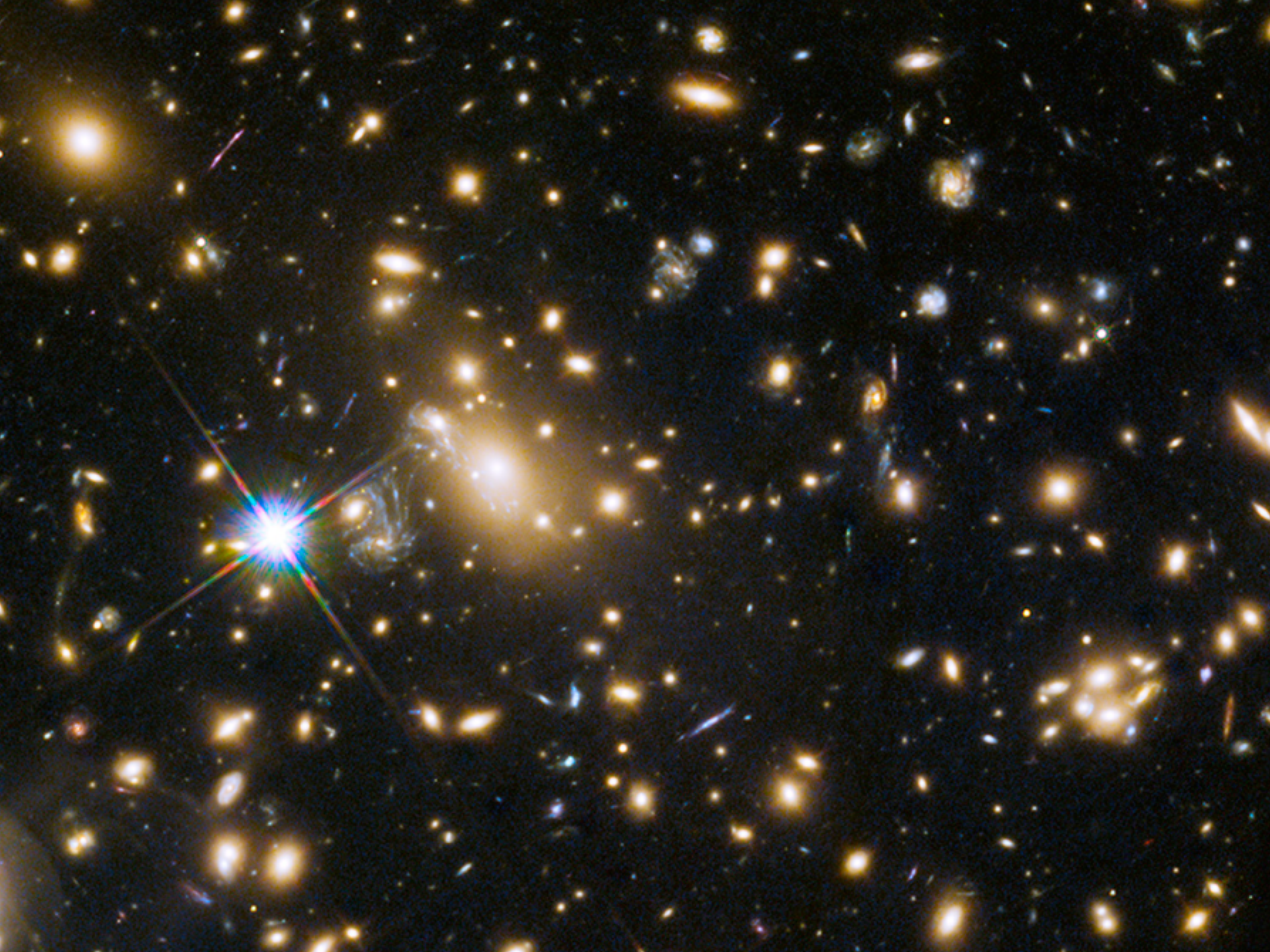
UNIVERSITY OF COPENHAGEN

DAWN

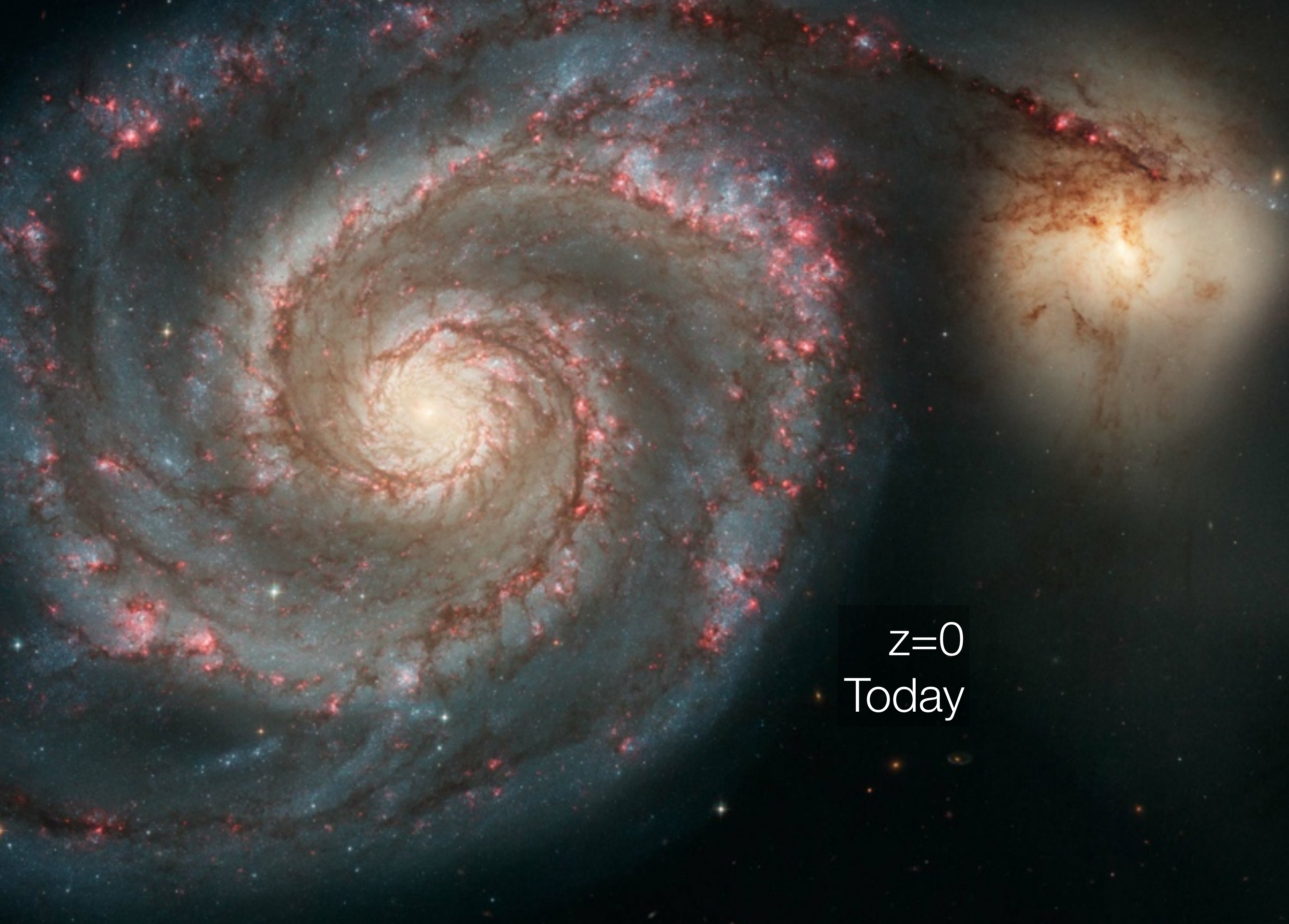
Danmarks Grundforskningsfond
Danish National Research Foundation

Applied Statistics: *Detection and characterization of the most distant galaxies*

Gabriel Brammer



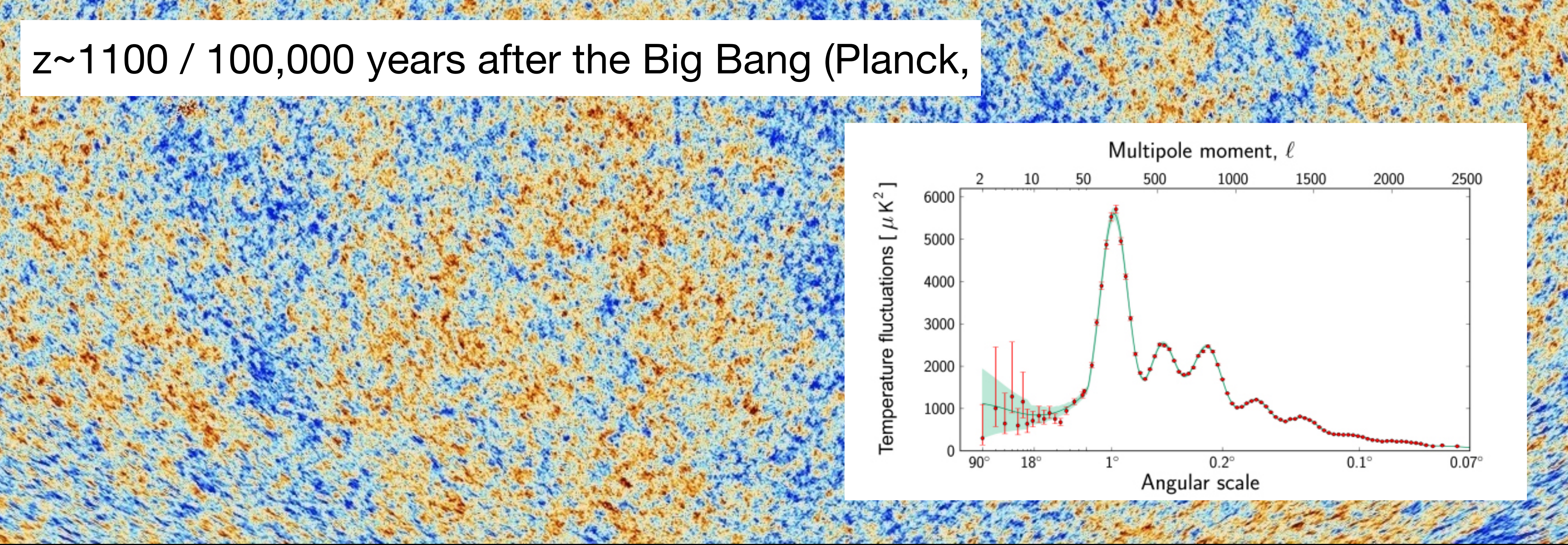

z~9.5
12 Billion years ago



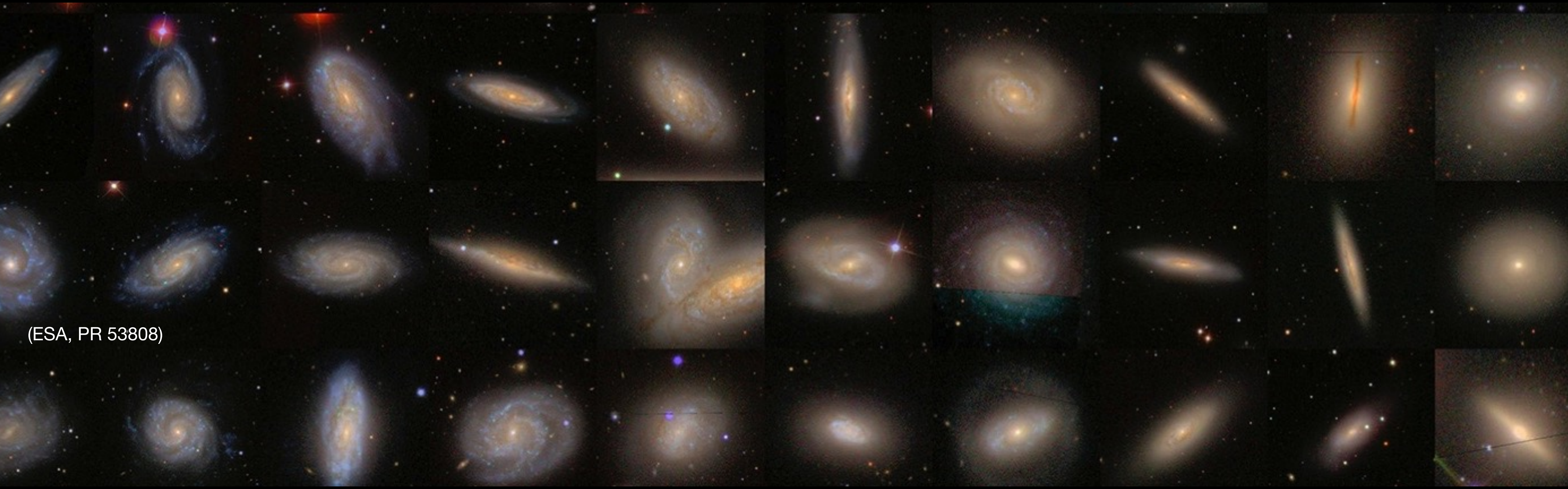
Expanding Universe:
"z" = "redshift" ≈ distance
≈ time in the past for finite ***c***

