Towards Better Planet Occurrence Rates from Kepler and K2

Andrew Vanderburg

NASA Sagan Fellow

The University of Texas at Austin

Sagan/Michelson Fellows Symposium

November 9, 2017

Collaborators

Andy Mayo

Harvard undergrad
—> Copenhagen Univ. Fulbright
—> UC Berkeley grad (Fall 2018)

Chris Shallue

Google Brain

Outline

- Planet occurrence rates and where we stand now

- Neural networks and how we can use them to identify planets

- Spectroscopy of K2 candidates to measure the planet radius distribution

Image: NASA

Planet Occurrence Rates

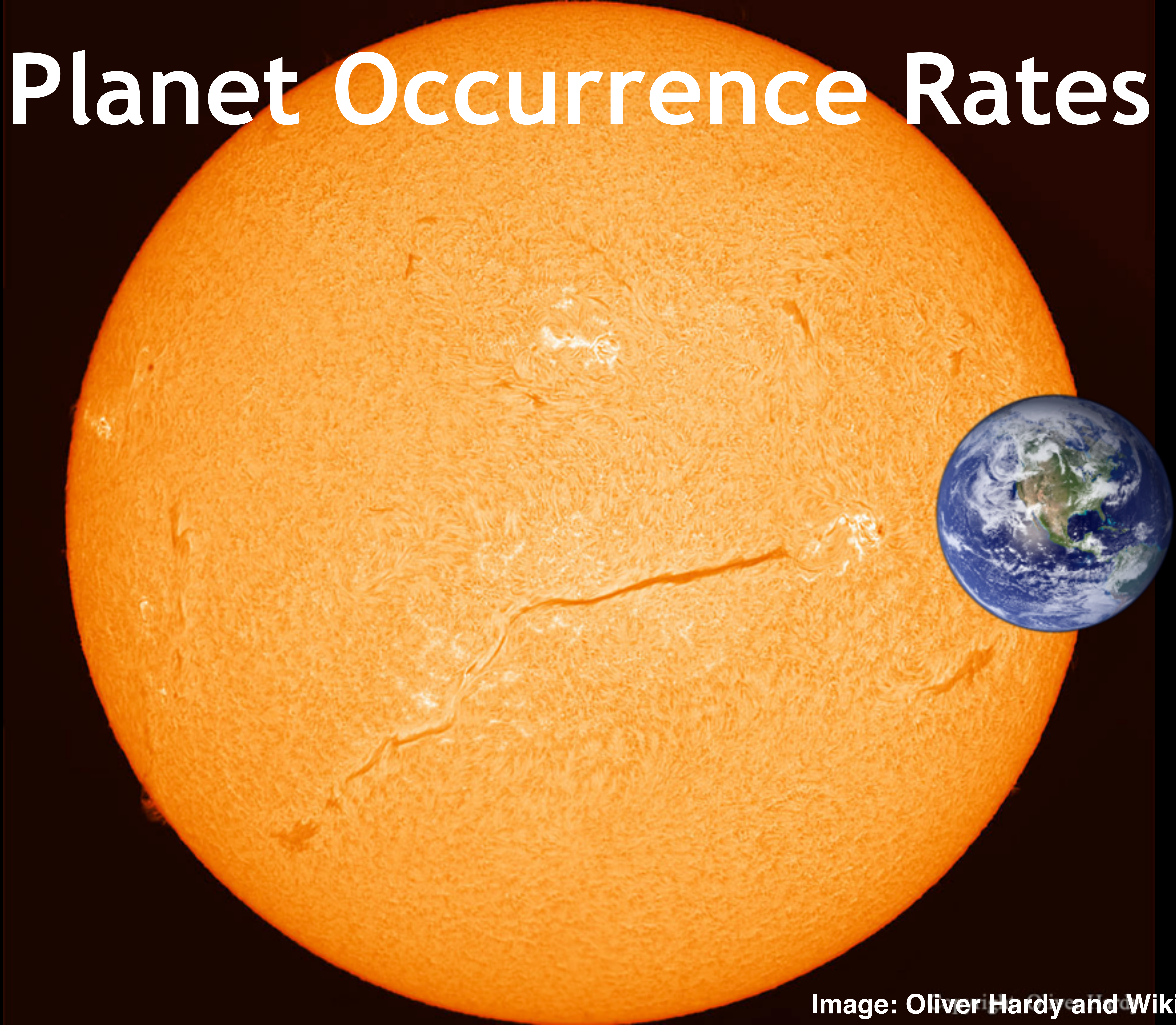
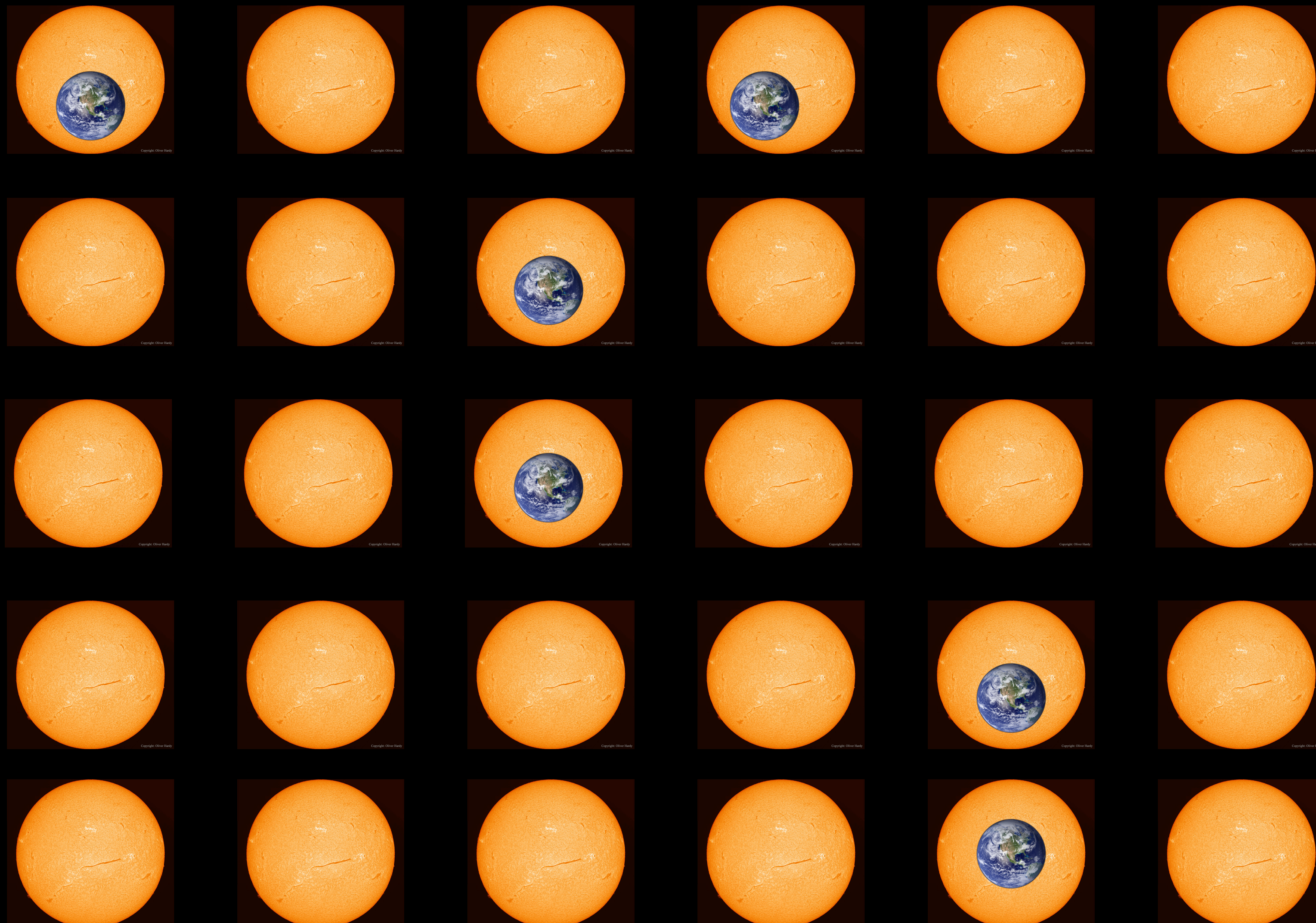
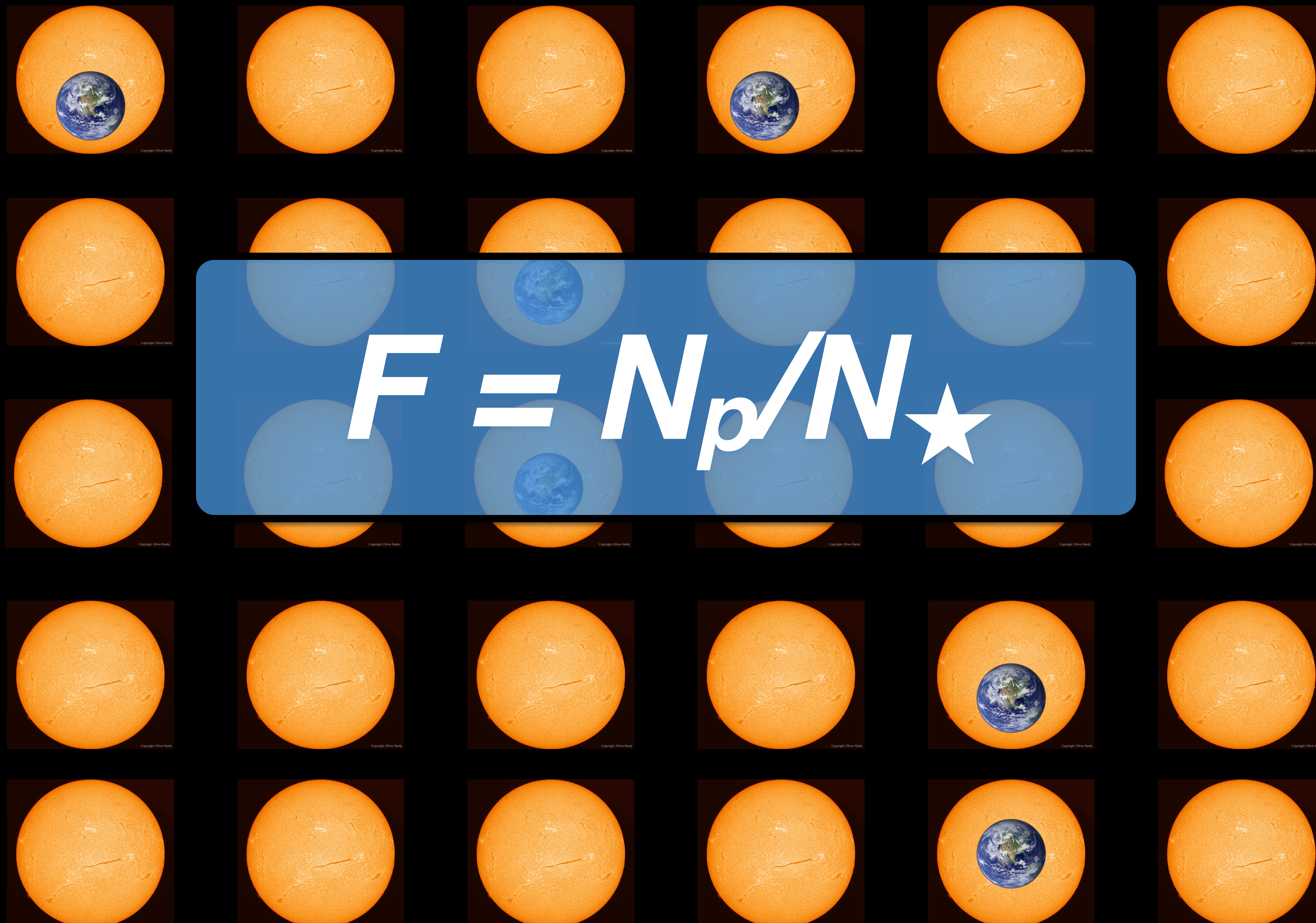


