THE EXTENDED LSST ASTRONOMICAL TIME SERIES CLASSIFICATION CHALLENGE (ELASTICC)

"THE STRETCH GOALS OF PLASTICC"

GAUTHAM NARAYAN

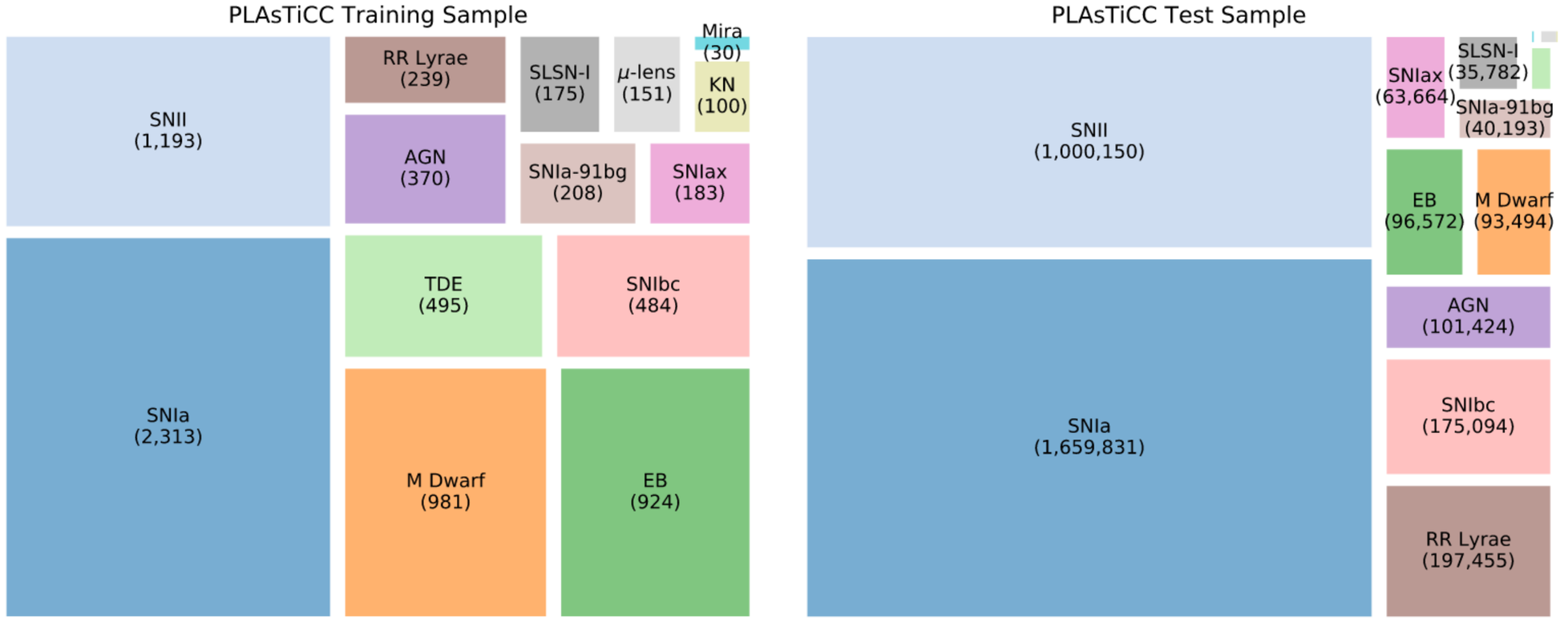
UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN

For the ELAsTiCC Team

LSST DESC Dark Energy Science Collaboration

PLAsTiCC (Dec. 2018 - Feb 2019)

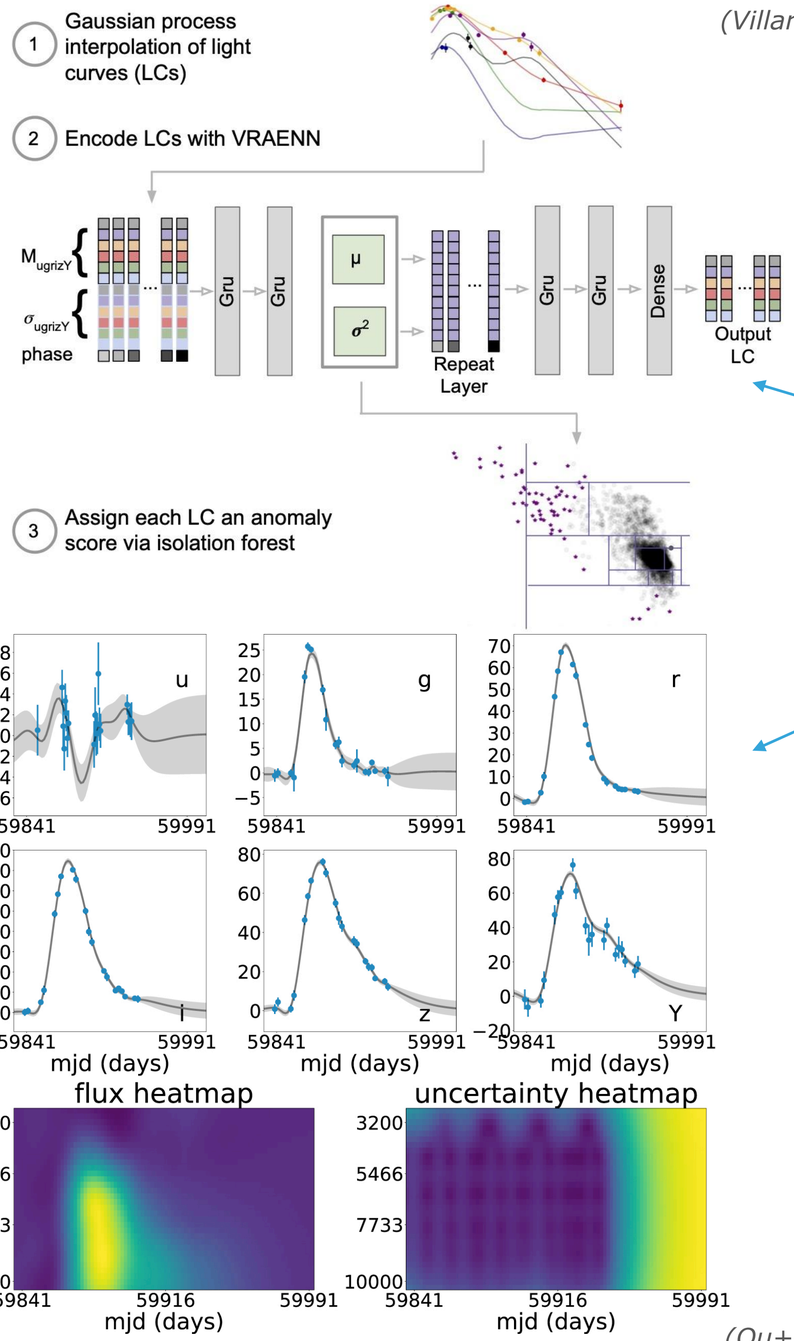
- Public $25k Kaggle challenge for photometric classification of time-domain sky (15 models, 1 million new SEDs, unrepresentative training sample)
- Joint effort between LSST DESC and LSST TVS SC
- Data: 3M VRO-simulated *ugrizY* lightcurves
- Primary goal: setup massive time-domain simulation infrastructure, **jump start ML photometric classification efforts**



(Hložek+2020)

PLAsTiCC has become a gold-standard time-domain dataset

(Villar+2021)



1 2022ApJS..258...23A 2022/02 cited: 4
Considerations for Optimizing the Photometric Classification of Supernovae from the Rubin Observatory
Alves, Catarina S.; Peiris, Hiranya V.; Lochner, Michelle *and 4 more*

2 2021AJ....162..275B 2021/12 cited: 3
ParSNIP: Generative Models of Transient Light Curves with Physics-enabled Deep Learning
Boone, Kyle

3 2021ApJS..255...24V 2021/08 cited: 10
A Deep-learning Approach for Live Anomaly Detection of Extragalactic Transients
Villar, V. Ashley; Cranmer, Miles; Berger, Edo *and 4 more*

4 2021AJ....162...67Q 2021/08 cited: 5
SCONE: Supernova Classification with a Convolutional Neural Network
Qu, Helen; Sako, Masao; Möller, Anais *and 1 more*

5 2021AAS...23820301N 2021/06
Deep Learning for Multimessenger Astrophysics
Narayan, G.

