Bridging Mesoscale Nucleosome Organization and Dynamics with Super Resolution Microscopy

Clayton W. Seitz

July 7, 2023

Introduction and Approach

Genome organization in eukaryotes

- The eukaryotic genome has hierarchical structure
- This structure is highly variable and often abberrant in disease

Finn et al., Science 365, 998 (2019)

A phase separation model for transcriptional control

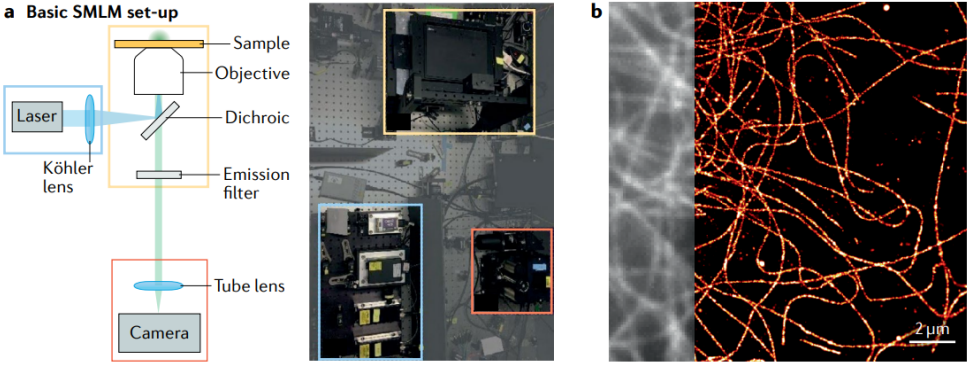
- Liquid-liquid phase separation (LLPS) is a major organizer of cellular biochemistry
- Recent work highlights the importance of CTCF-dependent transcriptional condensates in determining cell fates

Int. J. Mol. Sci. 2022, 23(14), 8039;

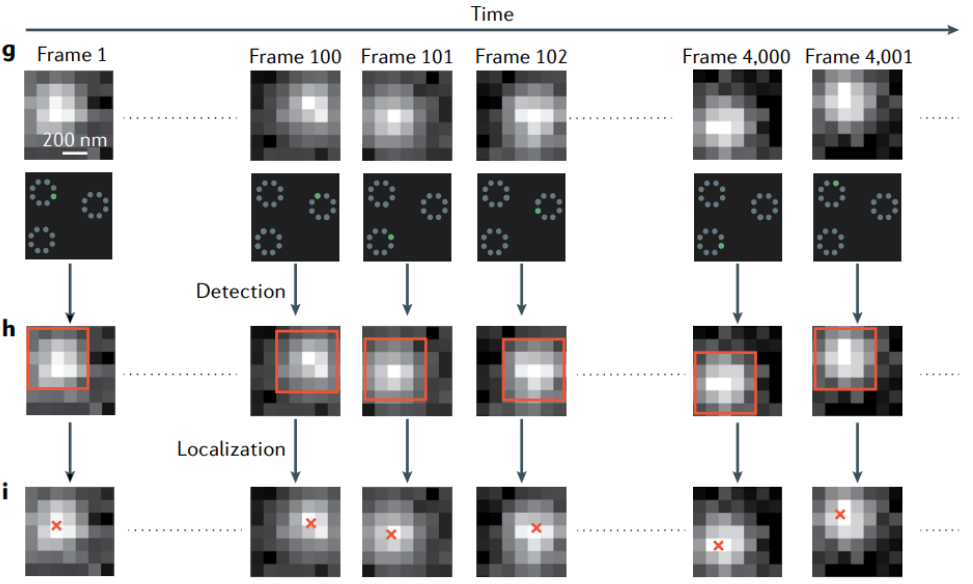
Formulate the basic research question and introduce the approach using major results from section 3