Image: Oliver Hardy and Wikimedia

Planet Occurrence Rates

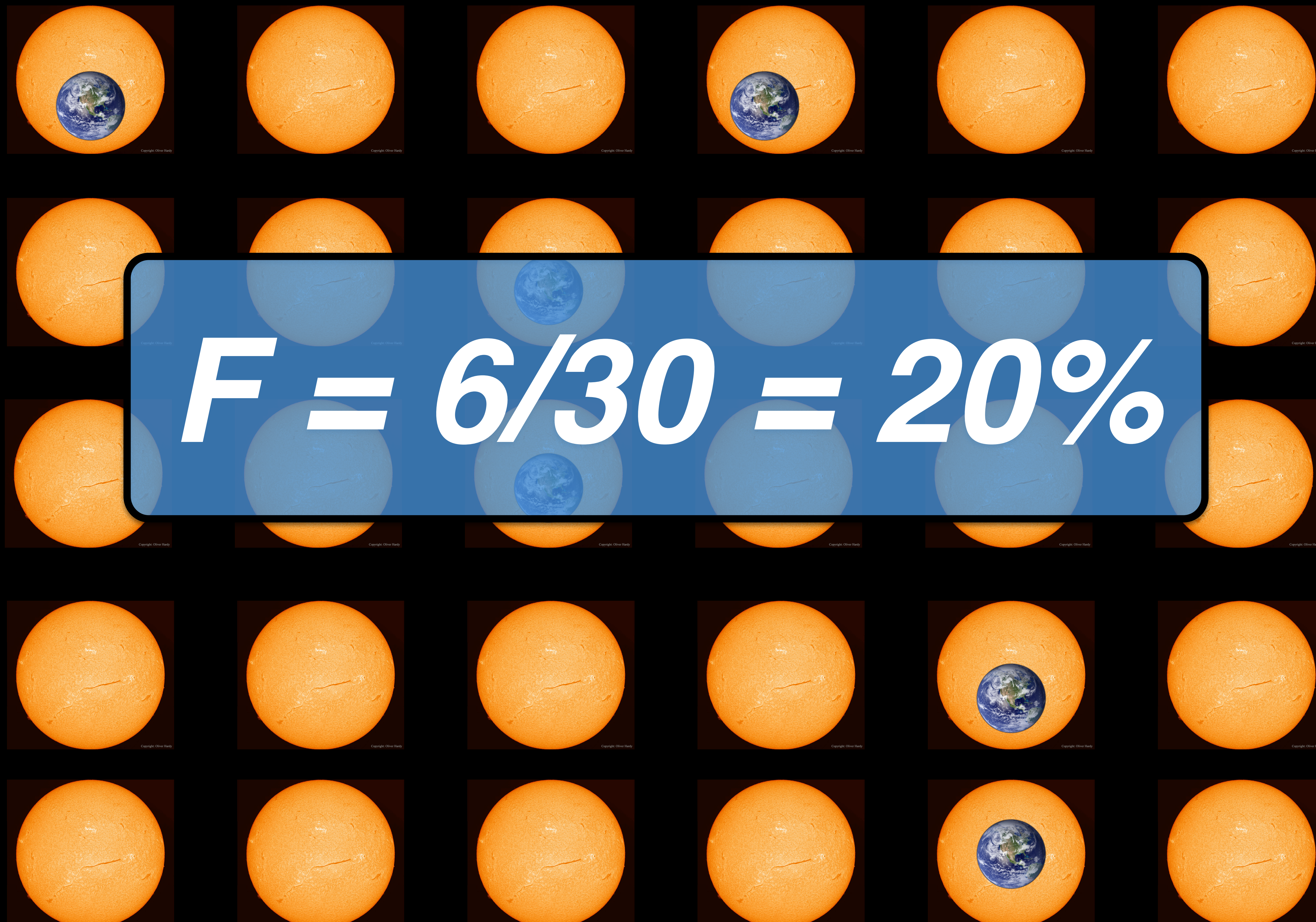


Planet Occurrence Rates

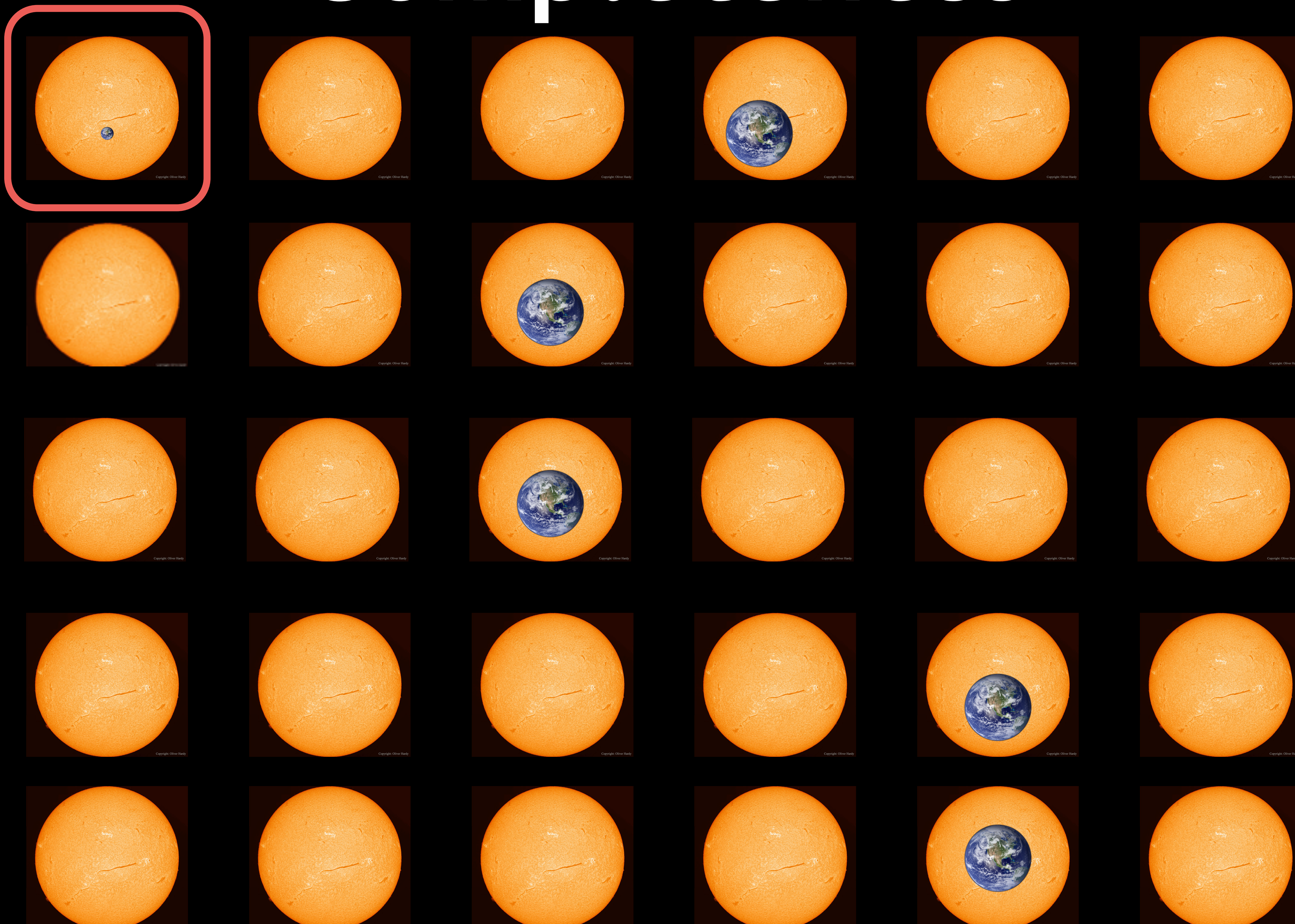


$$F = N_p / N_\star$$

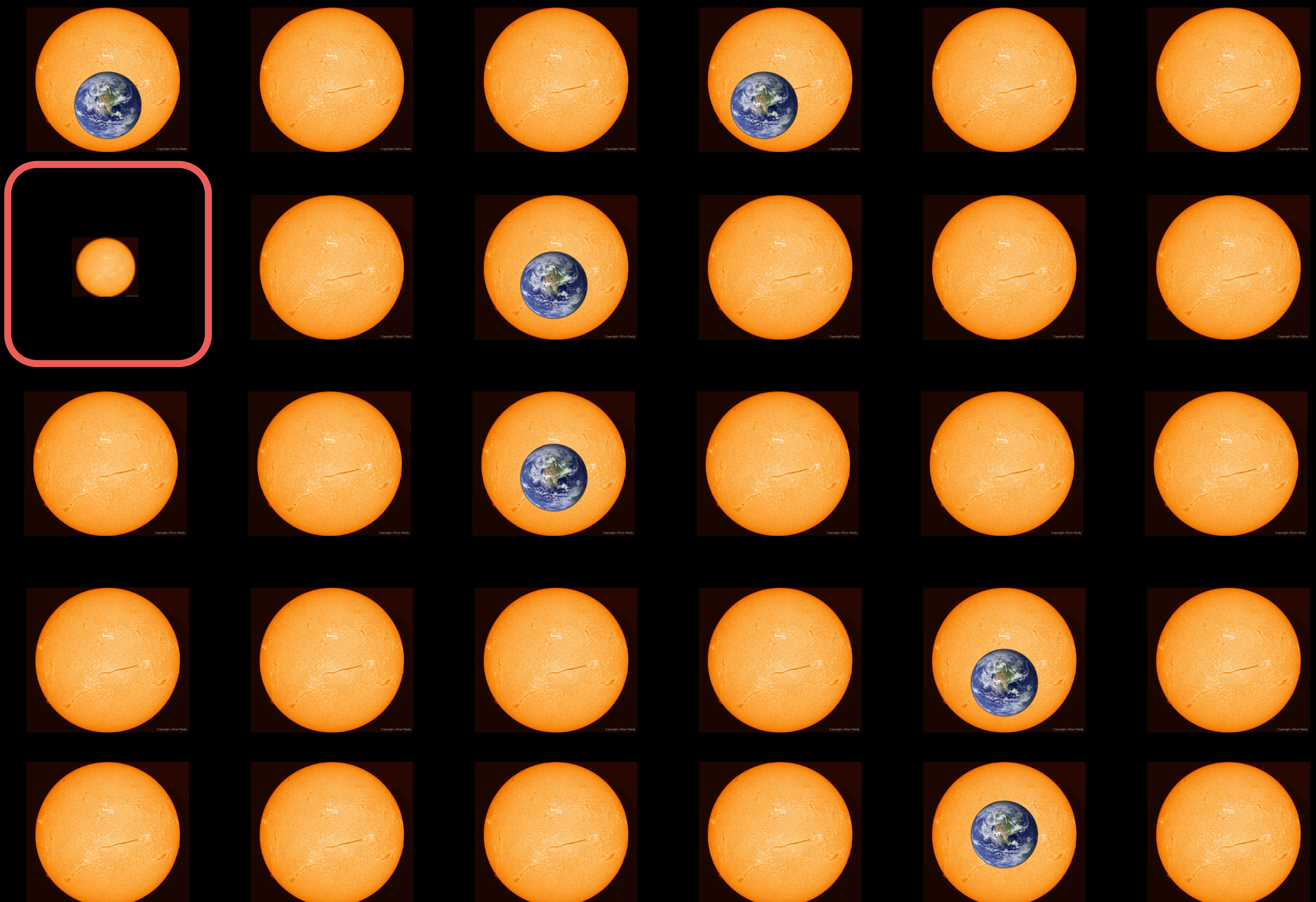
Planet Occurrence Rates



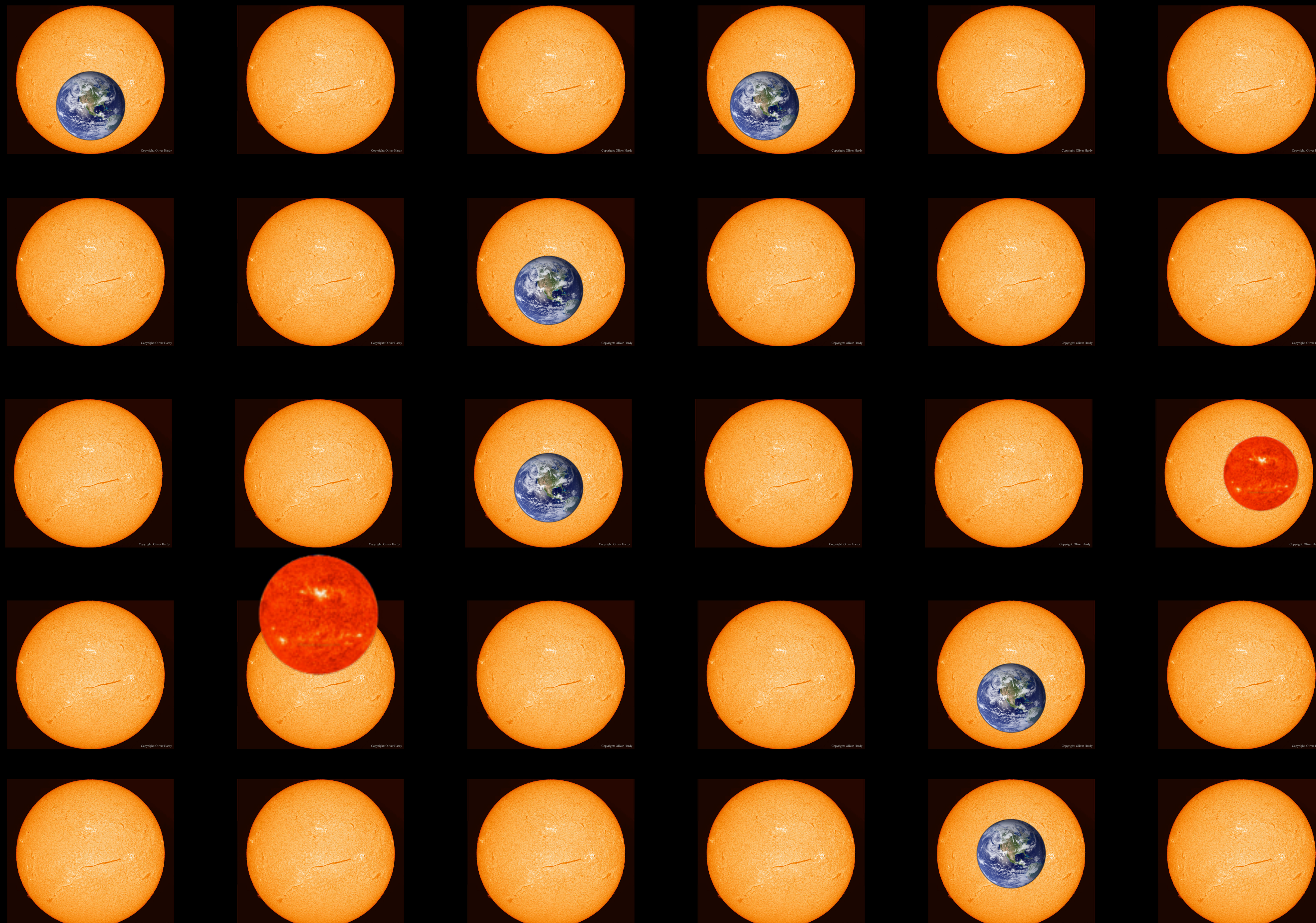
Completeness



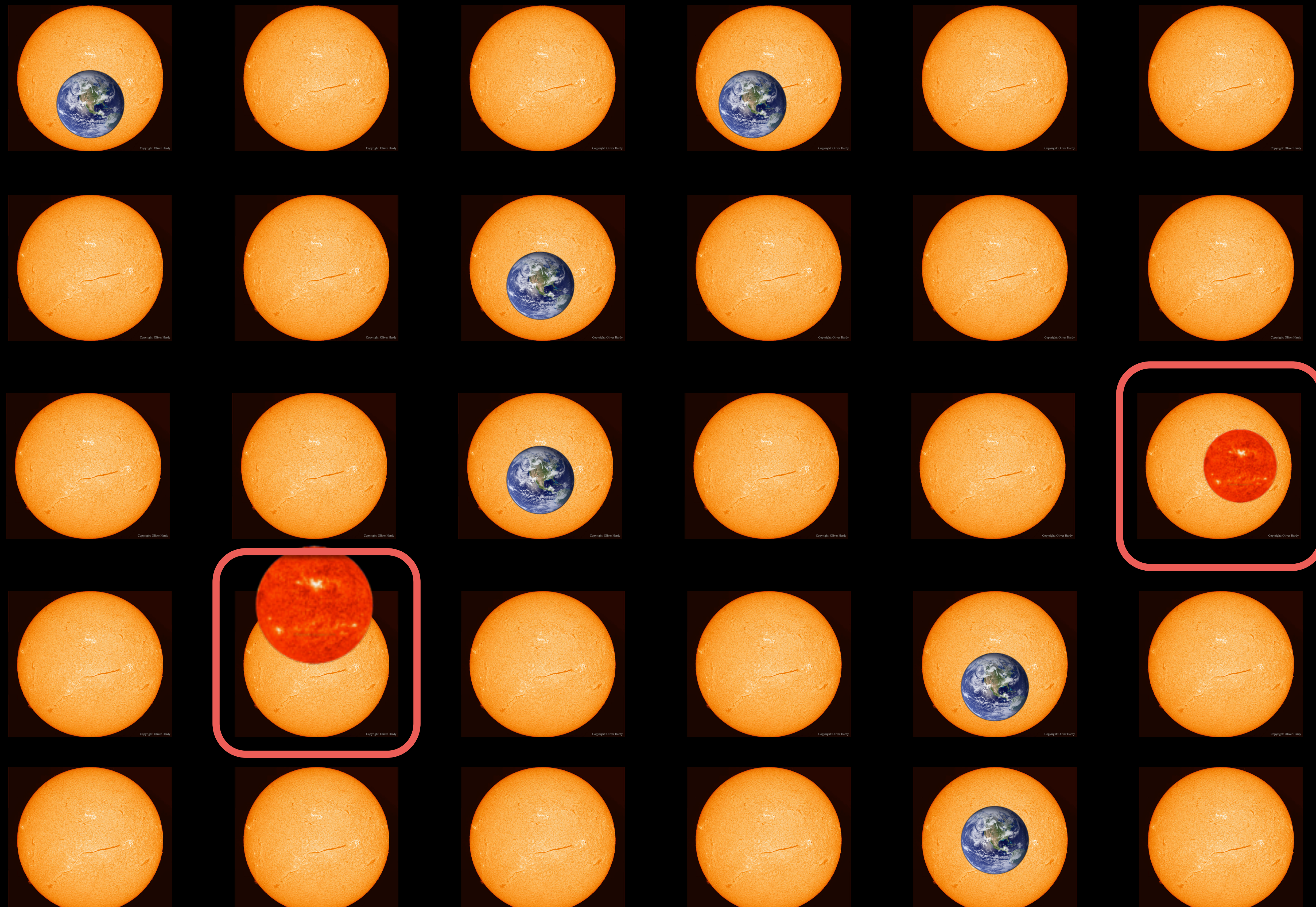
Completeness



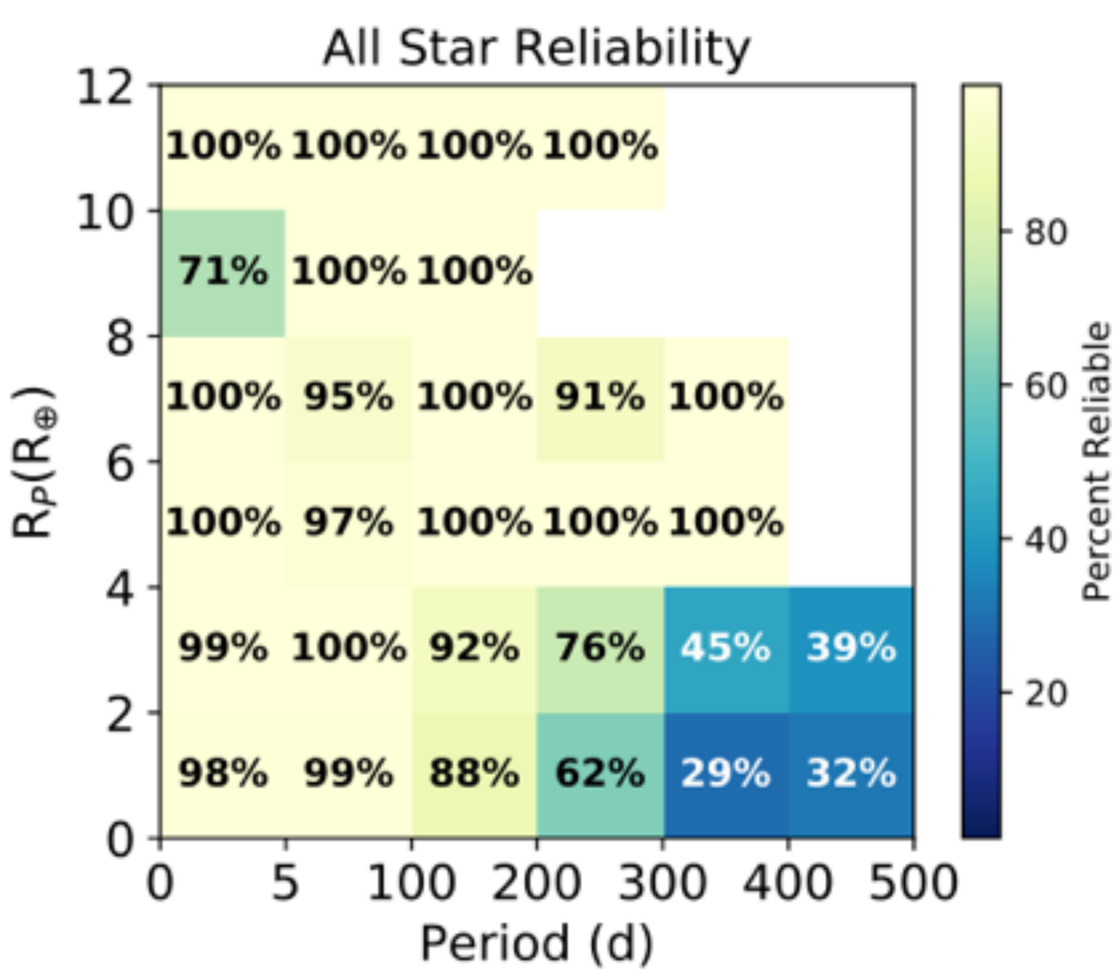
Reliability



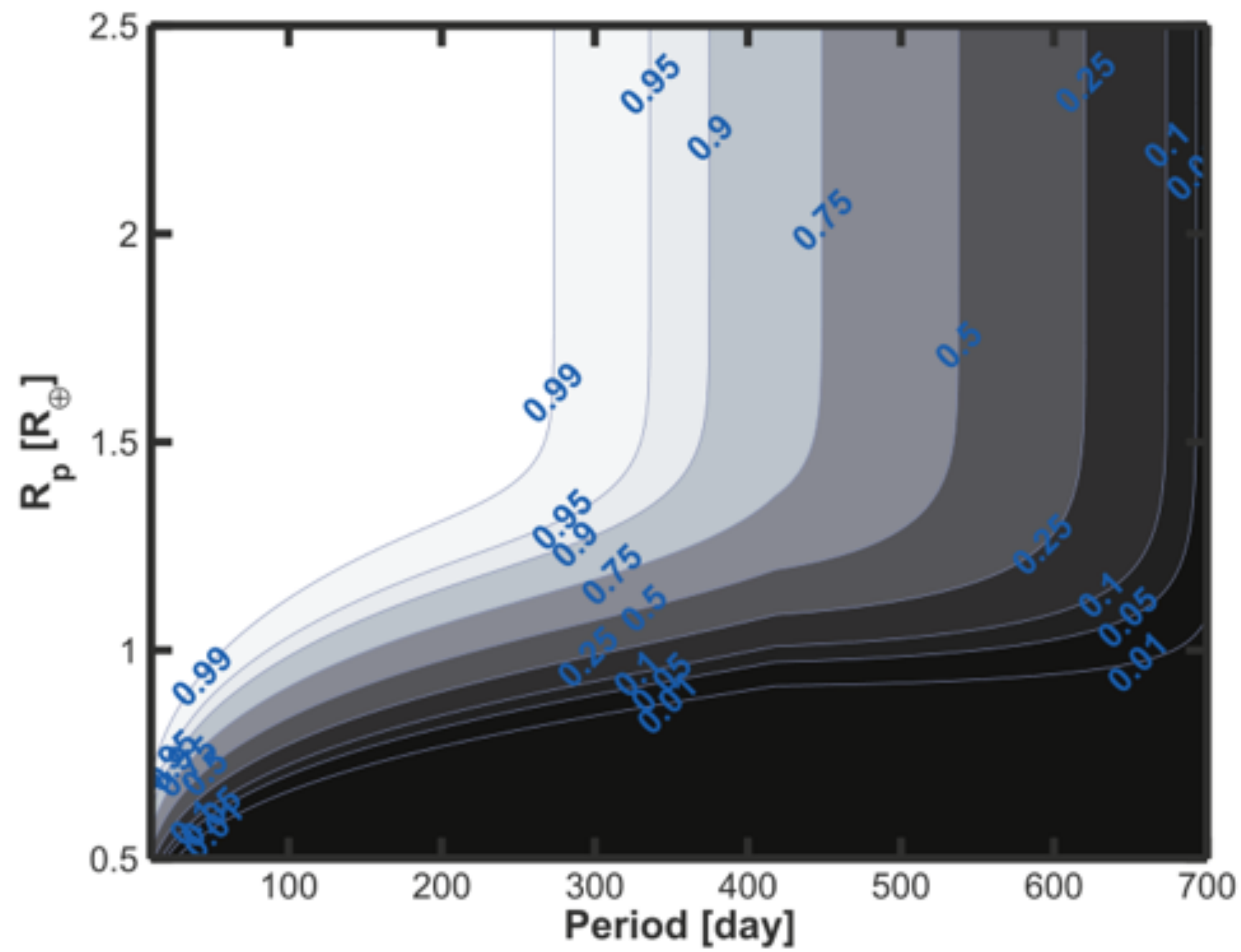
Reliability



Kepler is incomplete and unreliable at the limits of its sensitivity.