6 2021A&A...650A.195I 2021/06 cited: 20
Active anomaly detection for time-domain discoveries
Ishida, E. E. O.; Kornilov, M. V.; Malanchev, K. L. *and 7 more*

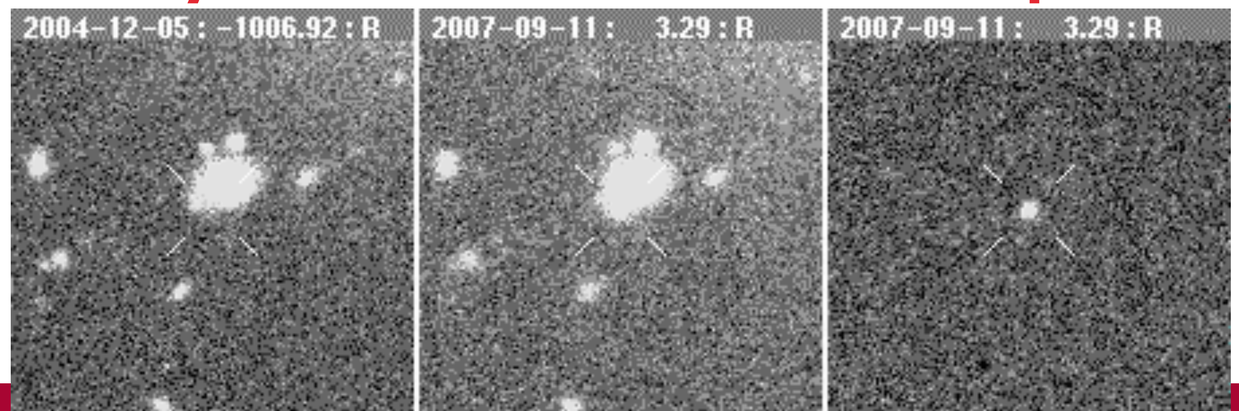
7 2021wems.confE..10T 2021/05
Time Series Classification: From Astronomical Transients to Human Heart Sounds
Torki, Motahare

8 2021arXiv210506178A 2021/05 cited: 4
Paying Attention to Astronomical Transients: Photometric Classification with the Time-Series Transformer
Allam, Tarek, Jr.; McEwen, Jason D.

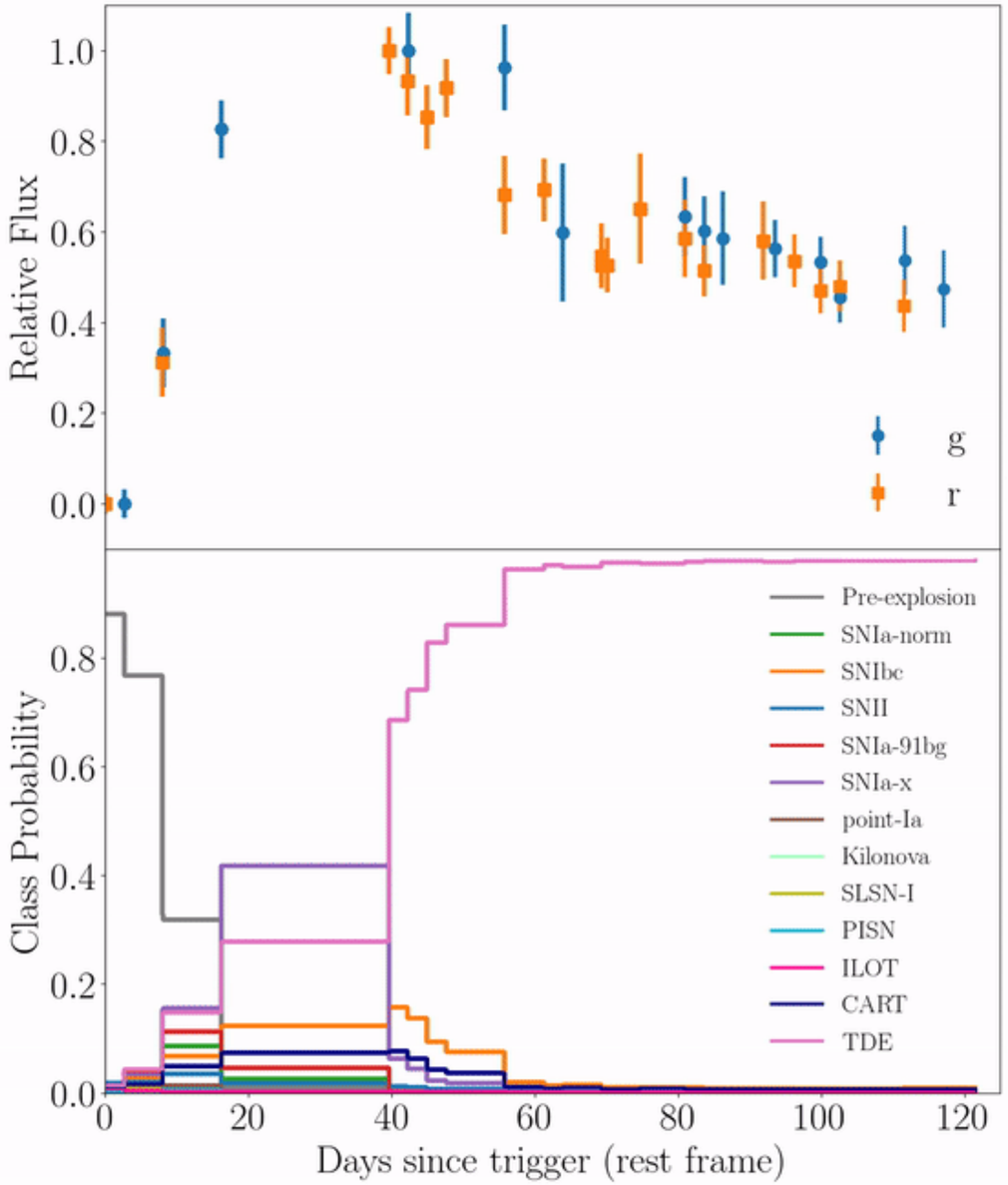
(Qu+2021)

PLAsTiCC has problems

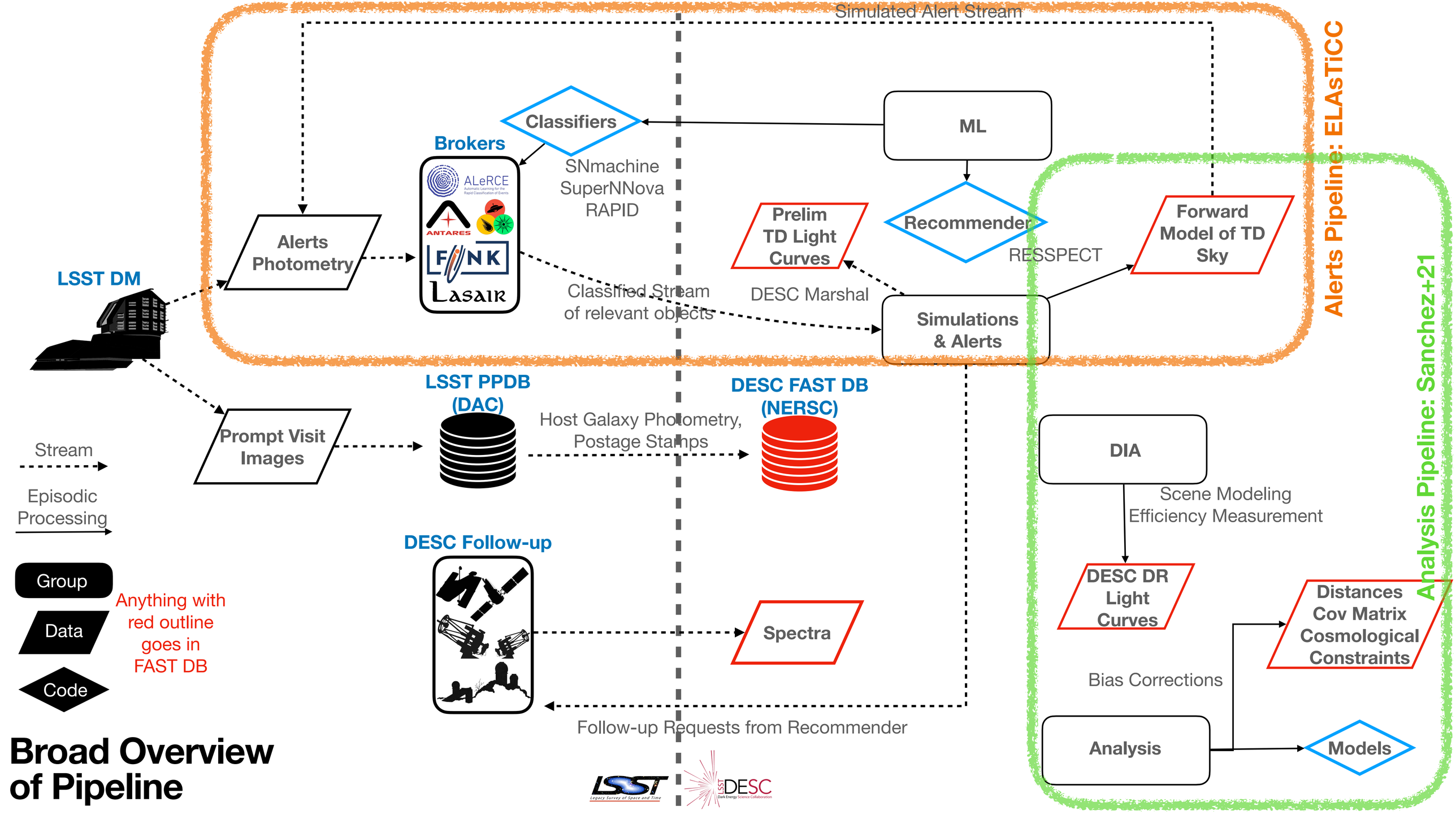
- Unrepresentative, imbalanced training sample **enforces strong implicit prior**
- You can guess SNIa ~70% of the time, and you'll get the right answer, but **what's verall accuracy is a boring, single question**
- You don't get light curves from LSST - **you get alerts, streamed in real-time, via 3rd party brokers**
- **No host galaxy information**
- **Models not diverse enough** - classifiers had it too easy
- **Surveys cannot afford to act independently**