Direct stochastic optical reconstruction microscopy

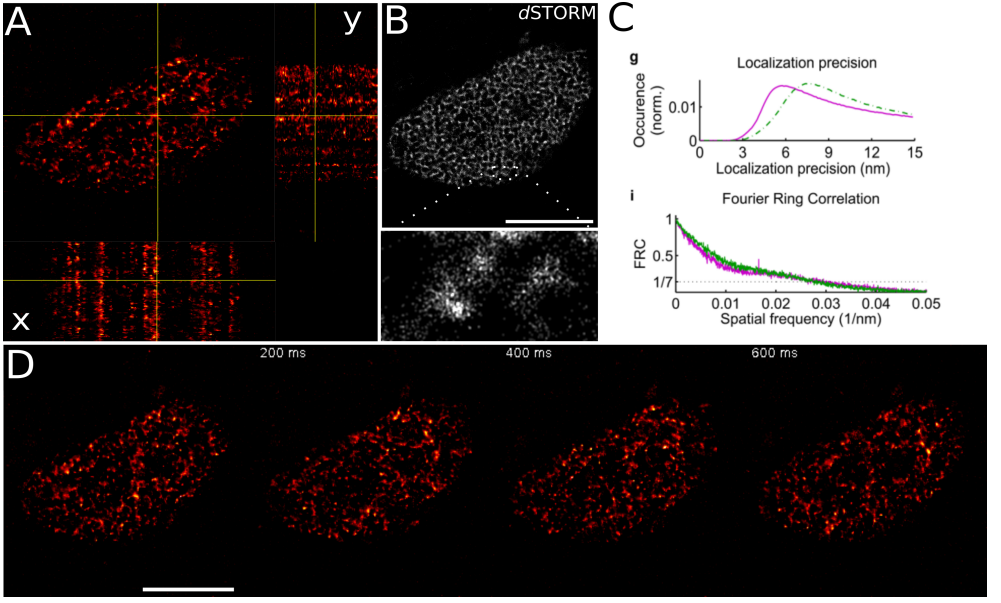
Direct stochastic optical reconstruction microscopy



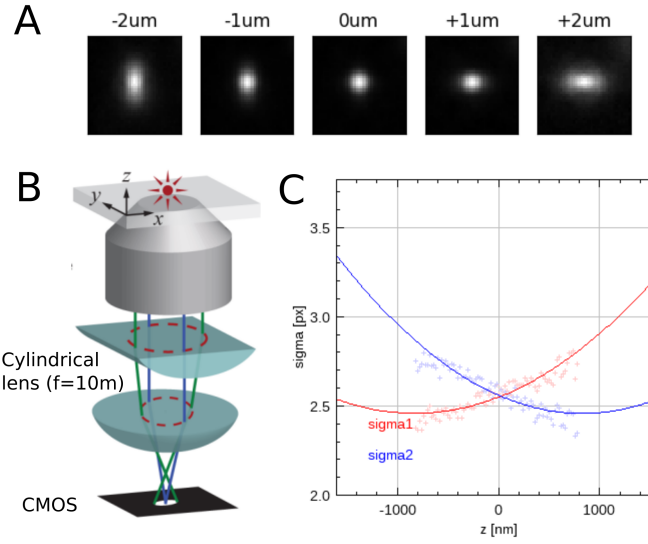
Direct stochastic optical reconstruction microscopy