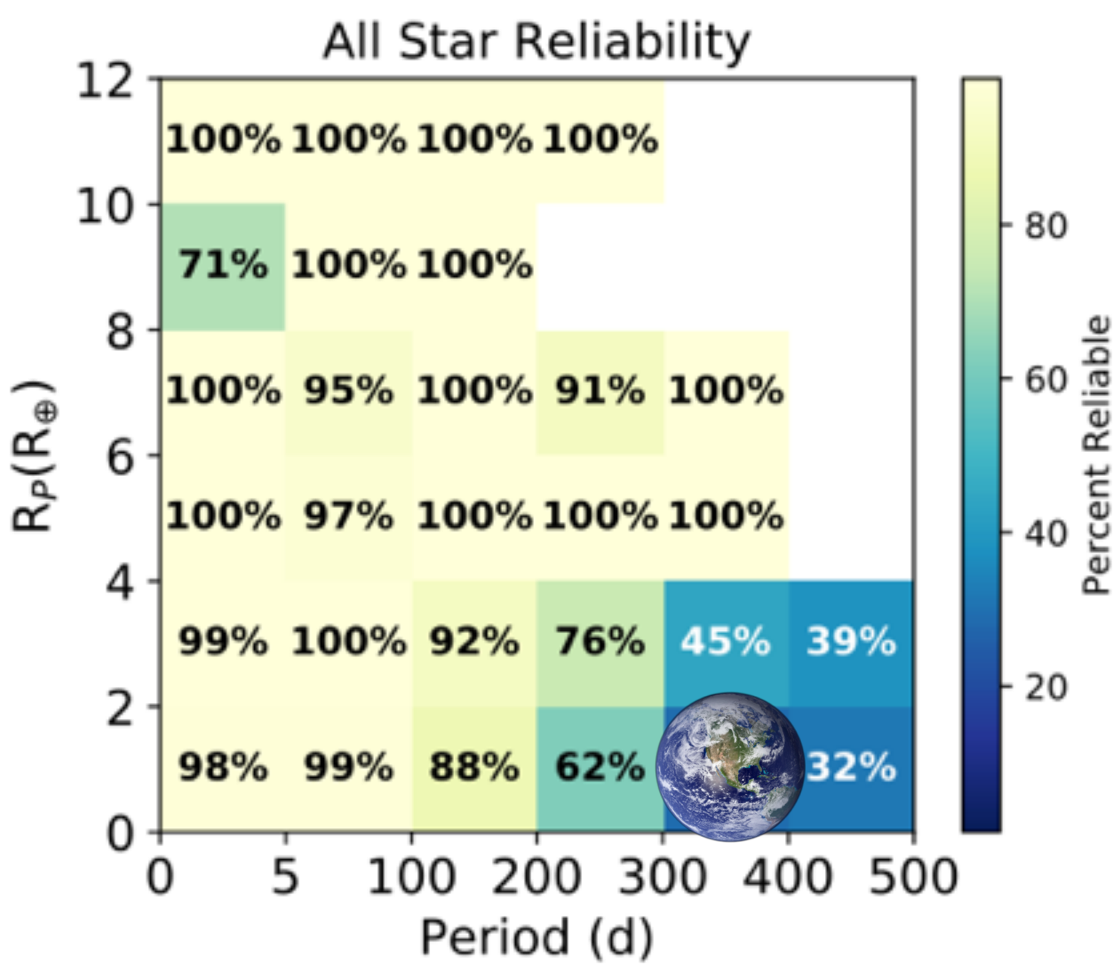


Thompson+2017

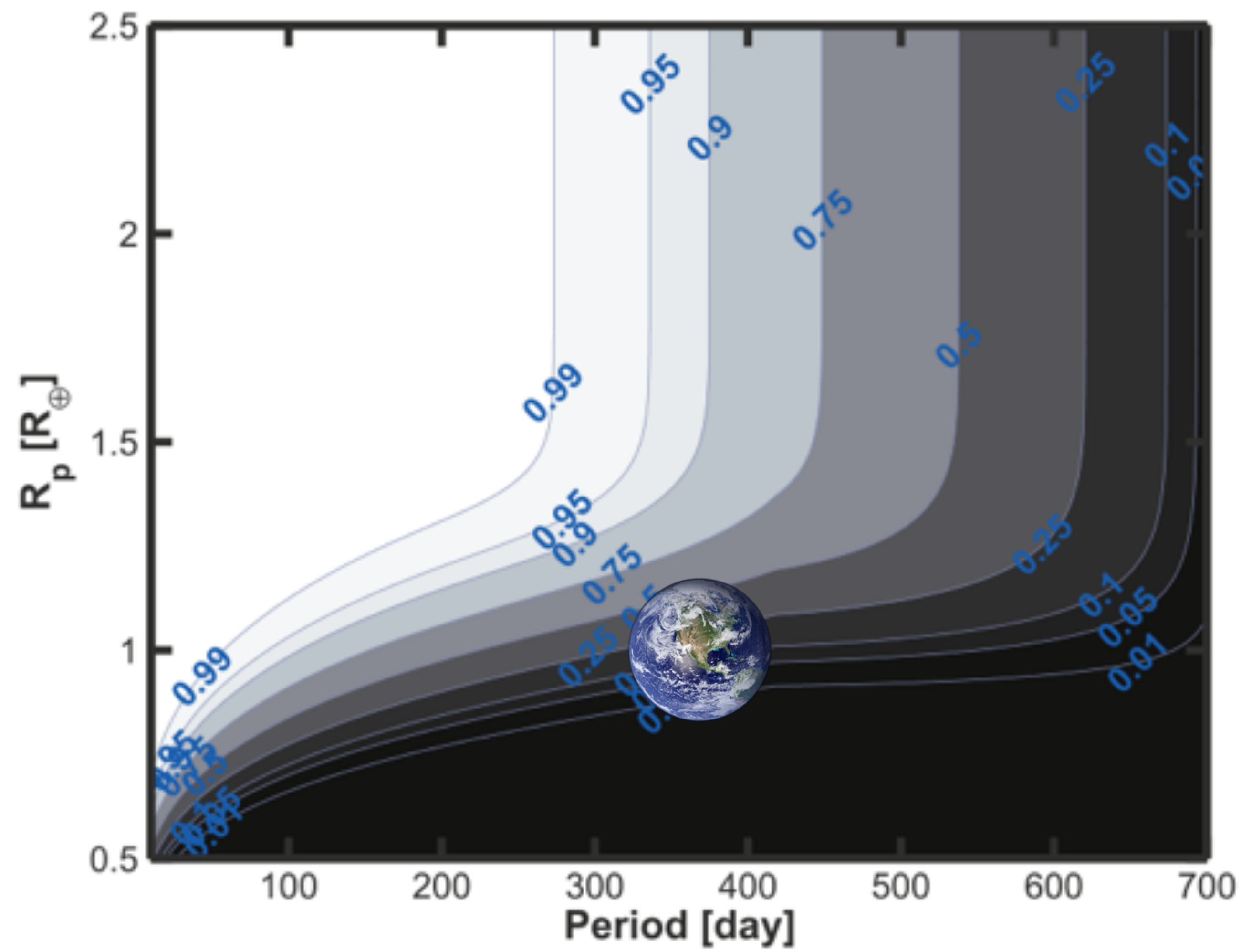


Burke+2015

Kepler is incomplete and unreliable for *Earth-sized planets in Earth-like orbits*.

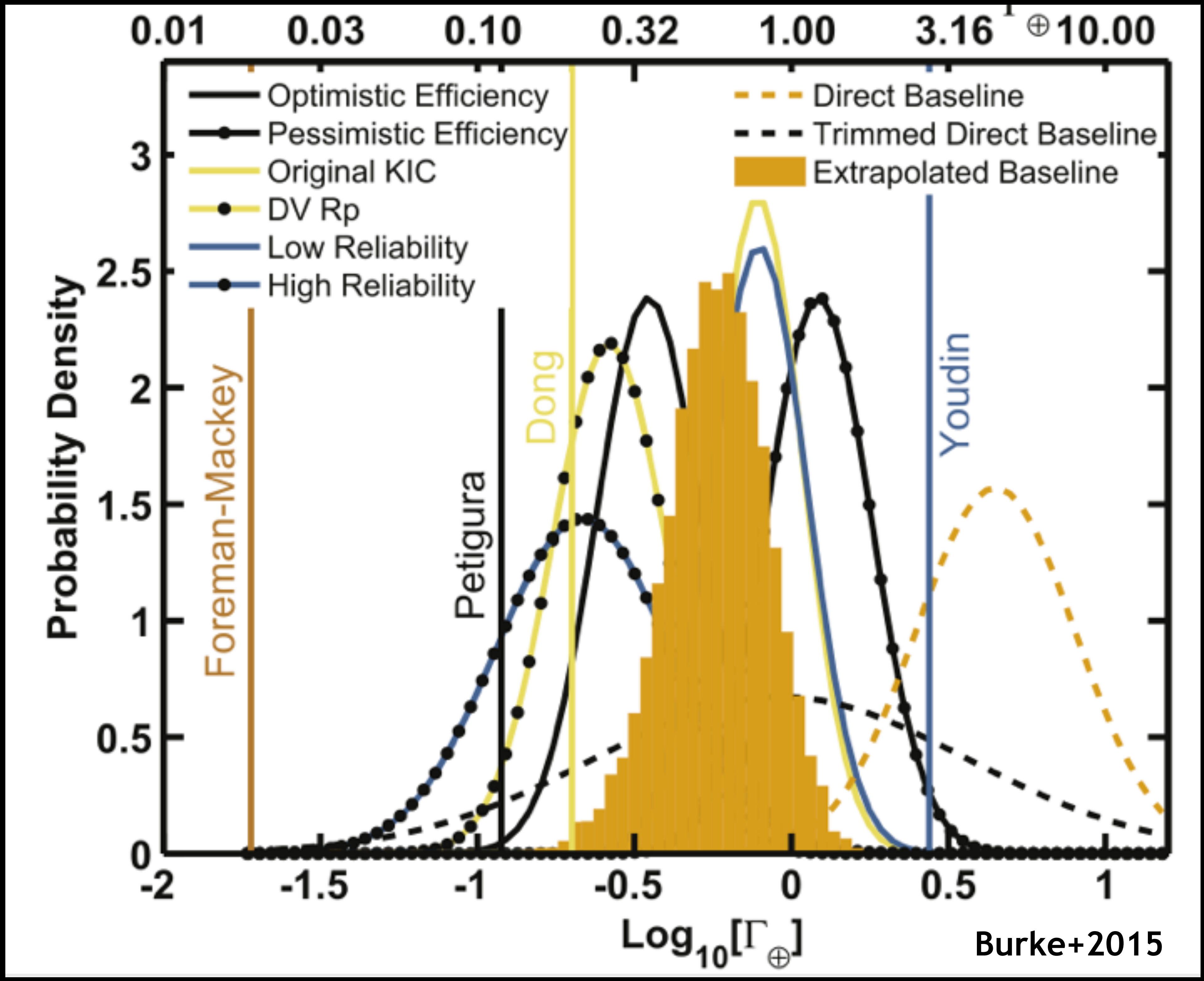


Thompson+2017

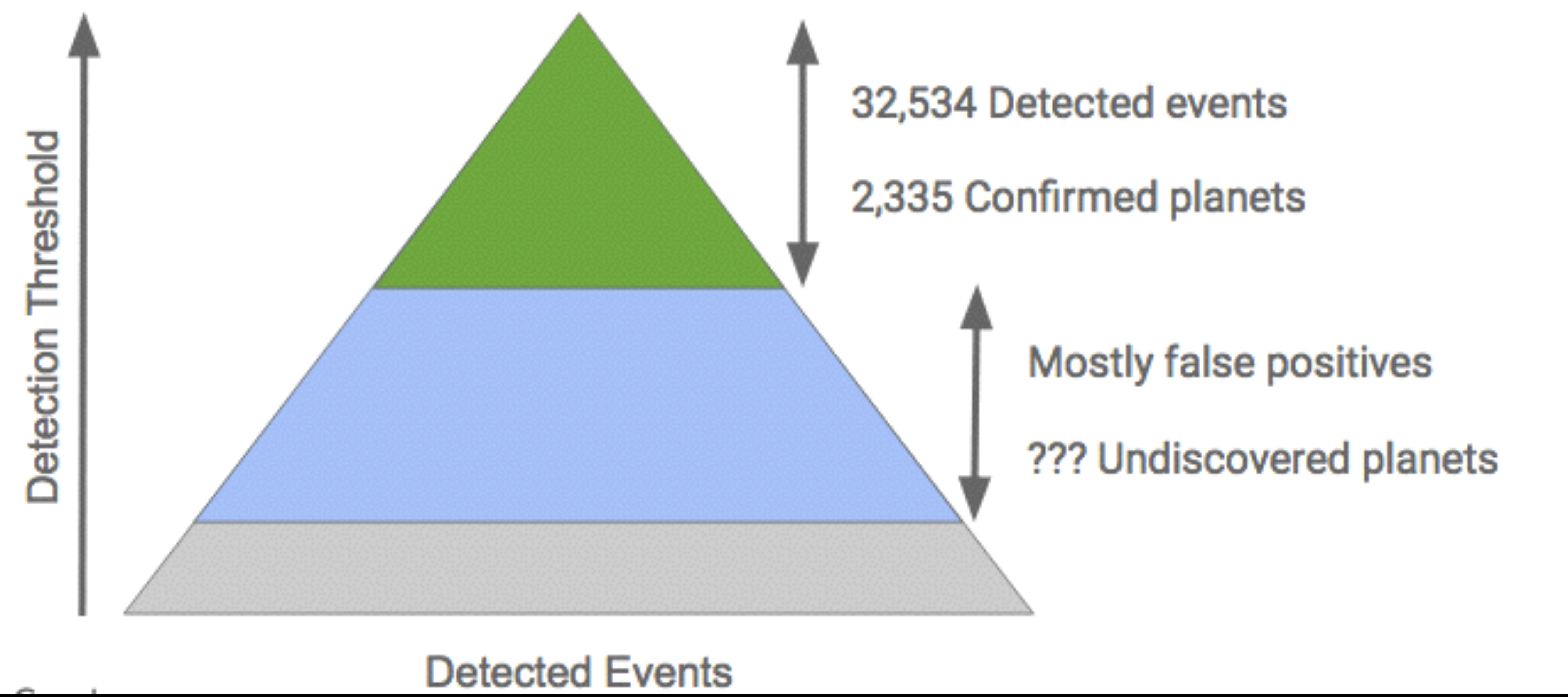


Burke+2015

There is an order of magnitude uncertainty on the occurrence of Earths around Sun-like stars



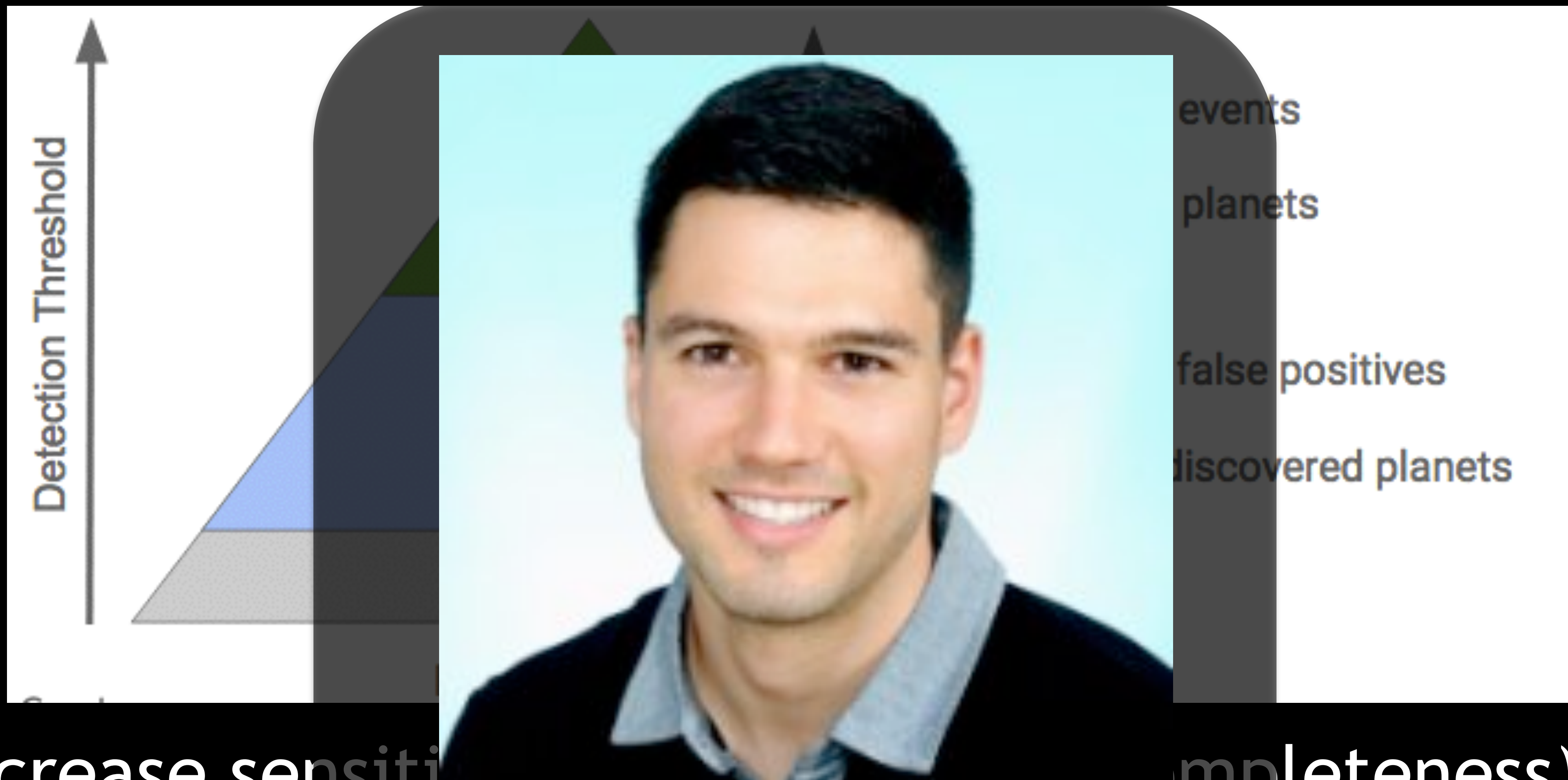
Our Approach



- 1. Increase sensitivity (and therefore completeness) by allowing weaker signals to be considered as planet candidates, at the cost of a higher false positive rate.

- 2. Use **deep learning** to more effectively distinguish real signals from false alarms and false positives, keeping reliability high.

Our Approach



Chris Shallue

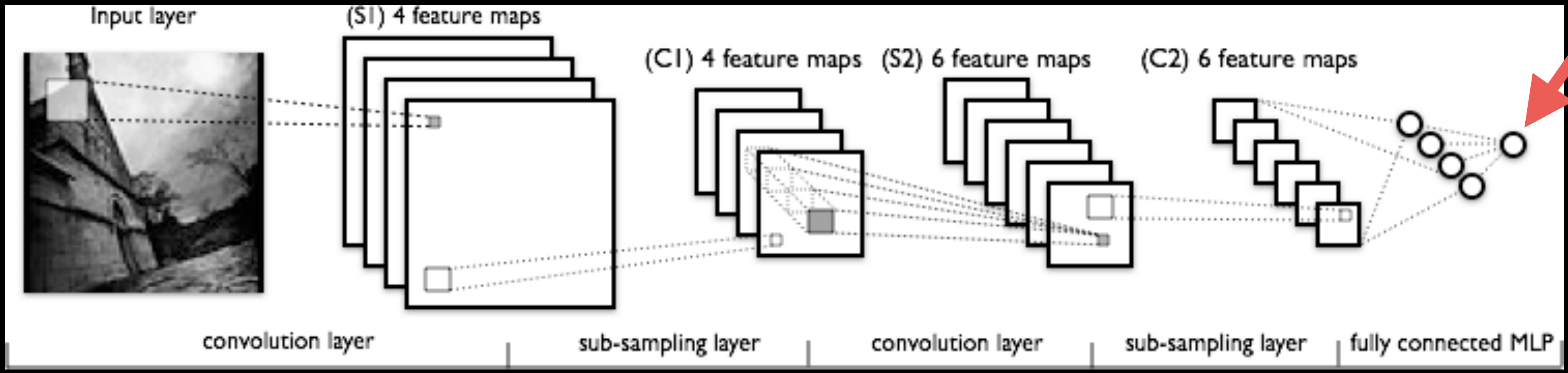
Google Brain

- 1. Increase sensitivity (and therefore completeness) by allowing weaker signals to be considered as planet candidates, at the cost of a higher false positive rate.
- 2. Use **deep learning** to more effectively distinguish real signals from false alarms and false positives, keeping reliability high.