(Muthukrishna+2020)

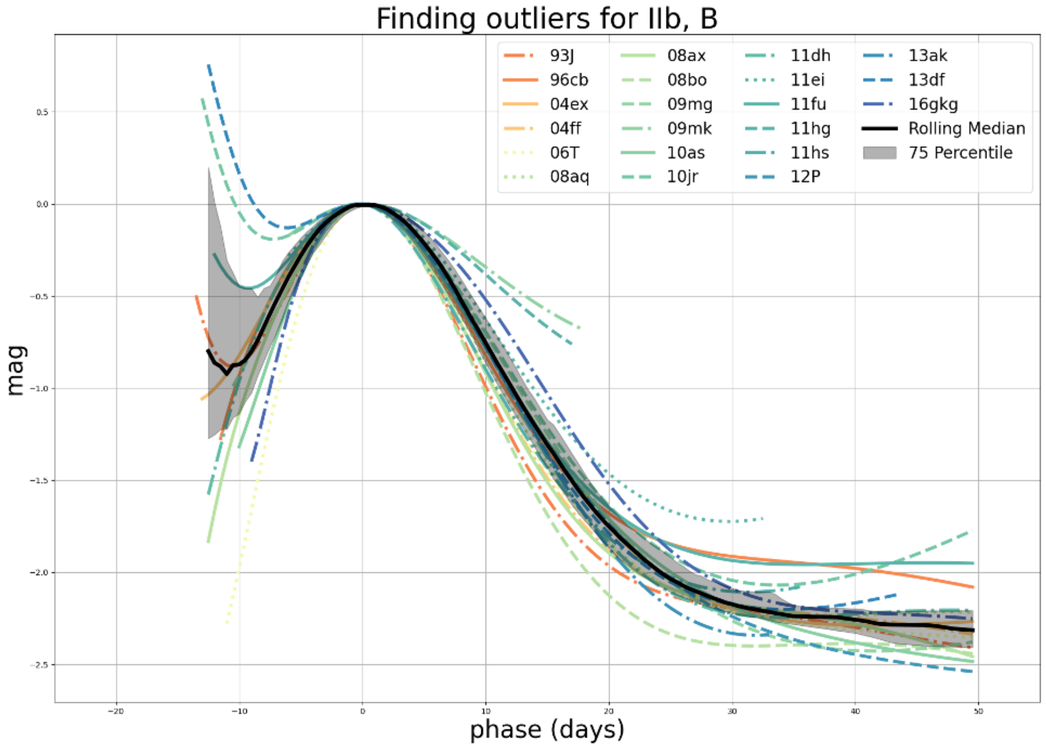


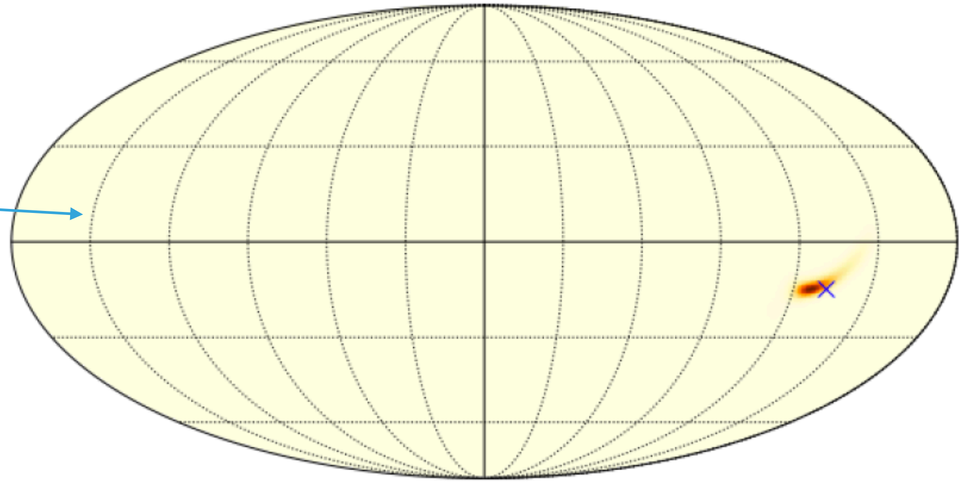
WHAT DOES A REALISTIC LSST TIME-DOMAIN AND COSMOLOGY WORKFLOW LOOK LIKE?


Simulated Alert Stream
Alerts Pipeline: ELAsTiCC
Analysis Pipeline: Sanchez+21
Classifiers
ML
Brokers
SNmachine
SuperNNova
RAPID
Alerts
Photometry
Prelim
TD Light
Curves
Recommender
RESSPECT
Forward
Model of TD
Sky
LSST DM
Classified Stream
of relevant objects
DESC Marshal
Simulations
& Alerts
LSST PPDB
(DAC)
DESC FAST DB
(NERSC)
Prompt Visit
Images
Host Galaxy Photometry,
Postage Stamps
DIA
Scene Modeling
Efficiency Measurement
Stream
Episodic
Processing
DESC Follow-up
DESC DR
Light
Curves
Distances
Cov Matrix
Cosmological
Constraints
Spectra
Bias Corrections
Group
Data
Code
Anything with
red outline
goes in
FAST DB
Follow-up Requests from Recommender
Analysis
Models
Broad Overview
of Pipeline

ELAsTiCC has a richer diversity of models

- *New:*
  - delta Scuti, Cepheids (K. Malanchev)
  - dwarf novae (Q. Cheng)
  - + others we've snuck into test data but not in training data
- *Updated:*
  - SNe Ib & c (M. Vincenzi)
  - M-dwarf flares (V. Shah)
  - KNe using Bulla et al. SEDs (D. Chatterjee)
    - **+ synthetic LVK O4 alert skymaps**



