuclear pre-60S intermediate population to a more mature intermediate after inhibiting Crm1-mediated nuclear export. This study demonstrates that relative 2DTM SNR ratios effectively distinguish related complexes and can identify changes to particle populations in cells.

Results:

2DTM identifies 60S in biologically relevant locations and orientations in FIB-milled lamellae

2DTM has been used to detect mammalian ribosomes in thin extensions of adherent cells (Rickgauer et al., 2020), and bacterial ribosomes in *Mycoplasma pneumoniae* cells (Lucas et al., 2021), both of which are sufficiently thin to permit imaging by transmission EM (TEM). Since most eukaryotic cells are too thick to image by TEM, focused ion beam (FIB)-milling is used to generate thin, electron transparent lamellae of cryogenically frozen cells (Marko et al., 2007; Rigort et al., 2012; Villa et al., 2013).

To evaluate the utility of 2DTM to locate molecules in FIB-milled lamellae, we collected 28 2D cryo-EM images of the nuclear periphery of lamellae generated from actively growing *Sacchromyces cerevisiae* cells (**Figure 1, Figure 1-figure supplement 1A-B**). We identified 4363 large ribosomal subunits by 2DTM using a template generated from a model representing the mature 60S (PDB: 6Q8Y) (Tesina et al., 2019) (**Figure 1A-C**). The peaks corresponding to significant detections were clearly distinguishable from background (**Figure 1E, Figure 1-figure supplement 1C**), enabling precise localization of mature 60S in the cell.

To assess the specificity of 60S detection, we identified regions of the images corresponding to the cytoplasm, nucleus and vacuole by visual inspection. Consistent with the expected high specificity of 2DTM, we did not observe any significant mature 60S-detected targets in regions of the image corresponding to the vacuole (**Figure 1C-D**). In contrast, 229 mature 60S-detected targets localized to the nucleus, representing ~5% of all mature 60S identified targets in these images (**Figure 1C-D**). In regions of the images corresponding to the cytoplasm we observe a median density of ~6500 60S/μm^3, which, assuming an average cell volume of ~42 μm^3 of which ~65% is cytoplasm, corresponds to a total of ~180,000 60S/cell (**Figure 1G**). This is consistent with prior estimates of 187,000 $\pm$ 56,000 ribosomes per yeast cell based on rRNA concentration (von der Haar, 2008).

Beyond the subcellular distribution of mature 60S-detected targets, we also confirmed that 2DTM identified specific 60S in biologically relevant locations and orientations. The nuclear envelope (NE) is contiguous with the endoplasmic reticulum and a known site for co-translational transport of transmembrane and secretory proteins, while the vacuole is not known to be a site of translation. We found that mature 60S-detected targets were oriented with their polypeptide exit tunnels facing the cytoplasmic surface of the NE but were depleted from within ~20 nm of the vacuole (**Figure 1C,F**). This indicates that the orientation of 60S identified by 2DTM is unlikely to be an artefact introduced by features of the membrane in the image. To confirm that the targets identified with the mature 60S template reflect ribosomes, we generated a 3D reconstruction using the locations and orientations of 3991 significant mature 60S-detected targets using standard single particle approaches as described previously (Lucas et al., 2021). In addition to the 60S the 10 Å-filtered reconstruction showed density consistent with the 40S small ribosomal subunit (**Figure 1H**). This is consistent with many of the mature 60S detected targets representing a population of 80S ribosomes. We conclude that 2DTM-identified locations and orientations in 2D cryo-EM images of FIB-milled lamellae reflect biologically relevant locations and orientations of ribosomes in the cell.

Relative 2DTM SNRs enable single particle classification **in situ**

The nuclear envelope (NE) creates a physical barrier that separates premature 60S in the nucleus from mature 60S in the cytoplasm and is easily distinguishable in many 2D images by its characteristic double membrane and by the more granular appearance of the cytoplasm vs the nucleus (e.g., **Figure 1B**). Our observation of a substantial population of mature 60S-detected targets in the nucleus, but not in the vacuole (**Figure 1C-D**), suggests that the nuclear 60S may result from cross-detection of nuclear precursors, which share part of their structure with mature 60S and therefore also produce significant correlations (**Figure 2A**). As a first step to differentiate between related 60S intermediates, we located precursor 60S by 2DTM searches using a template generated from a late nuclear intermediate (LN 60S, PDB: 6N8J) (Zhou et al., 2019) (**Figure 2A,B**), and annotated each target by its subcellular localization. The LN 60S was chosen because it represents the most mature nuclear intermediate for which there is a structure, and which retains ribosome biogenesis factors (RBFs) that are removed during nuclear and early cytoplasmic processing (**Figure 2A**). Thus, we expect that (1) the similarities between the

mature 60S and LN 60S structures will result in cross-detection of the respective other complex and (2) the cytoplasmic population will more closely resemble the mature 60S and nuclear population will more closely resemble the LN 60S resulting in a higher mature 60S / LN 60S 2DTM SNR ratio in the cytoplasm than the nucleus. In the 28 images of the nucleus and nuclear periphery we located 1651 significant LN 60S-detected targets of which 1382 (~84%) of the LN 60S-detected targets were cytoplasmic and 268 (16%) were nuclear (**Figure 2-figure supplement 1A**). We identified more cytoplasmic than nuclear targets in 2DTM searches with both mature and LN 60S templates because (1) the cytoplasm represented a larger area of our images and (2) the concentration of 60S is expected to be higher in the cytoplasm relative to the nucleus (e.g., (Delavoie et al., 2019)). Only one of the significant LN 60S-detected targets localized to the vacuole, which is below the expected false positive rate and further indicates the specificity of 2DTM.

As expected from the similarity between the mature and LN 60S templates, the locations of many of the targets identified in the two searches overlap (**Figure 2B,C**). We aligned the two sets of coordinates using the program *align_coordinates* (Lucas et al., 2021). Approximately one third of the mature 60S-detected targets overlapped with LN 60S-detected targets while 92% of the LN 60S-detected targets overlapped with mature 60S-detected targets (**Figure 2H**). Combining the results of both searches, only 0.5% of the cytoplasmic targets were LN 60S-detected only, compared to 30% of the nuclear targets (**Figure 2I**).

Consistent with their expected localizations, the median $\log_2$(mature 60S / LN 60S 2DTM SNR) values

of targets identified with both templates were significantly higher for the cytoplasmic population than the nuclear population ($p < 0.0001$, K-S. test) (**Figure 2D-G,J**). We classified each target as LN or mature 60S according to the highest 2DTM SNR (**Figure 1K**). Of the population detected with both mature and LN 60S templates. 94% of the 1361 cytoplasmic targets have a closer match (higher SNR) with the mature 60S and 88% of the 171 nuclear targets have a closer match with the LN 60S (**Figure 1J**). Combining all 60S-detected targets, the nuclear 60S targets are now more clearly distinguished from the cytoplasmic population with 98% of the cytoplasmic targets annotated as mature 60S, and 60% of the nuclear targets annotated as pre-60S (**Figure 1K,L**). The ~40% of nuclear targets that more closely resemble the mature 60S likely reflect nuclear intermediates different from the LN 60S (see below) and thus

do not perfectly match either template. We conclude that comparing 2DTM SNRs can effectively differentiate populations of related particles *in situ*.

***Defining a confidence metric for single particle classification* in situ**

To gain an understanding of cell biology at molecular resolution it is necessary to be able to confidently assign particle identity to individual targets. We show above that the nuclear and cytoplasmic 60S populations were significantly different with respect to their relative similarity to the LN and mature 60S (**Figure 2**). We also show that classifying targets by their highest 2DTM SNR effectively separates the nuclear from the cytoplasmic population (**Figure 2**). However, a single threshold does not fully capture the differences between the nuclear and cytoplasmic populations and for an individual particle the confidence of classification is unclear.

To assign a confidence in the class assignments of detected particles we developed a maximum likelihood-based approach to infer the probability of a particle deriving from one of a given number of populations. We sought to classify each of the 1531 LN and mature 60S-detected targets by their relative similarity to the LN 60S or mature 60S templates. We restricted our analysis to the targets that were detected by both templates to limit the contribution from noise. We made the initial simplifying assumption that: 1) each 60S identified more closely reflects either LN or mature 60S, i.e., the number of classes needed to describe all detected targets is two; 2) the nuclear targets more closely resemble the LN 60S and the cytoplasmic targets more closely resemble the mature 60S. We therefore define the prior probability that a randomly selected detected target belongs to a specific population according to the number of targets detected in the nucleus and cytoplasm, respectively (**Figure 2J, Figure 2-figure supplement 1A**).:

$$\text{P(targets being LN 60S)} = P(Nucleus) = 0.11 \text{ and}$$

$$\text{P(targets being mature 60S)} = p(Cytoplasm) = 0.89.$$

We used a maximum likelihood-based approach to model the $\log_2$(mature / LN 60S 2DTM SNR) values as a mixture of two Gaussians (**Figure 3A**, $R^2 = 0.993$). The fit suggests a major population that more closely reflected the mature 60S and a smaller population that more closely reflected the LN 60S (**Figure 3A**). Using the Gaussian distribution model (see Materials and Methods), we calculate the probability that a LN and mature 60S-detected target with a given $\log_2$(mature / LN 60S SNR) value more closely resembles the LN 60S than the mature 60S

via Bayes rule (**Figure 3B-C**). This analysis could easily be extended to cases where more than two templates are used in the search (see Materials and Methods). A confidence threshold of 95% assigns 27% of the nuclear targets and only ~0.2% of the cytoplasmic targets to the LN 60S population (**Figure 3C**). Defining a threshold at 50% classifies ~75% of the nuclear targets as LN 60S and 92% of the cytoplasmic targets as mature 60S (**Figure 3C**). The relative probability of each detected 60S belonging either to the LN or mature 60S population can be readily visualized (**Figure 3D**). This shows that the 2DTM SNR ratio can effectively delineate populations of related particles in cells with a specified confidence for each particle assignment.

Ribosome biogenesis factors differentiate nuclear from cytoplasmic 60S

Most of the mass difference between the LN and mature 60S templates results from proteins in the LN 60S that are absent in the mature 60S (**Figure 4A-C**). Notable exceptions are the proteins on the P-stalk are present only on the mature 60S (**Figure 4A-C, Figure 3A**). Additionally, several rRNA helices on the intersubunit interface are in different conformations, specifically the L1 stalk, helix 38 and helix 89, which undergo conformational changes during maturation (**Figure 4C**). To identify which of these features distinguish nuclear from cytoplasmic 60S, we investigated the relative dependence of the 2DTM SNRs on the rRNA and proteins of the LN 60S template. We generated truncated LN 60S templates containing either rRNA or protein only and calculated the change in the 2DTM SNR for each template at each target relative to the full-length template (**Figure 4D**). The rRNA contributed 1.5 and 1.8-fold more to the 2DTM SNR of the nuclear and cytoplasmic targets, respectively, despite comprising only 1.25-fold more of the template mass (1004 and 800 kDa, respectively), than the proteins (**Figure 4D**). Indeed, 60% of the cytoplasmic targets and 34% of the nuclear targets were no longer significant when searching with the proteins alone. Comparing the nuclear and cytoplasmic populations shows that the 2DTM SNR of the LN 60S-detected cytoplasmic targets is less affected by the removal of the LN 60S proteins and more strongly affected by the removal of the rRNA (**Figure 4D**). This shows that the LN 60S proteins contribute more to the SNR of the nuclear targets than the cytoplasmic targets and are therefore more effective at differentiating the nuclear from the cytoplasmic 60S population.

Since the LN 60S represents a late intermediate of 60S maturation in which the rRNA is almost fully folded, RBF proteins on the LN 60S account for most of the difference with the

mature 60S by mass (**Figure 4A-D**). To confirm that the SNR difference of nuclear LN 60S-detected targets and cytoplasmic mature 60S-detected targets is primarily due to the RBF proteins, we removed the RBFs from the LN 60S template and recalculated the SNR for each target. The removal increased the 2DTM SNR ratio of the cytoplasmic targets, while decreasing the 2DTM SNR of the nuclear targets (**Figure 4E**), making the SNR values more similar. This is consistent with the nuclear population having these RBFs and the cytoplasmic population lacking the RBFs. We conclude that the differentiation of detected targets using the observed 2DTM SNRs reflects biologically relevant differences between them.