Deep Learning

Given some image/vector as input, perform math operations which convert the image to an output.

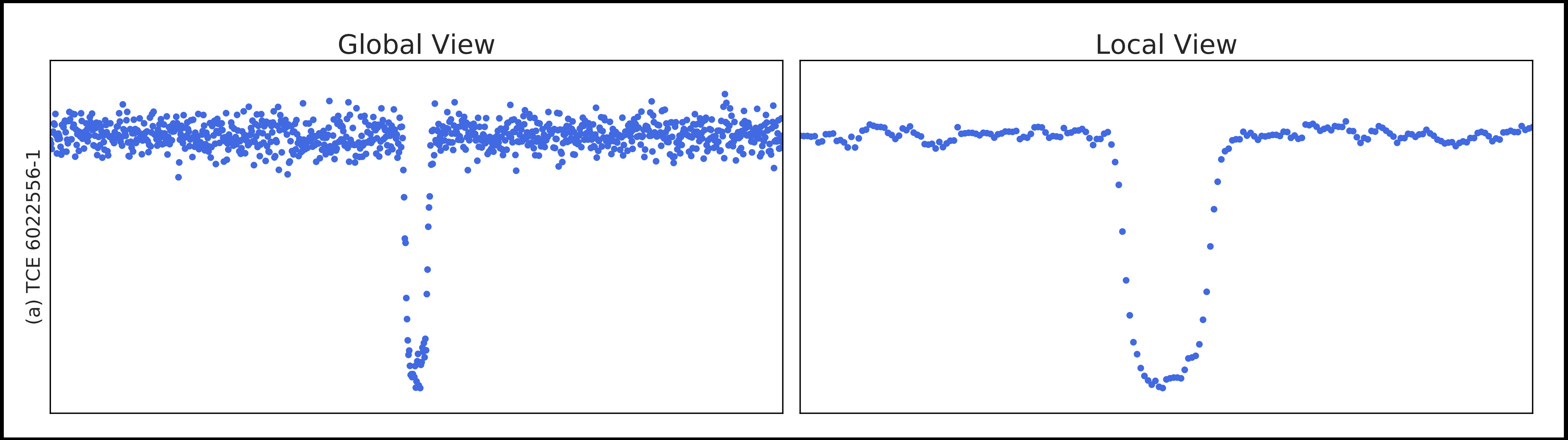
Church



The operations are "learned" by minimizing some cost function which compares data from a training set to their known classification.

Image: deeplearning.net

Inputs: binned and phase-folded *Kepler* light curves

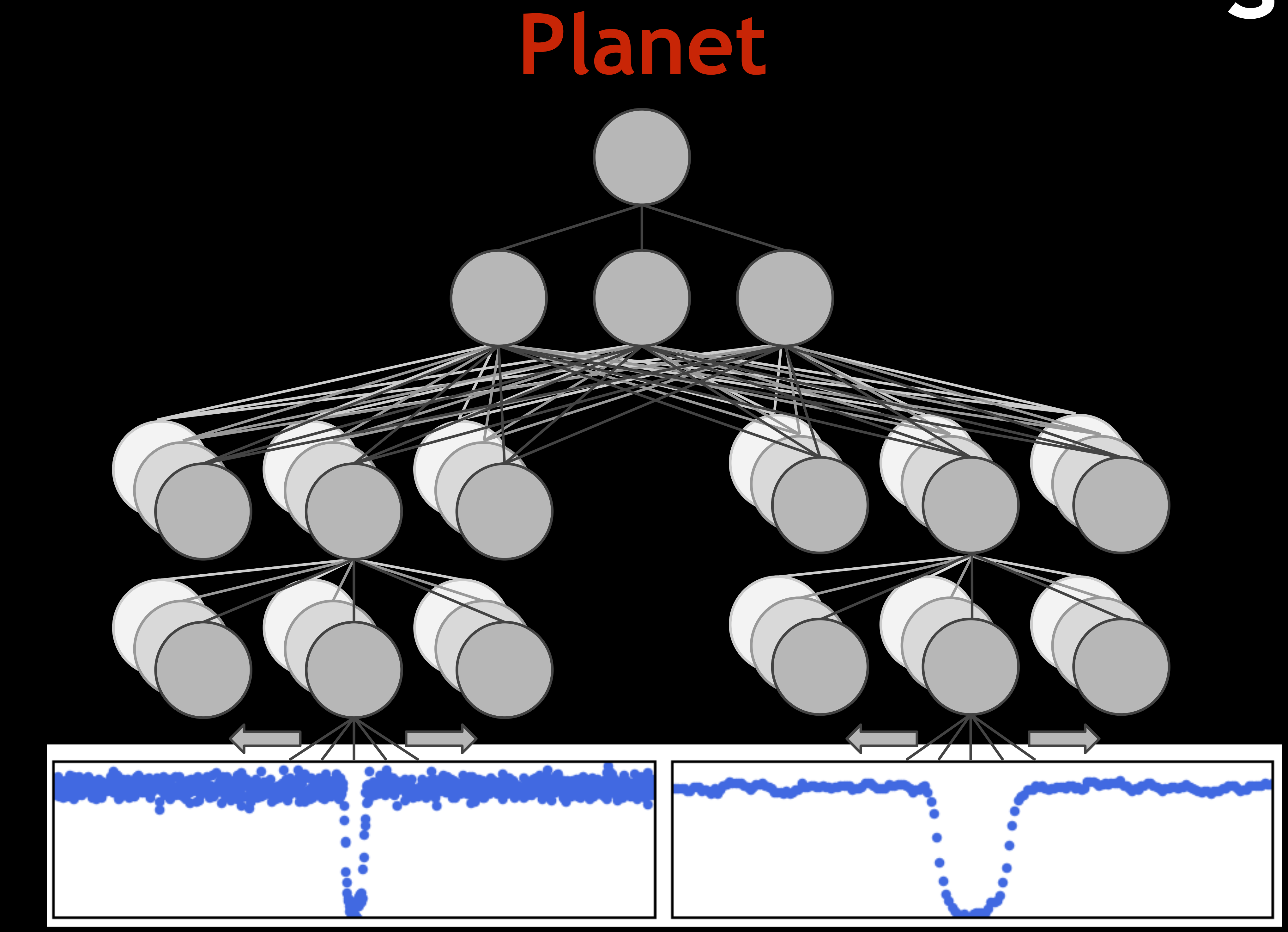


Global View:
shows phase-variations,
secondary eclipses, etc

Local View: close-up
look at the transit

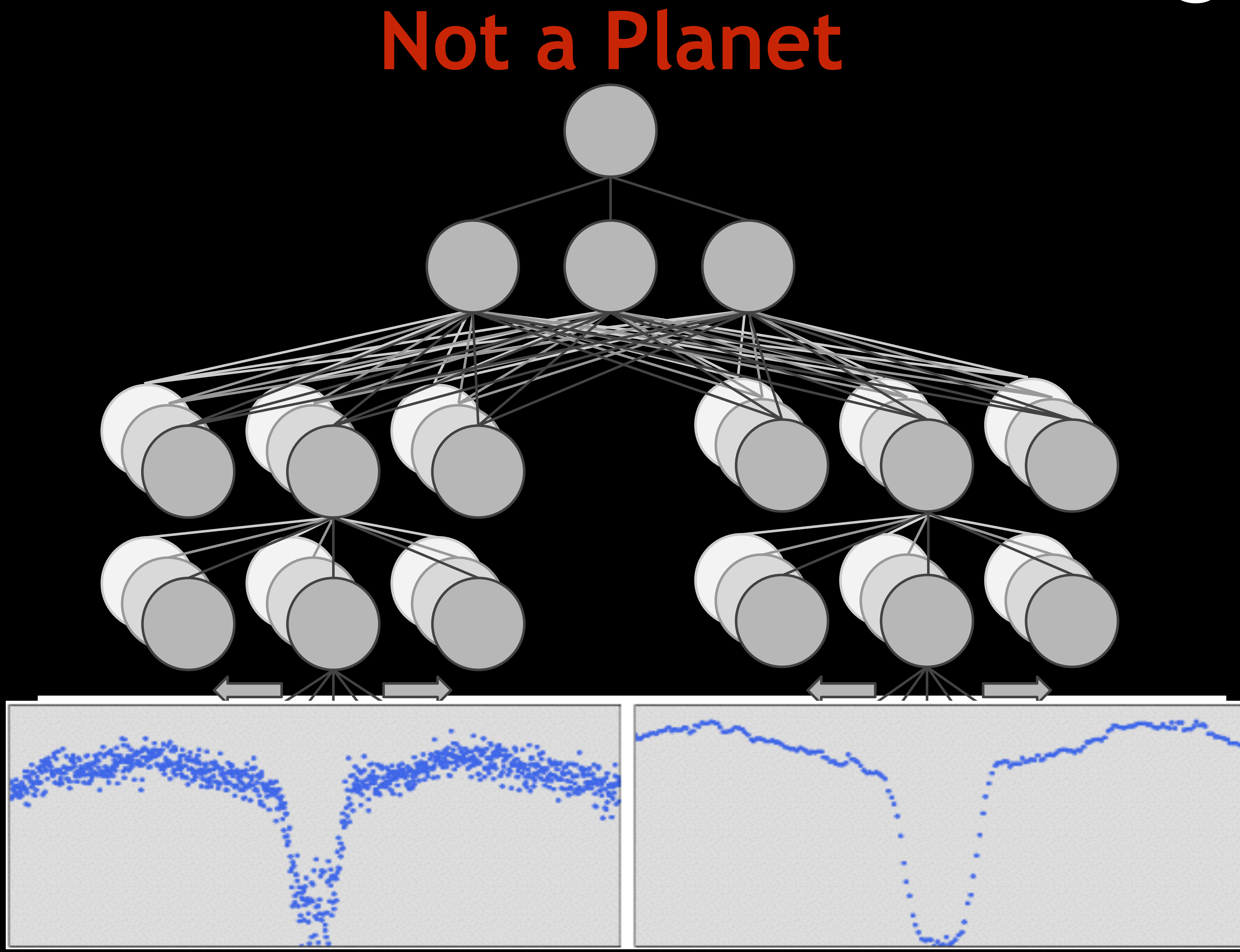
Shallue & Vanderburg (submitted)

Proof of Concept Neural Network for Planet Vetting



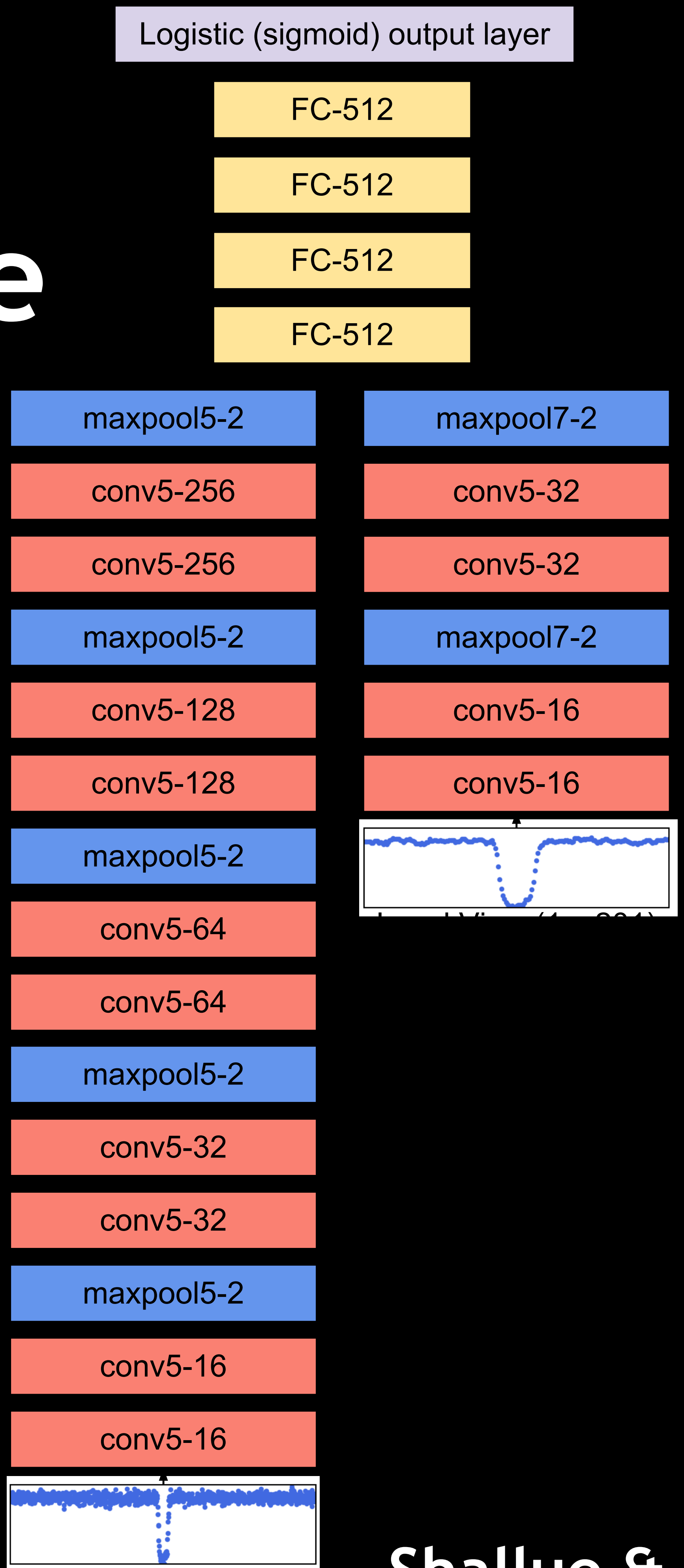
Shallue & Vanderburg (submitted)

Proof of Concept Neural Network for Planet Vetting



Shallue & Vanderburg (submitted)

Network Architecture



Shallue & Vanderburg (submitted)

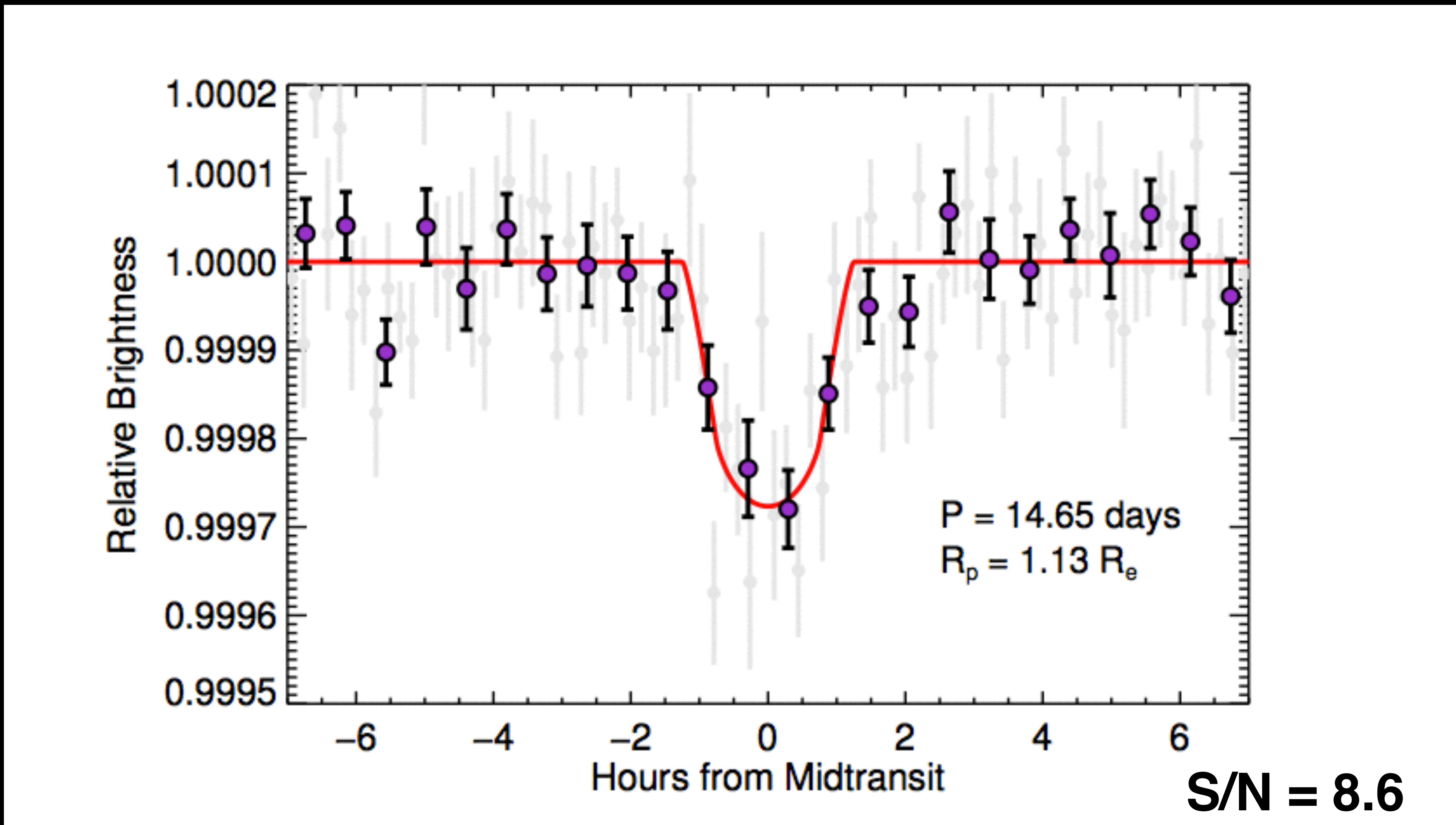
Results

	AUC	Accuracy
Our work	**0.988**	0.960
Coughlin et al. (Robovetter)	0.974	**0.974**
Armstrong et al.		0.87
McCauliff et al. (Autovetter)	(0.997)* not held-out results	(0.986)* not held-out results

Comparable to industry standards like Robovetter, but work necessary to incorporate other information as inputs (difference images, quarterly depth variations, etc.)