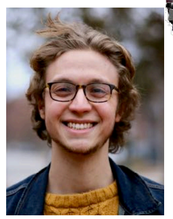




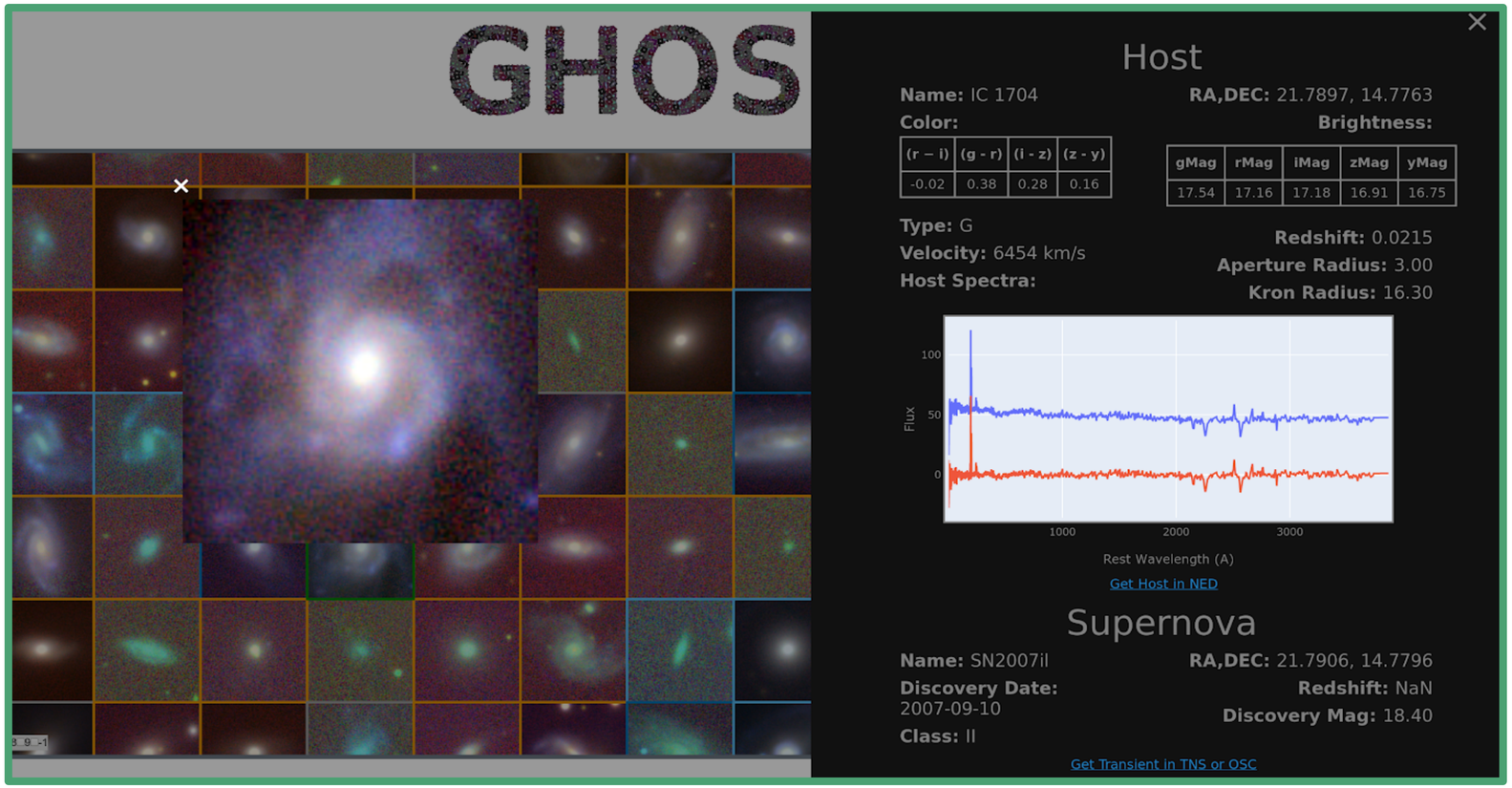


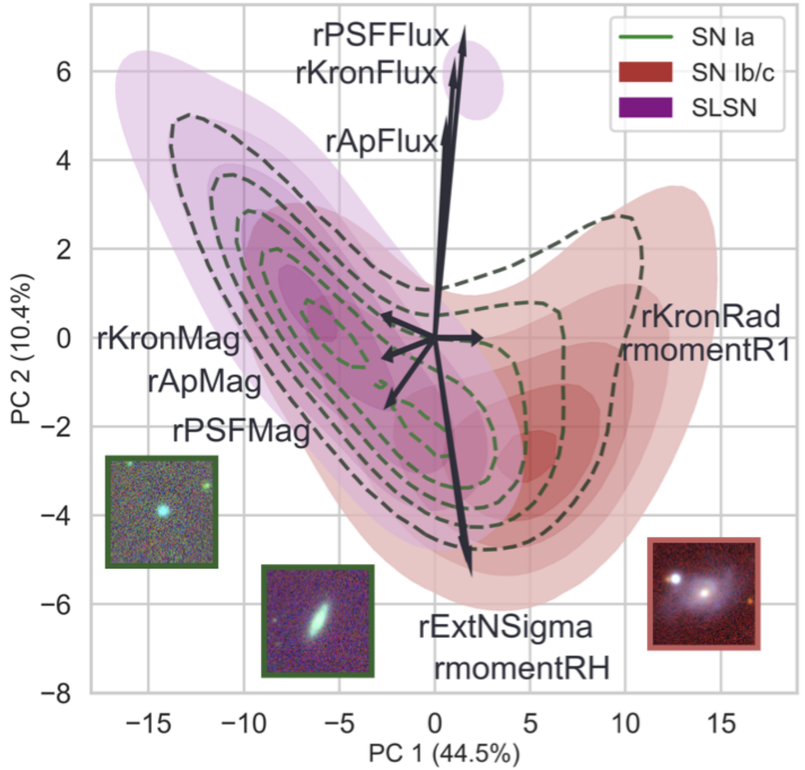
GHOST

Galaxies **HO**sting **S**upernovae and other **T**ransients



Alex Gagliano (UIUC, NSF Fellow) figures out correlations between transients and their hosts *http://ghost.rubin.science/*





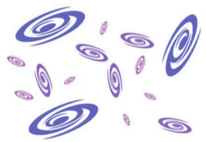
(Gagliano+2021)

- 16,228 SNe-host galaxy pairs: 78% of unique events reported on TNS/OSC.
- PS1,NED photometric & derived properties (color, redshift, radial moments)

Populating Hosts in ELAsTiCC uses Normalizing Flows

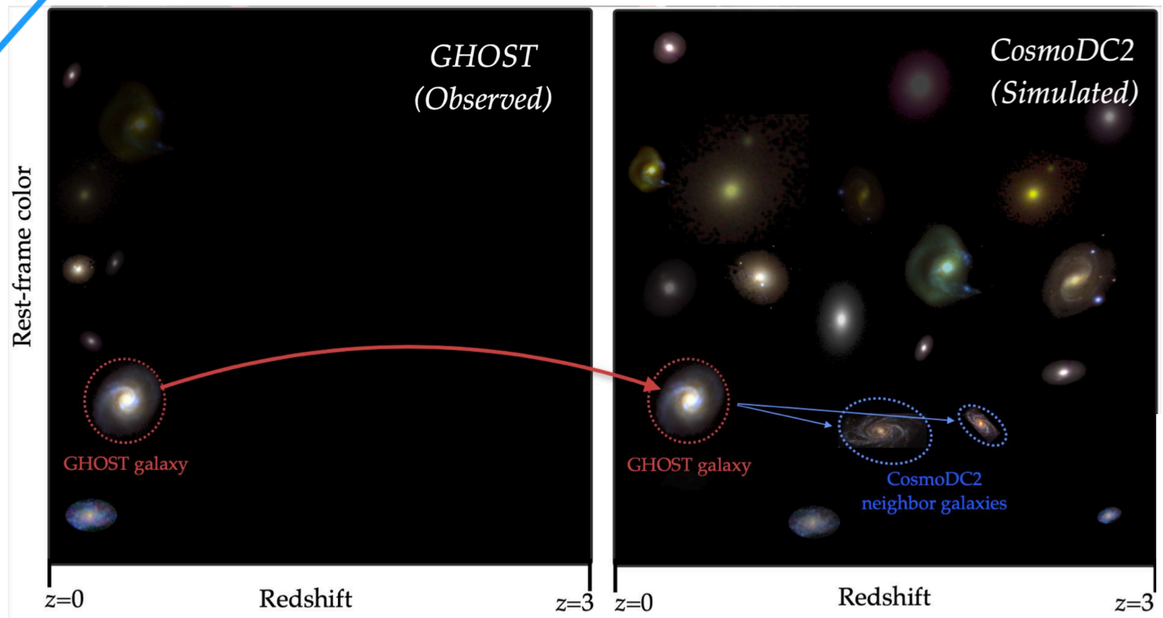
- Drawn from ***CosmoDC2*** matching photometry distributions from ***GHOST***

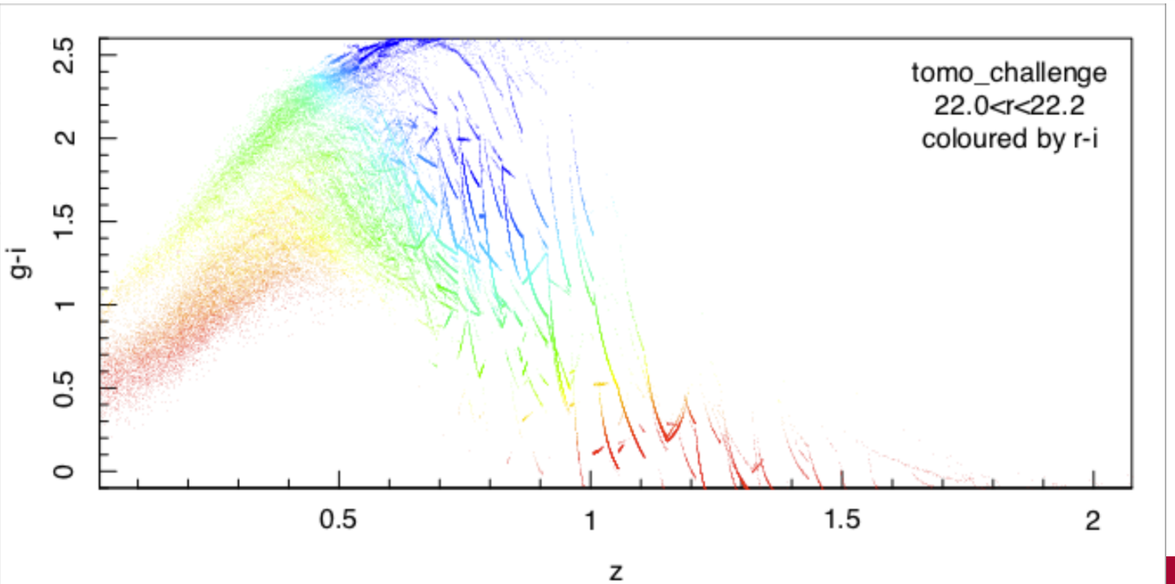
2B galaxies in 440 sq. deg simulation *(Korytov+2019)*

