Homework assignment #5
Due in class Mon 10/08

Measuring expression with single-molecule precision (20 points)

The techniques we are discussing in class this week (immunostaining, GFP fusion proteins…) infer relative levels of protein expression from the intensity of recorded fluorescence. A team of researchers at Fancy University designed an alternative approach. Focusing on mRNA instead of protein, they developed a labeling protocol producing images as on Fig. 1. Instead of simply summing the intensity of all pixels, they can now detect and count the **number** of bright spots. This, they claim, allows measuring the mRNA levels with incredible precision – by literally counting the molecules. Very exciting indeed!
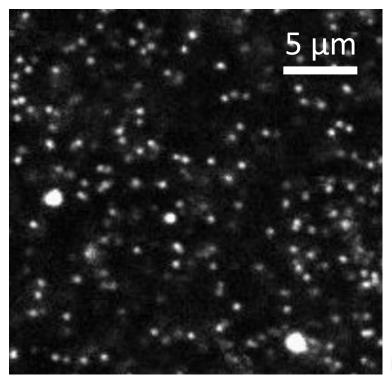

Figure 1

Spot counting: pros and cons

1. Name 3 advantages of counting spots compared to summing pixel intensities.
2. The pipeline "labeling protocol -> imaging -> automatic image processing & spot-counting" produces a number. The researchers claim this number is in fact the absolute number of mRNA molecules in the field of view. Name 3-5 potential sources of error – things which, one hopes, the authors of the work must have checked before making the bold claim.
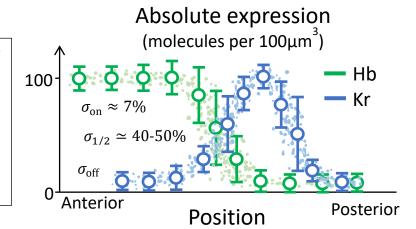3. Of the ones you named, which do you expect to be the most problematic? Why?

Expression noise

The researchers used their new method to measure concentration profiles of both Hunchback *Hb* (recall HW#3) and another gene called Krüppel (*Kr*), which is also involved in patterning. Figure 2 below shows data from 10 embryos all plotted together. The team made two very intriguing observations.

4. Curiously, the noise in the fully-on region (highlighted) is lower than at intermediate expression!
5. What's more, the fluctuations of Kr and Hb expression are markedly anti-correlated.
   Fluctuations are defined as deviation between the actual observed count, and the predicted count based on smooth interpolation of bin averages (circles on the plot).

Data points: molecule counts per nucleus, for each nucleus in the field of view, for 10 embryos.

Datapoints grouped into bins of 50 by antero-posterior position; circles and error bars show mean and standard deviation per bin.

The team's favorite theorist is very excited. The first observation supports a Fancy Theory about how the RNA polymerase recruits a co-factor with a complicated name, which should make fully-on genes less noisy. And the second is even more exciting, because large correlated fluctuations are evidence of a hyper-sensitive system poised at a critical point!

Can you think of simpler explanations of these observations, which would need to be excluded first?

Noise: Poissonian and sub-Poissonian [optional; up to 10 points extra credit]

Undeterred, the theorist runs the numbers and notices something else, truly fantastic. The expression noise in the fully-on region is not just low – it's sub-Poissonian!

6. [Warm-up] Consider a collection of freely diffusing non-interacting particles in a large volume V. Show that the number of particles in a given small test volume V_0 obeys Poisson statistics.

This calculation provides a useful rule of thumb for thinking about small-N events. To the extent that transcription and translation can be modeled as single-rate processes, this is a useful intuition for assessing the importance of fluctuations. Bringing noise below this level is not easy! Yet on the figure above, the Poisson expectation is 10%, but the measured σ_{on} is only 7%. And while there are lots of ways to experimentally overestimate fluctuations, a noise that is too *small* can't be an artifact – right?

7. Is there any chance this observation might still be an artifact of the method? *(Hint: Pay attention to the numbers on the Y axis. It may also be useful to note that both peaks are the same height...)*
8. If the sub-Poisson observation turns out to be real, any thoughts on what processes could result in such better-than-Poisson noise statistics? *[Intentionally very open-ended – only if you're curious.]*