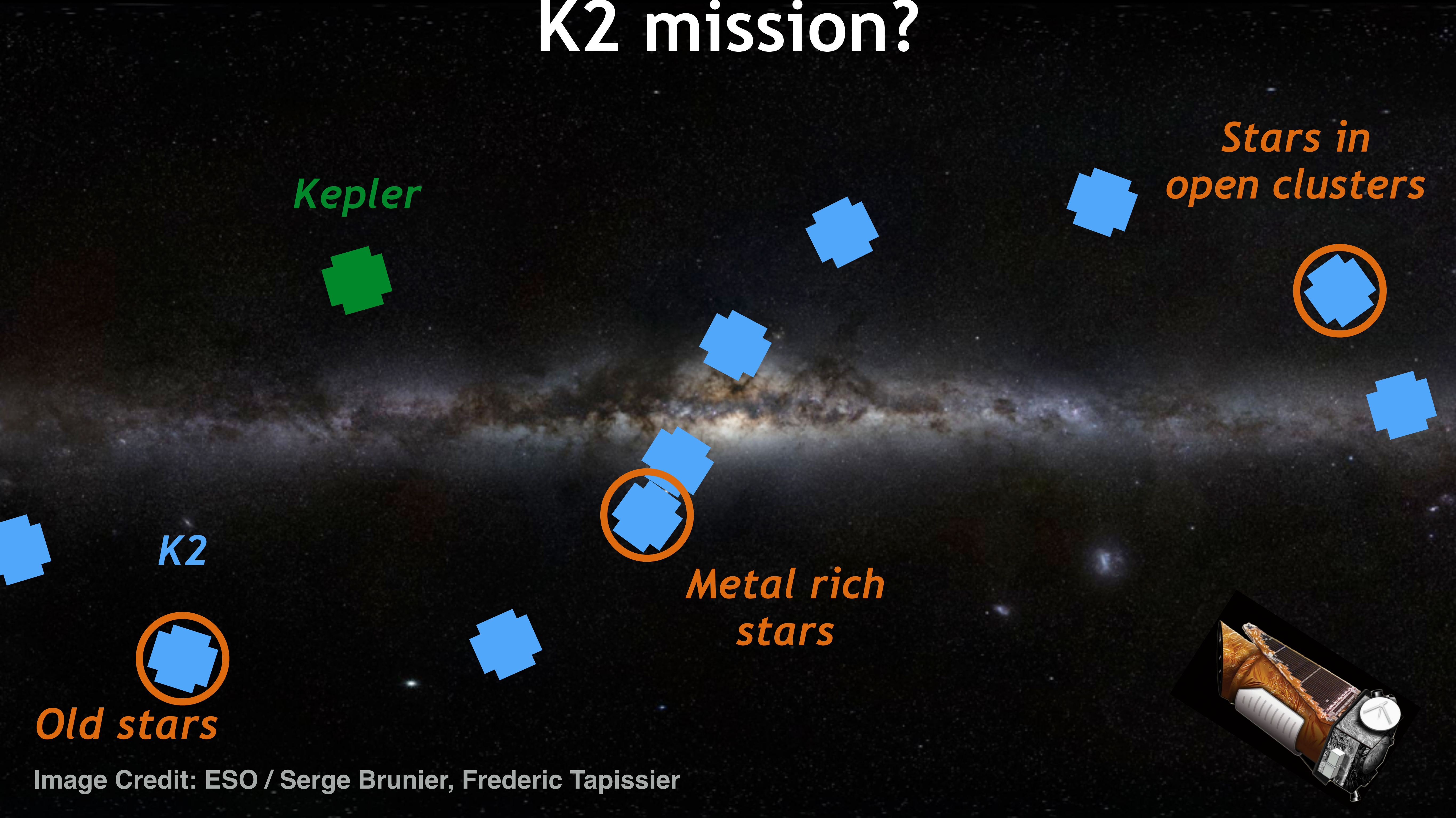
Shallue & Vanderburg (submitted)

Proof of concept search of Kepler multiplanet systems



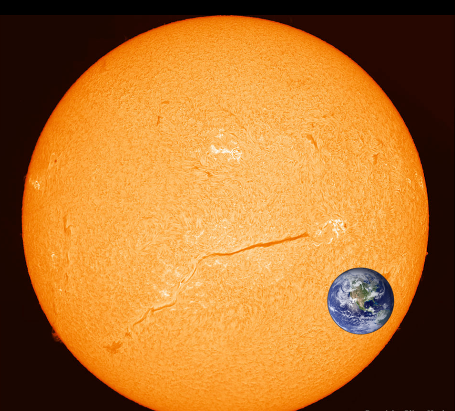
Shallue & Vanderburg (submitted)

Can we study planet occurrence with the K2 mission?



How do planetary systems vary in different galactic environments?

Occurrence rates with K2 are challenging for several reasons, including inhomogenous stellar parameters.



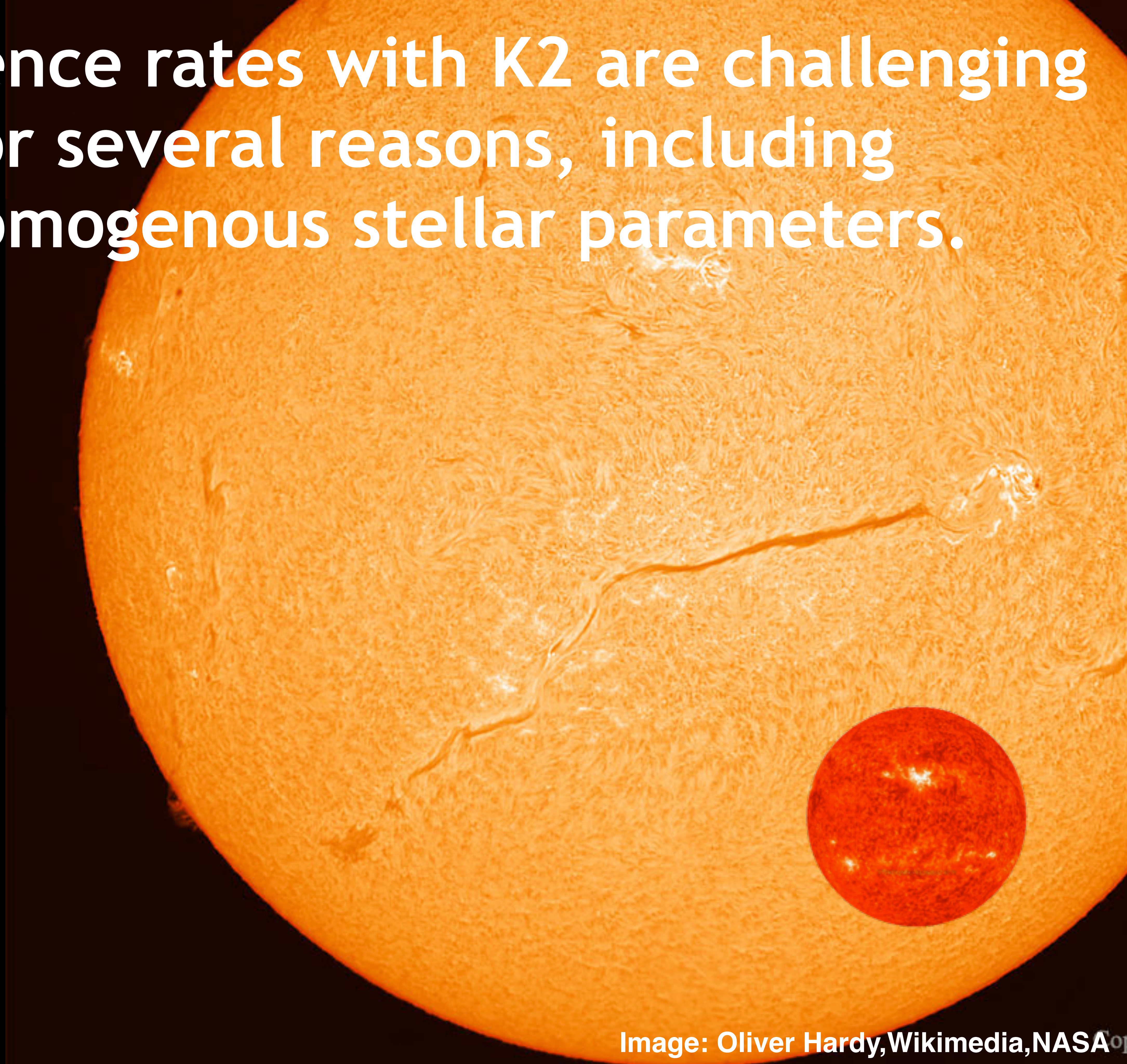


Image: Oliver Hardy,Wikimedia,NASA

Spectra of 275 K2 candidates from Whipple Observatory (Mt. Hopkins)

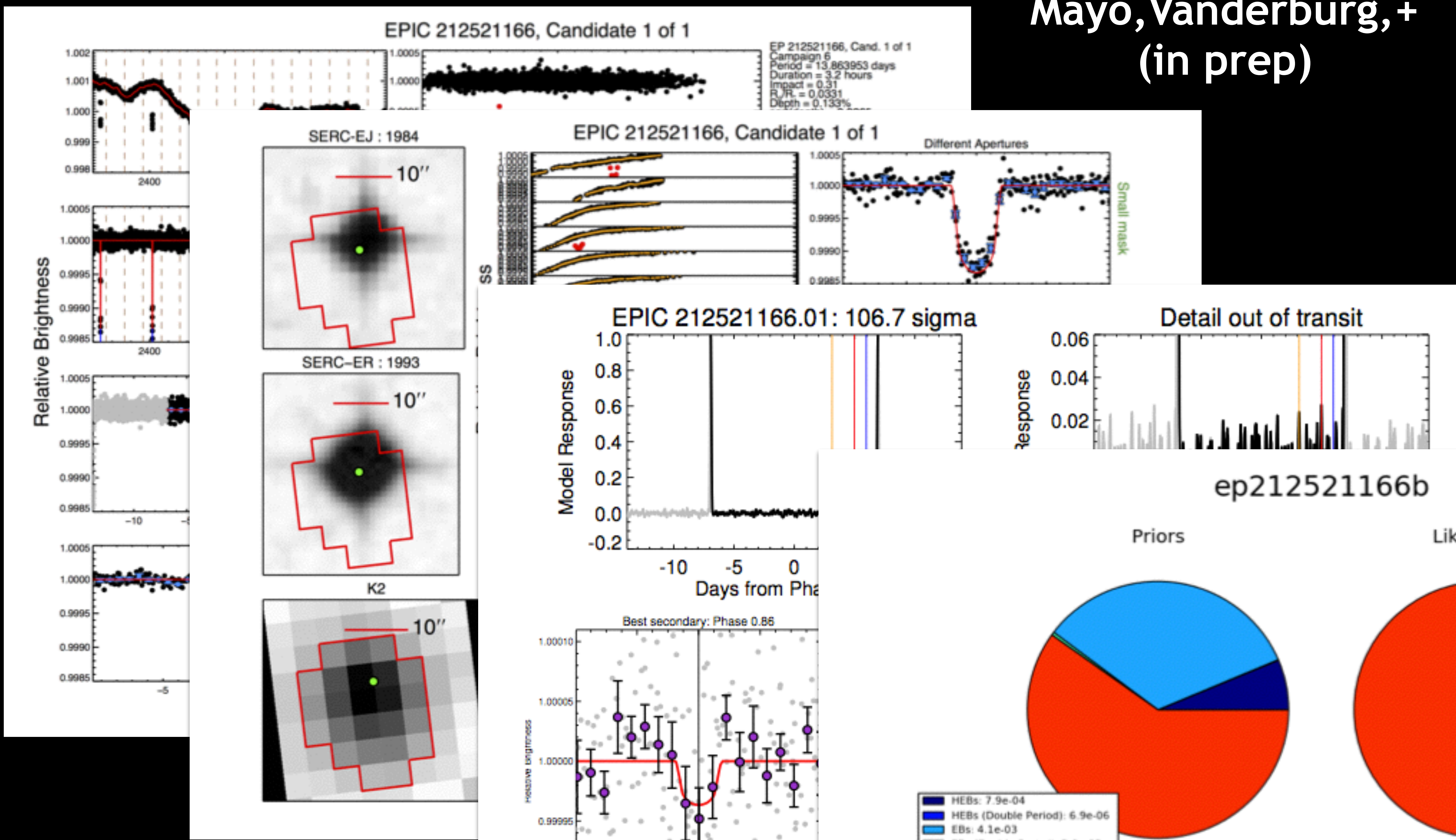
Spectra of 275 K2 candidates from Whipple Observatory (Mt. Hopkins)

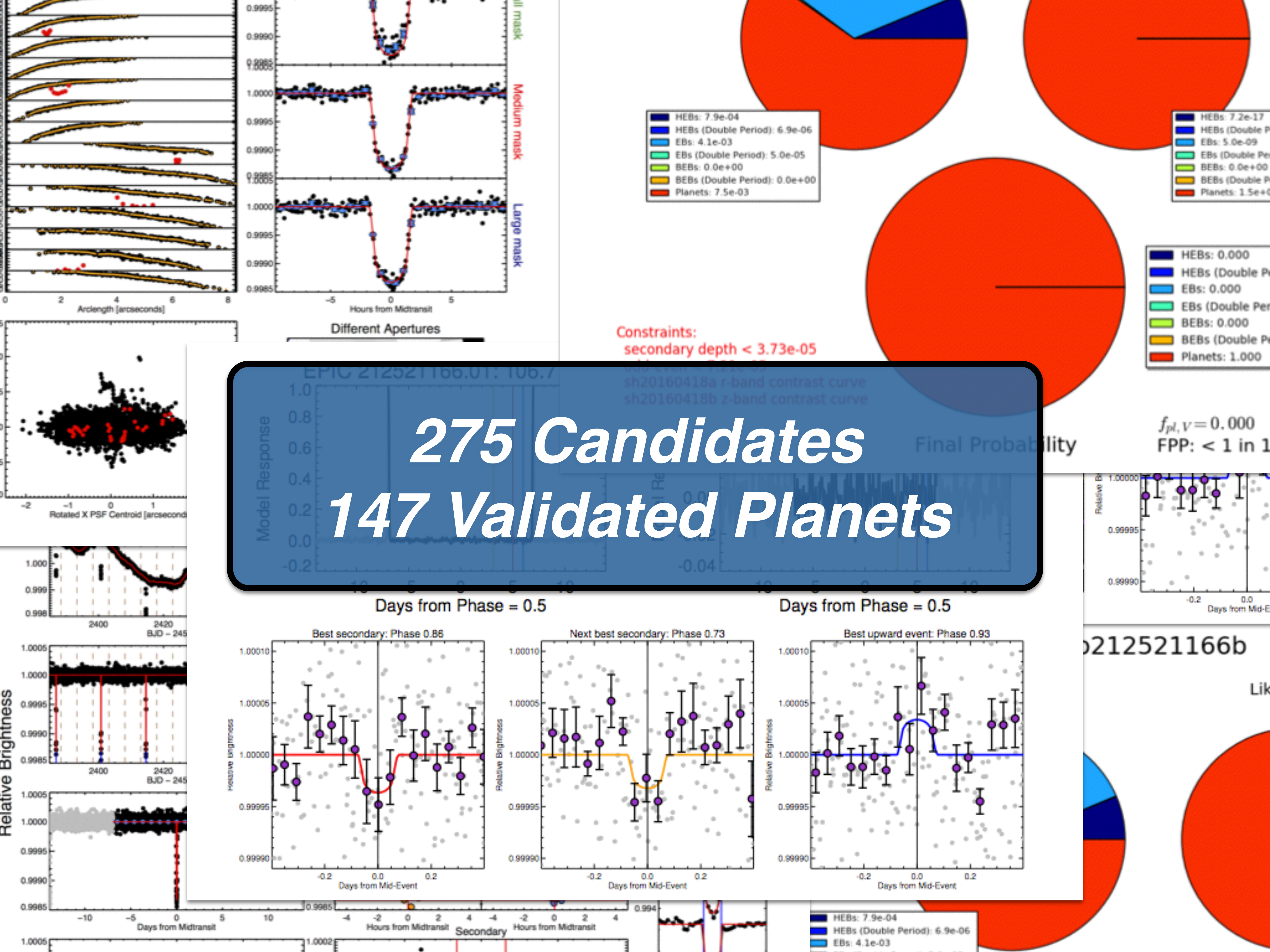
Andy Mayo

Harvard undergrad
—> Copenhagen Univ. Fulbright
—> UC Berkeley grad (Fall 2018)

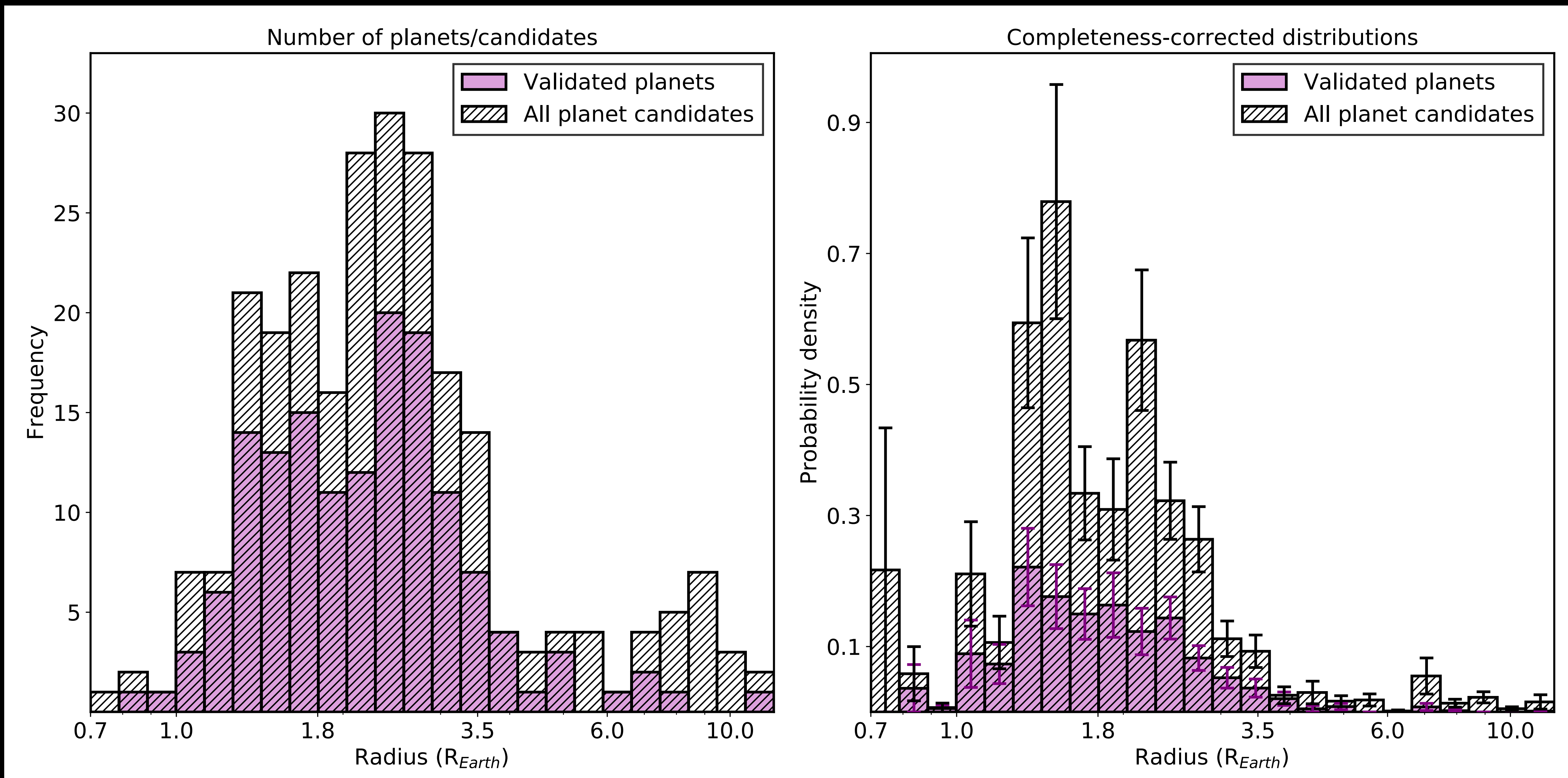
Planet Candidate Vetting and Validation

Mayo, Vanderburg, + (in prep)