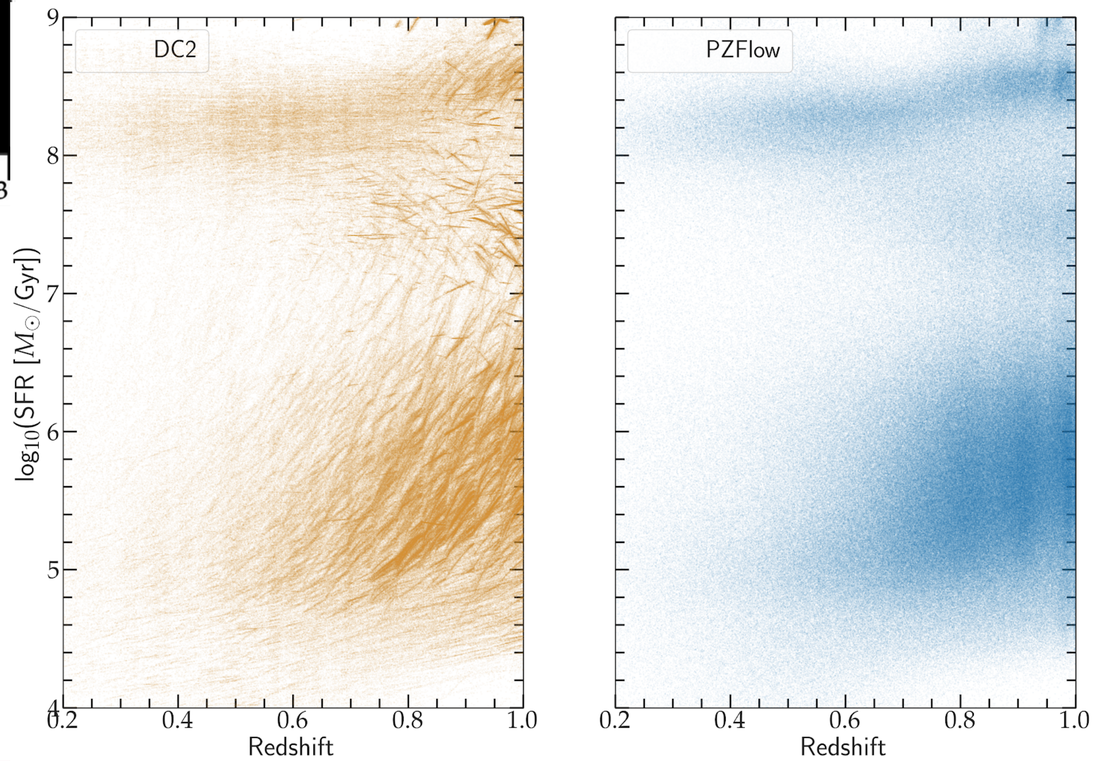


PS1 photometry for 16.5k observed SN host galaxies *(Gagliano+2021)*

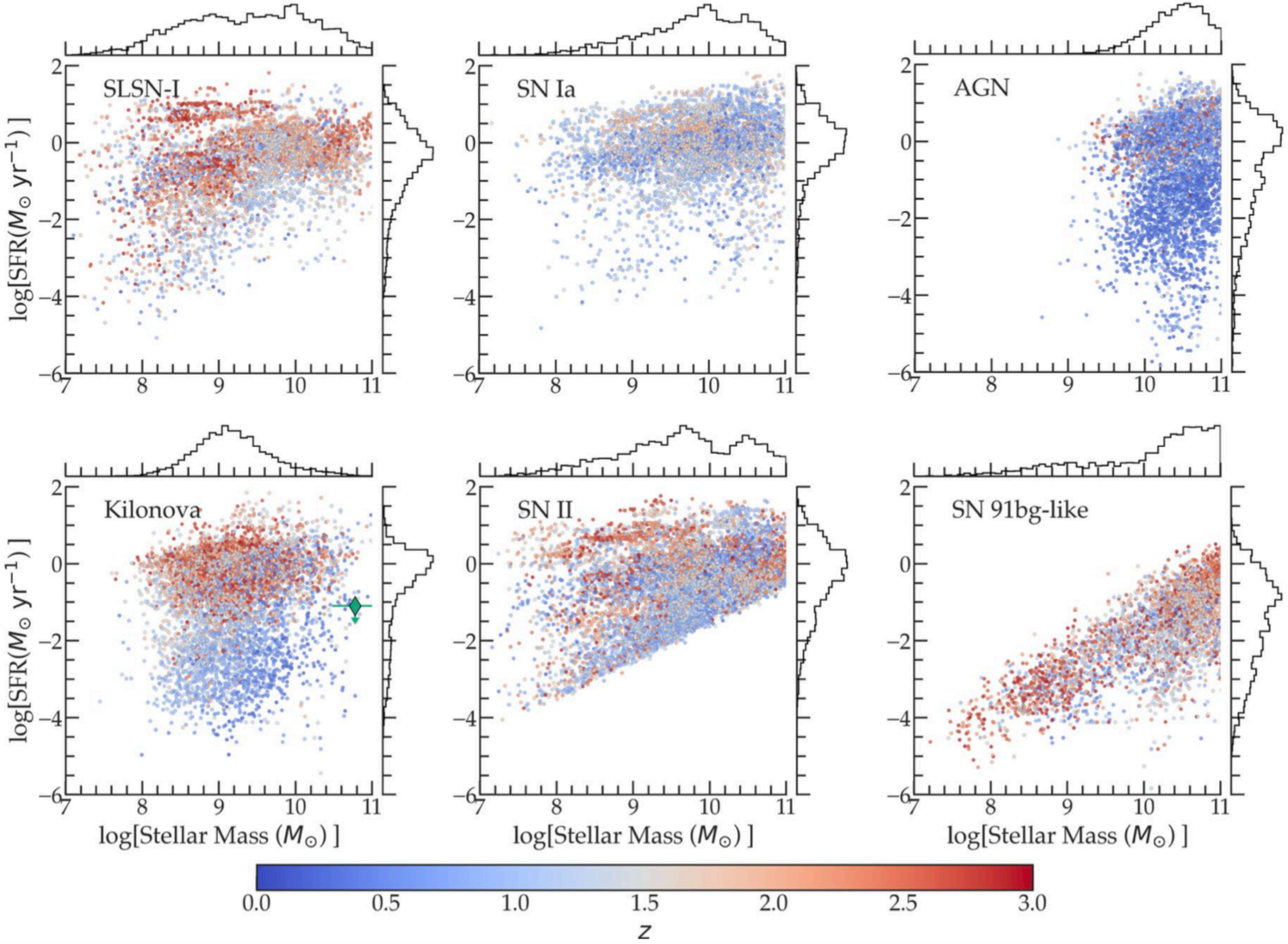
PZFlow (John Franklin Crenshaw) conditioned on DC2 color, magnitude, size, ellipticit, which we resample from







Hosts and Transients are correlated in ELAsTiCC



Host selection from *host library* weighted by class-specific *weight map*

Weight maps encode derived host correlations (SFR, M_*)

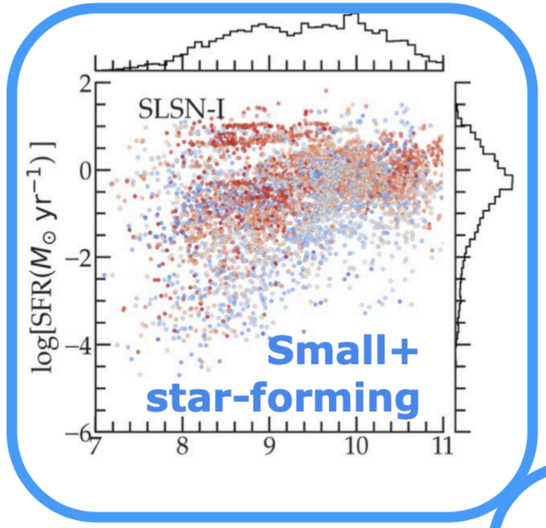
We used SGRBs as a proxy for KNe (L. Salo)