Super-Resolution imaging of H2B in living Hela cells

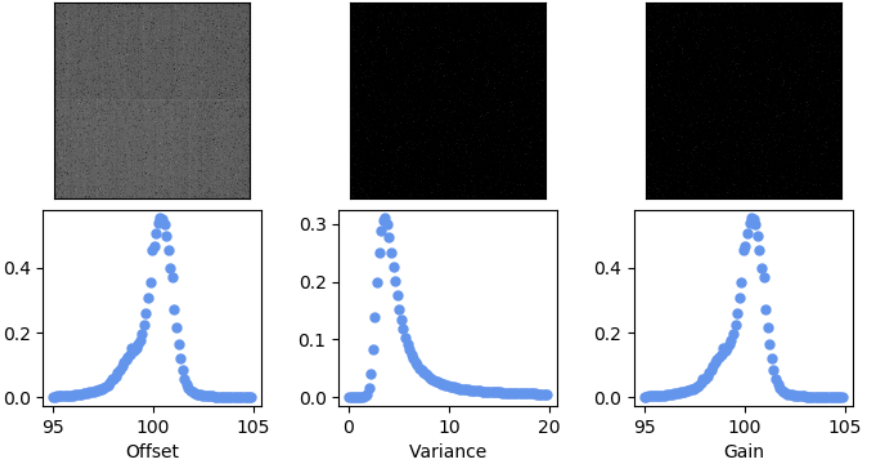


Point spread function engineering for three-dimensional imaging



Readout noise of sCMOS cameras

Hamamatsu ORCA v3 CMOS, air cooled to -10C



Measured signal: $H_k = S_k + \xi_k,\ S_k \sim \mathrm{Poisson}(\mu_k), \xi_k \sim \mathcal{N}(o_k, \sigma_k^2)$

Maximum likelihood localization of an isolated fluorescent emitter

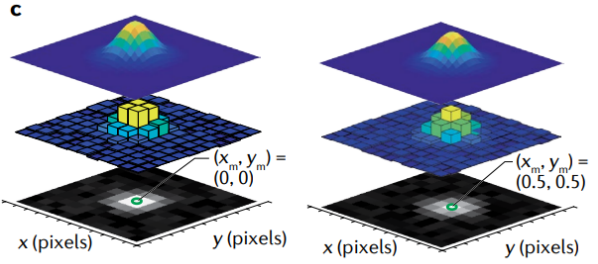
Localization: $\theta^* = \underset{\theta}{\operatorname{argmax}} \prod_k P(H_k|\theta) = \underset{\theta}{\operatorname{argmin}} - \sum_k \log P(H_k|\theta)$

$$\mu_k = g_k \eta N_0 \Delta \int_{\text{pixel}} G(x,y)dA$$

η – quantum efficiency

N_0 – emission rate

Δ – exposure time

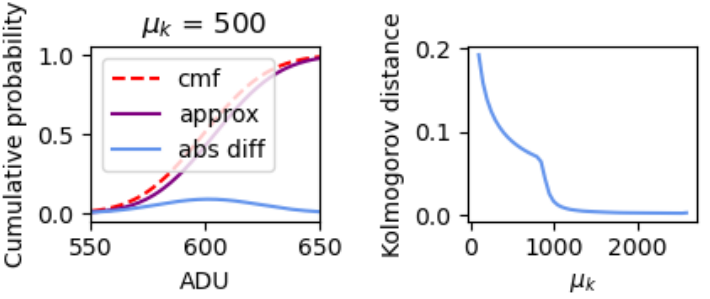


$$P(H_k|\theta) = A \sum_{q=0}^{\infty} \frac{1}{q!} e^{-\mu_k} \mu_k^q \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(H_k - g_k q - o_k)}{2\sigma_k^2}}$$

$P(H_k|\theta)$ can be approximated as Poisson at high signal-to-noise (SNR)

Quality of the Poisson approximation depends on SNR

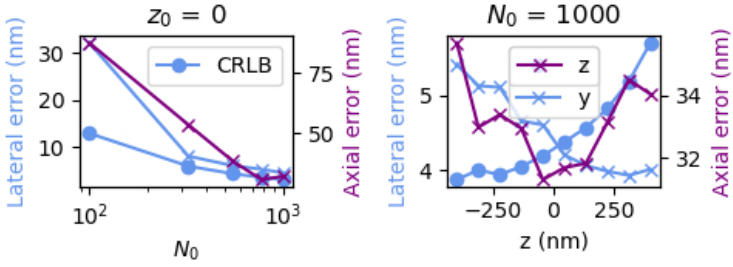
$P(H_k|\theta) \approx \mathrm{Poisson}(\mu_k + \sigma_k^2)$ for $N_0 > 500$ asssuming $\Delta = 100$ms



Using the approximation we can write

$$\ell(\vec{H}|\theta) = -\log \prod_k \frac{e^{-(\mu'_k)} (\mu'_k)^{n_k}}{n_k!} = \sum_k \log n_k! + \mu'_k - n_k \log (\mu'_k)$$