Nog2 lacking intermediates accumulate after inhibition of nuclear export

The two largest RBFs on the LN 60S are Nog1 and Nog2, together accounting for ~50% of the RBF mass (**Figure 4F,G**). During 60S maturation, Nog2 removal is required to permit binding of the nuclear export adaptor Nmd3 and Crm1-dependent export, and therefore Nog2 removal precedes nuclear export (Ho et al., 2000; Matsuo et al., 2014). In contrast, Nog1 is removed only upon export to the cytoplasm (Pertschy et al., 2007). In cells with active nuclear export, we find that removal of either Nog1 or Nog2 differentiates the nuclear from the cytoplasmic populations (**Figure 4F,G**, untreated cells). As a further test of differentiating different targets by 2DTM, we inhibited Crm1 mediated export by treating Leptomycin B (LepB) sensitive Crm1 (T539C) cells with LepB and located 60S targets with LN 60S and mature 60S templates in eight images of FIB-milled lamellae. To assess the relative occupancy of Nog1 and Nog2 after Crm1 inhibition, we measured the change in 2DTM SNR after removal of all RBFs, and Nog1 or Nog2 alone. Consistent with LepB inhibiting export of pre-60S from the nucleus, we detected a higher density of pre-60S in the nucleus than in cells with active Crm1 (**Figure 4-figure supplement 1A**). When nuclear export is inhibited, all RBFs (**Figure 4E**) and Nog1 alone (**Figure 4F**) still differentiate the nuclear from the cytoplasmic populations. In contrast, the occupancy of Nog2 is no longer significantly different between the nuclear and cytoplasmic populations (**Figure 4G**). This is consistent with a model in which, when Crm1-mediated export is active, nuclear intermediates are rapidly exported after removal of Nog2, depleting the Nog2-lacking population from the nucleus. In the presence of a Crm1-inhibitor, the late, export competent nuclear intermediate lacking Nog2 can no longer be exported and therefore accumulates. Since Nog1 is only removed after export, inhibition of export did not change the

occupancy of Nog1 on the maturing 60S. This demonstrates that comparing 2DTM SNRs is sufficiently sensitive to assess the occupancy of individual proteins *in situ*.

Classification of nuclear pre-60S intermediates

Ribosome biogenesis is a highly efficient molecular assembly line, and multiple intermediate states co-exist in the cell (Warner, 1999). Therefore, the nuclear population of pre-60S is unlikely to represent a single intermediate population. Accordingly, the distribution of the mature 60S / LN 60S SNR ratios of nuclear mature and LN 60S-detected targets fits a single Gaussian more poorly than the cytoplasmic targets (**Figure 2J**), suggesting that additional nuclear populations were identified with both 60S templates. To test this prediction and investigate the nuclear pre-60S population further, we generated a third template corresponding to an earlier nuclear intermediate (EN 60S). EN 60S (PDB: 3JCT) retains internally transcribed spacer RNA 2 (ITS2) and associated proteins and has 5S rRNP in a premature state rotated 180° relative to the LN and mature 60S (**Figure 5A**) (Wu et al., 2016). We identified 679 significant EN 60S-detected targets of which 545 (~80%) were also identified with the LN 60S template, and 489 (72%) were also identified with the mature 60S. All of the 489 EN 60S-detected targets identified with the mature 60S were also identified with the LN 60S (**Figure 5A**). 289 (43%) of the EN 60S-detected targets localized to regions of the images corresponding to the nucleus, similar to the 268 nuclear LN 60S-detected targets, while only 390 were cytoplasmic, >3-fold fewer than located with the LN 60S template, consistent with the EN 60S representing a less mature nuclear intermediate (**Figure 5B**). The number and localization of targets identified with 2DTM is consistent with their sequence in the maturation pathway, progressing from EN 60S to LN 60S in the nucleus to mature 60S in the cytoplasm.

Cross-detection of targets by different templates can be used to detect heterogeneity in target populations. When examining the SNR ratios of targets identified by both EN and LN 60S, the cytoplasmic targets display a distribution that is consistent with a single population that more closely resemble the LN 60S template (**Figure 5-figure supplement 1B, red**). The distribution of nuclear targets, however, was consistent with at least two populations (**Figure 5-figure supplement 1B, blue**), each of which is distinct from the cytoplasmic population. This indicated the presence of at least two nuclear populations that differ with respect to their relative similarity to the EN and LN 60S templates.

We next sought to classify the EN, LN and mature 60S-detected targets based on their relative similarity to the three 60S templates. For each target we calculated the $\log_2$(mature 60S / LN 60S SNR) and $\log_2$(EN 60S / LN 60S SNR) values. We used these values to classify each target based on the relative similarity to the three templates using the maximum-likelihood approach discussed above (**Figure 5C**). We found that, consistent with their expected subcellular distributions, targets assigned to the mature 60S population represented 315 (85%) of the cytoplasmic targets and only 1 (<1%) of the nuclear targets detected by all three templates (**Figure 5D**). In contrast, the EN 60S population represents 83 (70%) of the nuclear population and only 4 (~1%) of the cytoplasmic population detected with all three templates (**Figure 5D**). The LN 60S population was roughly evenly distributed between the nucleus and the cytoplasm, consistent with this structure representing a late maturation intermediate (**Figure 5D**).

The NE provides a convenient visual control for the classification of targets as LN / EN 60S or mature 60S (e.g., **Figure 1**). However, there are no clear features in the nucleoplasm that would enable visual separation of different populations of nuclear intermediates and thereby confirm their classification. To validate our classification of the nuclear pre-60S populations, we identified conditions wherein the relative occupancy of the two states would be expected to change. We show above that inhibiting Crm1-mediated export results in accumulation of nuclear intermediates that lack Nog2 (**Figure 4**). In cells with active Crm1, 57% of the nuclear 60S targets are assigned to the EN 60S population (**Figure 5E**). After inhibition of Crm1-mediated export, the EN 60S population is mostly depleted, and >90% of targets are assigned to the LN 60S population (**Figure 5E**). This confirms that 2DTM SNR ratios can be used to effectively classify mixed populations of particles in cells.

Discussion:

The immense potential for cryo-EM to reveal the molecular detail of biological processes in cells is currently largely unrealized. One of the major bottlenecks is the lack of reliable, quantitative methods to locate and characterize molecules in cells. Here we describe the application of 2DTM to *in situ* particle classification. By considering the relative 2DTM SNRs of alternate templates at a single location and orientation, we separate 60S precursors in the nucleus from mature 60S in the cytoplasm. We also show that a maximum likelihood approach effectively classifies a mixed population of nuclear pre-60S into at least two maturation states

with a specified confidence for each particle. We show that 2DTM can be used to probe the composition of complexes *in situ* by modifying 2DTM templates. In this study we extend the utility of 2DTM beyond a binary indicator of detection to provide a quantitative assessment of particle identity.

2DTM enables specific molecule localization in the dense interior of cells

Cryo-FIB milled eukaryotic cells are sufficiently well preserved to allow imaging with cryo-ET (Mahamid et al., 2016) and subtomogram averaging to yield 3D reconstructions at resolutions of >~12Å, e.g., (Schaffer et al., 2019). However, before the present work it was unclear if the milling preserves the high-resolution signal in these samples sufficiently well to allow for particle detection with 2DTM. Our results clearly show that FIB-milling is compatible with molecule localization by 2DTM. This expands the application of 2DTM to previously inaccessible cell types and further demonstrates the utility of 2DTM for *in situ* structural biology.

In many images, 60S subunits detected by 2DTM also generate low-resolution contrast in the cytoplasm that is readily visible (**Figure 1B**, yellow arrows). In the nucleoplasm, the similar density of RNA and DNA impedes the visual identification of all but a few pre-60S (**Figure 1B**, blue arrows). However, the reduced low-resolution contrast does not preclude effective detection of pre-60S with 2DTM. This is in contrast to particle localization in tomograms, wherein detection depends more strongly on low-resolution contrast and recognizable shapes. The ability to distinguish particles in crowded molecular environments is a major advantage of 2DTM relative to cryo-ET, which currently suffers from strong attenuation of high-resolution signal (large B-factors) in the raw tomogram (Schur et al., 2016). 2DTM may enable localization of molecules in other dense environments such as liquid-liquid phase separated granules, which remains challenging for cryo-ET despite success in some cases (Erdmann et al., 2021). Our results confirm that 2DTM is an effective method to localize molecules in dense regions of the cell even when the molecules cannot be distinguished by eye.

***2DTM enables single particle classification* in situ**

In previous work we and others have demonstrated that, when comparing populations of molecules, the average 2DTM SNRs reflect the relative similarity of different templates to the target populations (Lucas et al., 2021; Rickgauer et al., 2020). In this study, we extend this

observation to show that the relative 2DTM SNRs of aligned templates *at a specific location and orientation* can be used to calculate the relative probabilities of a target belonging to a specific particle population.

Of the nuclear targets identified with the mature 60S, ~50% were also detected with the EN 60S, all of which were also detected with the LN 60S (**Figure 5B**). When calculating the relative similarity to the three 60S templates, the EN 60S and mature 60S population were clearly distinct, with mean 2DTM SNR ratios more than three standard deviations apart (**Figure 5C**). The maximum likelihood estimation of Gaussian distributions enables quantitative classification even when particle populations are less distinct, by yielding relative probabilities for each detected target belonging to one of a given number of populations (e.g., **Figures 3&5**).

In this study, we effectively classify at least three populations of 60S maturation states from a population of <500 molecules (**Figure 5**). This means that given sufficient abundance of the target, it will be possible to distinguish populations based on data from a single image (**Figure 2-figure supplement 1D**). This contrasts with more traditional (reference-free) methods used to classify subtomograms and single particles, which require hundreds to thousands of particles to generate the class averages needed for particle assignment. 2DTM allows single molecule classification from fewer images, and therefore enables more information to be extracted from images collected from cells and purified samples (single-particle cryo-EM).

Confidence metric for single particle classification **in situ**

Calculating the confidence in class assignment of individual particles will aid interpretation of the results of 2DTM *in situ*. One major difference between *in situ* cryo-EM and single-particle cryo-EM is the type of biological information that is obtained. In single-particle cryo-EM, the goal is to generate high-resolution maps and establish the arrangement of atoms within a complex in different functional states, and to use this information to discern its molecular mechanism. In this case, B-factors and other metrics can be used to indicate uncertainty about an atomic coordinate, which aids interpretation of the model built into the map. In the cell, each individual instance of a complex may be in a different context relative to other similar molecules. For example, particles might be in different subcellular compartments such as the nucleus or cytoplasm or, as a more extreme example, a single particle within a nuclear pore exists in a very different context than particles in the nucleoplasm. For structural cell biology applications,

405 therefore, it is useful to define a metric to establish the confidence of single particle
406 classification. In this study, we show that a maximum likelihood approach using Gaussian fits to
407 $\log_2$ 2DTM SNR ratios of alternate templates at a specific subcellular location and orientation
408 can be used to calculate the relative probability of a single particle deriving from one of a given
409 number of classes. This provides a quantitative metric to establish confidence in the assignment
410 of single particles that will aid in the biological interpretation of cellular cryo-EM maps.
411
412 ***2DTM templates as computational molecular probes***
413 A major challenge in biological cryo-EM the retrieval of detailed structural information of
414 inherently flexible and heterogeneous macromolecules from noisy images collected at low dose
415 to limit radiation damage. In single particle cryo-EM, this problem is addressed by averaging
416 images of thousands of purified molecules to identify different structural states at high
417 resolution. By averaging images of many identical copies of a particle, novel structures can be
418 discovered, and this is a clear strength of this approach. However, since most complexes have a
419 low abundance in the cell, the utility of this approach for *in situ* structural biology is limited to
420 all but the most abundant complexes.
421 2DTM presents an alternate approach to using the signal in noisy images to gain insight into
422 the structural states of molecules. In this approach, a noise-free template represents a hypothesis
423 that a particle of a given conformational and compositional state is present in the image, and this
424 hypothesis can be tested by searching the image with the template, independent of how many
425 particles the image contains. We demonstrate that by generating modified templates representing
426 different hypotheses, we can directly assess the compositional and conformational states of
427 ribosomal subunits in cells.
428 Provided the templates have similar molecular mass and shape and are aligned with each
429 other, probing with multiple templates requires only a single initial exhaustive search with one of
430 the templates. This can be followed by a simple evaluation of the cross-correlation coefficient for
431 each additional template at locations and orientations of the detected targets in the initial search
432 (**Figure 4**), thereby avoiding time-consuming searches for all templates. In future studies, this
433 approach could be extended to assess the relative similarity of a target with respect to a library of
434 alternate structures. Alternate templates could be generated in multiple ways, depending on the
435 biological hypothesis being tested. To reveal compositional heterogeneity *in situ*, alternate