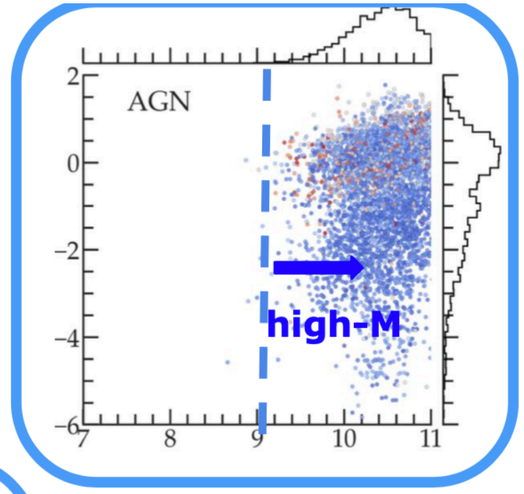
Hosts and Transients are correlated in ELAsTiCC

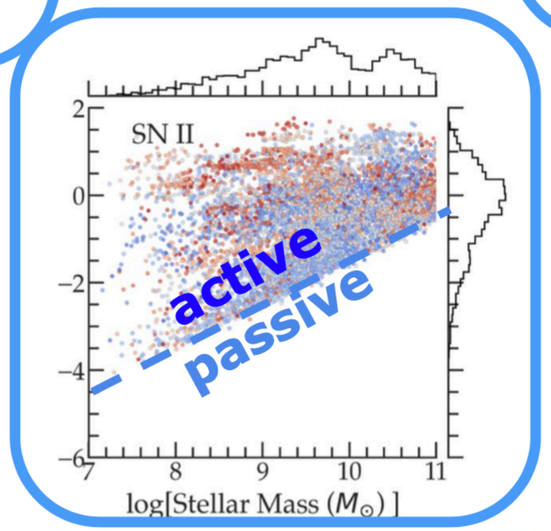
Need to roll this for a different survey - we made it possible to!

The Simulated Catalog of Optical Transients and Correlated Hosts (SCOTCH)

Lokken, Gagliano, Narayan et al. 2206.02815







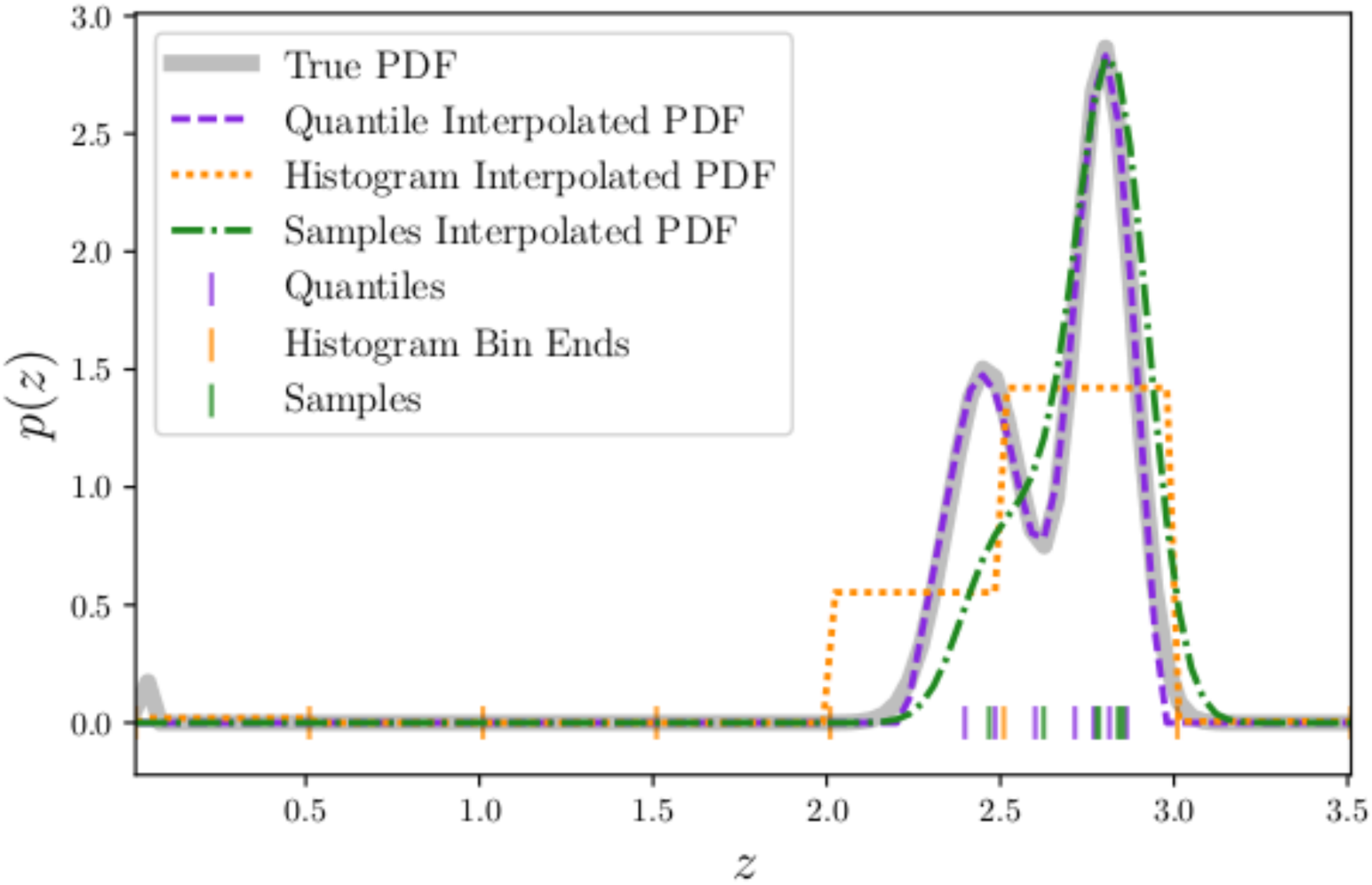
Host selection from *host library* weighted by class-specific *weight map*

Weight maps encode derived host correlations (SFR, M_*)

SLSNe-I occur in *compact blue galaxies*, SNe II (core-collapse) in *active galaxies;* AGN in *massive galaxies.*

Hosts and Transients are correlated in ELAsTiCC

- We also use PZFlow to mock host-galaxy photometric redshift PDFs
- Each ELAsTiCC transients comes with up to two possible hosts (both of which might be wrong!)
- Each host comes with 10 quantiles and an interpolation scheme to restore the full PDF
- Future: Expect transients & hosts to live in non-LCDM cosmology + simulated postage stamps