Estimator precision sets the resolution limit in localization microscopy

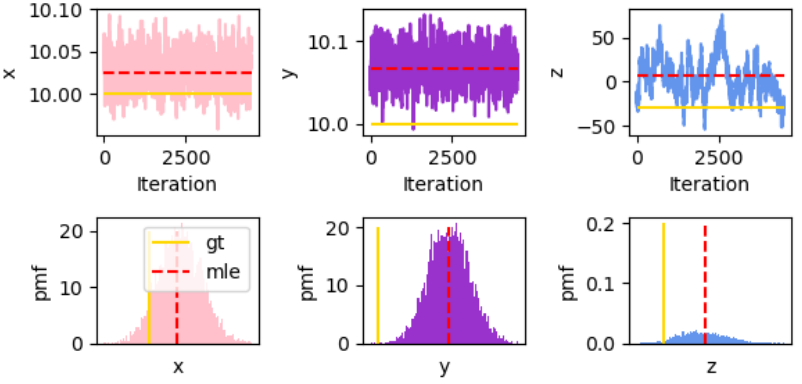


- MLE can approach the CRLB on simulated isolated emitter data
- MLE does not generalize well to dense time-series

Estimating uncertainty with gradient-based MCMC

Stochastic gradient langevin dynamics (SGLD):

$$dw = -\nabla \ell(\theta) dt + \epsilon \sqrt{\eta dt}, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \eta \propto dt$$



The diffusion samples from the posterior $P(\theta|\vec{H})$ as $t \to \infty$

Photoswitching kinetics of Janelia-Fluor 646 in thiol buffer

Estimator precision is dependent on imaging conditions

We have shown how localization precision is affected by SNR, which is affected by the frame rate. However, localization precision is also independently determined by the density, which is affected by photoswitching rate constants. True density is unknown apriori, so, as a rule, we strive for the lowest density possible for a fixed set of rate constants. Lateral precision is less sensitive to SNR, so it is probably possible with our setup.

Pointillist SMLM data is quantified with Ripley's K-function $K(r)$, which counts events within a radius r. For $K(\sigma) > 0$, SMLM data can be crowded for some $\vec{k}$. Suppose a particular fluorophore is ON for a sufficient amount of time during Δ to be detected. Now on average $\langle\xi\rangle\lambda$ signal photons are collected and $\langle\xi\rangle\lambda p_{\mathrm{ON}} K(\sigma)$ "background" photons will be collected where $\xi \sim P(\xi|\vec{k})$. P might be a beta-distribution or normal approximating binomial for $N \to \infty$. Define $\zeta = 1/p_{\mathrm{ON}}K(\sigma)$ which we attempt to maximize with $K(\sigma)$ fixed and under the constraint that p_{ON} also achieves a desirable SR frame duration e.g., 1 second

The lifetime distribution of the ON state is trivial. We can use the probability current to get the lifetime of the OFF state

$$J(t) = P(0)Ge^{Gt} \quad h_{\mathrm{OFF}}(t_n) \approx \epsilon J(t_n)$$

Thus we have the optimization problem

$$\mathcal{L} = \zeta(\vec{k}) + \lambda(t(\vec{k}) - \Delta_{\mathrm{SR}})^2$$

where Δ_{SR} is the desired SR frame length. $t(\vec{k}; N)$ is the amount of time to collect N molecules.

Deep learning enables accurate 3D localization and single molecule tracking

Precise localization in dense images is intractable

$$\mathcal{L} = \sum_{i,j} \log p_{ij}(\tilde{x}) = \sum_{i,j} \log \frac{\exp(-s_{ij}(\tilde{x}))}{\sum_{x \in \chi} \exp(-s_{ij}(\tilde{x}))}$$

p_{ij} is the probability the model assigns a pixel to the true class $\tilde{x} \in \{0, 1\}$

Mesoscale nucleosome organization and dynamics

Dirichlet process Gaussian mixture model (DPGMM)

GMM cluster analysis of H2B

Number of clusters is unknown apriori - Bayesian nonparametrics

Mesoscale nucleosome organization and dynamics

Mesoscale nucleosome organization and dynamics

BRD4 associates with the small clusters

Besag's L-Function

Besag's L-Function and 3D diffusion