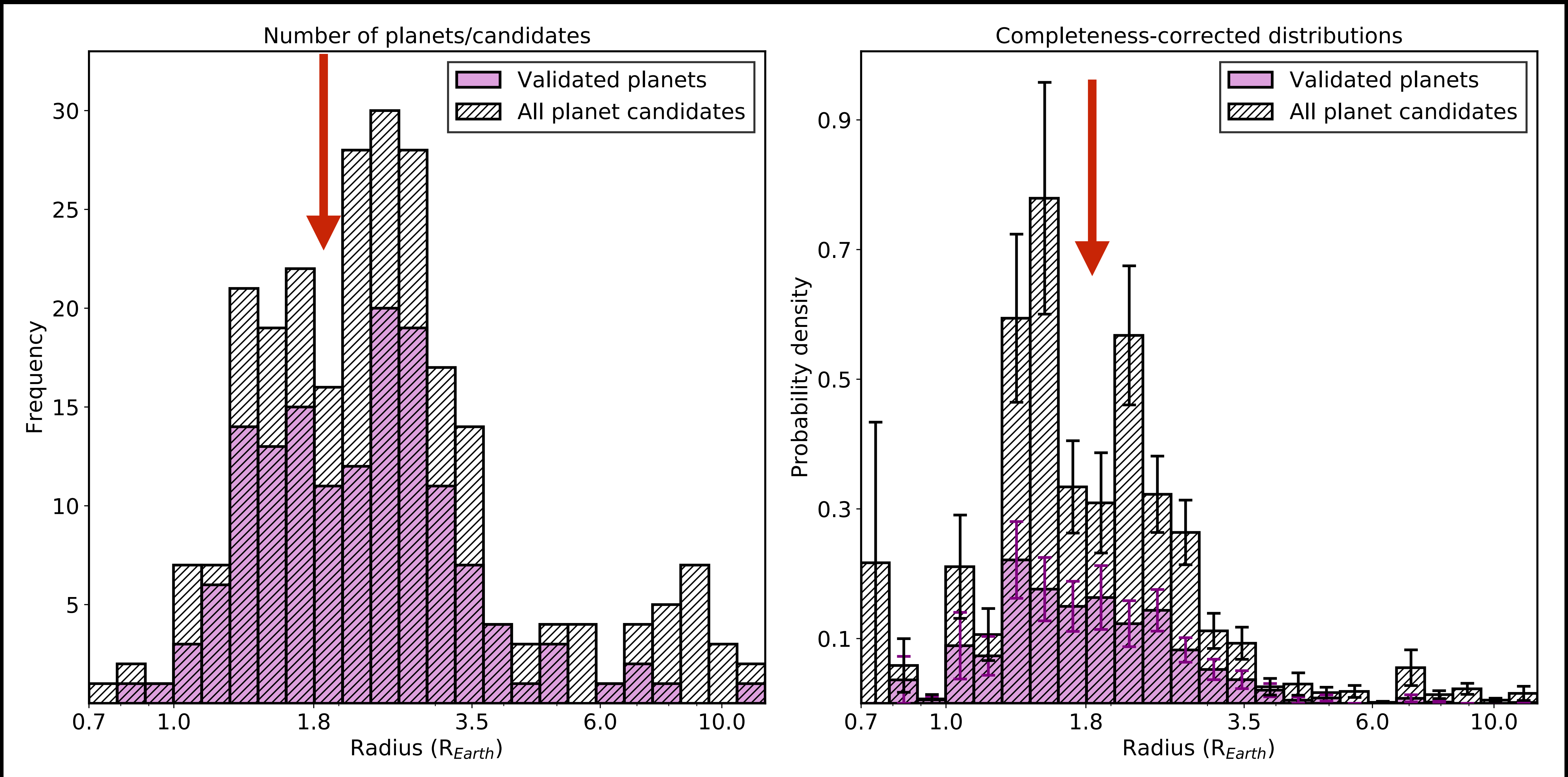

275 Candidates
147 Validated Planets

Planet Radius Distribution



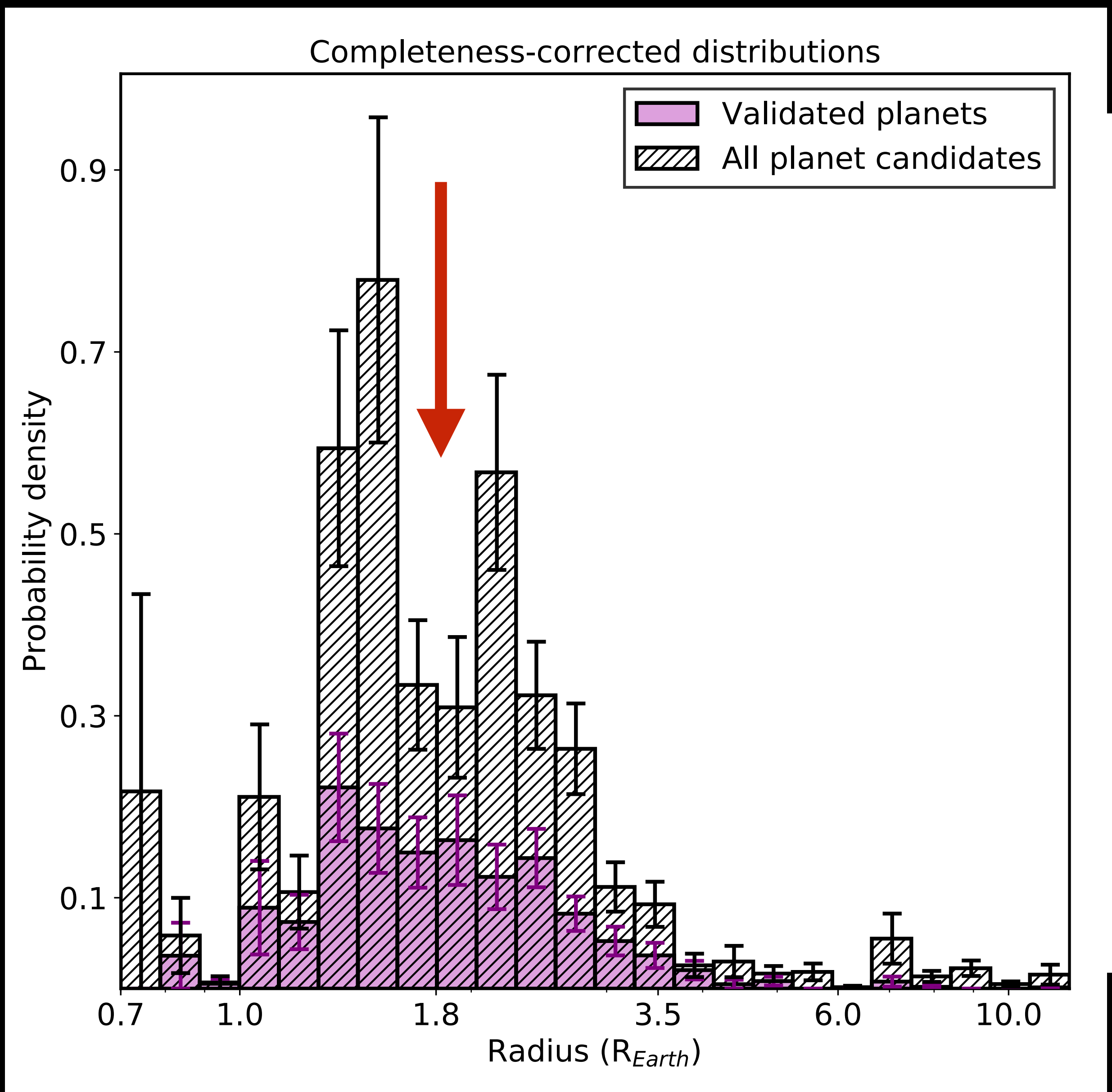
Mayo,Vanderburg+ (in prep)

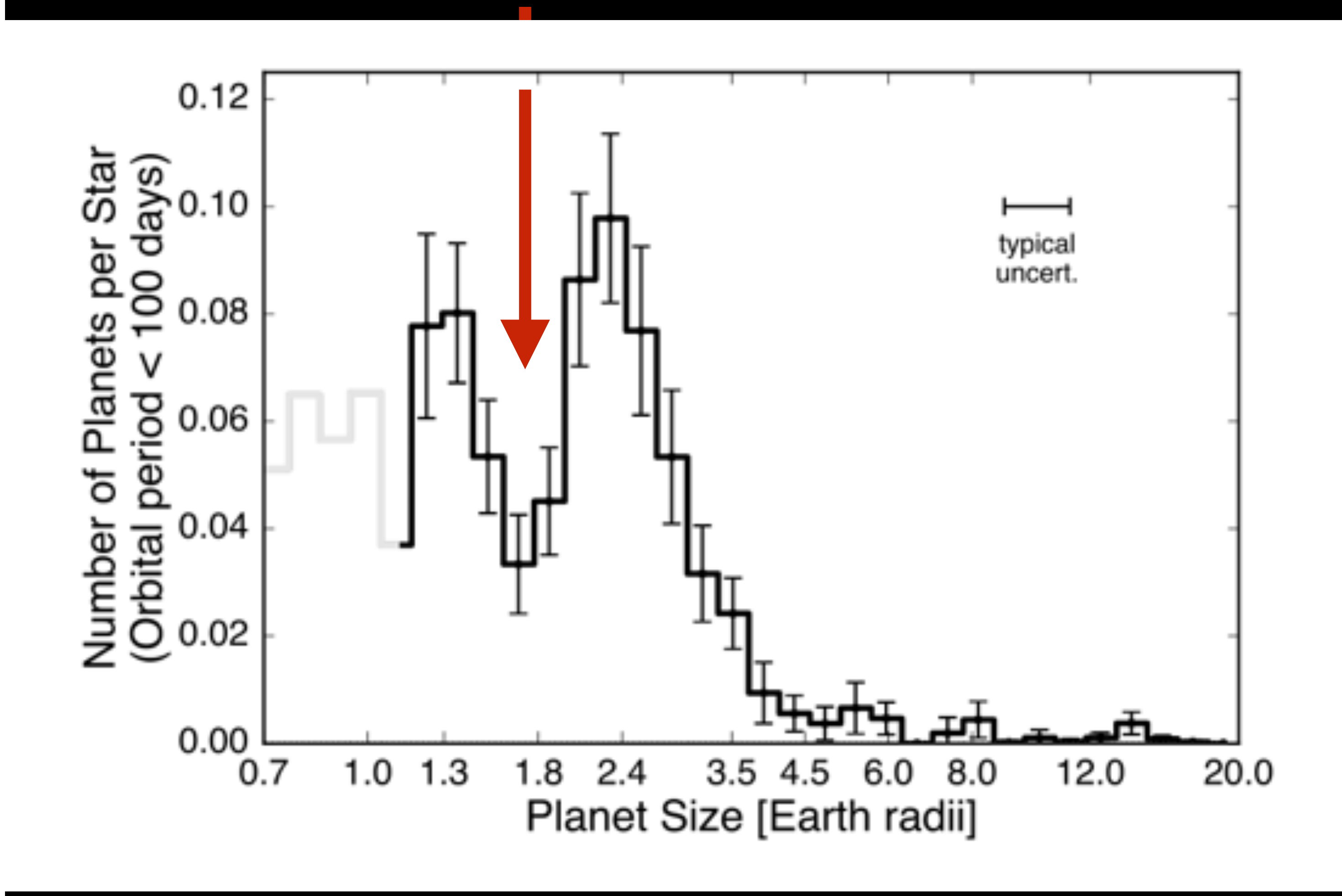
Planet Radius Distribution



Mayo, Vanderburg+ (in prep)

Planet Radius Distribution





Mayo,Vanderburg+ (in prep)

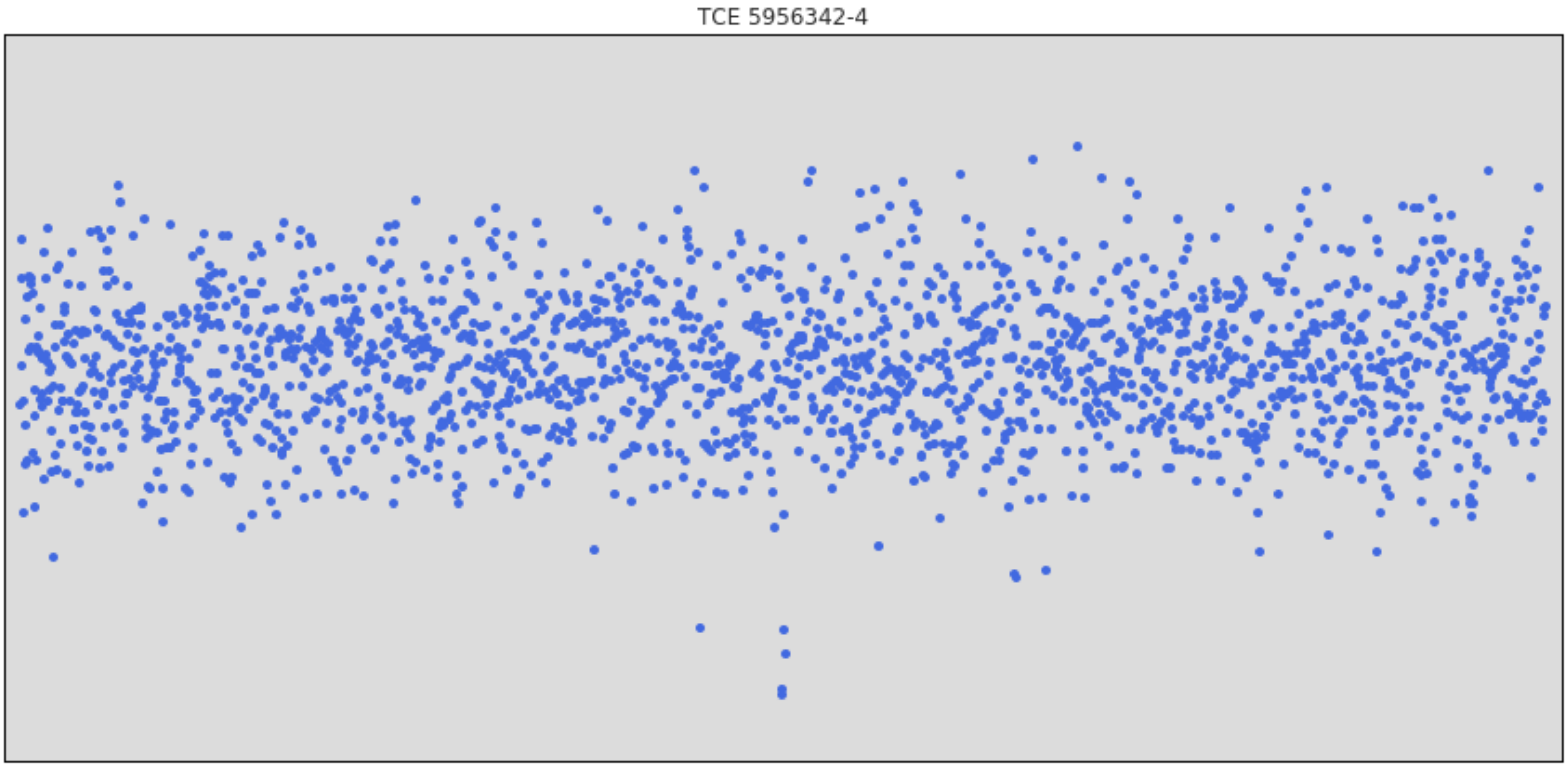
Fulton+2017

Conclusions

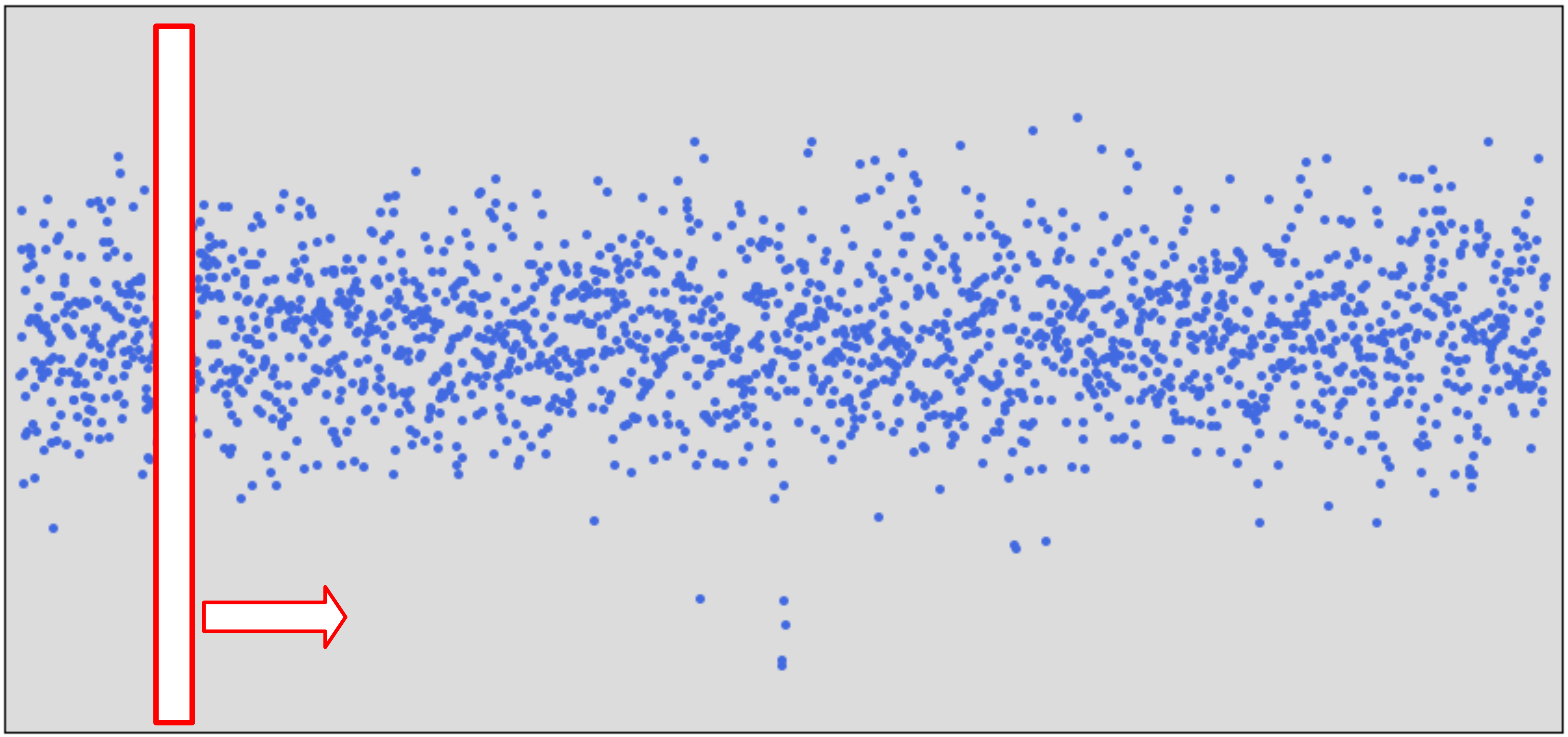
- We have used deep neural networks to vet Kepler planet candidates and achieved good results in a proof-of-concept test.
- By lowering detection thresholds and compensating for the increased number of false positives, we can improve Kepler's sensitivity to weak transit signals. We have detected new planets in Kepler multi-planet systems using this strategy.
- We are working towards occurrence studies with K2 data by carefully characterizing and vetting planet candidates and their host stars. We can see the first hint of the gap in the planet radius distribution in K2 data.