ELASTICC MAPS
TAXONOMY ONTO A HIERARCHICAL TREE

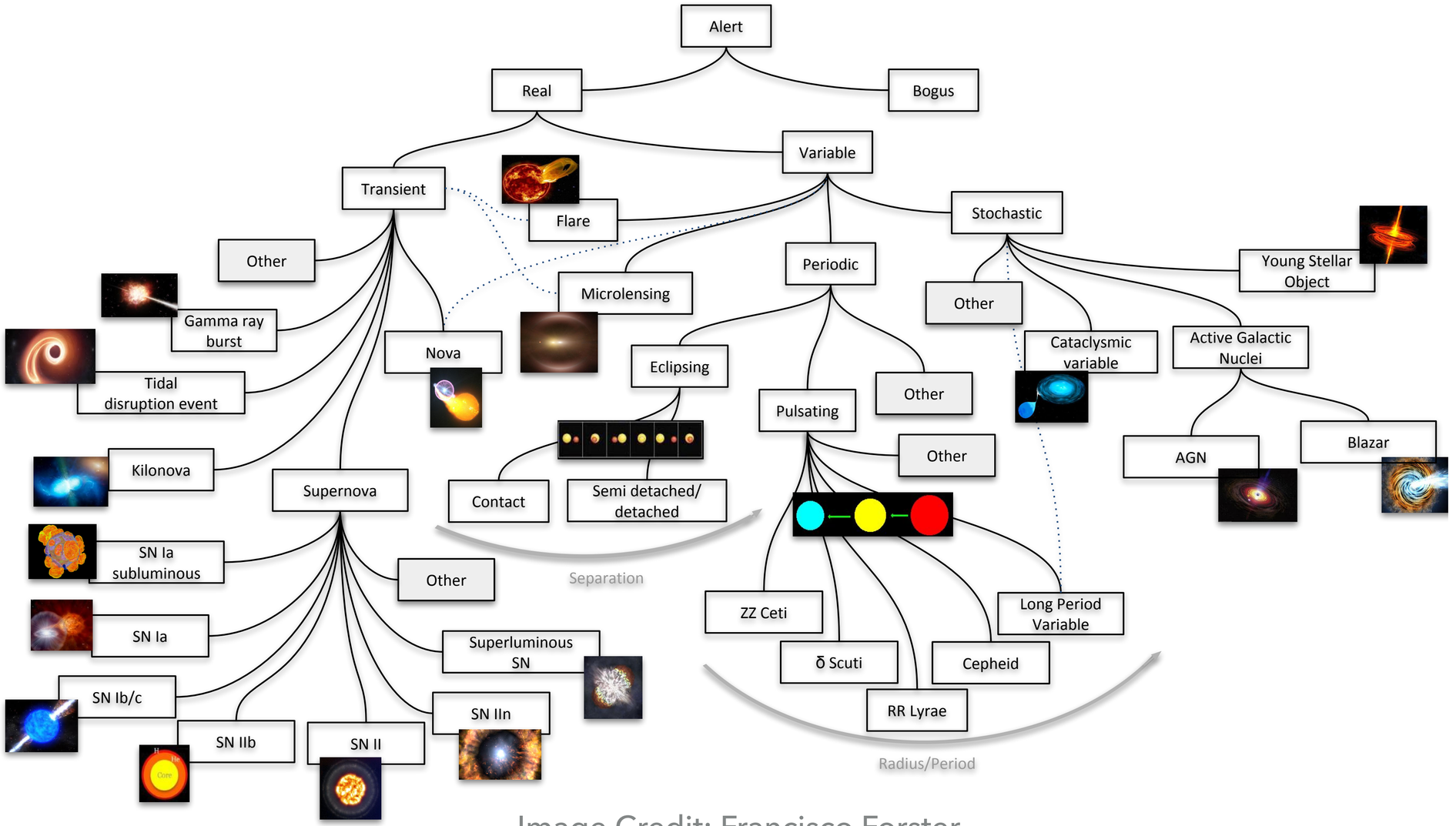


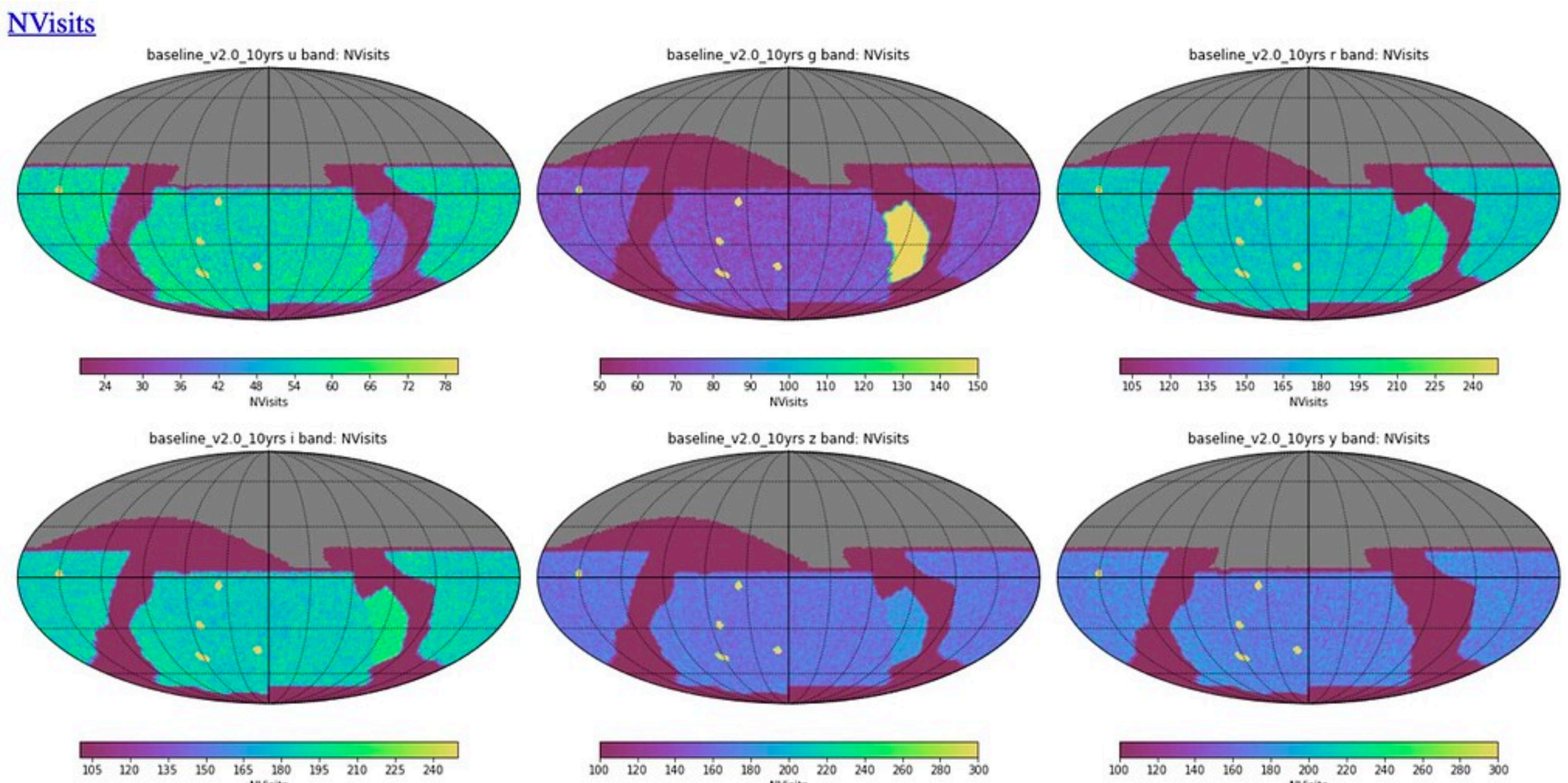
Image Credit: Francisco Forster

```
Alert/Real/Static
├── 0 Static/Other
├── 1 Non-Recurring
│   ├── 10 Non-Recurring/Other
│   ├── 11 SN-like
│   │   ├── 110 SN-like/Other
│   │   ├── 111 Ia
│   │   ├── 112 Ib/c
│   │   ├── 113 II
│   │   ├── 114 Iax
│   │   └── 115 91bg
│   ├── 12 Fast
│   │   ├── 120 Fast/Other
│   │   ├── 121 KN
│   │   ├── 122 M-dwarf Flare
│   │   ├── 123 Dwarf Novae
│   │   └── 124 uLens
│   └── 13 Long
│       ├── 130 Long/Other
│       ├── 131 SLSN
│       ├── 132 TDE
│       ├── 133 ILOT
│       ├── 134 CART
│       └── 135 PISN
└── 2 Recurring
    ├── 20 Recurring/Other
    ├── 21 Periodic
    │   ├── 210 Periodic/Other
    │   ├── 211 Cepheid
    │   ├── 212 RR Lyrae
    │   ├── 213 Delta Scuti
    │   ├── 214 EB
    │   └── 215 LPV/Mira
    └── 22 Non-Periodic
        ├── 220 Non-Periodic/Other
        └── 221 AGN
```

LSST Surveys

WFD:
5 sigma point-source detection depths of

u	g	r	i	z	y	
23.9,	25.0,	24.7,	24.0,	23.3,	22.1	single image
26.1,	27.4,	27.5,	26.8,	26.1,	24.9	stacked



ELAsTiCC is Live

The DESC ELAsTiCC Challenge

The purpose of ELAsTiCC ("Extended LSST Astronomical Time-series Classification Challenge") is to spur the creation and testing of an end-to-end real-time pipeline for time-domain science. The challenge starts with a simulation of ~5 million detected events that includes ~50 million alerts. These alerts will be streamed from LSST to brokers, who will classify the events and send new alerts with classifications back to DESC. A talk about ELAsTiCC given at the LSSTC Enabling Science Broker Workshop in 2021 can be found on YouTube.

For discussion or questions about the challenge, use the **`#elasticc-comms`** channel on the DESC Slack.

There is a new github repository for ELAsTiCC-related code and information: LSSTDESC/elasticc.

- Current Status
- Timeline
- Training Set
- Classification codes and models
- Participants

Current Status

`#elasticc-comms`

This channel on the LSST Slack is where you can contact the ELAsTiCC team and discuss the campaign. If you are not on the LSST Slack, you can join this one channel using Slack Connect. (That link expires every 14 days, and we need to renew it, so if you find the link expired, please email `raknop@lbl.gov` and me to update the link.)