z~9.5
12 Billion years ago

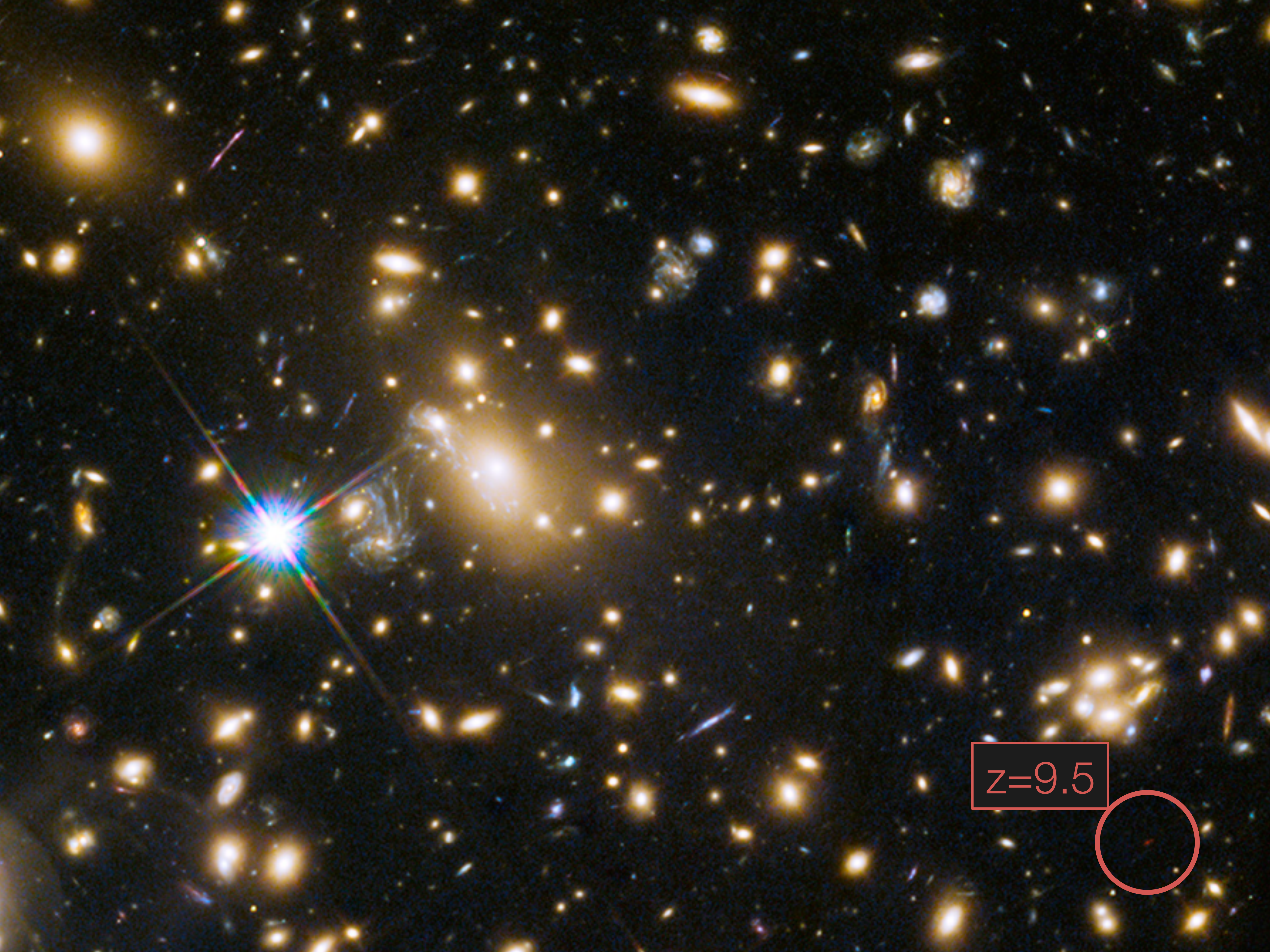
z~1100 / 100,000 years after the Big Bang (Planck,

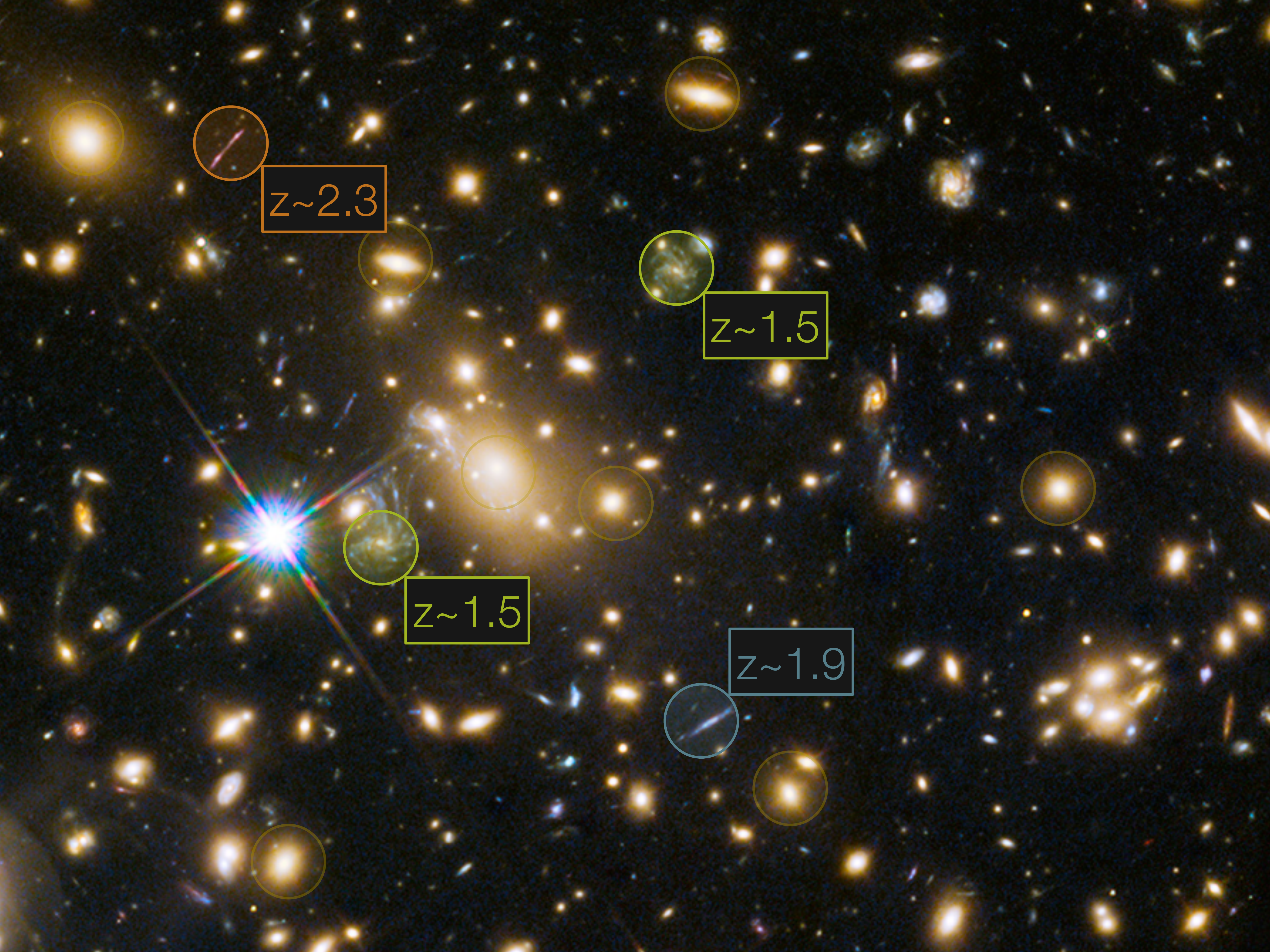


13.7 gigayears after the Big Bang (Sloan Digital Sky Survey)



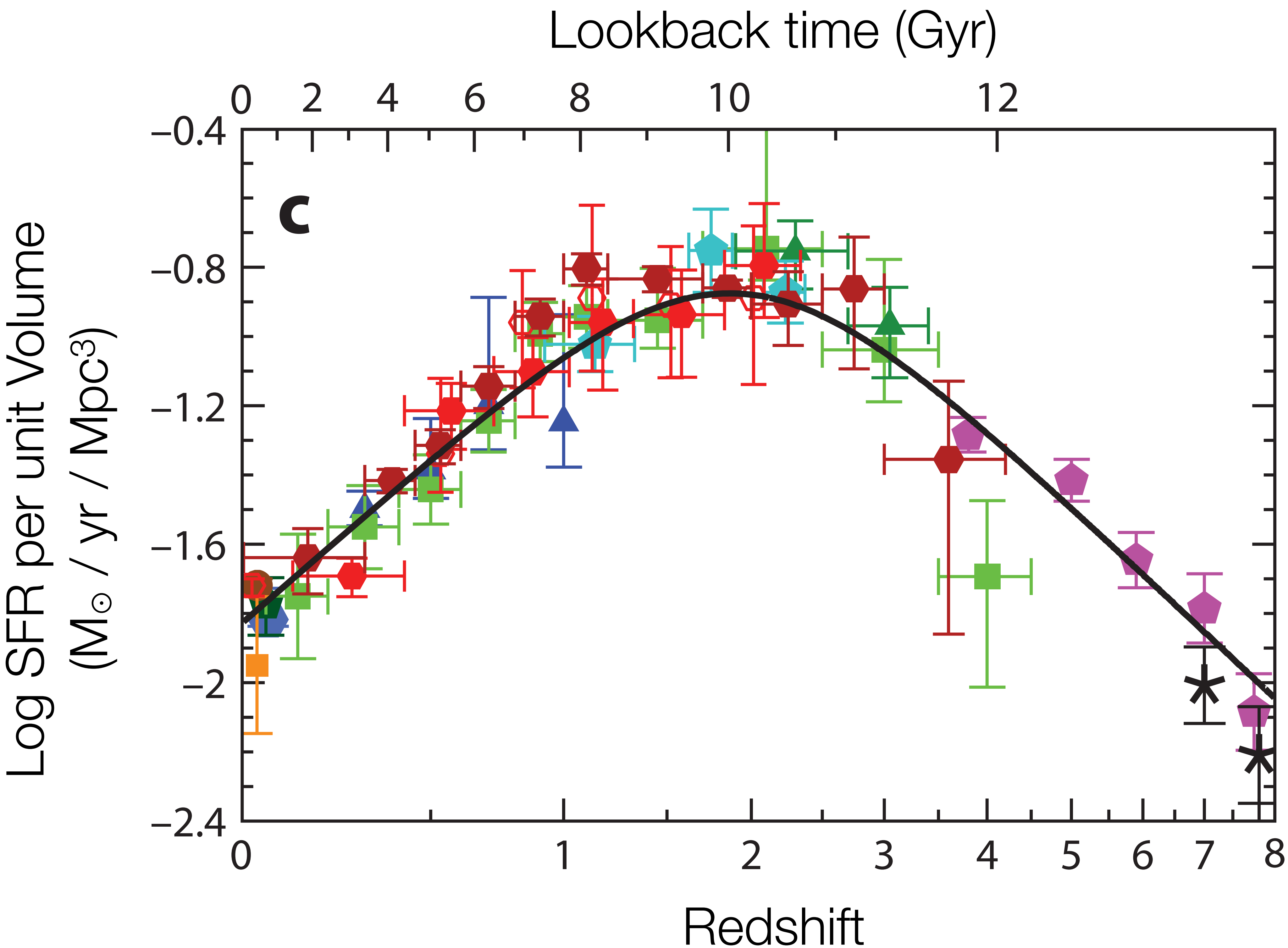
(ESA, PR 53808)


z=9.5


z~2.3
z~1.5
z~1.5
z~1.9

Cosmic Star Formation History

Madau & Dickinson +14



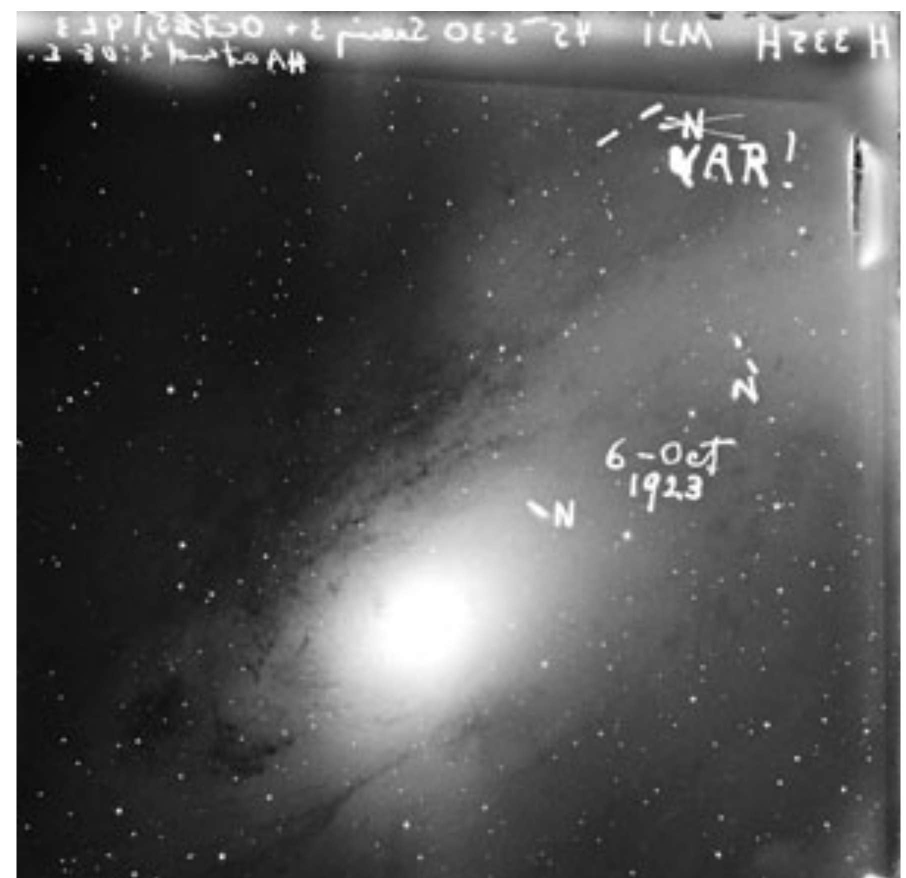
Under the hood - standard imaging

- Images from a telescope sample the brightness distribution of the sky
- Calibration, background, noise model
- Typically sources of interest are much fainter than the background, so take many exposures and average them (**central limit theorem**)
  - **S/N ~ sqrt(t)**