436 structures could be generated that lack specific subunits of interest as shown in **Figure 4**.
437 Additionally, to interrogate *in situ* conformational heterogeneity, templates could be generated
438 from time points of molecular dynamics simulations.
439
440 ***Addressing potential sources of error***
441 In our study, we used the physical separation of nuclear and cytoplasmic 60S populations
442 to develop and test *in situ* classification of targets by 2DTM. We found that there are several
443 requirements to permit classification of related molecules by 2DTM. First, the molecular models
444 must be aligned relative to one another resulting in a correlation peak at the same pixel in the
445 image. Comparing SNR values resulting from global searches with different templates may be
446 lowered by imperfect, off-grid rotational matches, potentially affecting 2DTM SNR ratios and
447 hence, target classification. Differences in model quality may also affect the 2DTM SNR ratios,
448 masking other differences of interest. In this study, the mature 60S template was generated using
449 the atomic coordinates of the large subunit of the ribosome built into a map with an overall
450 resolution of 3.1 Å (PDB: 6Q8Y) (Tesina et al., 2019). The large subunit of the ribosome is
451 structurally less variable than the small subunit and local resolution estimates suggest that parts
452 of the LSU map extend to ~2.5 Å (Tesina et al., 2019). The maps used to build the EN 60S and
453 LN 60S subunits were reconstructed at 3.08 Å and at 3.5 Å resolution, respectively. The
454 accuracy of the atomic coordinates of a model will depend on the resolution of the underlying
455 density map. Moreover, the greater number of mature ribosome structures, relative to maturation
456 intermediate structures, may provide more confidence in the atomic coordinates of the mature
457 60S. We expect that more accurate coordinates will result in higher 2DTM SNR values, which
458 may affect target classification.
459 The dependence of 2DTM SNR values on the quality of the atomic model presents the
460 possibility to use 2DTM to refine atomic models directly against 2D images of purified samples,
461 and *in situ* against targets detected in images of cells. This approach may bypass some of the
462 difficulties associated with the use of intermediary 3D reconstructions in atomic model
463 refinement, such as inaccurate representation of the full extent of heterogeneity in a dataset and
464 loss of resolution in flexible parts of a molecule. Further development is required to address the
465 potential of overfitting when refining against noisy 2D images, and to detect and quantify errors
466 in the refined models.

The classification of structurally similar targets could be further improved by identifying and controlling the factors that affect the distribution of observed 2DTM SNR ratios for a given set of templates. Ideally, the mean ratio of SNR values for a set of templates and given target depends only on the structural differences between the templates, while the distribution of observed ratios is solely a function of the noise and background in the images. However, factors that contribute to loss of signal such as sample thickness, radiation damage, beam induced motion, charging and movie frame alignment errors due to sample deformation all result in loss of high-resolution signal, making the 2DTM SNR ratios less sensitive to structural differences in the templates and biasing their $\log_2$ values towards 0. Additionally, structural variability in the targets that is not captured by the templates, as well as different degrees of overfitting during 2DTM (Lucas et al., 2021) and a target orientation dependence of the SNR values may lead to a wider spread of observed 2DTM SNR ratios. Further research is required to account for these factors and reduce the variance in 2DTM SNR ratios, thereby enabling classification of targets with smaller structural differences.

Additional intermediate populations

In the present study, we only considered three alternate 60S templates. We note that the Gaussian fits to the 2DTM SNRs of mature 60S and LN 60S-detected nuclear targets is imperfect, potentially indicating additional pre-60S populations (**Figure 2-figure supplement 1C**). Further examination of the observed 2DTM SNR ratios revealed the presence of at least one additional pre-60S population (**Figure 5**). We also observed a small population of cytoplasmic 60S targets with higher SNR values against the LN 60S template than against the mature 60S (**Figure 5D**). 60S maturation intermediates exit the nucleus in an immature form and complete maturation in the cytoplasm. Whether the cytoplasmic 60S with higher SNR values against the LN 60S template represent cytoplasmic intermediates or reflect the limits of our classification strategy requires further investigation. Future work using additional templates representing other intermediates of 60S maturation will reveal further details about the spatiotemporal organization of pre-60S intermediates in cells.

In this study, we identified an EN 60S population of nuclear 60S with the 5S rRNP in a premature state rotated 180° relative to the mature 60S, consistent with *in vitro* determined structures (Leidig et al., 2014). The presence of this complex during maturation *in vivo* has been

difficult to establish. Our observation that this population accounts for more than half of the 60S identified in the nucleus argues that this is an on-pathway assembly intermediate. We also identified a nuclear LN 60S population. This population reflects a late intermediate that has already undergone 5S rotation and ITS2 removal, implying a temporal lag after 5S rotation and/or ITS2 removal, and subsequent export from the nucleus. To test these possibilities more thoroughly, future studies establishing the flux through the assembly pathway are needed. By freezing cells at different time points after inhibition of specific maturation steps, 2DTM could be used to study the kinetics of assembly and the flux through the assembly pathway.

Materials and Methods:

Yeast cell culture and plunge freezing

Saccharomyces cerevisiae strains BY4741 (ATCC), or MNY8 (a gift from Michael Rosbash) colonies were inoculated in 20mL of YPD, diluted 1/5 and grown overnight to an OD_{600} of ~0.5 to 1. The cells were then diluted to 10,000 cells/mL and 3uL applied to a 2/1 or 2/2 Quantifoil 200 mesh Cu grid, allowed to rest for 15 seconds, back-side blotted for 8 seconds and plunged into liquid ethane at -184°C using a Leica EM GP2 plunger. Frozen grids were stored in liquid nitrogen until FIB-milled. When indicated Crm1(T539C) (MNY8 cells, a gift from Michael Rosbash, Brandeis) were additionally incubated at 30°C with shaking in the presence of 200 nM Leptomycin B (Cell Signaling Technologies) for 30 min before applying to grids and plunge freezing.

FIB milling

Grids were transferred to an Aquilos cryo-FIB SEM, sputter coated with metallic Pt for 15s then coated with organo-Pt for 10s and milled in a series of sequential milling steps using a 30kV Ga+ beam using the following protocol: rough milling 1: 0.1 nA rough milling 2: 50 pA lamella polishing: 10 or 30 pA at a stage tilt of 15° (milling angle of 8°).

Cryo-EM data collection

Lamellae were imaged using a Titan Krios 300 keV cryo-TEM (Thermo Fisher) equipped with a K3 direct detector (Gatan) and an energy filter (Gatan) at a sample pixel size of 1.06 Å. Movies were collected at an exposure rate of 1 $e^-/Å^2$ to a total dose of 30 $e^-/Å^2$.

Image processing

Images were processed using *cis*TEM (Grant et al., 2018) as described previously (Lucas et al., 2021), and using sample tilt determination implemented in a modified version of CTFFIND4 (Rohou and Grigorieff, 2015) to estimate sample defocus to account for the ~8° tilt of the lamella introduced during FIB-milling. Images of 3D densities and 2DTM results were prepared in ChimeraX (Pettersen et al., 2021).

2DTM

The molecular models noted in the text were aligned to one another to have the same origin using their 28S rRNA using the MatchMaker function in UCSF Chimera (Meng et al., 2006; Pettersen et al., 2004) and 2DTM templates were generated by simulating 3D densities (Himes and Grigorieff, 2021). 2DTM was performed using the program *match_template* in the *cis*TEM GUI (Lucas et al., 2021) using the default parameters. The coordinates were refined using the program *refine_template* (Lucas et al., 2021) in rotational steps of 0.1° and a defocus range of 200Å with a 10Å step.

3D reconstruction using mature 60S 2DTM coordinates

We used the program *prepare_stack_matchtemplate* (Lucas et al., 2021) to generate a particle stack using the locations and orientations of the significant mature 60S-detected targets after refinement as described above. We then used *cis*TEM to generate a 3D reconstruction from 3991 mature 60S targets detected in 28 images of the nuclear periphery, only including targets with a 2DTM SNR of >8. The reconstruction had a nominal resolution of 3.5 Å using an Fourier Shell Correlation (FSC) threshold of 0.143 (**Figure 1-figure supplement 1D**) (Rosenthal and Henderson, 2003) that is expected to overestimate the resolution due to overfitting (Lucas et al., 2021). To limit the noise due to overfitting, we low-pass filtered the reconstruction to 10 Å, representing an FSC of 0.9.

Calculating 2DTM SNR values and ratios of SNR values

Targets identified in two or more searches with aligned templates were identified using the program *align_coordinates* (Lucas et al., 2021). The 2DTM SNRs of targets identified in two or

more searches were compared by taking the $\log_2$ of the SNR ratio. The $\log_2$ was used in place of the direct ratio because, the shape of the distribution is independent of the order of comparison, except for a mirror around 0, while the distribution of the direct ratios shows more complicated behavior. Histograms of both the $\log_2$ values and direct ratios of the cytoplasmic 60S population have approximately Gaussian distributions with fits characterized by the coefficient of determination R^2=0.993 and R^2=0.991 respectively. To calculate the change in the 2DTM SNR with modified templates, the program *refine_template* (Lucas et al., 2021) was used to calculate 2DTM SNRs for additional templates using the locations and orientations from a previous exhaustive search with an initial template, without performing a rotational search by specifying the rotational step as 360°. To obtain consistent ratios of 2DTM SNRs, the 2DTM SNR values for both the initial template and the additional templates were calculated.

Calculating relative probabilities

Histograms were generated (bin 0.05) of the calculated $\log_2$ 2DTM SNR ratios and Gaussians were fitted using GaussianMixture in sklearn (Pedregosa et al., 2011). Based on the shape of the histogram, we model the $\log_2$ 2DTM SNR ratios as a mixture of K-component multivariate Gaussian distributions, when K templates are used in the search. We fit Gaussians to the $\log_2$ SNR ratios of any two selected templates. Each target i is then associated with $K-1$ such SNR ratios x_i. For example, for $K=4$, we can define the following:

$$X_i = \begin{bmatrix} log_2(SNR_{i,k=1}/SNR_{i,k=2}) \\ log_2(SNR_{i,k=1}/SNR_{i,k=3}) \\ log_2(SNR_{i,k=1}/SNR_{i,k=4}) \end{bmatrix} \quad (1)$$

For particles belonging to the same population (class), the $\log_2$ SNR ratio can be described by the multivariate Gaussian probability density function (PDF):

$$P(X_i|\Theta_k, z_i = k) \sim \mathbb{N}(\mathrm{M}_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_k|^{\frac{1}{2}}} \exp\left(-\frac{(X_i-\mathrm{M}_k)^T{\Sigma_k}^{-1}(X_i-\mathrm{M}_k)}{2}\right) \quad (2)$$

$$P(z_i = k) = \pi_k \quad (3)$$

where X_i is a vector of $K-1$ $\log_2$ SNR ratios, z_i indicates the identity of the target ($k = 1,2,\dots,K$), and $\Theta_k = \{\mathrm{M}_k, \Sigma_k, \pi_k\}$ is the set of parameters of the Gaussian PDF $\mathbb{N}$ and the prior probability that a detected target belonging to class k. The total joint likelihood for N detected targets is then

$$L(\Theta; \mathrm{X}) = P(\mathrm{X}|\Theta) = \prod_{i=1}^{N} P(X_i|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j \mathbb{N}(\mathrm{M}_j, \Sigma_j) \quad (4)$$

with $\Theta = \{\Theta_1, \Theta_2 \dots \Theta_K\}$ and $\mathrm{X} = \{X_1, X_2 \dots X_N\}$. We use an expectation-maximization (EM) algorithm to iteratively calculate the maximum likelihood estimates of the model parameters where the E-step calculates the posterior probability via Bayes rule,

$$P(z_i = k|X_i, \Theta) = \frac{\pi_k \mathbb{N}(\mathrm{M}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathbb{N}(\mathrm{M}_j, \Sigma_j)} \quad (5)$$

and the M-step updates the model parameters for each class,

$$\pi_k = \frac{\sum_{i=1}^{N} P(z_i = k|X_i, \Theta)}{N} \quad (6)$$

$$\mathrm{M}_k = \frac{\sum_{i=1}^{N} X_i \cdot P(z_i = k|X_i, \Theta)}{\sum_{i=1}^{N} P(z_i = k|X_i, \Theta)} \quad (7)$$

$$\Sigma_k = \frac{\sum_{i=1}^{N} P(z_i = k|X_i, \Theta)(X_i - \mathrm{M}_k)(X_i - \mathrm{M}_k)^T}{\sum_{i=1}^{N} P(z_i = k|X_i, \Theta)} \quad (8)$$

Prior probabilities (π) can be set by subjective assessment based on the experiment, or set to $1/K$ where all classes have equal probability. For example, to determine the relative probability that an LN 60S-detected nuclear target belongs to the LN 60S or EN 60S class, we assume that their relative frequencies are the same and therefore the prior probability of the two intermediates in the nucleus is equal: $P(LN\ 60S) = P\ (EN\ 60S) = 0.5$.