10% Test Stream

Starting sometime between 20:00 and 21:00 UTC (1-2PM PDT) today (June 17), we will be streaming alerts to **`public.alerts.ztf.uw.edu`** in the topic **`elasticc-mid-june-test`**. The rate at which alerts will be added to this topic is similar to the rate of alerts that will come during the actual elasticc campaign. However, there are only about 10% the total number of alerts, so the whole thing will be streamed over 8–10 days. The purpose of this test stream is for bulk testing, of the plumbing, and of the mechanics of the classifiers running on the brokers.

https://portal.nersc.gov/cfs/lsst/DESC_TD_PUBLIC/ELASTICC/

TAKEAWAYS

- ELAsTiCC goes far beyond PLAsTiCC

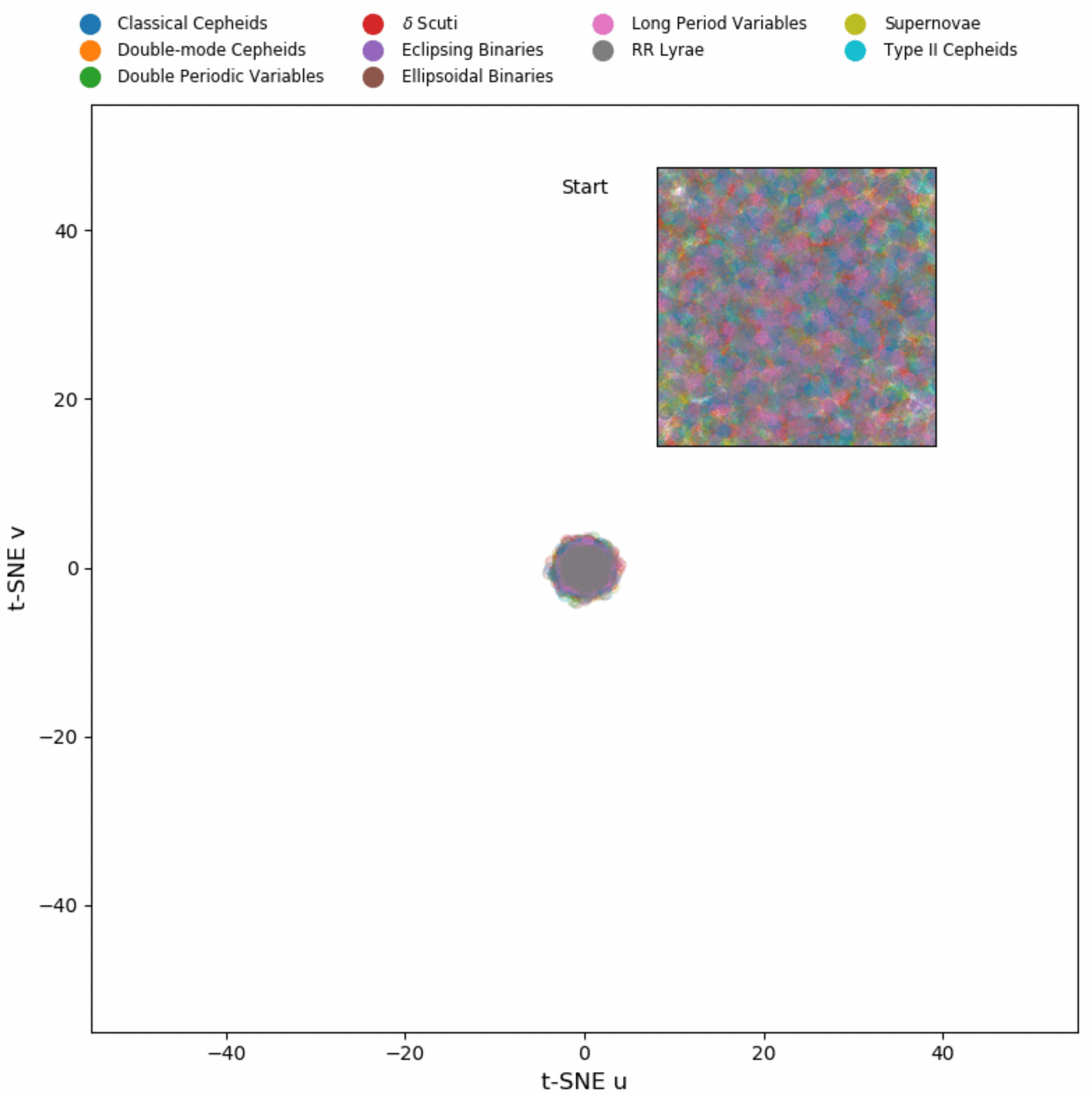
  5M light curves, over 50M alerts, with hosts streamed over 3 months through real LSST brokers with real LSST format, using real LSST infrastructure, with best available LSST properties

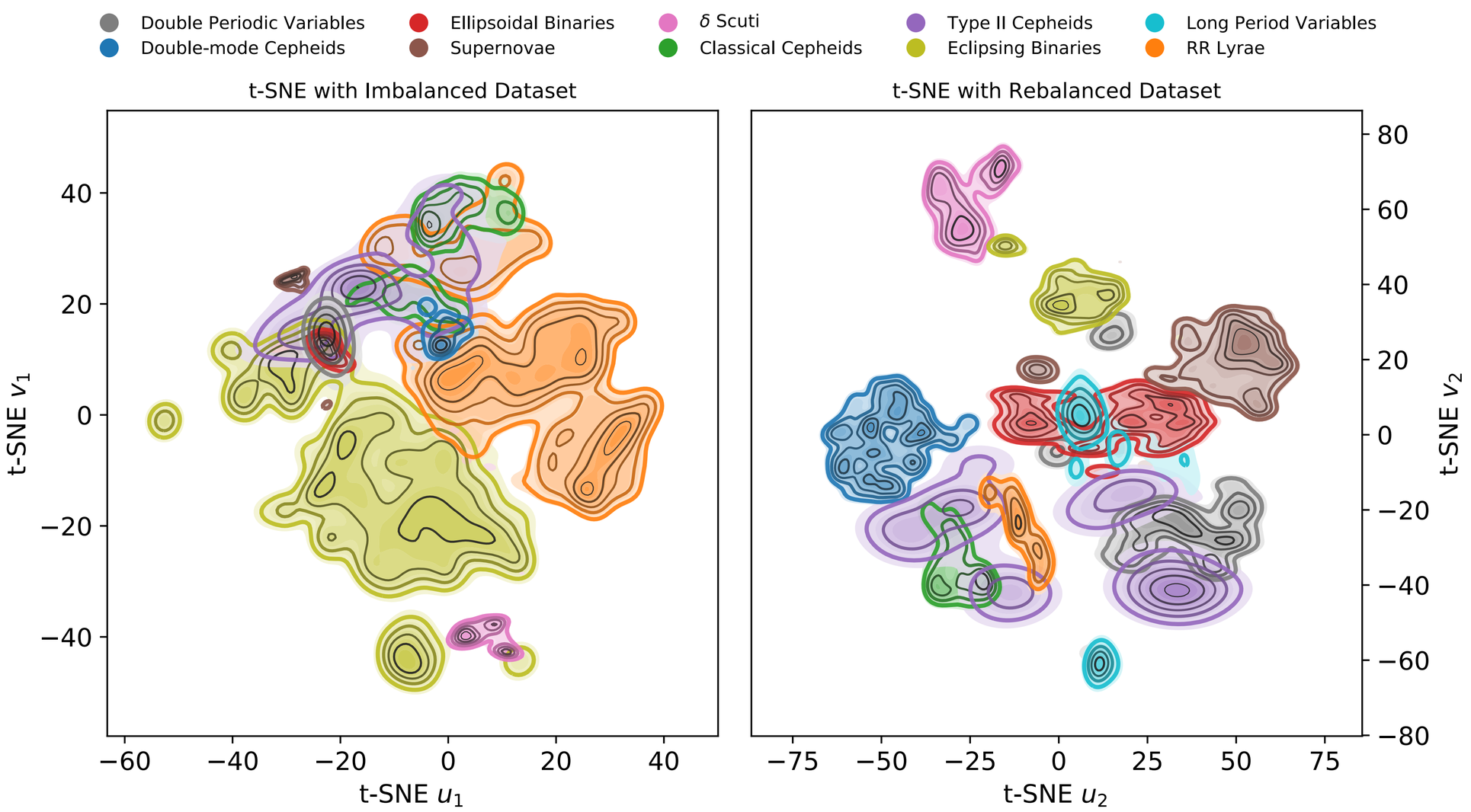
- designed to test your AI/ML mettle with richer questions, a more complex dataset,

- Beginning of an iterative process - DL methods should respond to real-data and get better over LSST operations

- **Next step: make this first step of a full DESC injection pipeline - i.e. AI/ML methods will play a key role in our cosmological constraints with LSST**

Use Unsupervised learning/ clustering techniques





Account for vastly different rates of events by intelligently oversampling minority classes - dramatically improves classification accuracy

t-SNE FEATURE VISUALIZATION

t-SNE: t-distributed Stochastic Neighbor Embedding
SMOTE: Synthetic Minority Oversampling TEchnique