Under the hood - standard imaging

- Basic "astronomy" done with 2D images that sample the (projected) brightness distribution of sources on the sky

1923
E. Hubble



Under the hood - standard imaging

- Modern digital detectors are ~**linear** photon-counting devices that can be **calibrated** to an absolute scale (e.g. W / m^2)

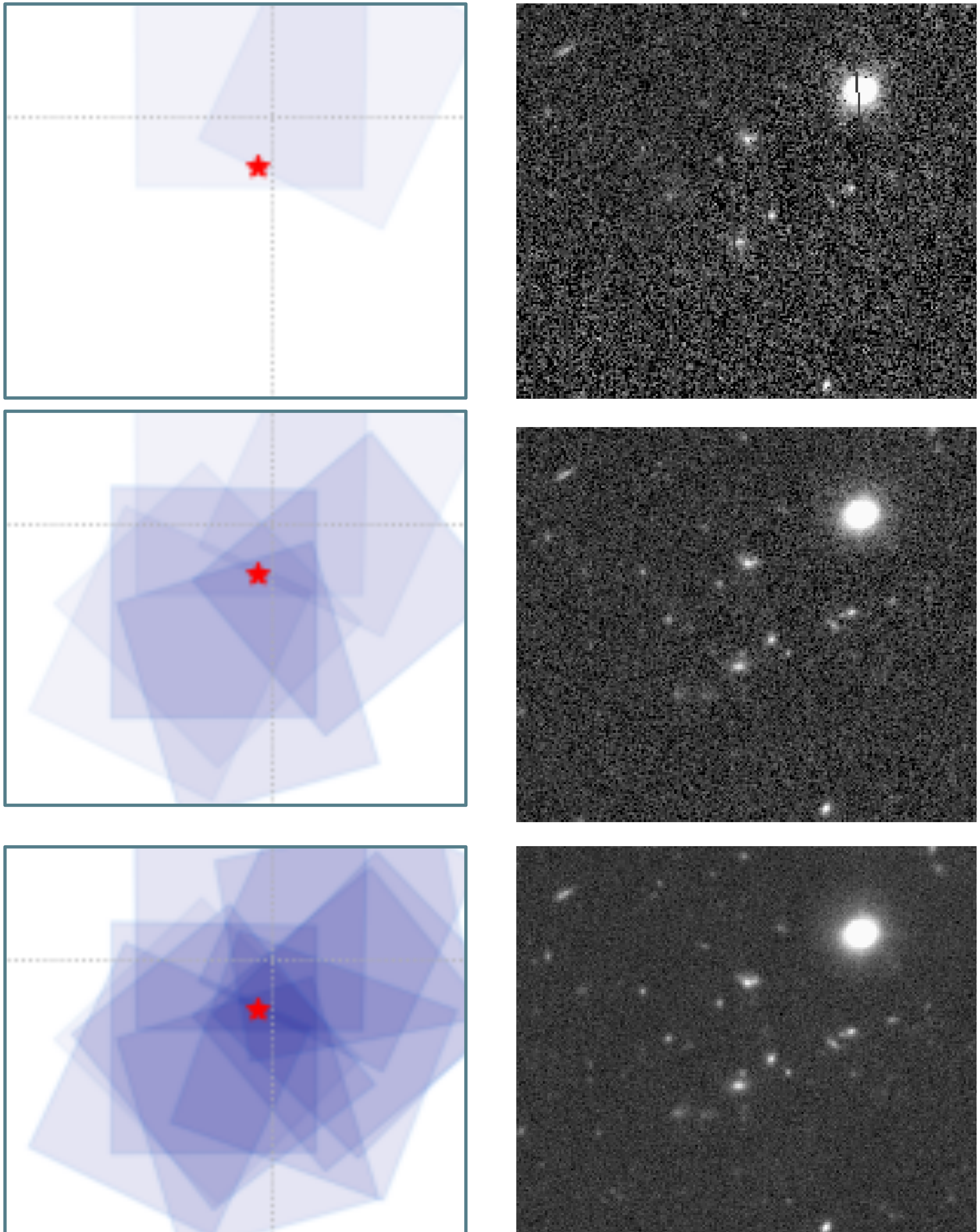
2004, Hubble Space Telescope

Imaging statistics

- ***S*** - Signal rate from source of interest (e.g., photons / s)
- ***B*** - "background" rate
- $N\epsilon^2$ - Random noise reading the charge on the detector N times
- ***t*** - "open shutter" integration time
- Variance $Var = (S + B)\, t + N\epsilon^2$

- For B >> S, $N\epsilon^2$ (faint sources, expensive detectors),
  Signal-to-noise $\approx t \,/\, t^{-1/2} \approx t^{1/2}$
  - Increasing signal-to-noise by a factor of **two** requires **four** times the integration time

- **Central limit theorem** provides approximately Gaussian statistics, but this should be verified and preserved!



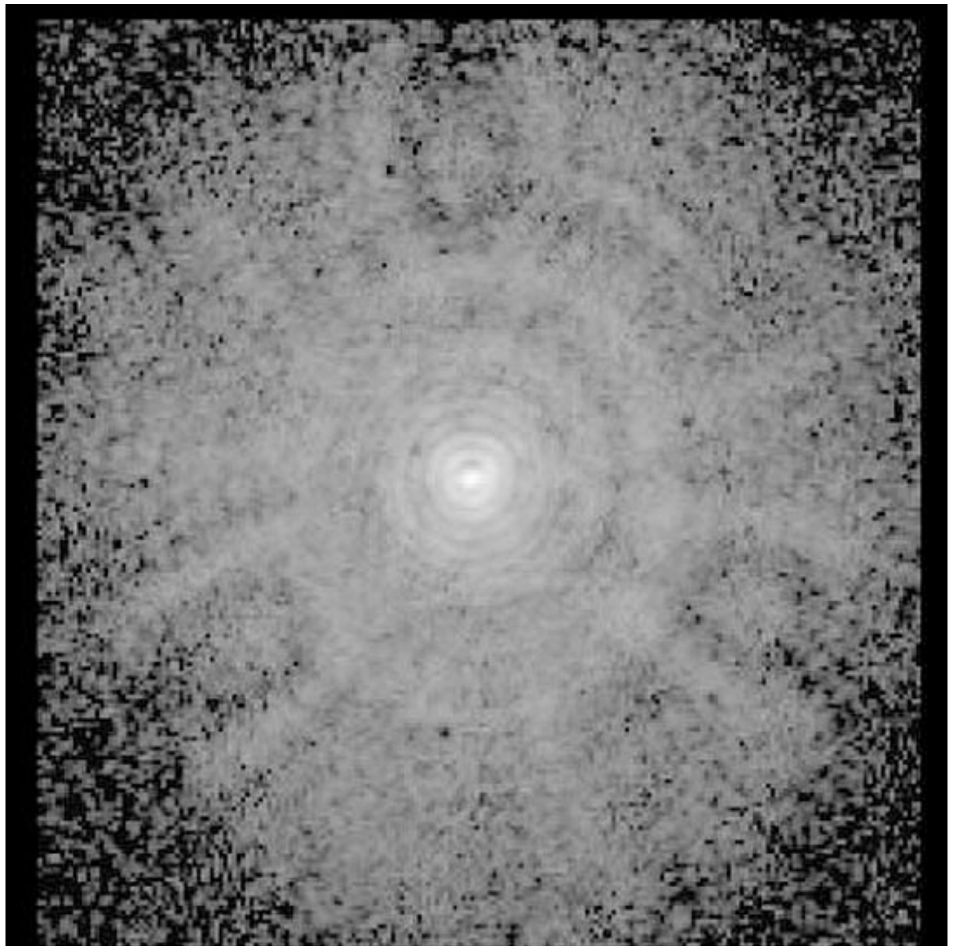
$t \sim 10$ minutes

$t \sim 2$ hours

$t \sim 9$hours

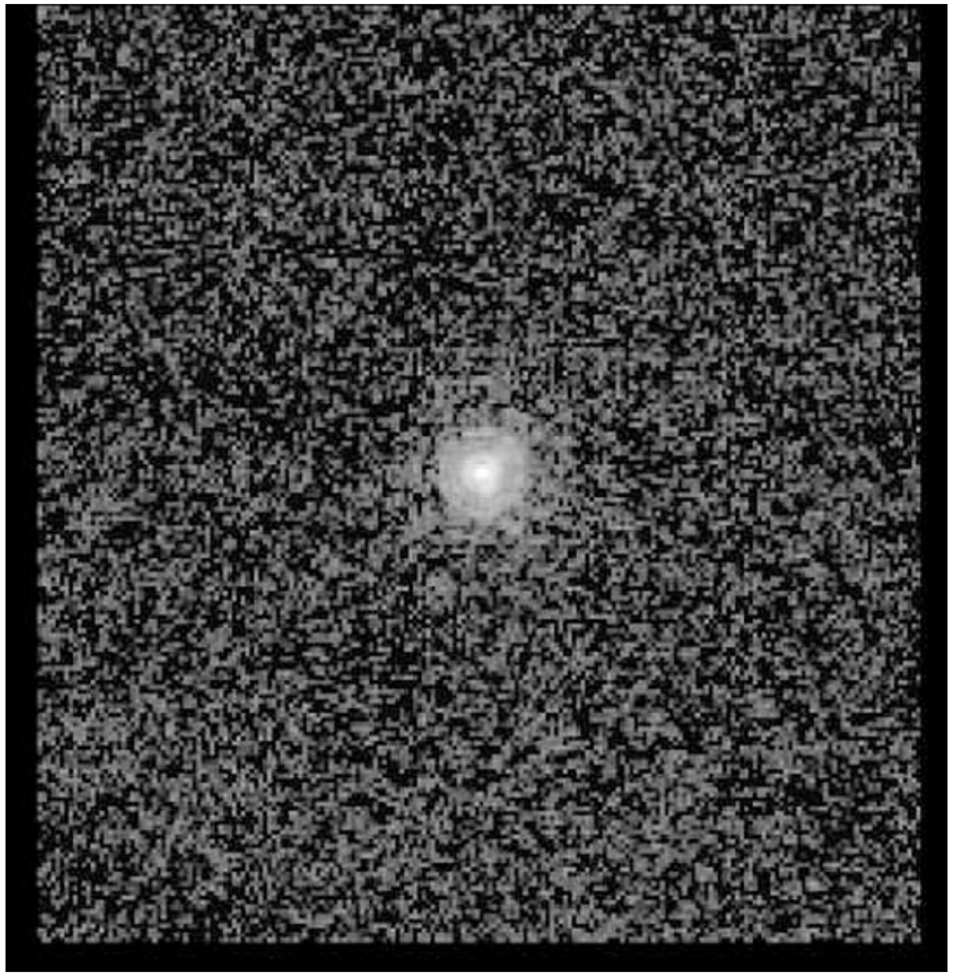
Source detection and characterization

- Point sources in an image (stars) have a finite size set by the telescope diameter and optics
  - "Point spread function"



Source detection and characterization

- Point sources in an image (stars) have a finite size set by the telescope diameter and optics
  - "Point spread function"



Source detection and characterization

- Point sources in an image (stars) have a finite size set by the telescope diameter and optics
  - "Point spread function"

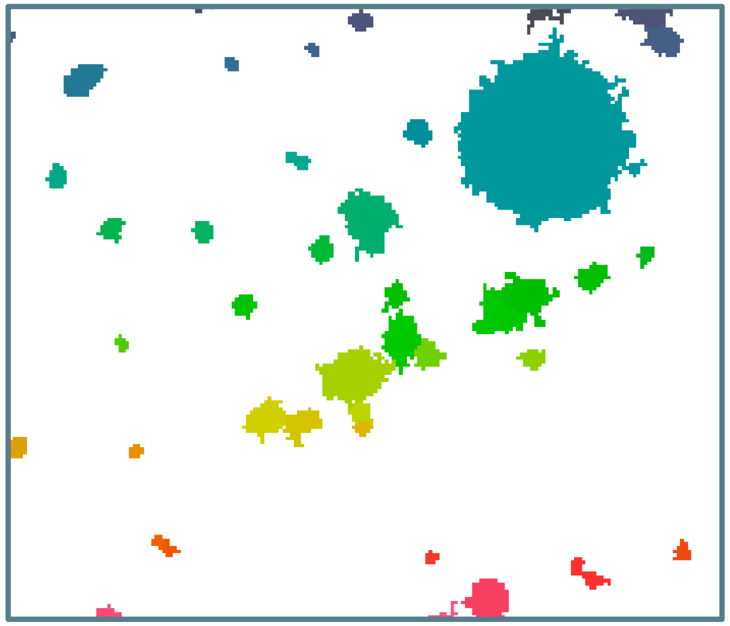
- Basic weighted source detection
  - $I = (S \,/\, Var * \mathrm{PSF}) \,/\, (1 \,/\, Var * \mathrm{PSF}^2)$
  - $\mathrm{Var}(I) = 1 \,/\, (1 \,/\, Var * \mathrm{PSF}^2)$
- Essentially, a least squares fit for the intensity of a point source anywhere in an image

Source detection and characterization

- **But be careful!** Large images can be provide many "trials" for detecting sources, so "3σ" can be risky.
- E.g., relatively small HST images are 2400x2400 "PSFs" in size, so *p*=0.01 can still be a very big number.
- Worse still in presence of uncontrolled systematics.