Data availability:

Micrographs, templates and scaled maximum intensity projections (MIPs) in this study are accessible with the following public access code: EMPIAR-10998

Acknowledgements:

We are grateful to Michael Rosbash (Brandeis) for providing MNY8 cells, Xiaowei Zhao for help trouble-shooting FIB-milling, and Peter Rickgauer, Tim Grant and Ben Himes for helpful comments and suggestions. BL and NG acknowledge support from the Chan Zuckerberg Initiative (VPI-0000000116).

References:

Delavoie F, Soldan V, Rinaldi D, Dauxois JY, Gleizes PE. 2019. The path of pre-ribosomes through the nuclear pore complex revealed by electron tomography. *Nature Communications* **10**:1–12. doi:10.1038/s41467-019-08342-7

Erdmann PS, Hou Z, Klumpe S, Khavnekar S, Beck F, Wilfling F, Plitzko JM, Baumeister W. 2021. In situ cryo-electron tomography reveals gradient organization of ribosome biogenesis in intact nucleoli. *Nature Communications* **12**:1–9. doi:10.1038/s41467-021-25413-w

Grant T, Rohou A, Grigorieff N. 2018. CisTEM, user-friendly software for single-particle image processing. *Elife* **7**:e35383. doi:10.7554/eLife.35383

Gubins I, Chaillet ML, van der Schot G, Veltkamp RC, Förster F, Hao Y, Wan X, Cui X, Zhang F, Moebel E, Wang X, Kihara D, Zeng X, Xu M, Nguyen NP, White T, Bunyak F. 2020. SHREC 2020: Classification in cryo-electron tomograms. *Computers & Graphics* **91**:279–289. doi:10.1016/J.CAG.2020.07.010

Himes B, Grigorieff N. 2021. Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *urn:issn:2052-2525* **8**:943–953. doi:10.1107/S2052252521008538

Himes BA, Zhang P. 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* **15**:955–961. doi:10.1038/s41592-018-0167-z

Ho JHN, Kallstrom G, Johnson AW. 2000. Nmd3p Is a Crm1p-Dependent Adapter Protein for Nuclear Export of the Large Ribosomal Subunit. *The Journal of Cell Biology* **151**:1057. doi:10.1083/JCB.151.5.1057

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Kürner J, Frangakis AS, Baumeister W. 2005. Cryo-electron tomography reveals the cytoskeletal structure of Spiroplasma melliferum. *Science* **307**:436–438. doi:10.1126/science.1104031

Leidig C, Thoms M, Holdermann I, Bradatsch B, Berninghausen O, Bange G, Sinning I, Hurt E, Beckmann R. 2014. 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nature Communications* **5**:3491. doi:10.1038/ncomms4491

660 Lucas BA, Himes BA, Xue L, Grant T, Mahamid J, Grigorieff N. 2021. Locating
661 macromolecular assemblies in cells by 2D template matching with cisTEM. *Elife* **10**.
662 doi:10.7554/elife.68946
663 Lučić V, Rigort A, Baumeister W. 2013. Cryo-electron tomography: The challenge of doing
664 structural biology in situ. *Journal of Cell Biology* **202**:407–419.
665 doi:10.1083/JCB.201304193
666 Mahamid J, Pfeffer S, Schaffer M, Villa E, Danev R, Cuellar LK, Förster F, Hyman AA, Plitzko
667 JM, Baumeister W. 2016. Visualizing the molecular sociology at the HeLa cell nuclear
668 periphery. *Science* **351**:969–972. doi:10.1126/science.aad8857
669 Marko M, Hsieh C, Schalek R, Frank J, Mannella C. 2007. Focused-ion-beam thinning of
670 frozen-hydrated biological specimens for cryo-electron microscopy. *Nature Methods* **4**:215–
671 217. doi:10.1038/nmeth1014
672 Matsuo Y, Granneman S, Thoms M, Manikas RG, Tollervey D, Hurt E. 2014. Coupled GTPase
673 and remodelling ATPase activities form a checkpoint for ribosome export. *Nature* **505**:112–
674 116. doi:10.1038/nature12731
675 Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. 2006. Tools for integrated sequence-
676 structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**:339. doi:10.1186/1471-
677 2105-7-339
678 Moebel E, Martinez-Sanchez A, Lamm L, Righetto RD, Wietrzynski W, Albert S, Larivière D,
679 Fourmentin E, Pfeffer S, Ortiz J, Baumeister W, Peng T, Engel BD, Kervrann C. 2021.
680 Deep learning improves macromolecule identification in 3D cellular cryo-electron
681 tomograms. *Nature Methods* 1–9. doi:10.1038/s41592-021-01275-4
682 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
683 P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,
684 Duchesnay E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine*
685 *Learning Research* **12**:2825–2830.
686 Pertschy B, Saveanu C, Zisser G, Lebreton A, Tengg M, Jacquier A, Liebminger E, Nobis B,
687 Kappel L, van der Klei I, Högenauer G, Fromont-Racine M, Bergler H. 2007. Cytoplasmic
688 Recycling of 60S Preribosomal Factors Depends on the AAA Protein Drg1. *Molecular and*
689 *Cellular Biology* **27**:6581–6592. doi:10.1128/mcb.00668-07
690 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004.
691 UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of*
692 *Computational Chemistry* **25**:1605–1612. doi:10.1002/JCC.20084
693 Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE.
694 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers.
695 *Protein Science : A Publication of the Protein Society* **30**:70. doi:10.1002/PRO.3943
696 Rickgauer JP, Choi H, Lippincott-Schwartz J, Denk W. 2020. Label-free single-instance protein
697 detection in vitrified cells. *bioRxiv* 2020.04.22.053868. doi:10.1101/2020.04.22.053868
698 Rickgauer JP, Grigorieff N, Denk W. 2017. Single-protein detection in crowded molecular
699 environments in cryo-EM images. *Elife* **6**:e25648. doi:10.7554/elife.25648
700 Rigort A, Bäuerlein FJB, Villa E, Eibauer M, Laugks T, Baumeister W, Plitzko JM. 2012.
701 Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proc*
702 *Natl Acad Sci U S A* **109**:4449–4454. doi:10.1073/pnas.1201333109
703 Rohou A, Grigorieff N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron
704 micrographs. *Journal of Structural Biology* **192**:216–221. doi:10.1016/J.JSB.2015.08.008

Rosenthal PB, Henderson R. 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology* **333**:721–745. doi:10.1016/J.JMB.2003.07.013

Schaffer M, Pfeffer S, Mahamid J, Kleindiek S, Laugks T, Albert S, Engel BD, Rummel A, Smith AJ, Baumeister W, Plitzko JM. 2019. A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue. *Nature Methods* **16**:757–762. doi:10.1038/s41592-019-0497-5

Schur FKM, Obr M, Hagen WJH, Wan W, Jakobi AJ, Kirkpatrick JM, Sachse C, Kräusslich HG, Briggs JAG. 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**:506–508. doi:10.1126/science.aaf9620

Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J. 2021. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* **18**:186–193. doi:10.1038/s41592-020-01054-7

Tesina P, Heckel E, Cheng J, Fromont-Racine M, Buschauer R, Kater L, Beatrix B, Berninghausen O, Jacquier A, Becker T, Beckmann R. 2019. Structure of the 80S ribosome–Xrn1 nuclease complex. *Nature Structural & Molecular Biology* **26**:275–280. doi:10.1038/s41594-019-0202-5

Vilas JL, Oton J, Messaoudi C, Melero R, Conesa P, Ramirez-Aportela E, Mota J, Martinez M, Jimenez A, Marabini R, Carazo JM, Vargas J, Sorzano COS. 2020. Measurement of local resolution in electron tomography. *Journal of Structural Biology* **4**:100016. doi:10.1016/J.YJSBX.2019.100016

Villa E, Schaffer M, Plitzko JM, Baumeister W. 2013. Opening windows into the cell: focused-ion-beam milling for cryo-electron tomography. *Current Opinion in Structural Biology* **23**:771–777. doi:10.1016/J.SBI.2013.08.006

von der Haar T. 2008. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Systems Biology* **2**:87. doi:10.1186/1752-0509-2-87

Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**:437–440. doi:10.1016/S0968-0004(99)01460-7

Wu S, Tutuncuoglu B, Yan K, Brown H, Zhang Y, Tan D, Gamalinda M, Yuan Y, Li Z, Jakovljevic J, Ma C, Lei J, Dong MQ, Woolford JL, Gao N. 2016. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**:133–137. doi:10.1038/nature17942

Xue L, Lenz S, Zimmermann-Kogadeeva M, Tegunov D, Cramer P, Bork P, Rappsilber J, Mahamid J. 2021. Visualizing translation dynamics at atomic detail inside a bacterial cell. *bioRxiv* 2021.12.18.473270. doi:10.1101/2021.12.18.473270

Zhou Y, Musalgaonkar S, Johnson AW, Taylor DW. 2019. Tightly-orchestrated rearrangements govern catalytic center assembly of the ribosome. *Nature Communications* **10**. doi:10.1038/s41467-019-08880-0

746 **Figure Legends:**

747

748 **Figure 1: Detection of cytoplasmic mature 60S and mitochondrial ribosomes in 2D images**
749 **of FIB-milled yeast lamella.** A) Cryo-EM like density generated using the atomic coordinates of
750 PDB:6Q8Y that correspond with the mature 60S. B) TEM image of the nuclear periphery from a
751 FIB-milled yeast lamella. Yellow arrows indicate low-resolution features in the cytoplasm that
752 may indicate the presence of ribosomes. Blue arrows indicate regions of similar size and contrast
753 in the nucleoplasm. NE: nuclear envelope; NPC: nuclear pore complex. C) Cryo-EM micrograph
754 of yeast nuclear periphery from FIB-milled lamella with the results from a 2DTM search using
755 the mature 60S template. Significant targets are indicated by mapping the template in the best
756 matching locations and orientations (shown in yellow). The red box indicates the regions
757 highlighted in E) and F). Scale bar = 50 nm. D) Bar chart indicating the number of mature 60S-
758 detected targets identified in the indicated subcellular compartments in 28 images of the nuclear
759 periphery. E) Scaled maximum intensity projection (MIP) showing the results of 2DTM using
760 the template in A) in the region of C) indicated in red. F) 3D slab indicating the locations and
761 orientations of mature 60S-detected targets in the indicated region of C). The red polypeptide
762 indicates the location of the polypeptide exit tunnel on each 60S. G) Plot showing the density of
763 mature 60S in the regions of the images corresponding to the cytoplasm. Each dot represents a
764 different image. The solid black bar indicates the median. H) 10 Å filtered 3D reconstruction
765 calculated from 3991 60S subunits at the locations and orientations detected in 28 images,
766 showing clear density for the 40S small subunit. The molecular model of the 60S used to
767 generate the template in A) is shown in yellow.