Backup Slides

Visualizing the Network’s Decisions

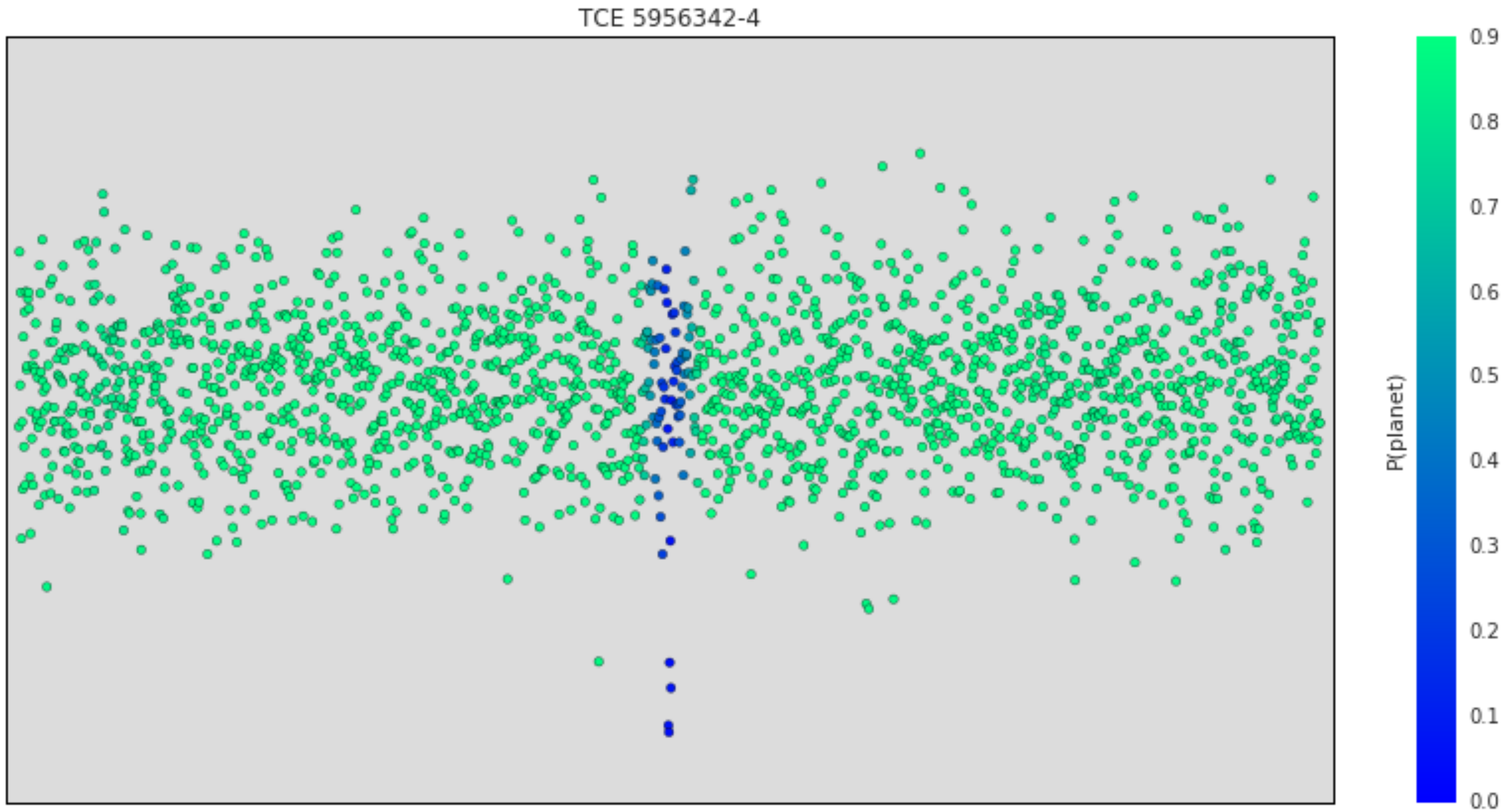


Visualizing the Network's Decisions



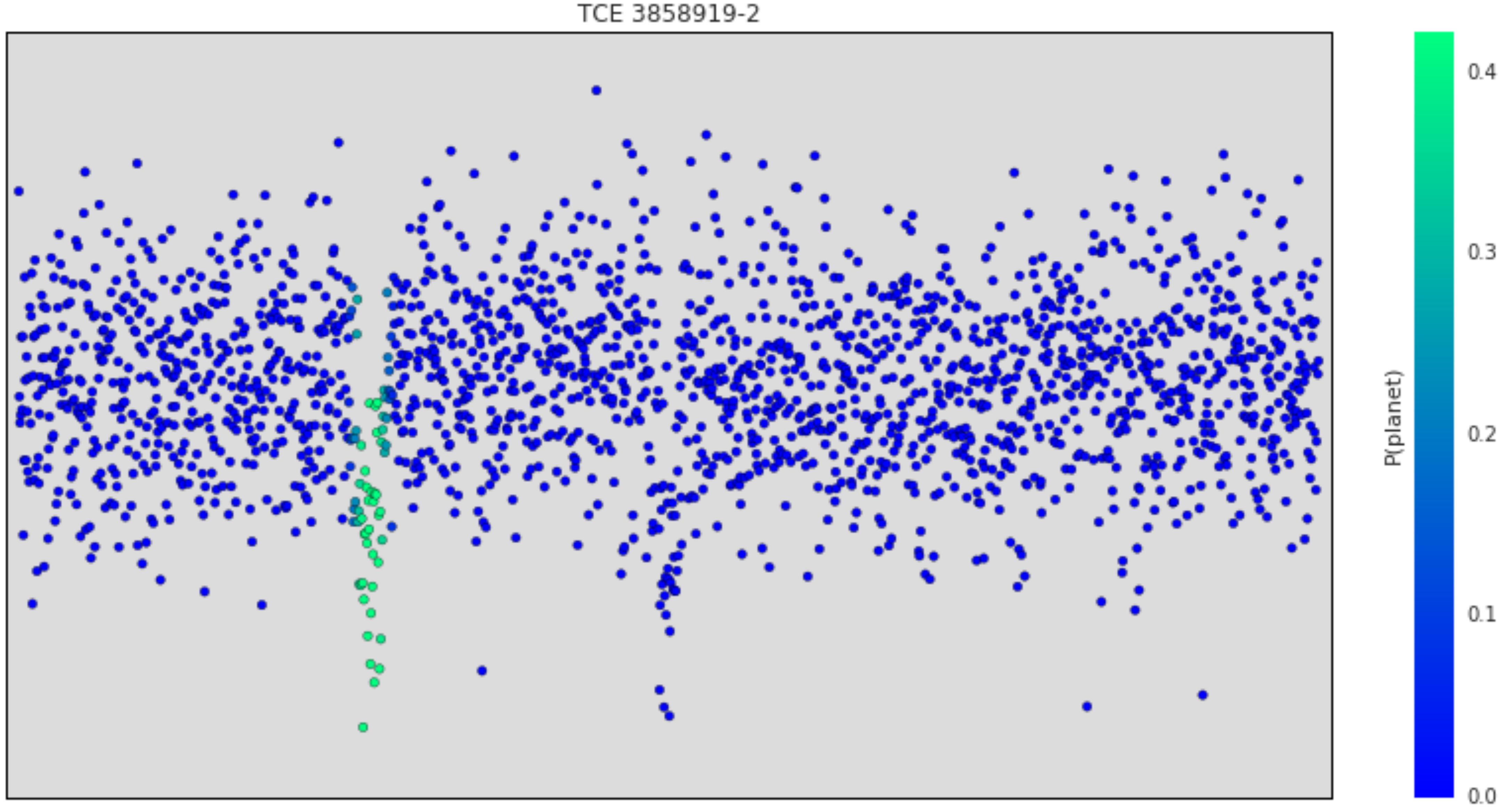
Google

Visualizing the Network's Decisions



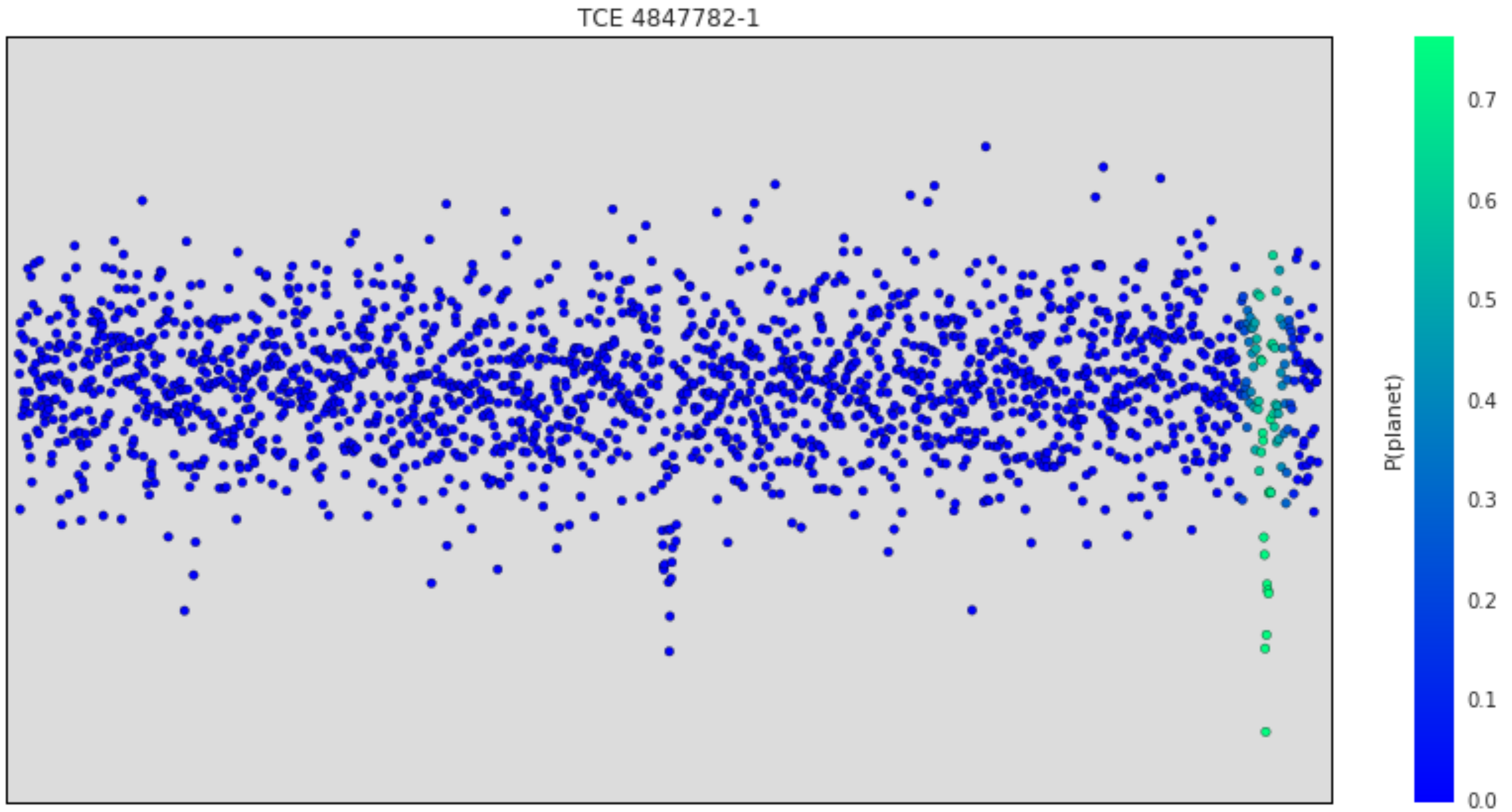
Google

Visualizing the Network's Decisions



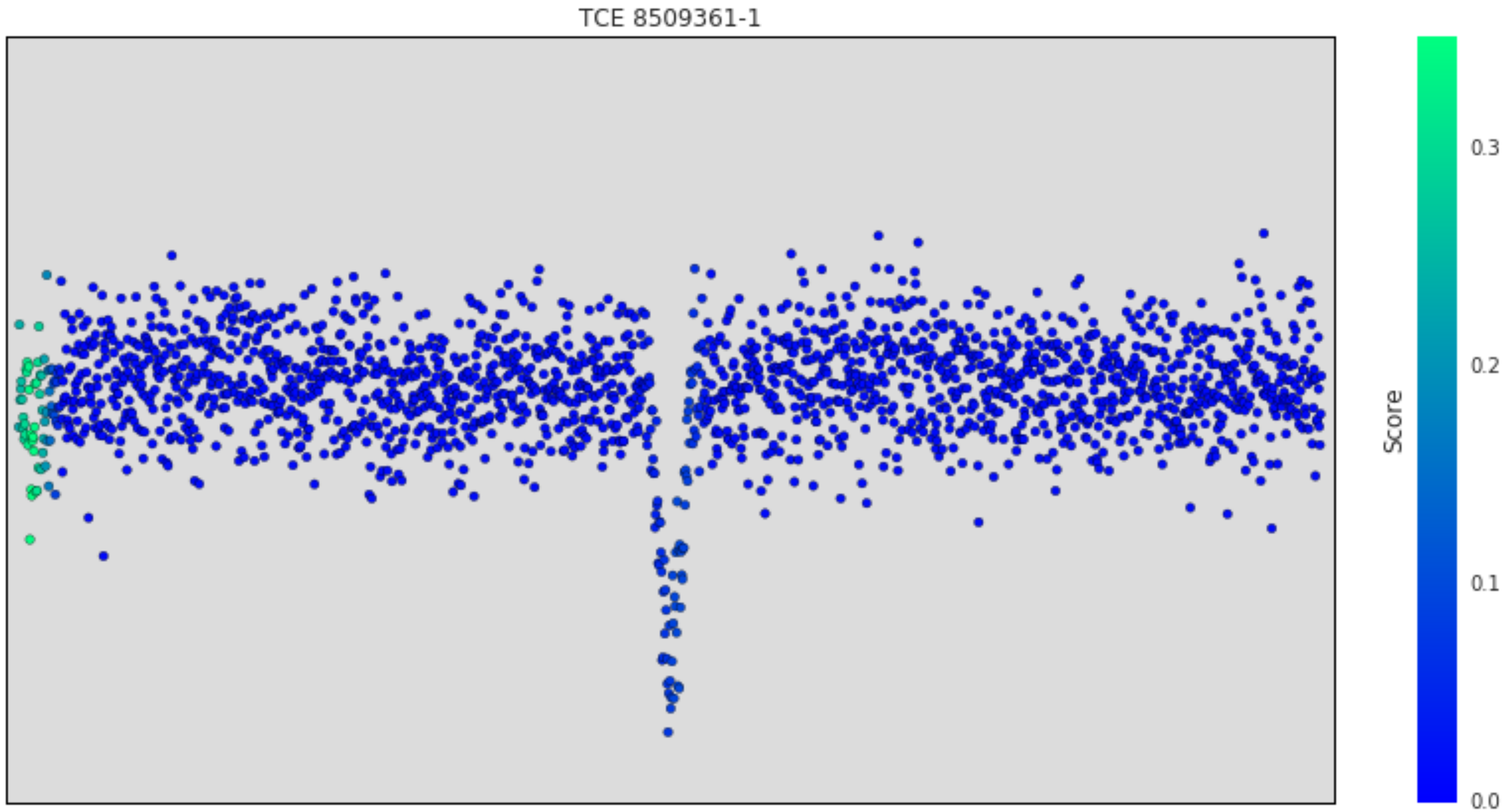
Google

Visualizing the Network’s Decisions



Google

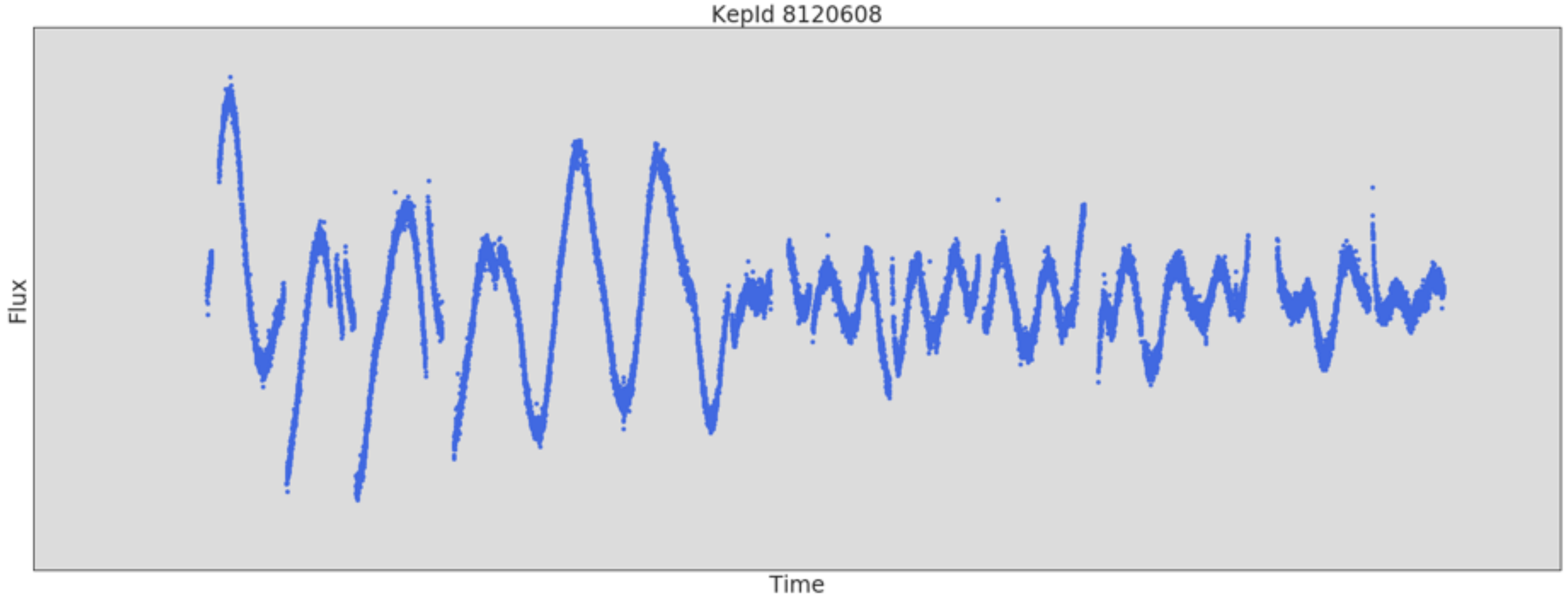
Visualizing the Network's Decisions



Google

Producing Input Vectors

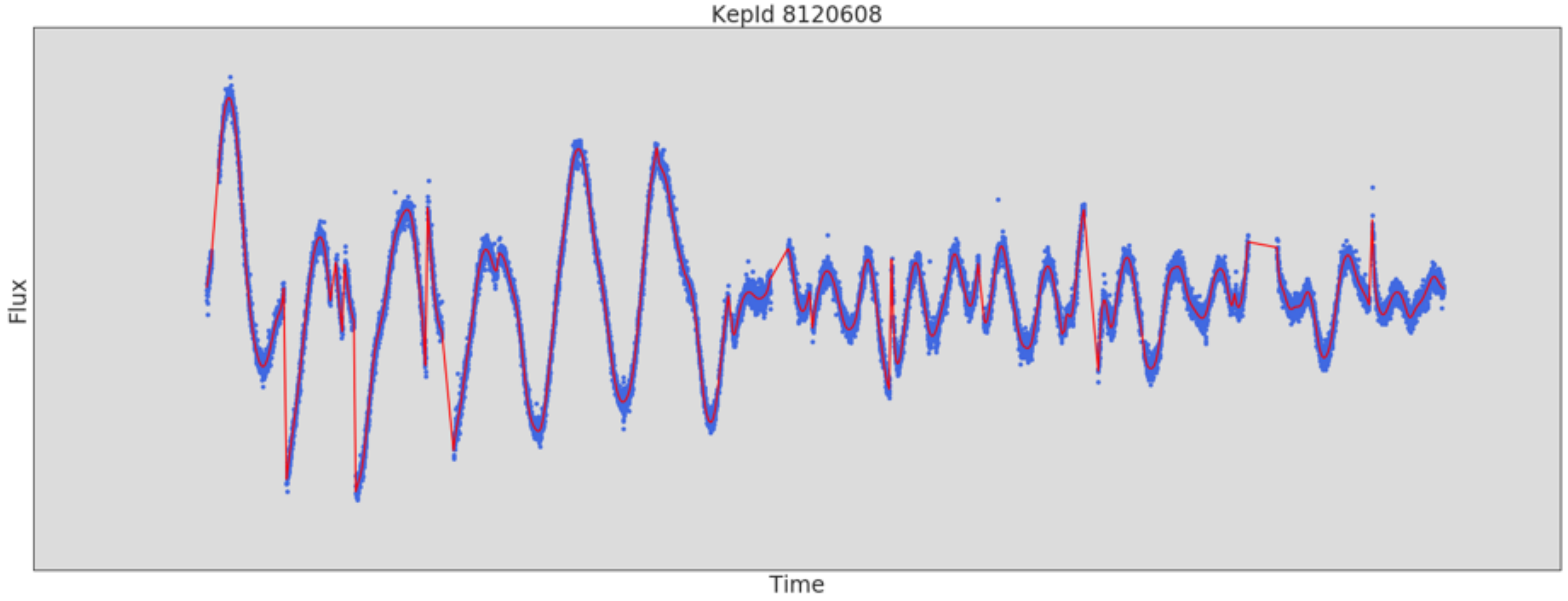
Raw Light Curve



Google