Source detection and characterization

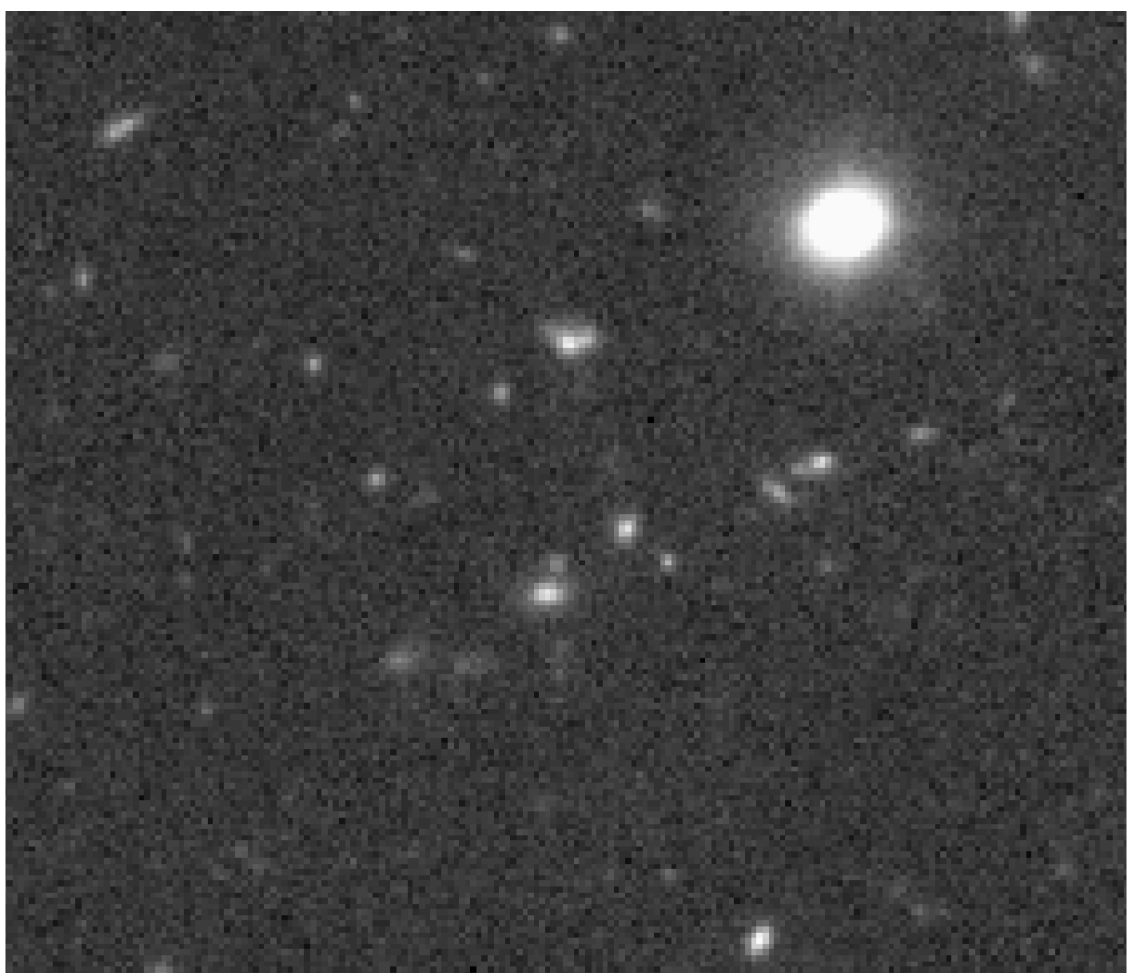
- Image **segmentation** assigns regions of the image to discrete sources (e.g., `scikit-image`)



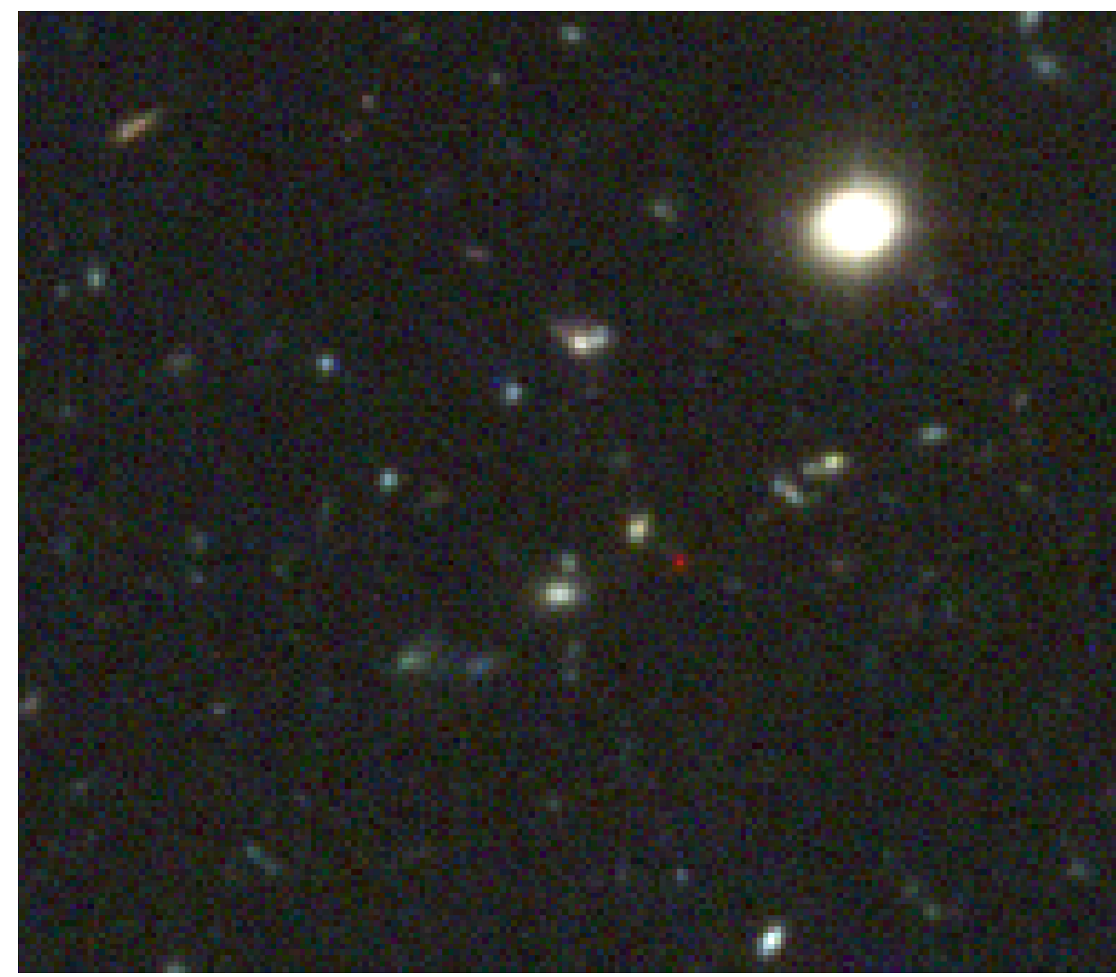
- Measure **moments** of the light distribution for discrete sources, e.g., sum, FWHM

A third dimension: spectral shape $I(\lambda)$

- Imaging: bandpass filters
- (Much more complicated optics can provide full spectra across the 2D spatial field)

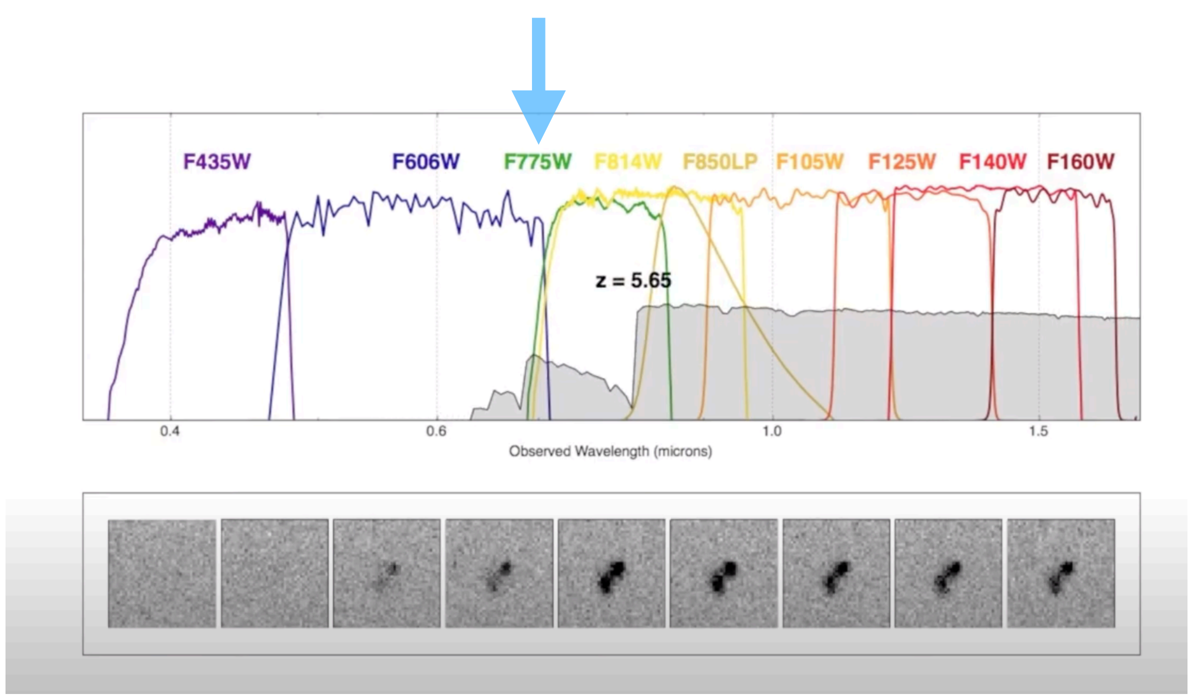


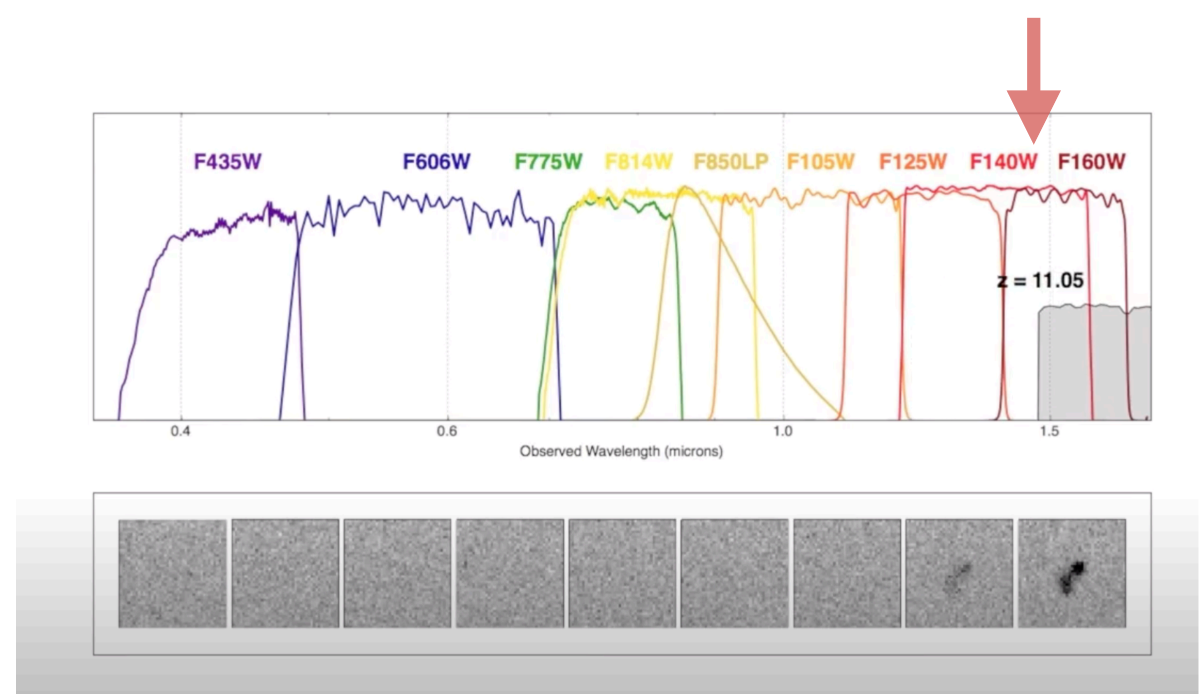
λ
1.0 μm
1.3 μm
1.6 μm



Crude distance estimates: "dropouts"

- Neutral Hydrogen along the line-of-sight to distant galaxies absorbs all light below 1216 Å
- Creates a step-function signal that can be an effective distance indicator ("Lyman break")
    - Simply observe an object in multiple filters and the bluest wavelength in which that object is detected ~1216Å





(firstgalaxies.org, Magee et al.)