768

769 **Figure 2: 2DTM SNRs differentiate cytoplasmic mature 60S from nuclear pre-60S in 2D**
770 **images of FIB-milled yeast lamella.** A) Diagram showing the compositional changes that
771 accompany the maturation from the late nuclear (LN) 60S (PDB: 6N8J), shown in blue, to the
772 mature 60S (PDB: 6Q8Y), shown in yellow, in the cytoplasm. B) Cryo-EM micrograph of yeast
773 nuclear periphery from FIB-milled lamella with the results from a 2DTM search using the LN
774 60S template. Significant targets are indicated by mapping a projection of the template in the
775 best matching locations and orientations (shown in blue). Scale bar = 50 nm. C) As in B),
776 showing the results from a 2DTM search of the indicated image using the mature 60S as a

777 template (yellow). D) Maximum intensity projection showing the results of a 2DTM search with
778 the LN 60S template in the region of the image in B) highlighted in orange. Orange circles
779 indicate two targets identified by both LN 60S and mature 60S. E) As in D) showing the results
780 of a 2DTM search with the mature 60S template in the region of the image in C) highlighted in
781 orange. F) As in D) corresponding to the region of B) highlighted in blue and circles indicating
782 two LN 60S-detected targets. G) As in D) showing the results of a 2DTM search with the mature
783 60S template in the region of the image in C) highlighted in blue. H) Diagram indicating the
784 number of mature 60S (yellow) and LN 60S (blue)-detected targets identified in 2DTM searches
785 of 28 images of the nuclear periphery. The overlap of the Venn diagram indicates the number of
786 targets identified in both searches. I) Bar chart indicating the number of targets detected by the
787 mature 60S (yellow), the LN 60S (blue), and by both (black) in regions of the images
788 corresponding to the nucleus or cytoplasm. J) Plot showing the $\log_2$ 2DTM SNR ratios for LN
789 and mature 60S-detected targets grouped by subcellular compartment. Each dot indicates a 60S
790 detected in both searches. ****: $p<0.0001$. K) Image showing the identified targets color-coded
791 by the best-matching template (blue: LN 60S, yellow: mature 60S) as determined by the higher
792 2DTM SNR at each overlapping location. Scale bar = 50 nm. L) Pie chart indicating the
793 proportion of all nuclear (left) and cytoplasmic (right) 60S targets that more closely resemble the
794 mature 60S (yellow) or LN 60S (blue) template, respectively, as determined by the highest
795 2DTM SNR at each identified location and orientation.
796
797 **Figure 3: Relative probability of detecting mature or LN 60S.** A) Histogram showing the
798 distribution of the mature 60S / LN 60S 2DTM SNR ratios for each LN and mature 60S-detected
799 target fit with two Gaussians indicating populations 1 (blue dashed line) and 2 (red dashed line).
800 The black line indicates the sum of the two Gaussians, $R^2= 0.993$. B) Line graph showing the
801 probability that a given target belongs to the LN 60S population (blue) line, or mature 60S
802 population (red), as a function of $\log_2$ 2DTM SNR ratio. C) Line graph showing the fraction of
803 nuclear (blue) and cytoplasmic (red) targets classified as LN 60S, at the indicated confidence
804 intervals determined using Eq (5). D) Heat map showing the probability of each LN and mature
805 60S-detected target belonging to either the LN or mature 60S populations. Each row indicates a
806 detected target, and the rows are sorted by their subcellular distribution. The targets assigned to

807 the mature 60S population are indicated in yellow and the targets assigned to the LN 60S
808 population are indicated in blue.
809
810 **Figure 4: Classification of cytoplasmic mature 60S and nuclear pre-60S by 2DTM**
811 **corresponds with biologically relevant differences in the templates.** A) The LN 60S (blue)
812 and mature 60S (yellow) 2DTM templates aligned in UCSF Chimera. B) LN 60S with difference
813 map calculated using UCSF Chimera showing the density in the LN 60S template that is not
814 present in the mature 60S template (red, transparent). C) As in B), showing the mature 60S with
815 density that is not in common with the LN 60S template (red, transparent). D) Boxplots showing
816 the change in 2DTM SNR when only RNA (left) or protein (right) components of the LN 60S
817 template are included, relative to the full-length template for each significant target. The targets
818 are grouped by their subcellular localization. E) Upper: LN 60S template with all ribosome
819 biogenesis factors (RBFs) indicated in red. Lower: Boxplot showing the change in the 2DTM
820 SNR of the nuclear (blue) and cytoplasmic (red) targets when all RBFs are removed, relative to
821 the full-length LN 60S template in untreated cells, and when Crm1-mediated nuclear export is
822 inhibited by treating Crm1(T539C) cells with Leptomycin B (LepB). Box width indicates the
823 interquartile range, the central line indicates the median and the whiskers indicate the range of
824 95% of the targets. F) As in E), for RBF Nog1. G) As in E), for RBF Nog2. ****: $p<0.0001$, ns:
825 not significant ($p>0.05$).
826
827 **Figure 5: Classification of nuclear targets by relative similarity to early or late nuclear**
828 **intermediates.** A) Venn diagram showing the number of significant targets detected in 2DTM
829 searches with the indicated templates. Overlap indicates targets identified in two or more
830 searches. B) Venn diagrams showing the number of significant targets detected in 2DTM
831 searches with the indicated templates in the nucleus (left) and cytoplasm (right). C) Scatterplot
832 showing the EN 60S / LN 60S 2DTM SNR ratios relative to the mature 60S / LN 60S 2DTM
833 SNR ratios for each EN, LN and mature 60S-detected target. Ellipses indicate the fits of three
834 Gaussians and each concentric ellipse indicates one standard deviation from the mean. Each
835 target is colored according to its most likely class membership. D) Heat map showing the
836 probability of each of the targets examined in C) belonging to one of the populations, EN, LN or
837 mature 60S. Targets are grouped by their subcellular localization, followed by their classification

as EN 60S (purple), LN 60S (light blue), or mature 60S (yellow). E) Bar chart showing the proportion of the LN 60S-detected targets in the indicated cells that are classified as LN 60S (blue), mature 60S (yellow) or EN 60S (purple). F) Cryo-EM micrograph of the yeast nuclear periphery from a FIB-milled lamella shown in in **Figure 1,** displaying the results of 2DTM searches, colored by their classification as mature 60S (yellow), LN 60S (blue) or EN 60S (purple) based on their relative 2DTM SNRs.

Figure 1 — figure supplement 1: A) FIB-image of two yeast cells frozen on a cryo-EM grid. B) FIB image of the lamella after milling the cells shown in A). C) Survival histogram showing the number of search locations with 2DTM SNR values above a given threshold from a 2DTM search using the mature 60S template in **Figure 1A**. D) FSC obtained for the 3D reconstruction shown in **Figure 1H** calculated using the targets identified by 2DTM.

Figure 2 — figure supplement 1: A) Venn diagrams showing the number of mature (yellow) and LN 60S (blue) detected targets in the indicated subcellular compartments. The overlap indicates targets detected in searches with both templates. B) Violin plot showing the kernelled distribution of 2DTM SNRs of mature 60S-detected targets (left) and LN 60S-detected targets (right) in the indicated subcellular compartments. ****: $p<0.0001$, ns: not significant, $p>0.05$. C) Histogram showing the relative frequency of mature 60S / LN 60S 2DTM SNR ratios grouped by subcellular localization. Gaussian fits are indicated by a solid line. D) Boxplot showing the mature 60S / LN 60S 2DTM SNR ratios of the nuclear (blue) and cytoplasmic (red) populations from each of the 28 images analyzed, indicating that the nuclear and cytoplasmic populations are distinct, even within single images.

Figure 3 — figure supplement 1: A) Scatterplot showing the 2DTM SNRs for nuclear (blue) and cytoplasmic (red) targets detected in searches with the LN and mature 60S templates. B) Scatterplot showing P(LN 60S) for nuclear (blue) or cytoplasmic (red) LN 60S-detected target as a function of the 2DTM SNR. Dotted line indicates the 2DTM threshold.

Figure 4 — figure supplement 1: A) TEM image of the nuclear periphery and vacuole in Crm1 (T539C) cells treated with Leptomycin B, overlaid with LN 60S-detected targets in blue. Scale bar: 50 nm. B) Bar chart showing the number of LN and mature 60S-detected targets in the indicated subcellular compartments. C) Violin plot showing the kernelled distribution of 2DTM SNRs for LN 60S-detected targets in the indicated subcellular compartment. ****: $p<0.0001$. D) As in C), showing mature 60S-detected targets. ns: not significant, $p>0.05$. E) Violin plot showing the kernelled distribution of 2DTM SNR ratios of targets identified as both LN and mature 60S-detected targets in the indicated subcellular compartment. ****: $p<0.0001$.

Figure 5 — figure supplement 1: A) Violin plots showing the kernelled distribution of 2DTM SNRs for EN 60S-detected targets in the indicated subcellular compartment. ****: $p<0.0001$. B) Histogram showing the distribution of EN 60S / LN 60S SNR ratios of EN and LN 60S-detected targets in untreated cells. Targets are grouped by their subcellular distribution. Gaussian fits are indicated in solid colors. C) TEM image of the nuclear periphery shown in **Figure 1**, overlaid

881 with EN 60S detected targets in purple. D) TEM image of the nuclear periphery. Scale bar
882 indicates 50 nm. E) The image in D) is shown overlaid with EN 60S-detected targets in purple,
883 F) LN 60S-detected targets in blue or G) mature 60S-detected targets in yellow. H) As in E)
884 showing the results of 2DTM searches, colored by their classification as mature 60S (yellow),
885 LN 60S (blue) or EN 60S (purple) based on their relative 2DTM SNRs
886