Producing Input Vectors

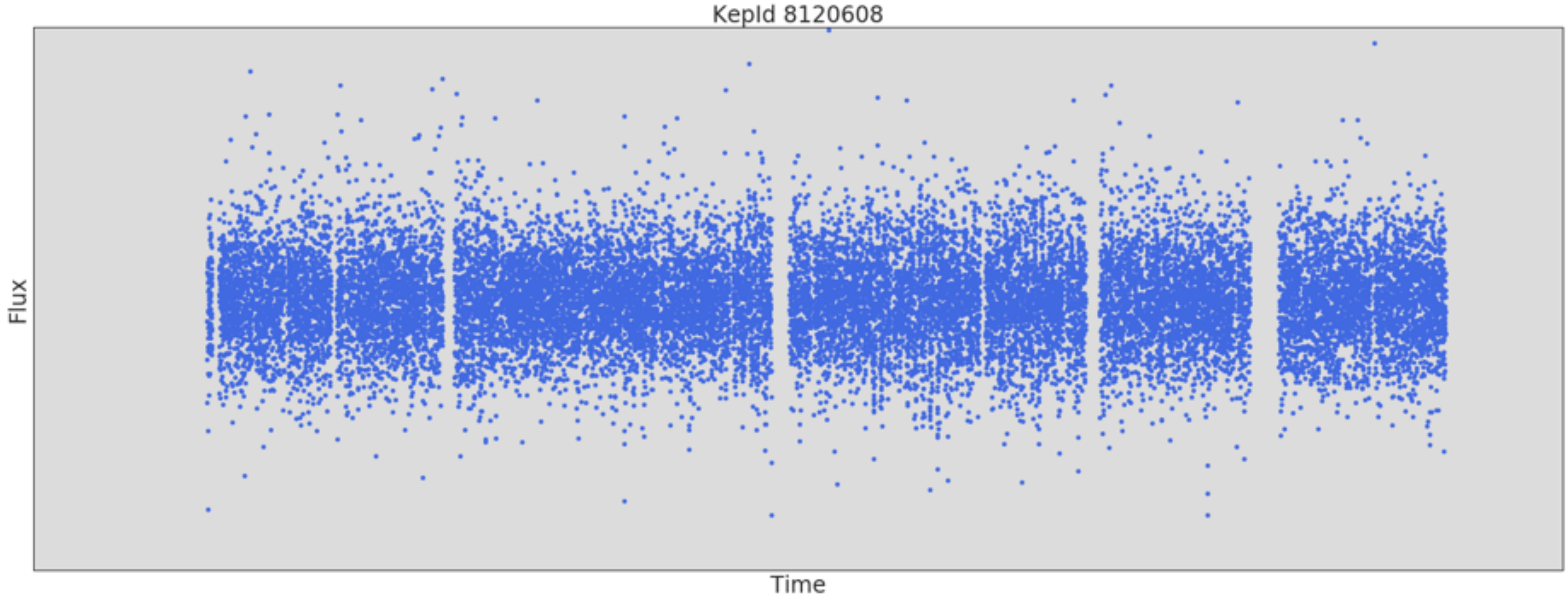
Fit Normalization B-Splines



Google

Producing Input Vectors

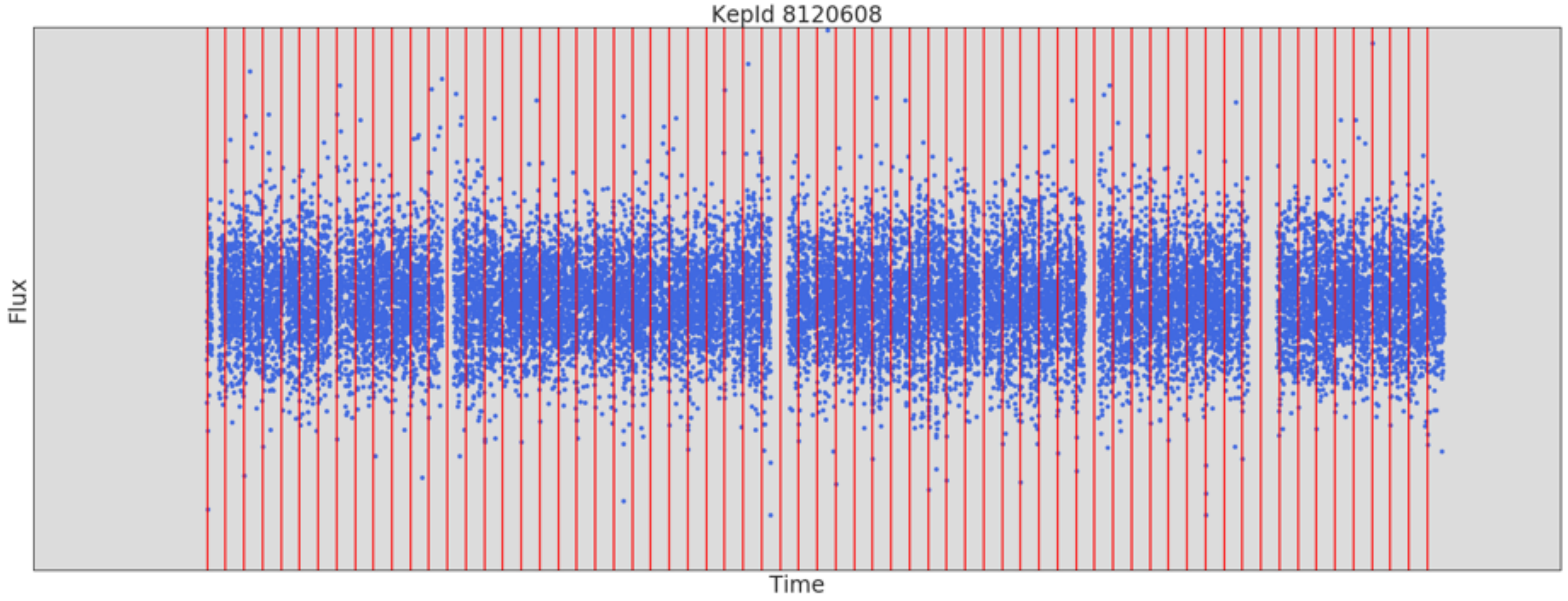
Normalized Light Curve



Google

Producing Input Vectors

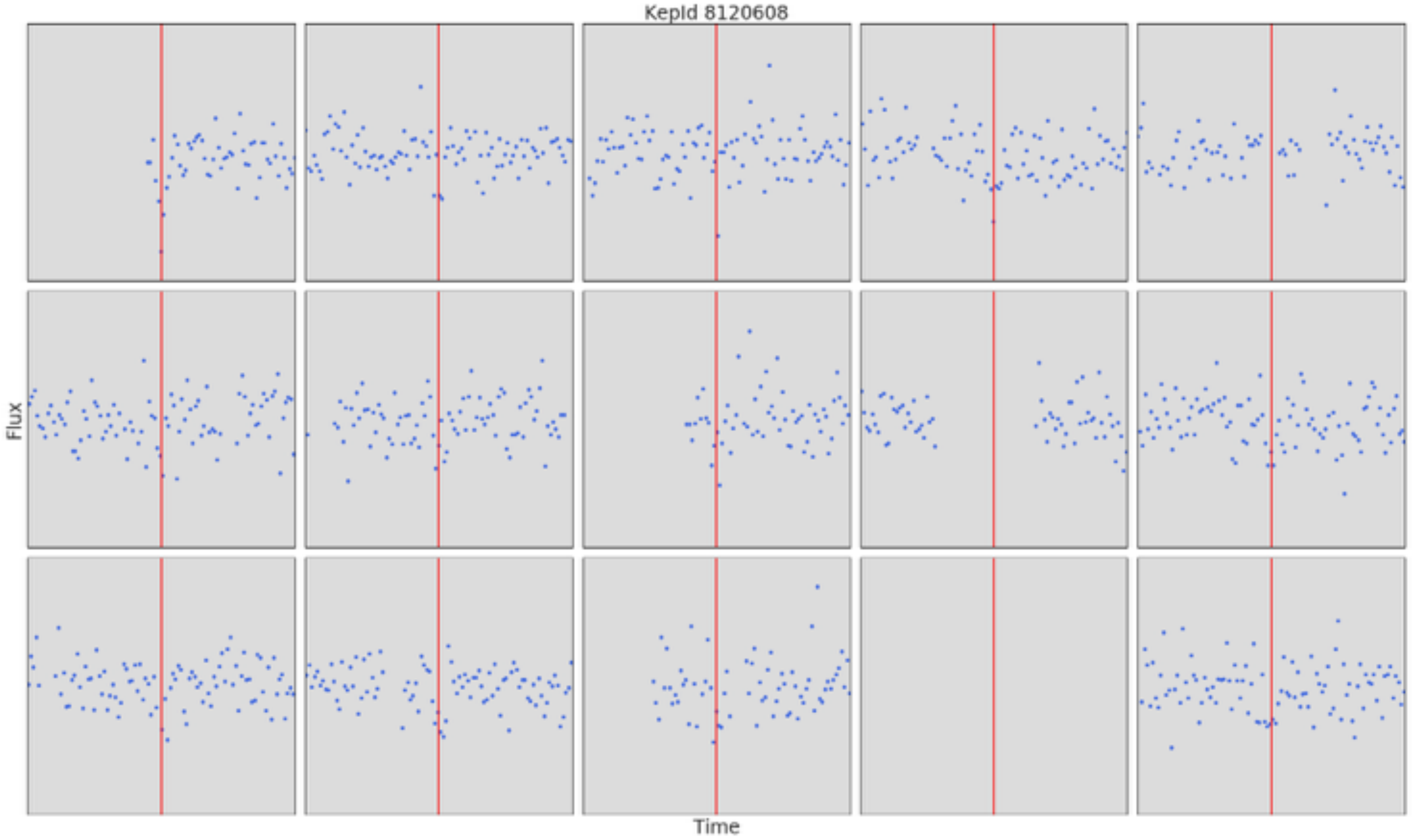
Planet Transits (7 day period)



Google

Producing Input Vectors

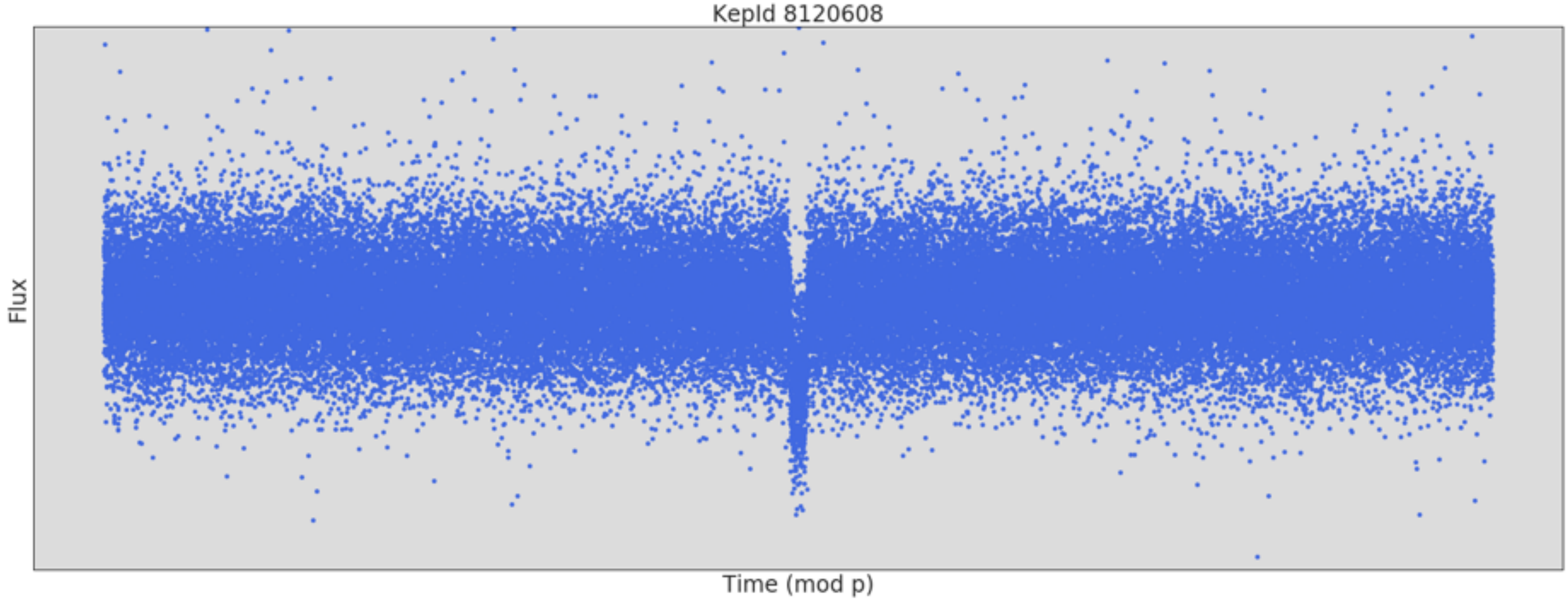
Individual Transits



Google

Producing Input Vectors

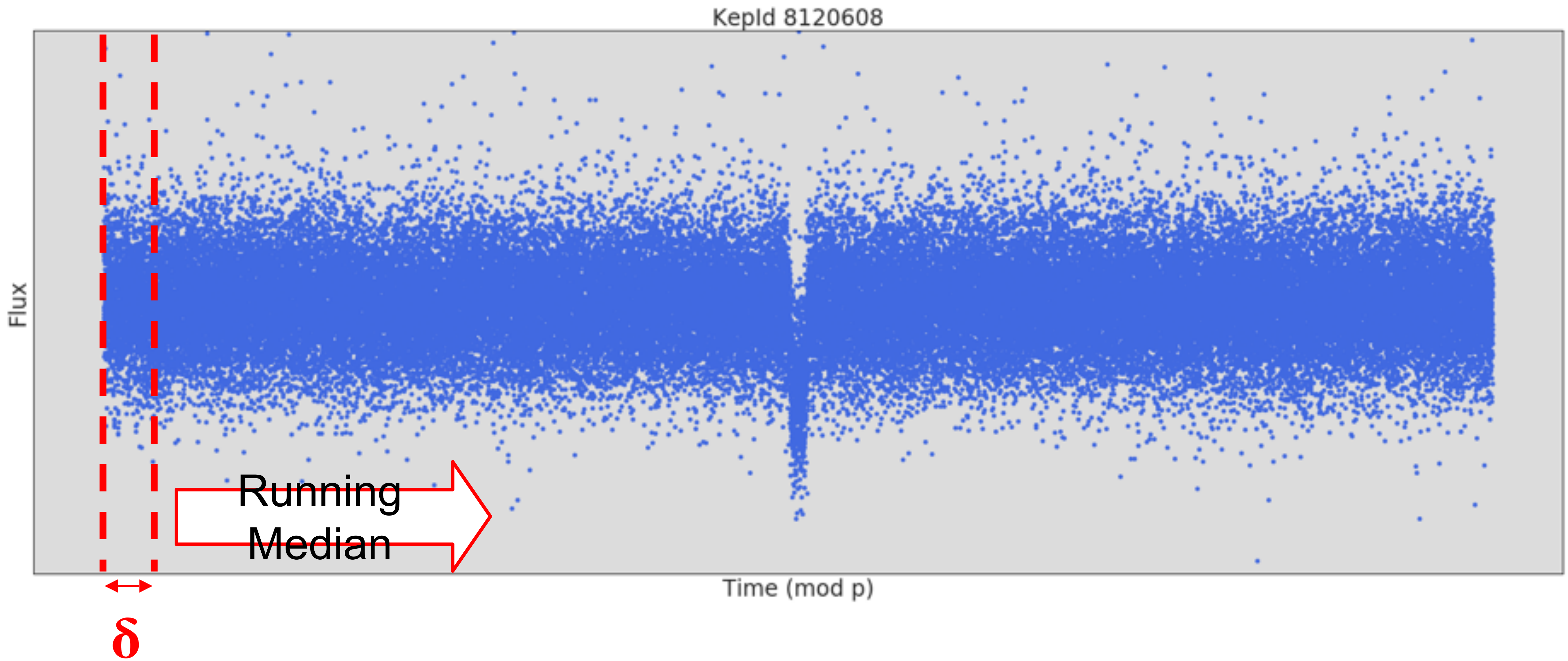
Phase Folded



Google

Producing Input Vectors

Running Median



Producing Input Vectors

Running Median