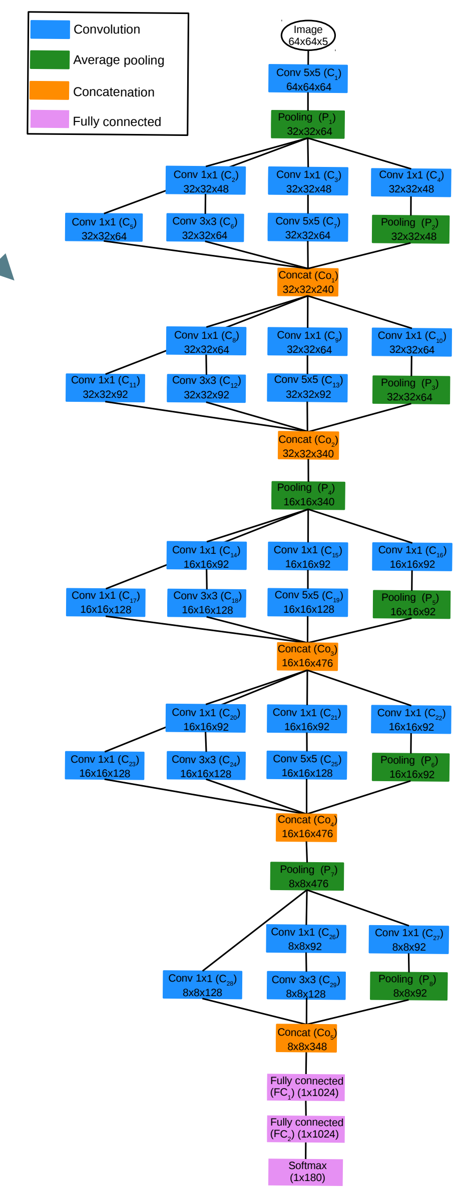
Data Reduction and Machine Learning

- The procedures above are done as objectively as possible, inevitably involve data compression and loss of information
- Image moments are essentially *features* a (semi-)intelligent machine—the researcher—has chosen as important
- Speed, efficiency vs. interpretability

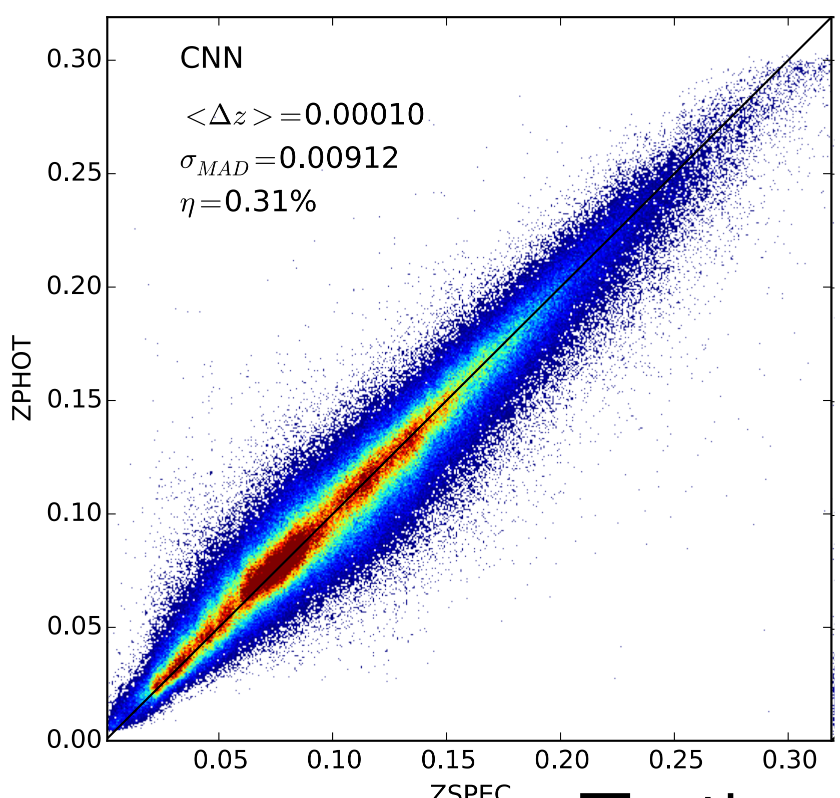
Data Reduction and Machine Learning

- Example, just give a machine the images themselves and let it figure out the mapping between Image ➜ redshift

z=0.258 z=0.039 z=0.094 z=0.163 z=0.078



Predicted



Truth

Pasquet et al. https://arxiv.org/abs/1806.06607

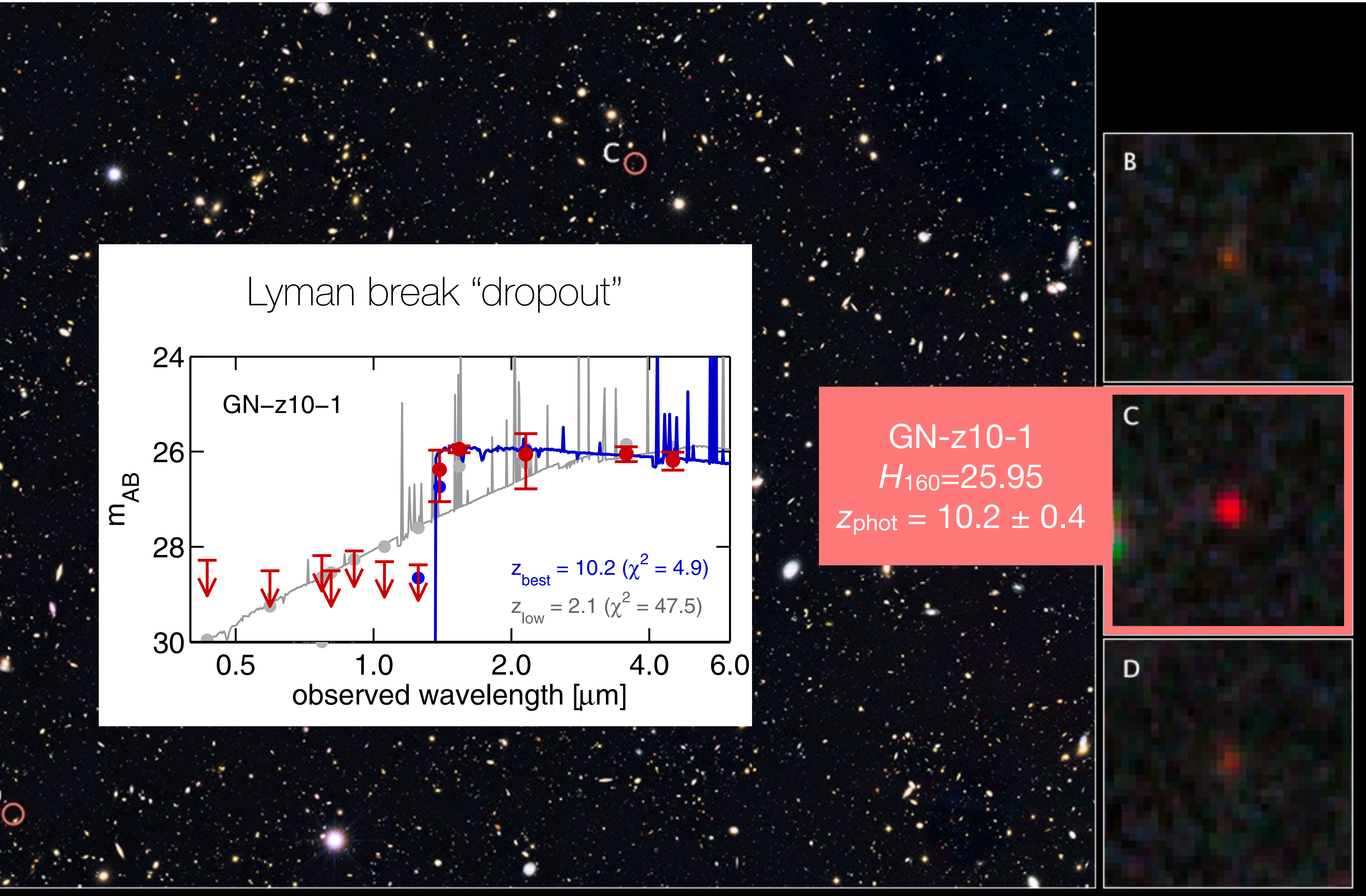
Candidate #2 (GN-z10): z~10?

Oesch+15

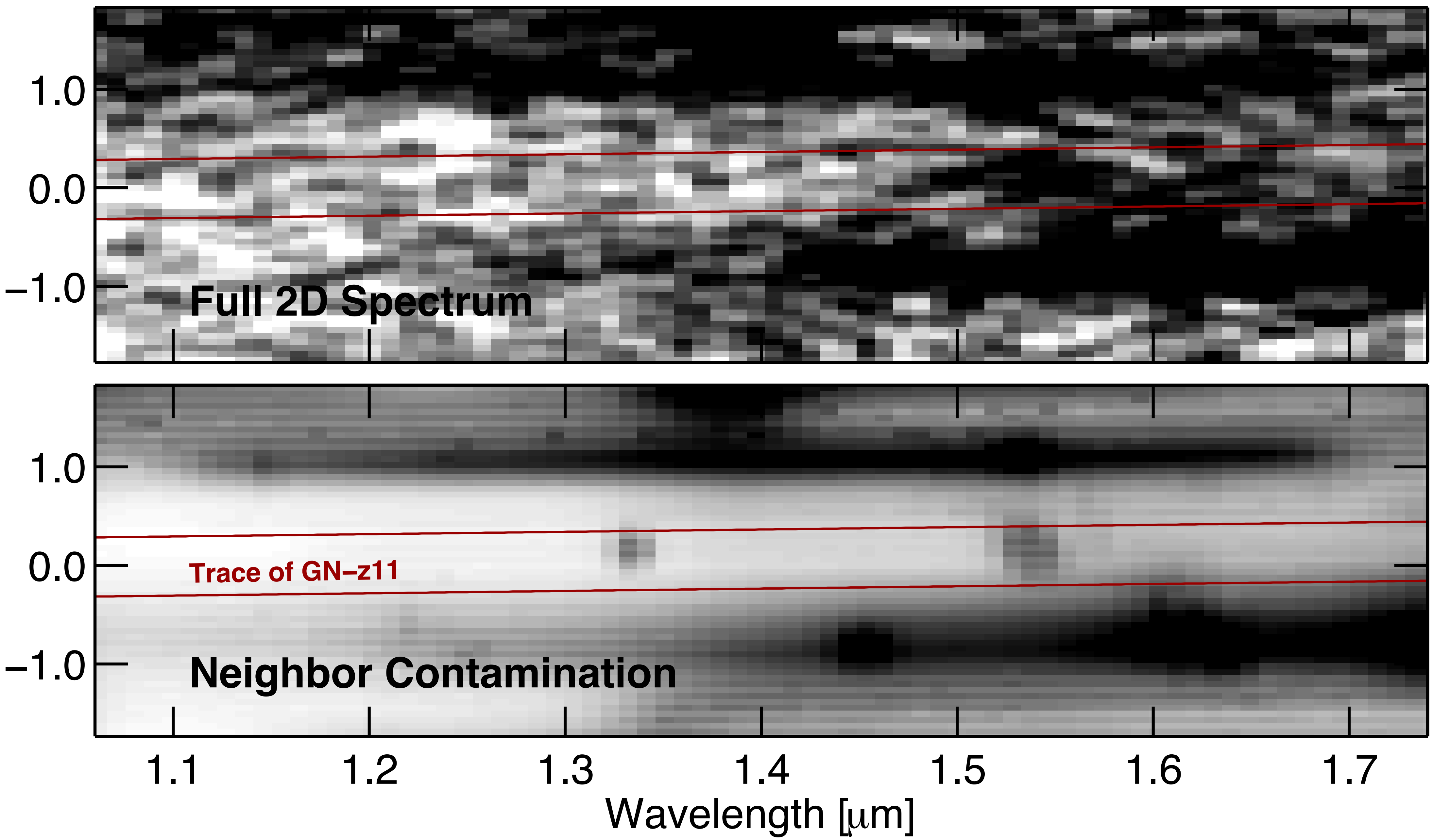
GN-z10-1
H_{160}=25.95
z_{phot} = 10.2 ± 0.4

Candidate #2 (GN-z10): z~10?