887

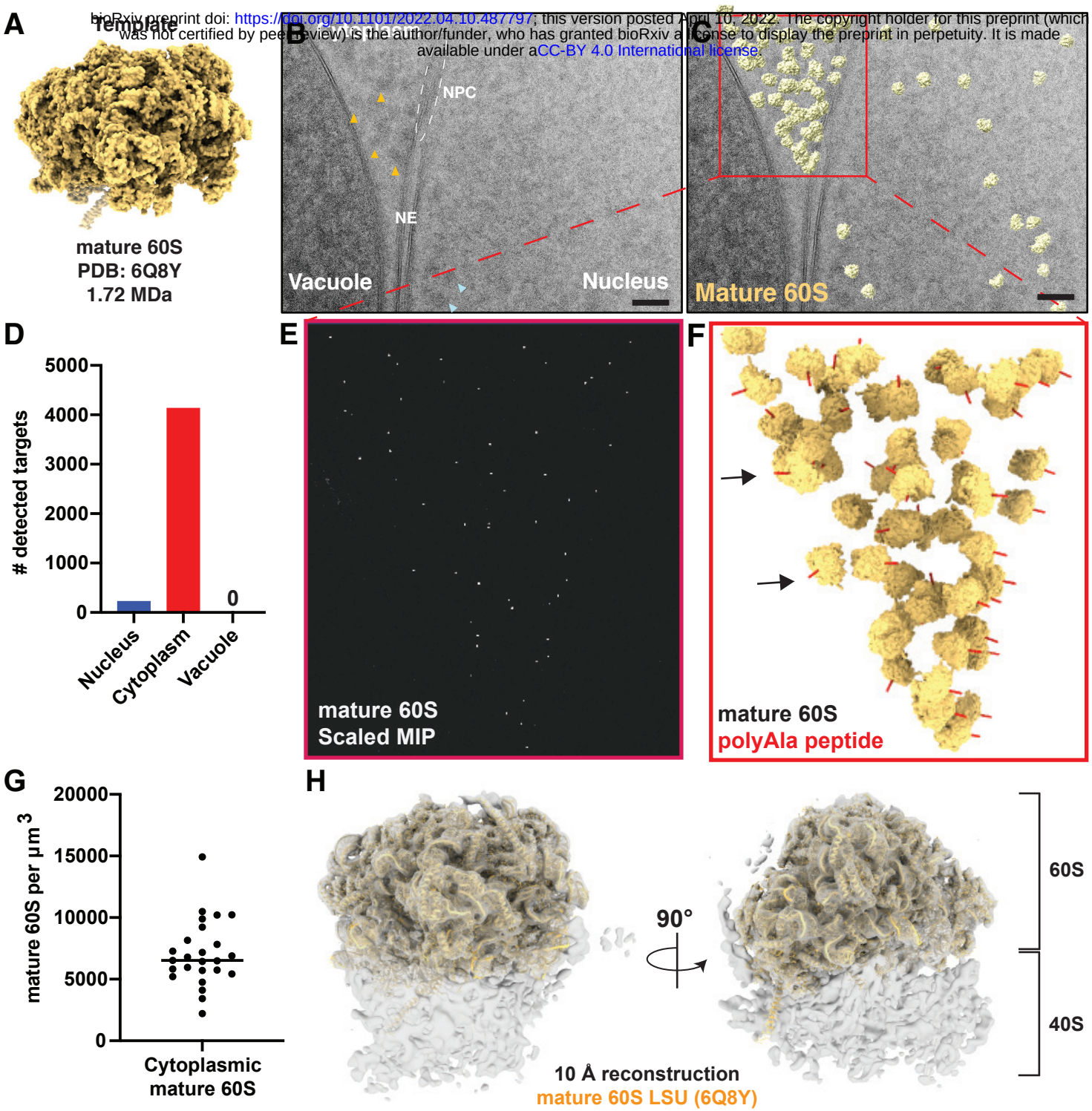


Figure 1

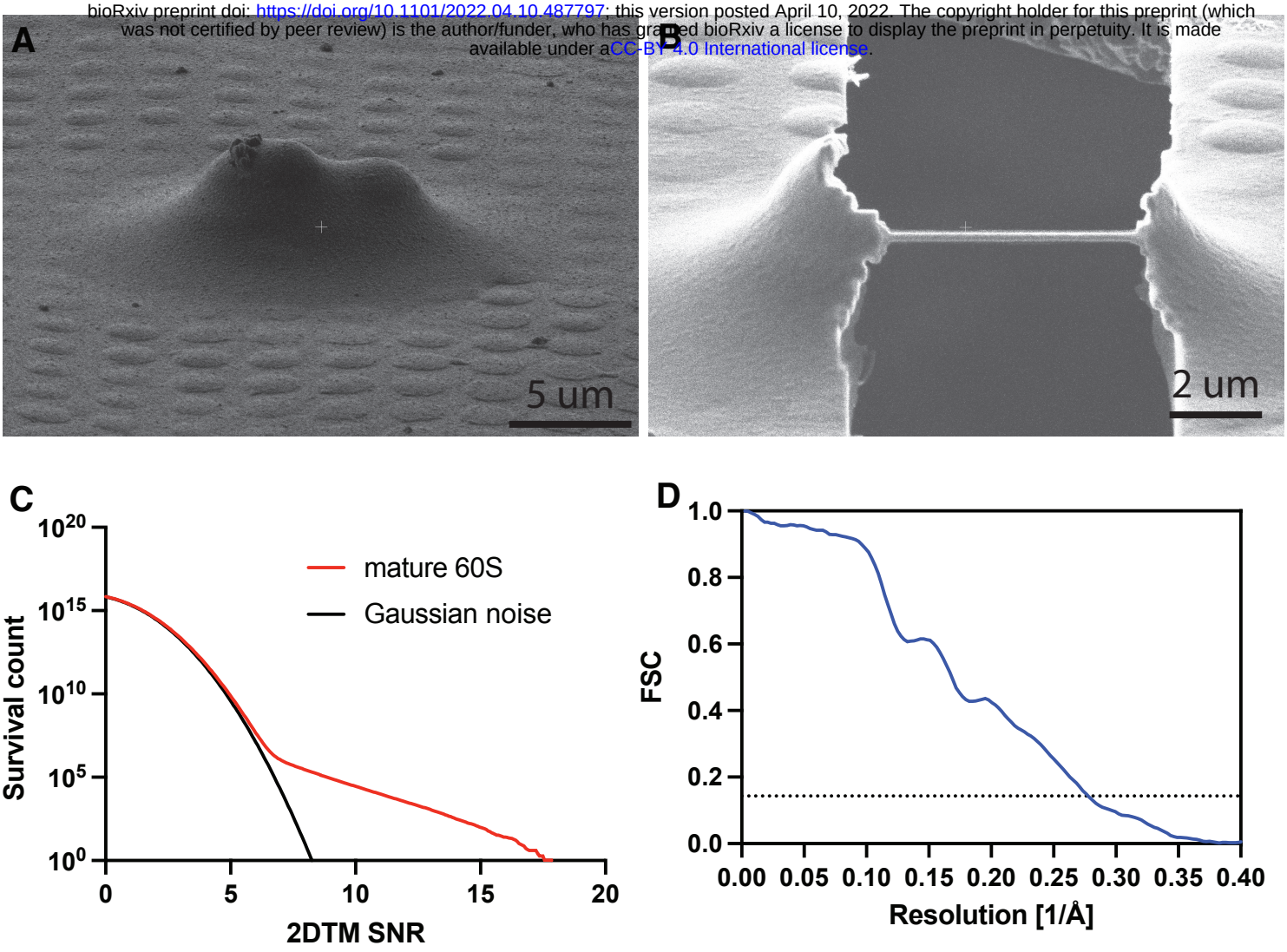


Figure 1 - figure supplement 1

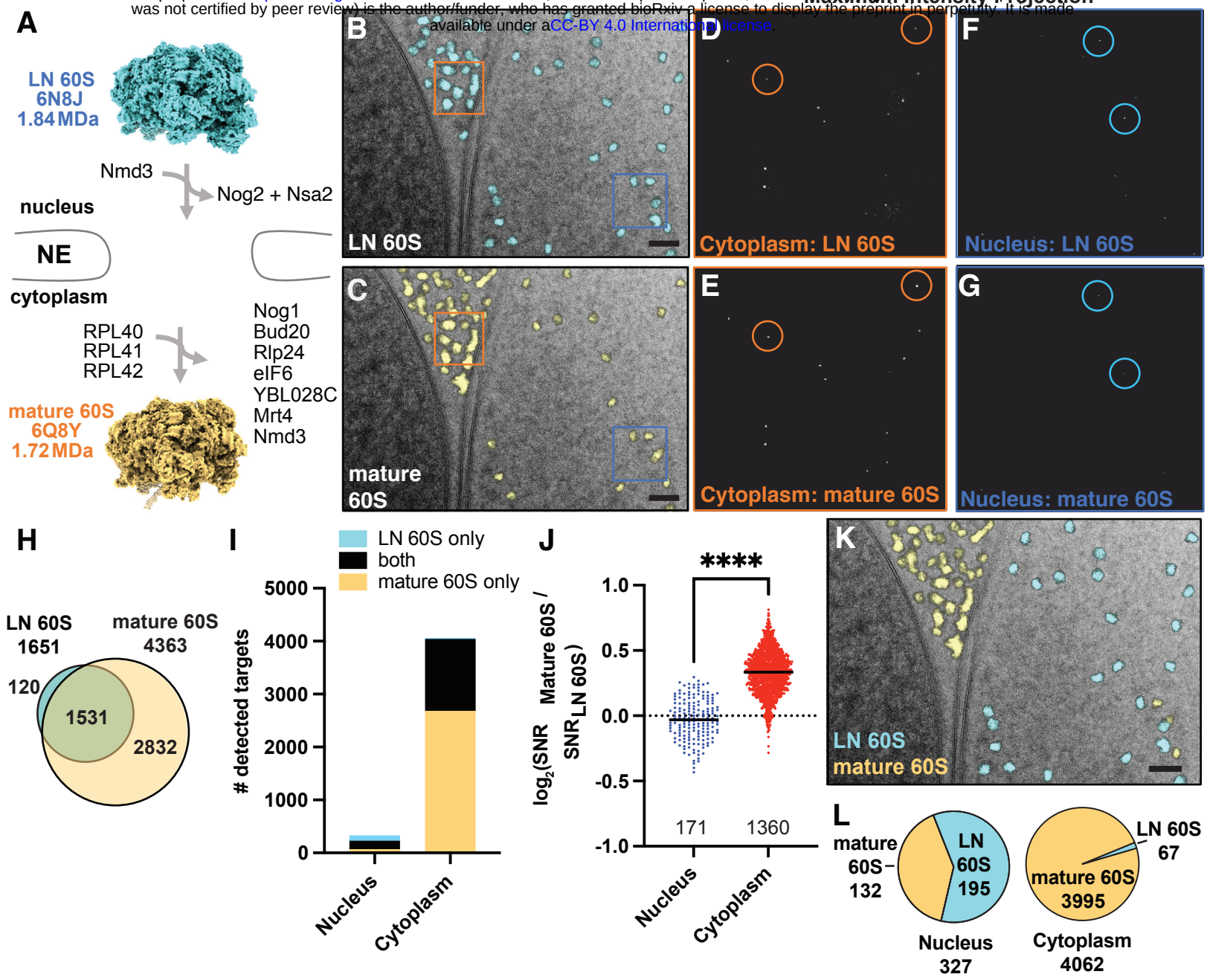


Figure 2

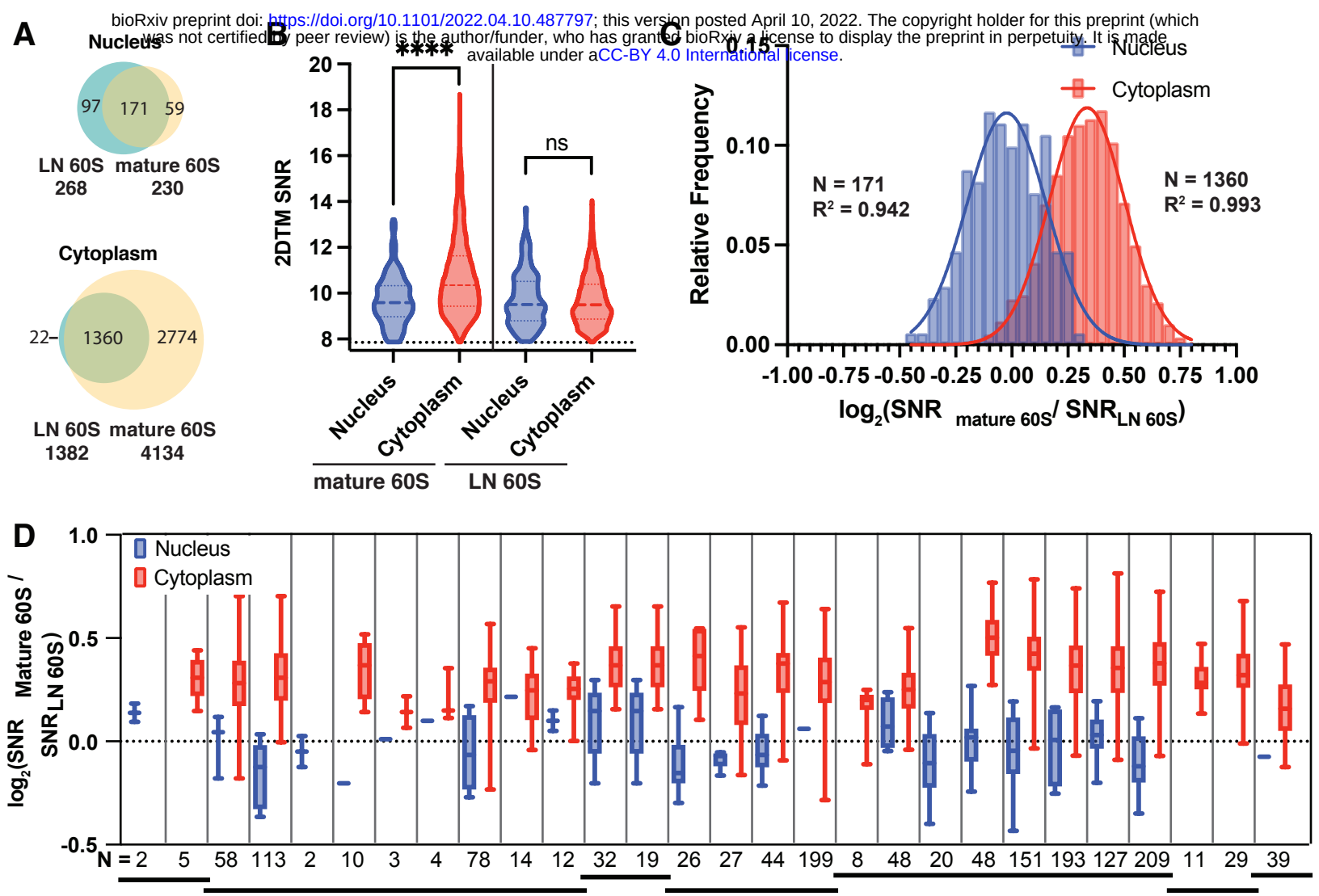


Figure 2 - figure supplement 1