Oesch+15



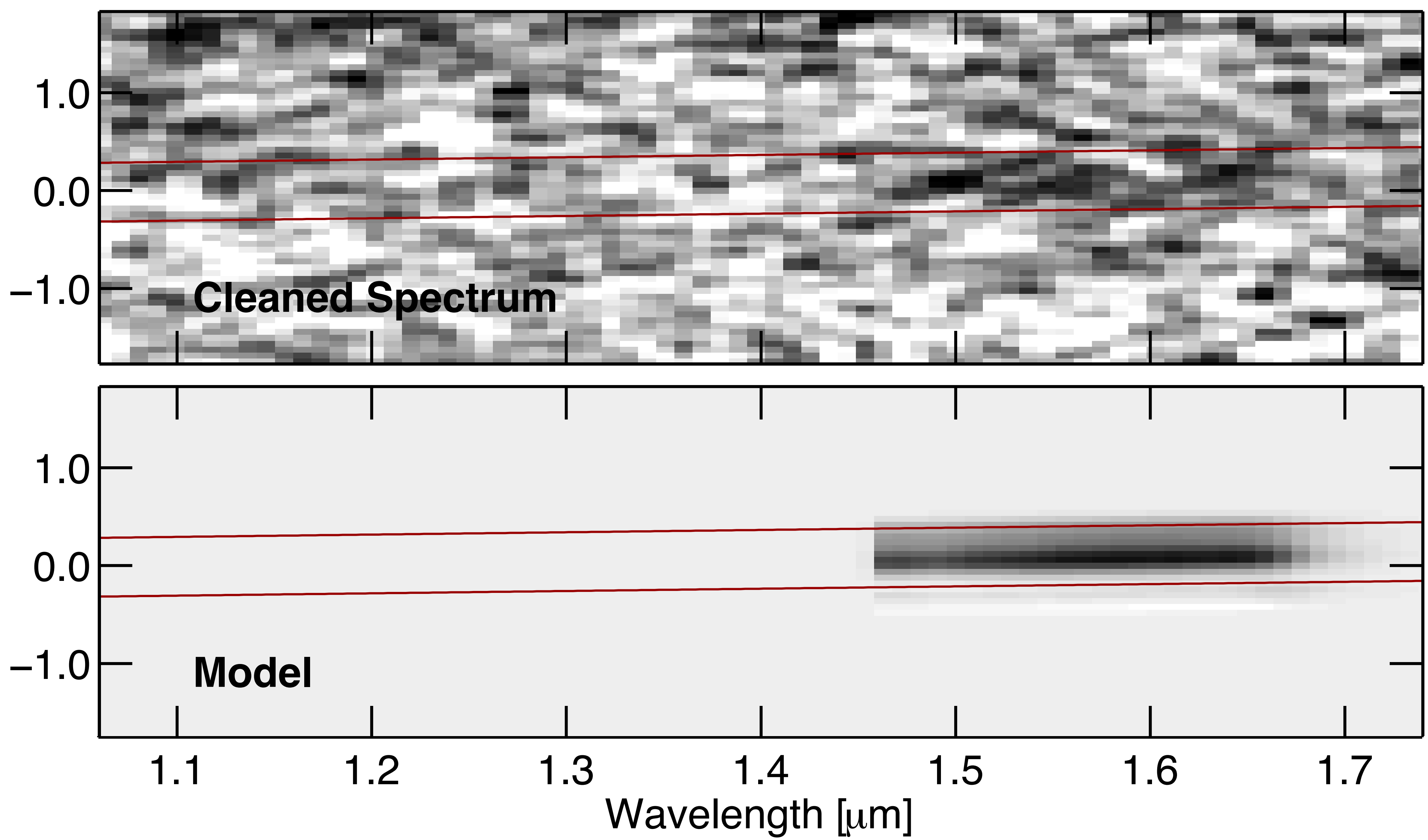
Candidate #2 (GN-z10): z~10?



Candidate #2 (GN-z11): z~10? z=11.1±0.1

Oesch,Brammer+16

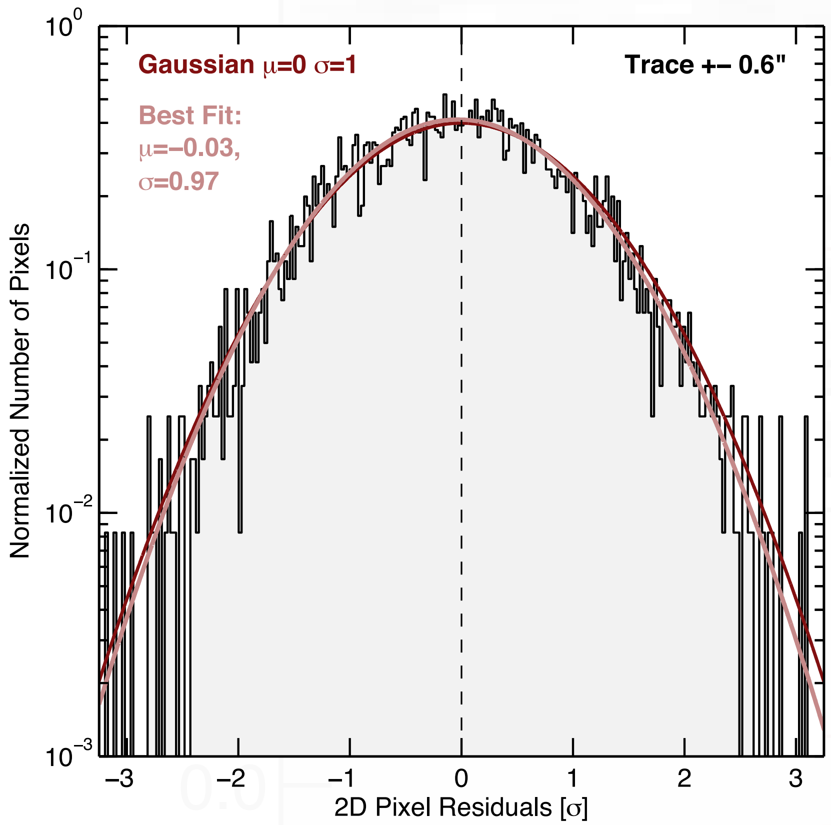
- Deep *HST* spectrum supports Lyman-break @ z=11

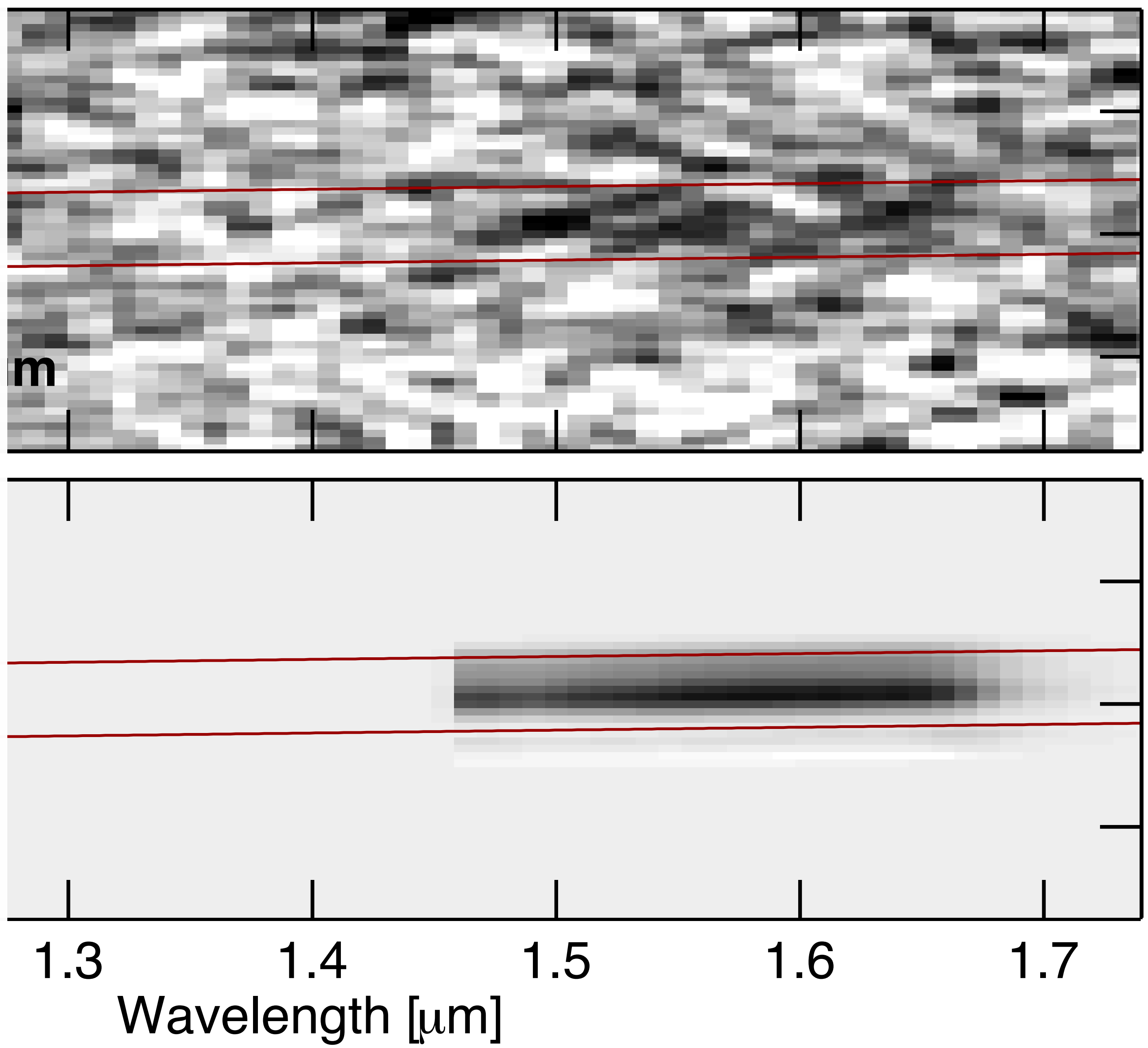


Candidate #2 (GN-z11): z~10? z=11.1±0.1

Oesch,Brammer+16

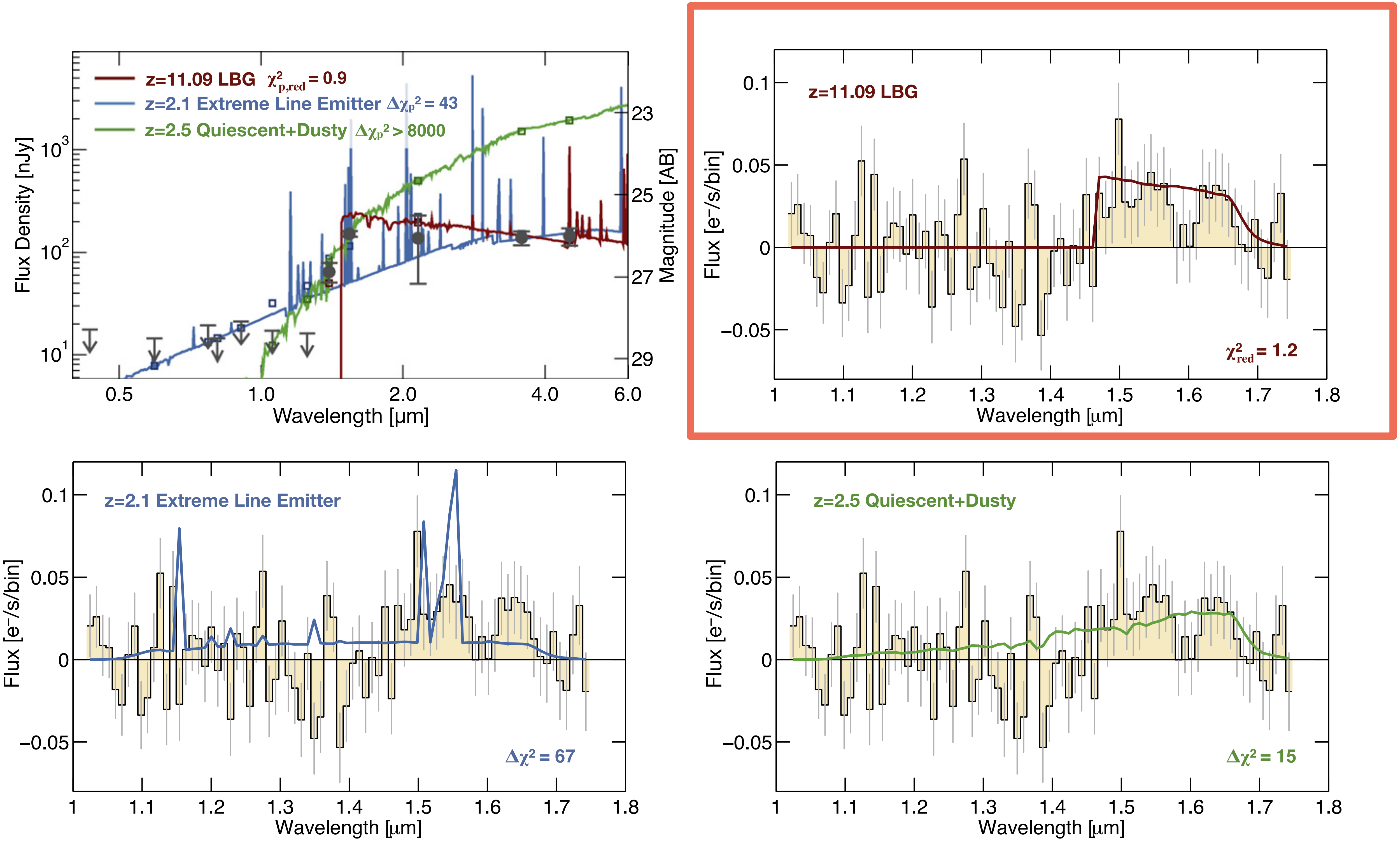
- Deep *HST* spectrum supports Lyman-break @ z=11





Uncertainties are ≈ Gaussian

Favored interpretation



Potential contaminants inconsistent with the spectrum

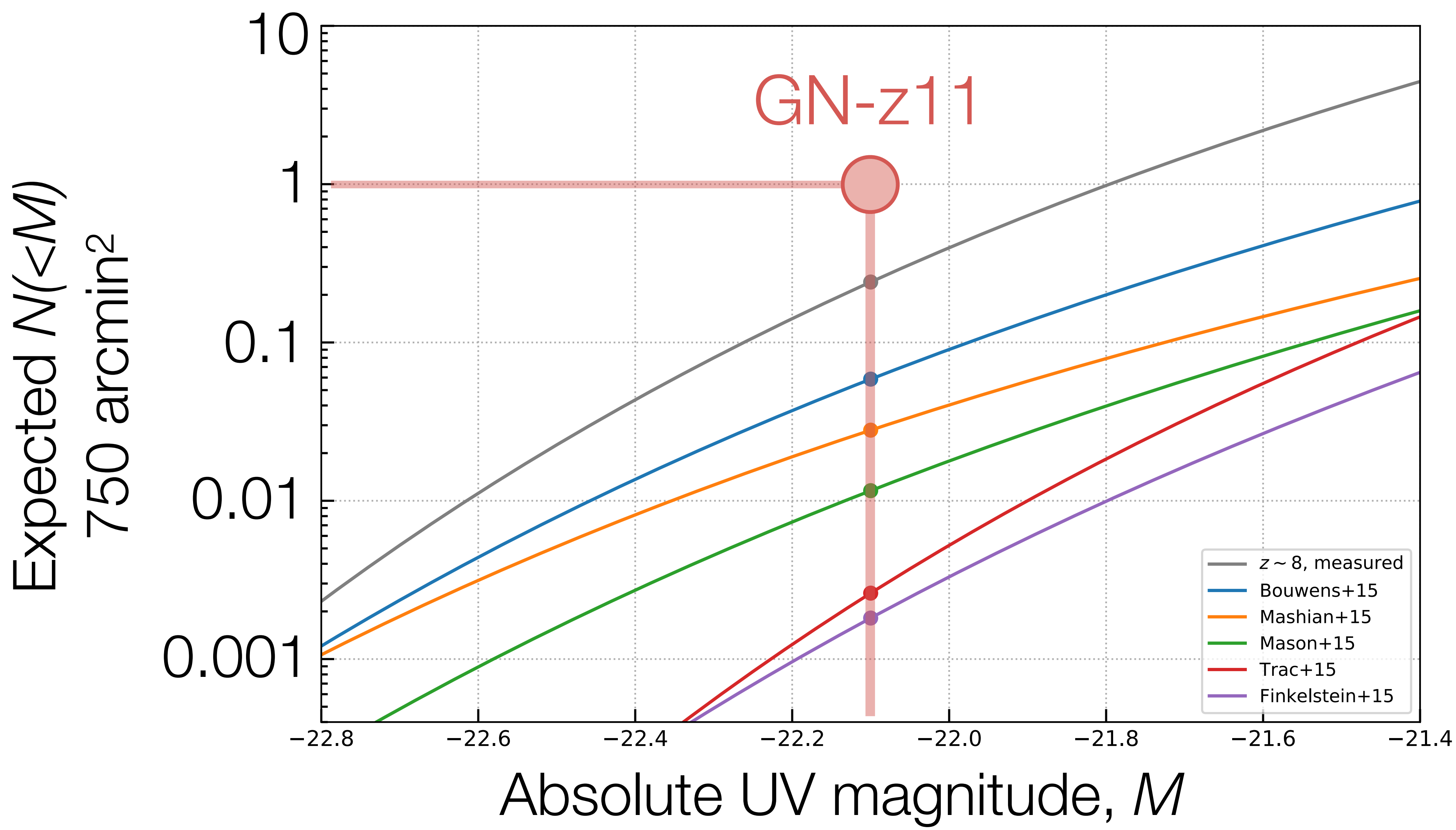
Candidate #2 (GN-z11): z~10? z=11.1±0.1

Oesch,Brammer+16

	GN-Z11	**Milky Way**
Distance:	13.4b light years	–
Radius:	2500 light years	25 × larger
Mass in stars:	$10^9\ M_\odot$	50 × larger
Star formation rate:	$25\ M_\odot$ / yr	50 × *lower*

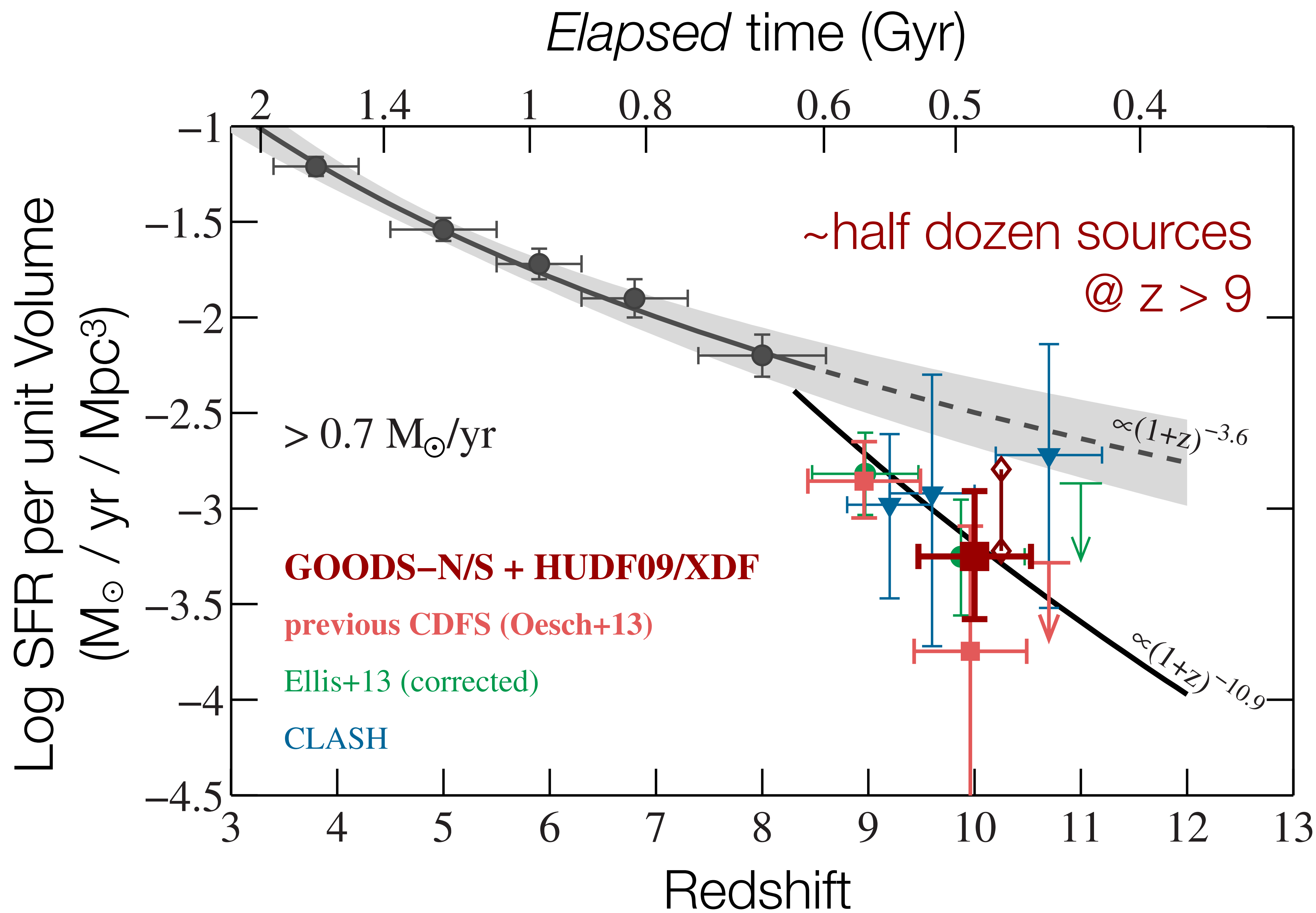
Expected counts at z>10

- Should have needed a survey >15x larger to find GN-z11 based on extrapolations from lower z!