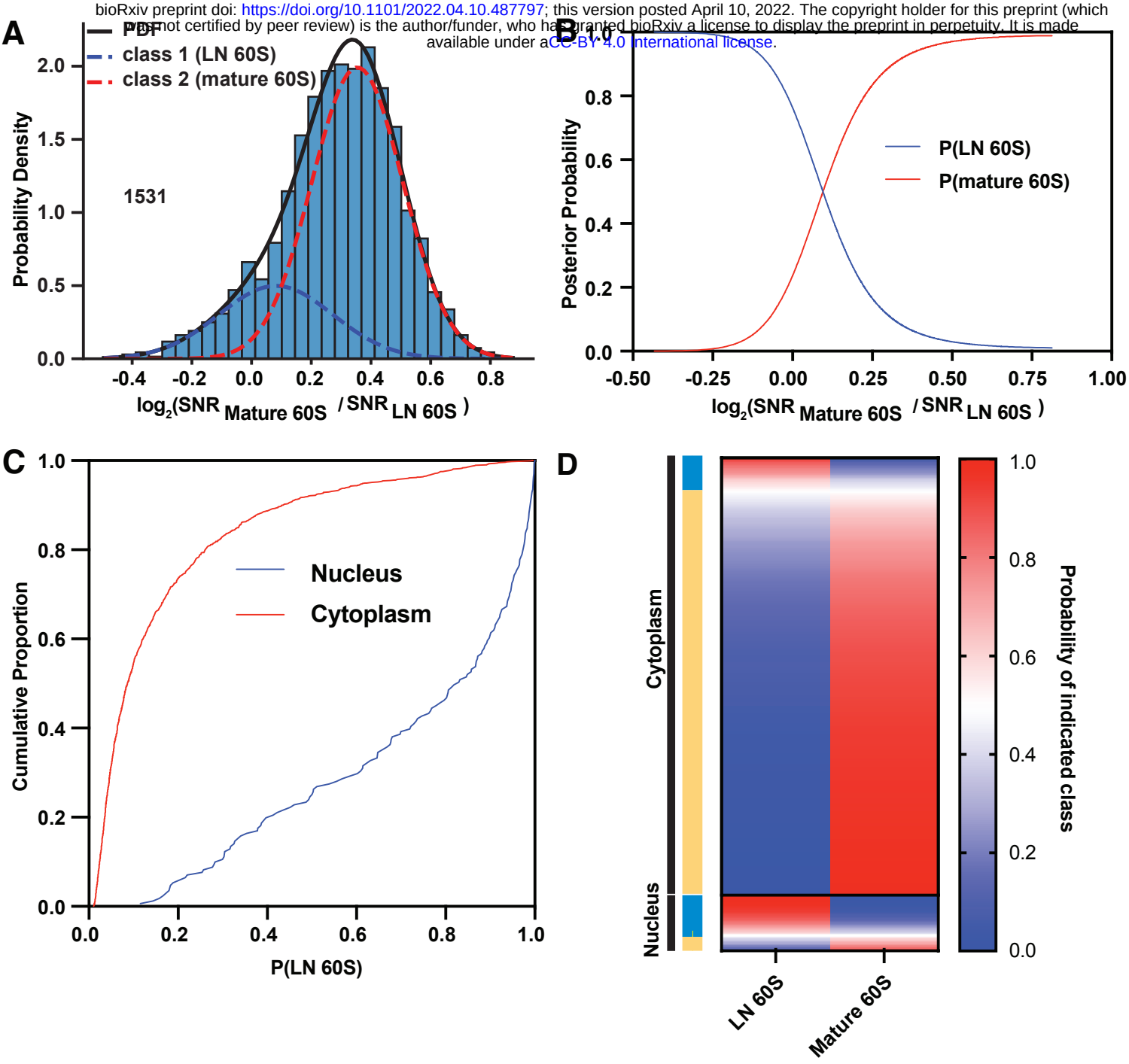


Figure 3

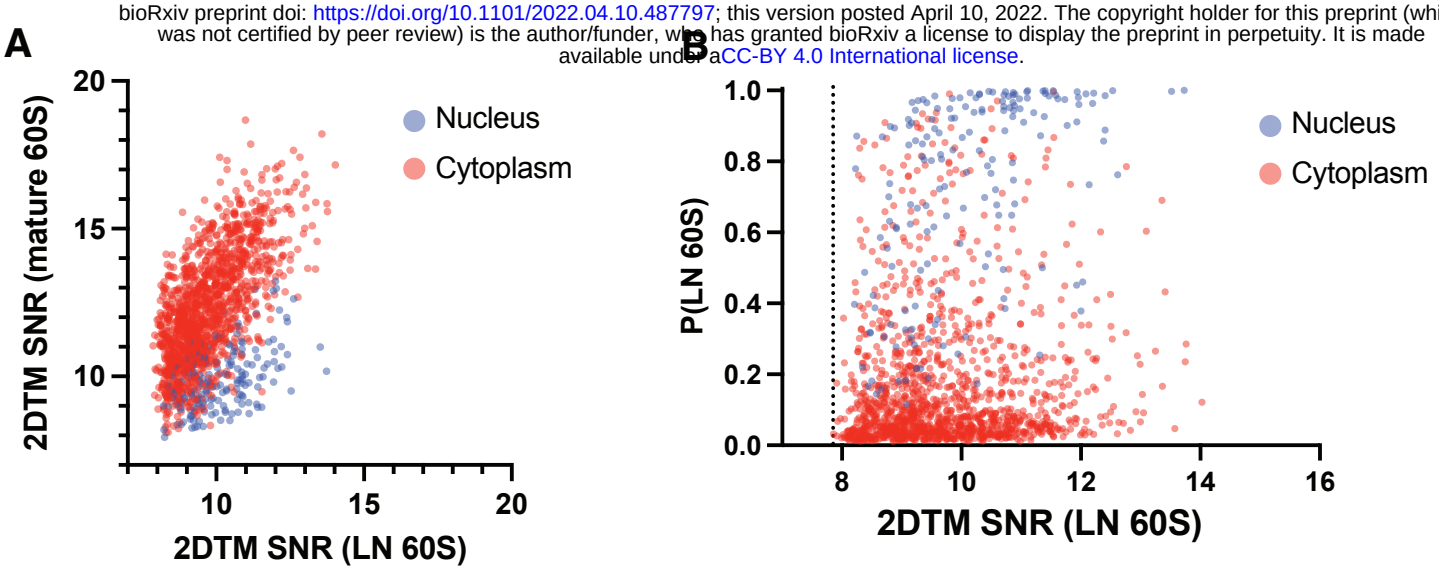


Figure 3 - figure supplement 1

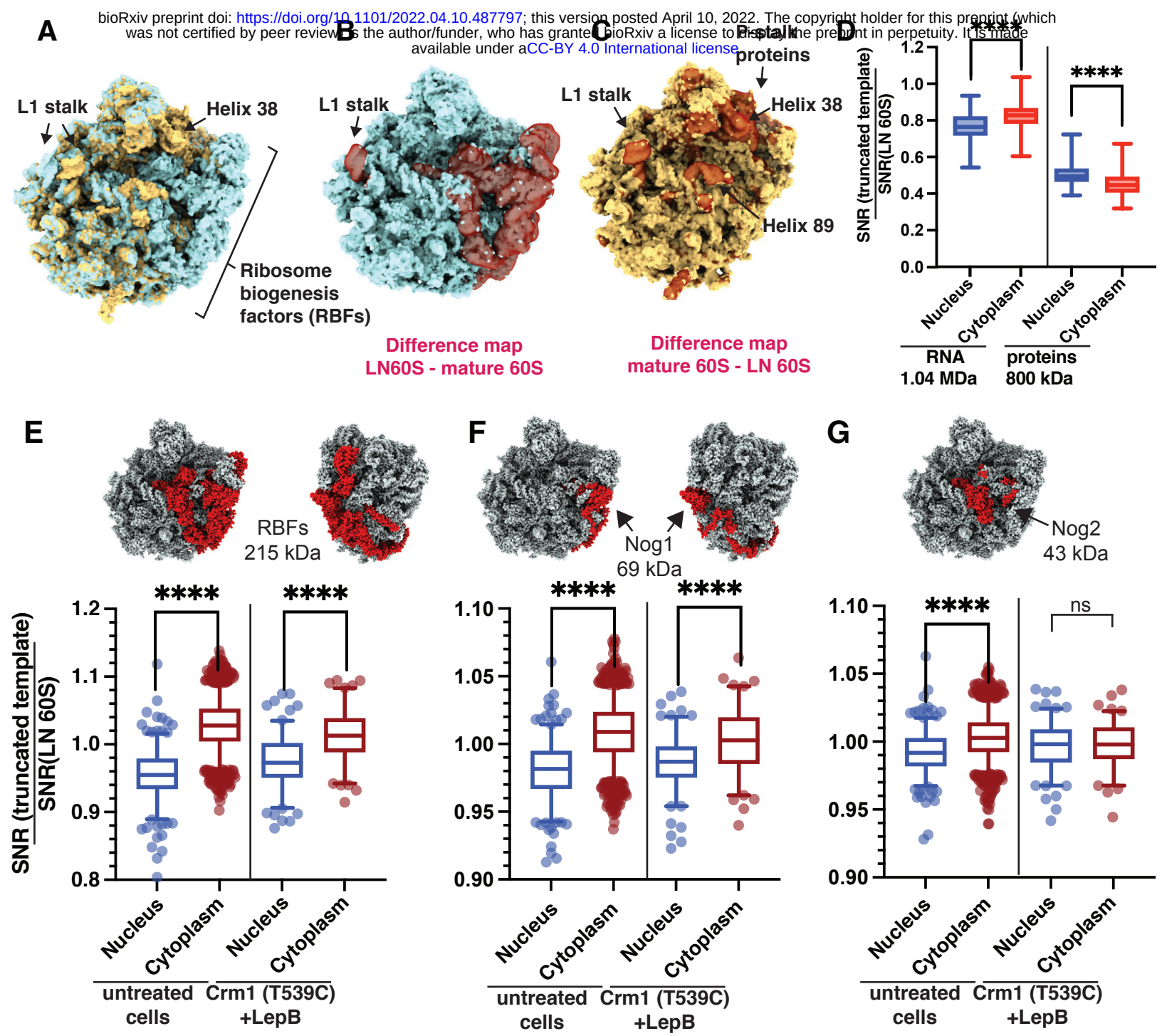
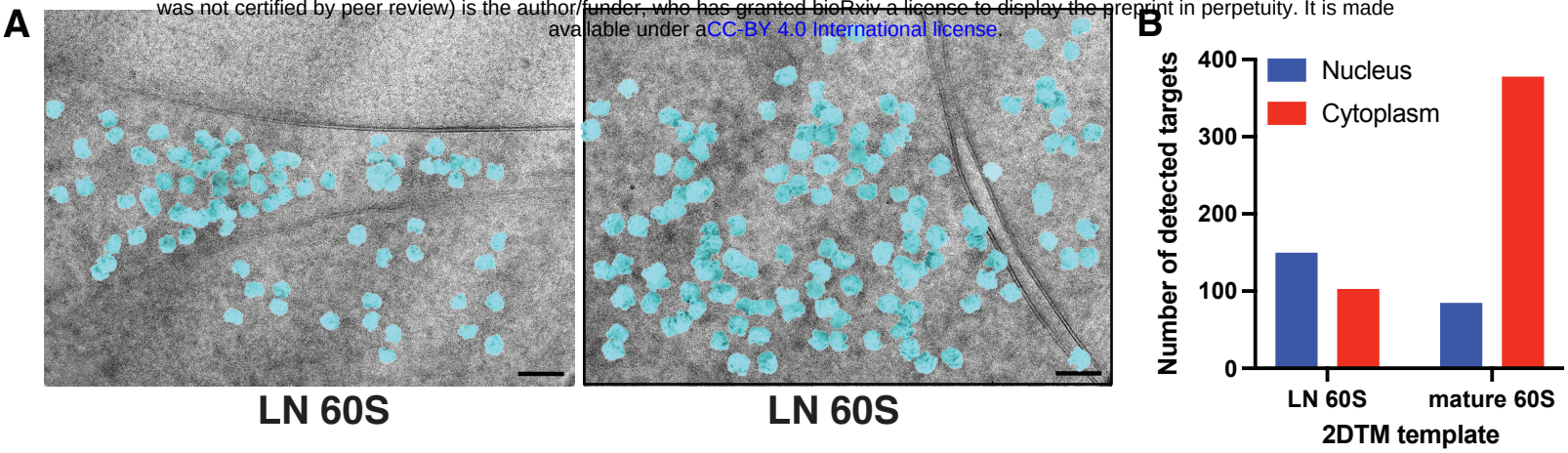
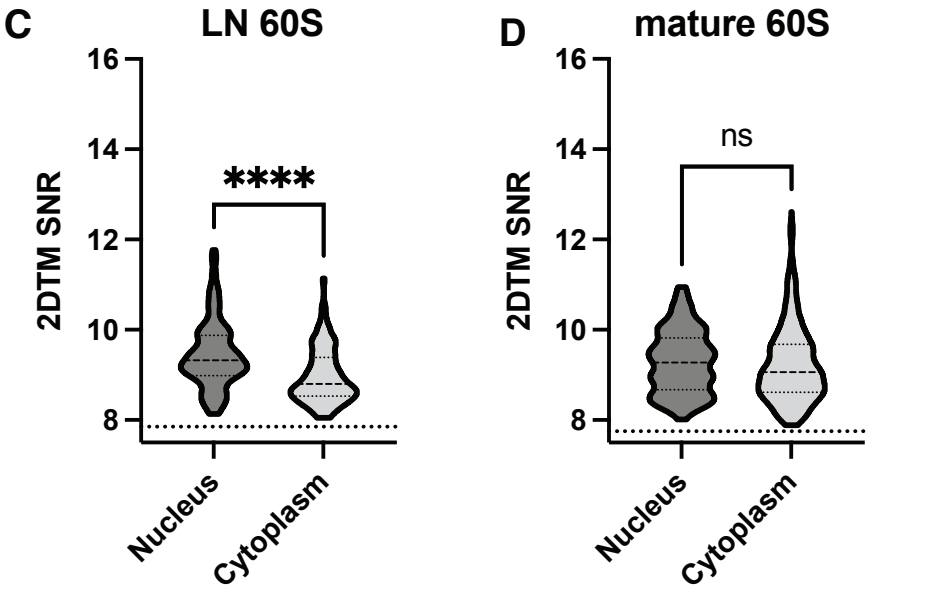