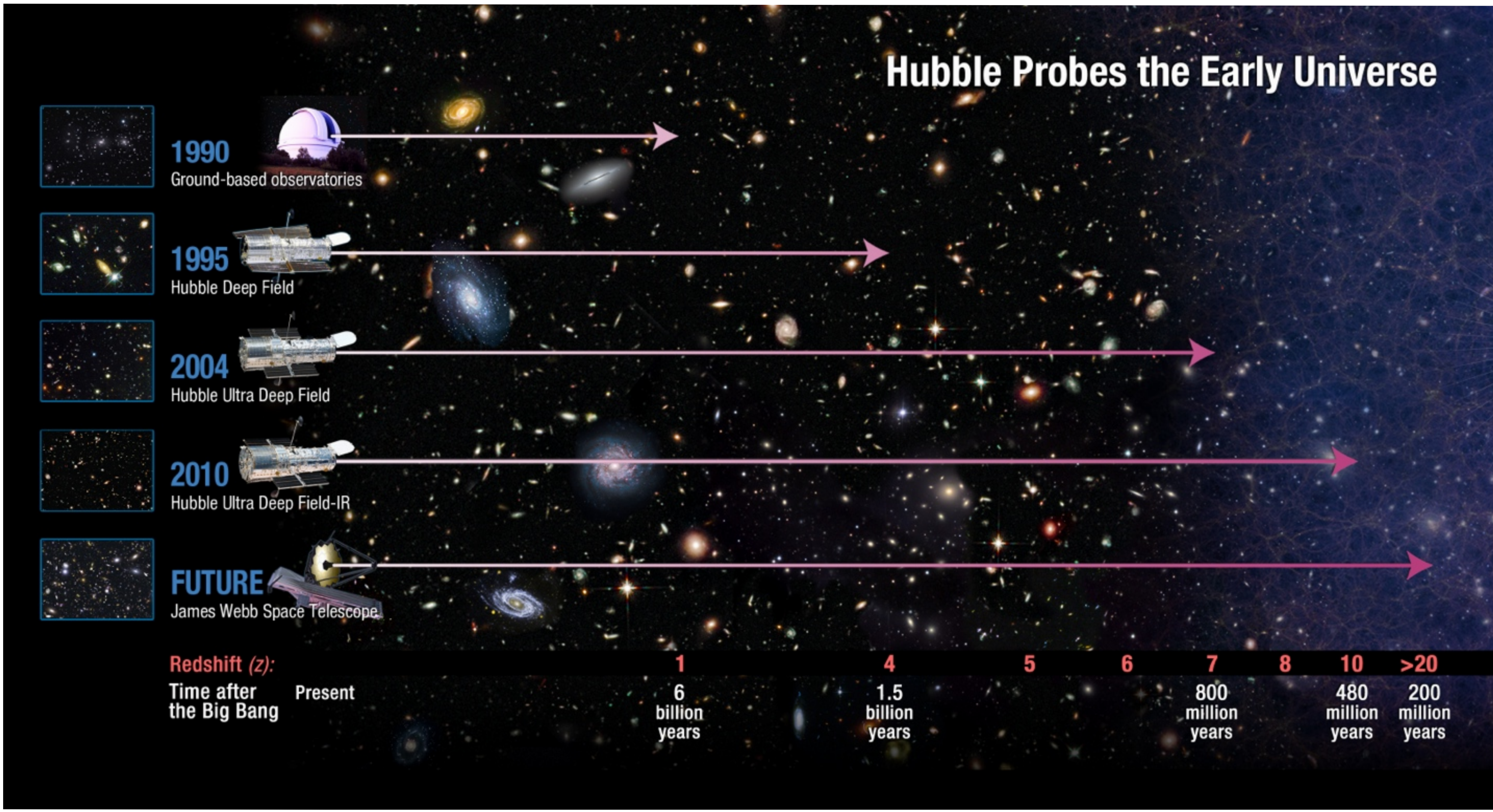


Cosmic Star Formation History: Cosmic Dawn

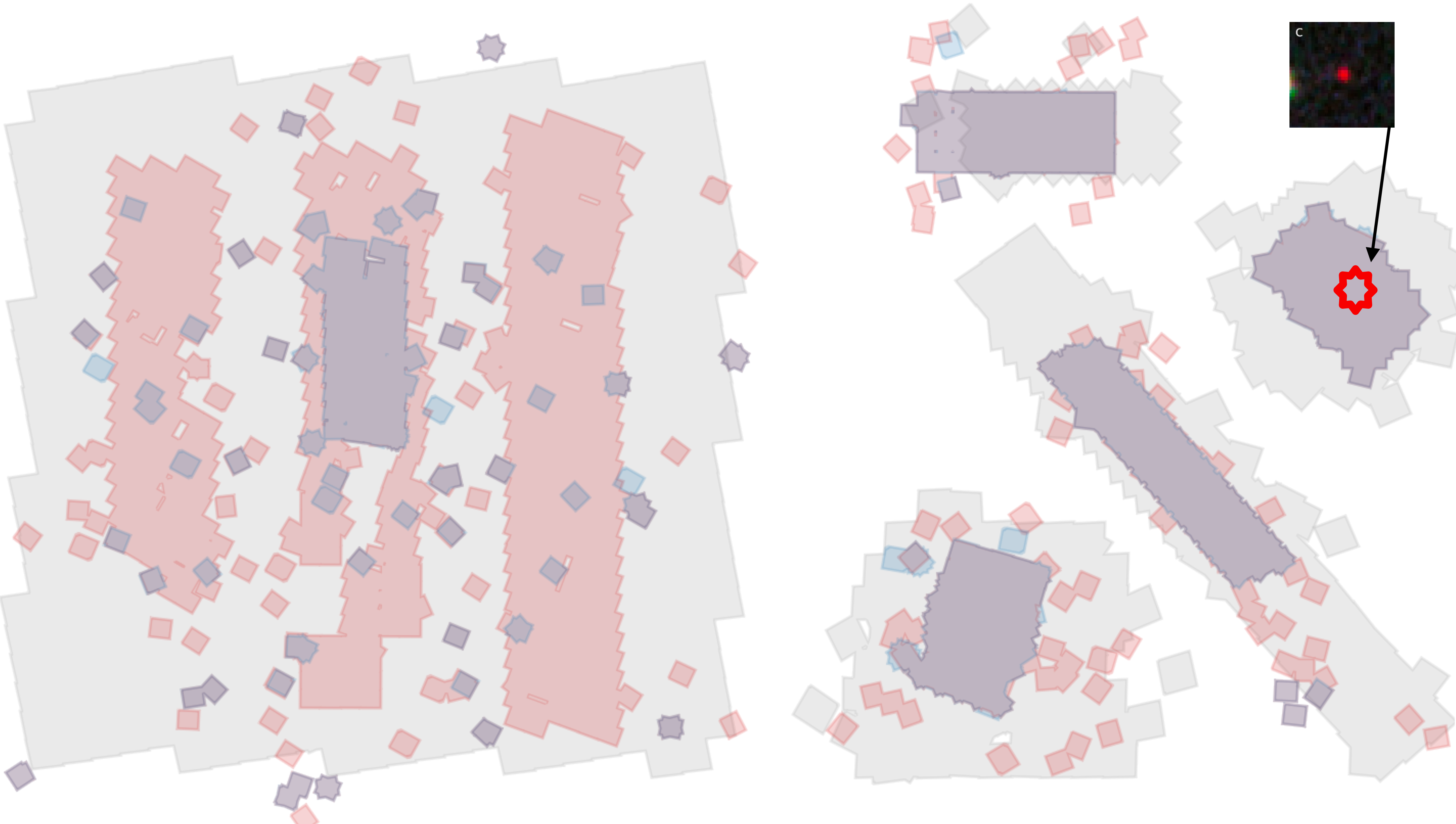
Oesch+15



Cosmic Star Formation History: Cosmic Dawn



Automated "pipeline" processing for more comprehensive search.



Aside: a UV burst in GNz11?

- Jiang et al. (Nature Astronomy, Dec. 2020, https://arxiv.org/abs/2012.06937) reported detection of a **transient burst** while they were observing GN-z11 in 2017!

NEWSLETTERS
Sign up to read our regular email newsletters

NewScientist

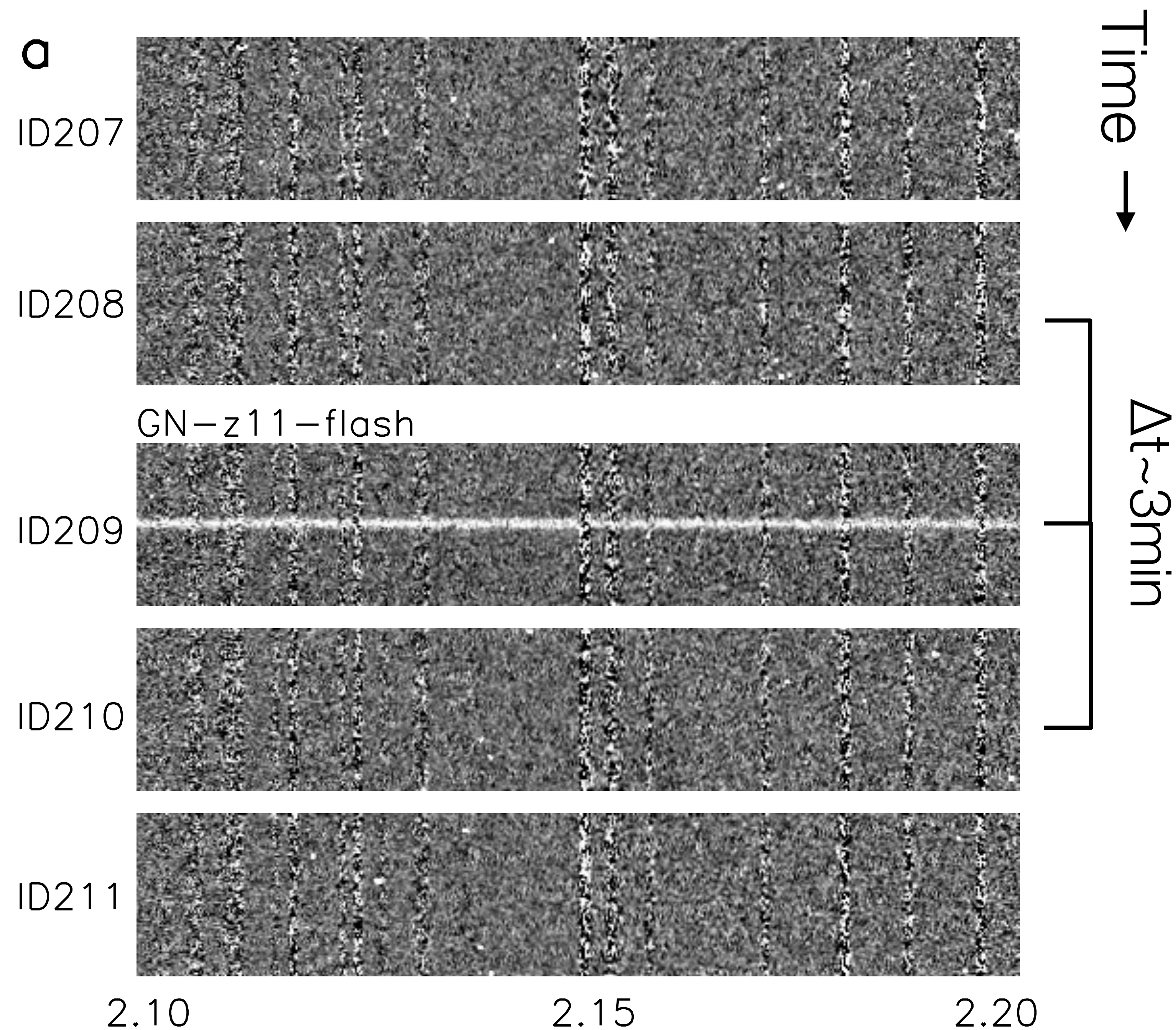
News Podcasts Video Technology Space Physics Health More Shop Courses Events

We may have seen a huge explosion in the oldest galaxy in the universe

SPACE 14 December 2020

Aside: a UV burst in GNz11?

- The data: time series of dispersed spectra from the *Keck* telescope.
- Burst appears in one ~2 minute exposure but is invisible in exposures immediately before and after



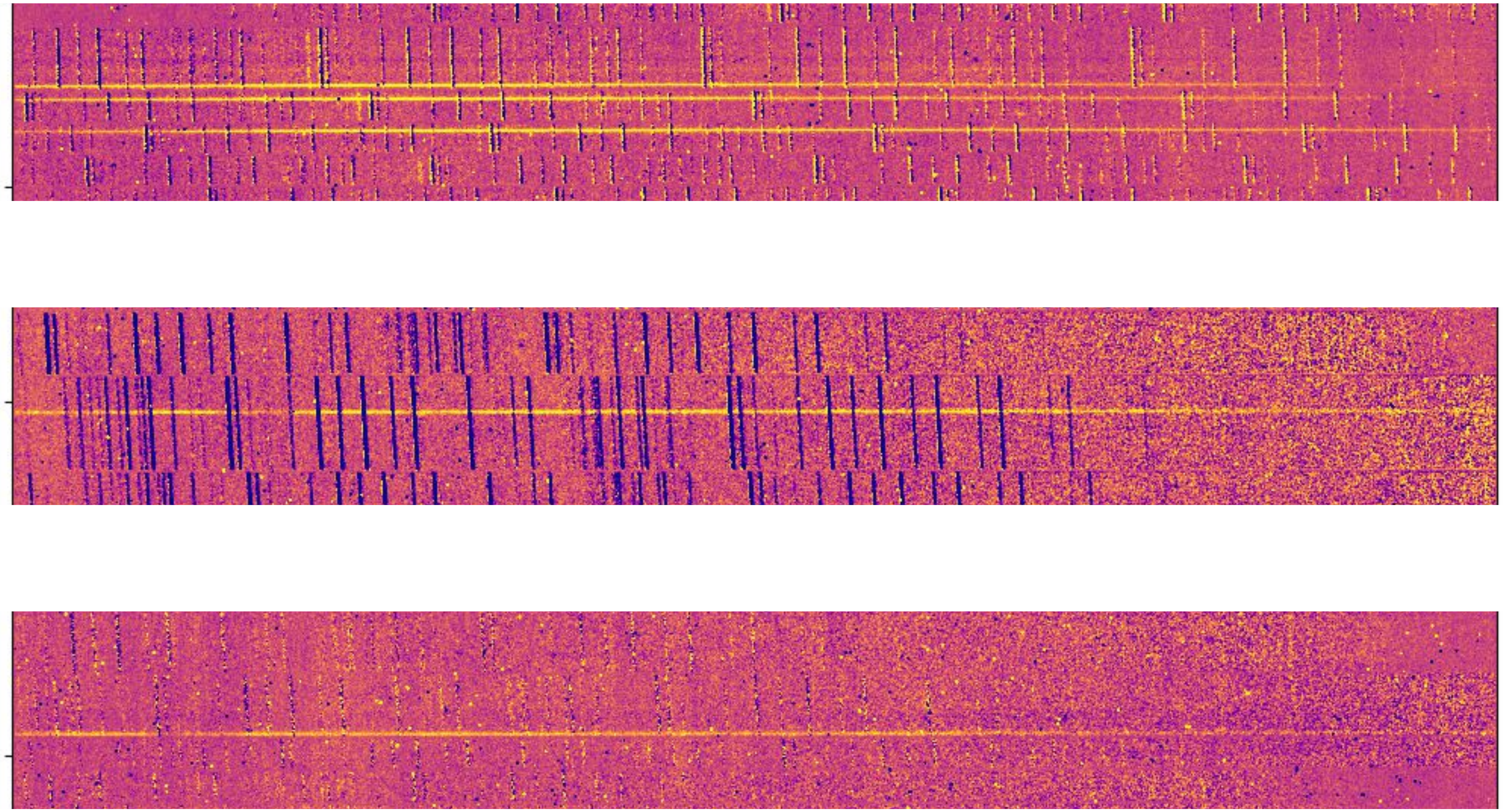
Aside: a UV burst in GNz11?

- Is it above the atmosphere? Yes
- Was it a known artificial satellite? No, but the website used to check is now down
- Was it a known asteroid or other natural body? No
- So it's a Gamma Ray Burst 400 Myr after the Big Bang?

Long GRBs reside in active star-forming galaxies. GN-z11 is a luminous star-forming galaxy with a UV star formation rate of ~26 solar masses per year[17]. During the observations of GN-z11, the chance probability of detecting one GRB as bright as GN-z11 in the UV/optical is estimated to be $(0.3 \sim 60) \times 10^{-10}$ (Methods). This probability is low, but is roughly 10^3 -10^5 times higher than the chance probability of detecting a random GRB, and is at least 2 orders of magnitude higher than the probabilities from other sources considered

However....

- At least 10 similar "flashes" observed in a search of archival datasets from the same instruments in different fields



Aside: a UV burst in GNz11?

- Is it above the atmosphere? Yes
- Was it a known artificial **satellite**? Probably
- Was it a known asteroid or other natural body? No
- ~~So it's a Gamma Ray Burst 400 Myr after the Big Bang?~~

Long GRBs reside in active star-forming galaxies. GN-z11 is a luminous star-forming galaxy with a UV star formation rate of ~26 solar masses per year[17]. During the observations of GN-z11, the chance probability of detecting one GRB as bright as GN-z11 in the UV/optical is estimated to be $(0.3 \sim 60) \times 10^{-10}$ (Methods). This probability is low, but is roughly 10^3 -10^5 times higher than the chance probability of detecting a random GRB, and is at least 2 orders of magnitude higher than the probabilities from other sources considered