Figure 4





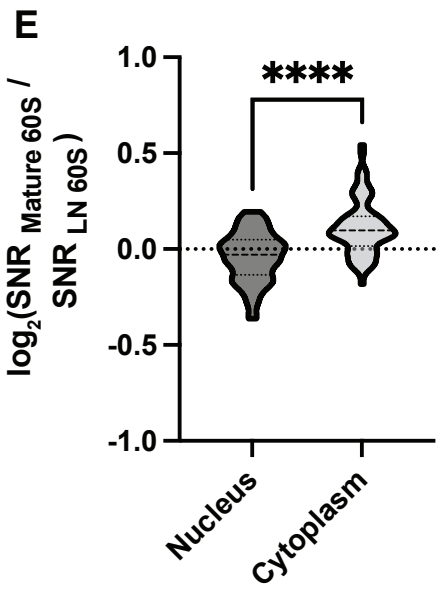


Figure 4 - figure supplement 1

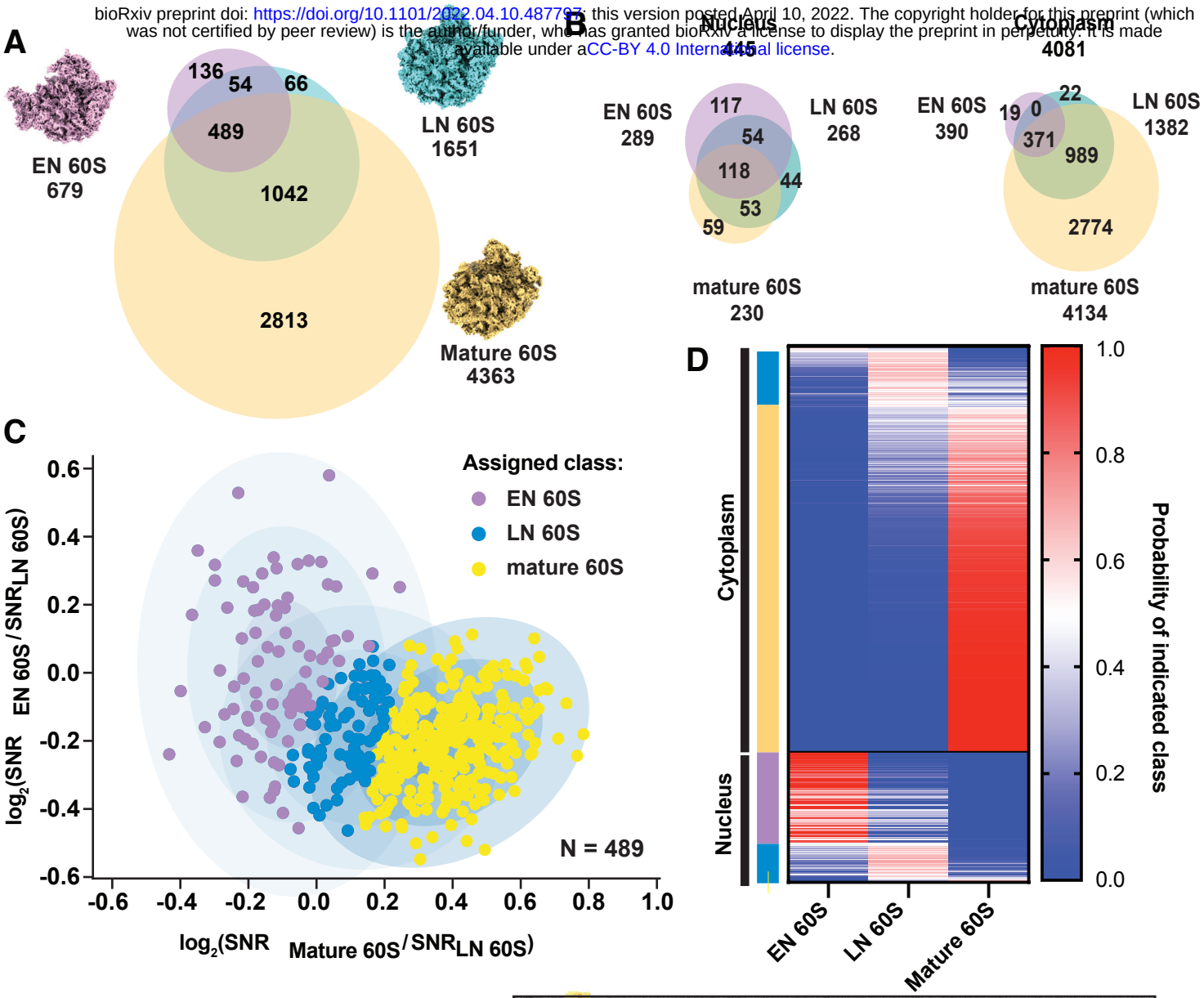


Figure 5

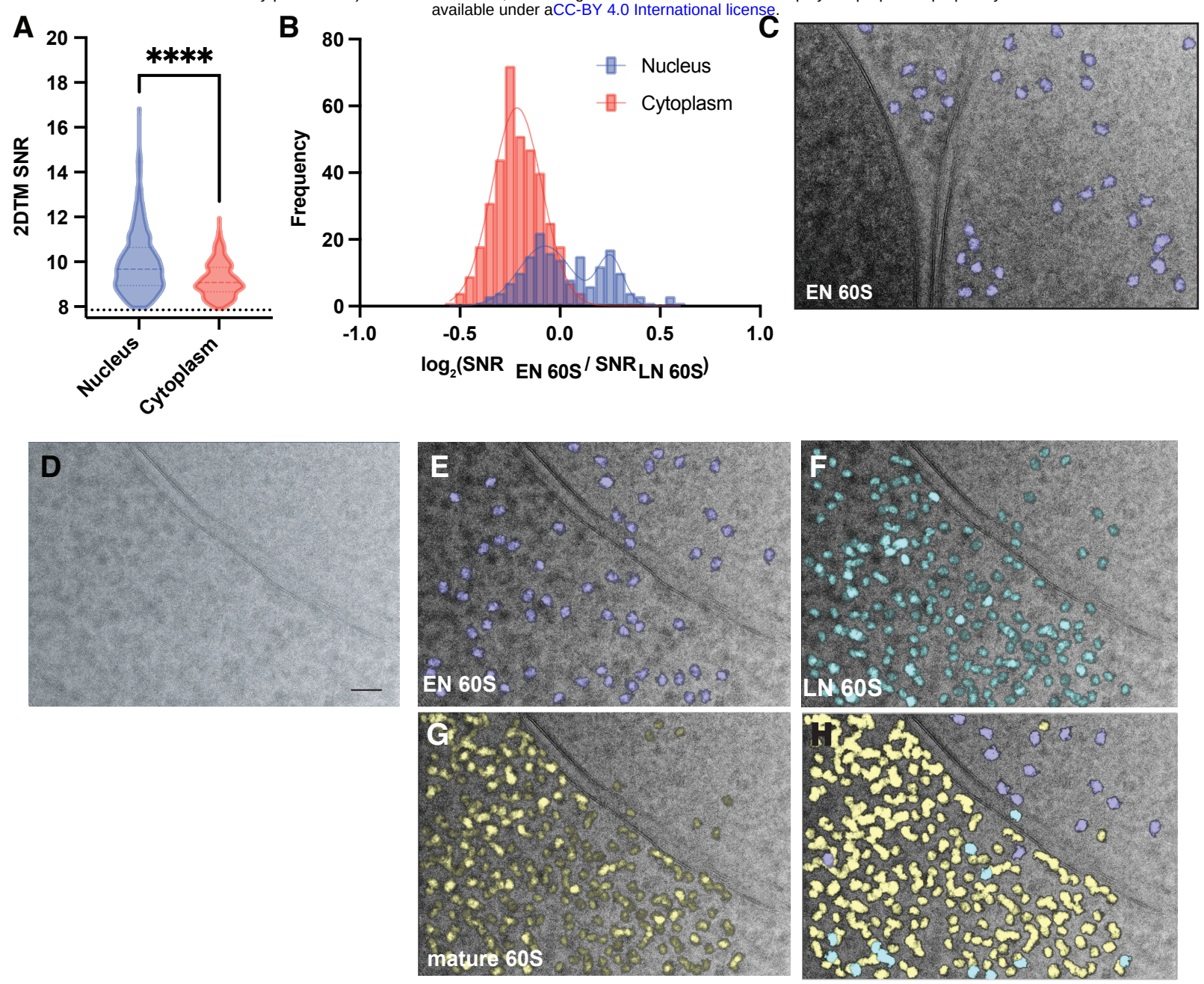


Figure 5 - figure supplement 1