UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Workflows e Plataformas Científicas: preparação para o LSST DR1

Prospects and challenges for the study of extragalactic globular clusters in the LSST-era

---

JAs: Nicholas Schweder and Pedro Lopes

PIs: Ana Chies Santos and Charles Bonatto

March 25th, 2025

THE TEAM

Nicholas Schweder

Pedro Lopes (UFRGS)

Stars, Milky Way, and Local Volume Science Collaboration (SWMLV)
and Galaxies Science Collaboration (GSC)

Star Clusters Working Group

Ana Chies-Santos (UFRGS)

Charles Bonato (UFRGS)

Juan Caso (UNLP)

Thayse Pacheco (UFRGS)

Rafael S. de Souza (Hertfordshire)

Pedro Floriano (UFRGS)

Kristen Dage (CIRA)

Paula Coelho (IAG)

Marco Canossa (UFRGS/Kapteyn)

Ana Ennis (WCA)

Bruno de Bértoli (UNLP)

STELLAR GLOBULAR CLUSTERS

- Globular clusters (GCs) are formed by tens of thousands to millions of stars;
- Spherical arrangement of stars;
- Among the oldest objects in the Universe >10 Gyrs;

Globular Cluster: M92

Image: JWST

see: Krumholz et al. (2019), Chies-Santos et al. (2011b) and Forbes & Bridges (2010)

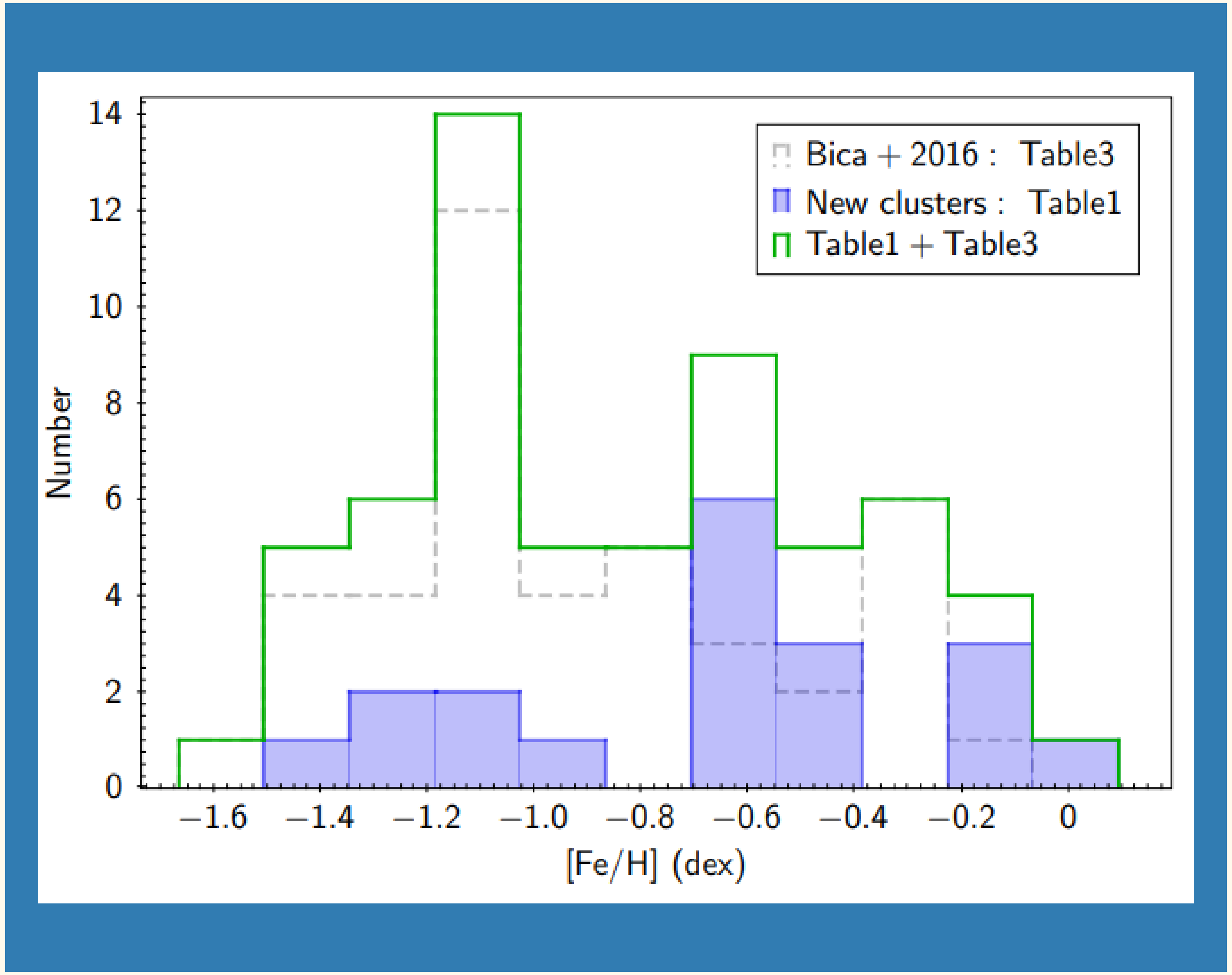
STELLAR GLOBULAR CLUSTERS

- GC system assembly: in-situ/ex-situ;

- Metal-rich GCs -> redder -> in-situ;
- Metal-poor GCs -> bluer -> ex-situ;

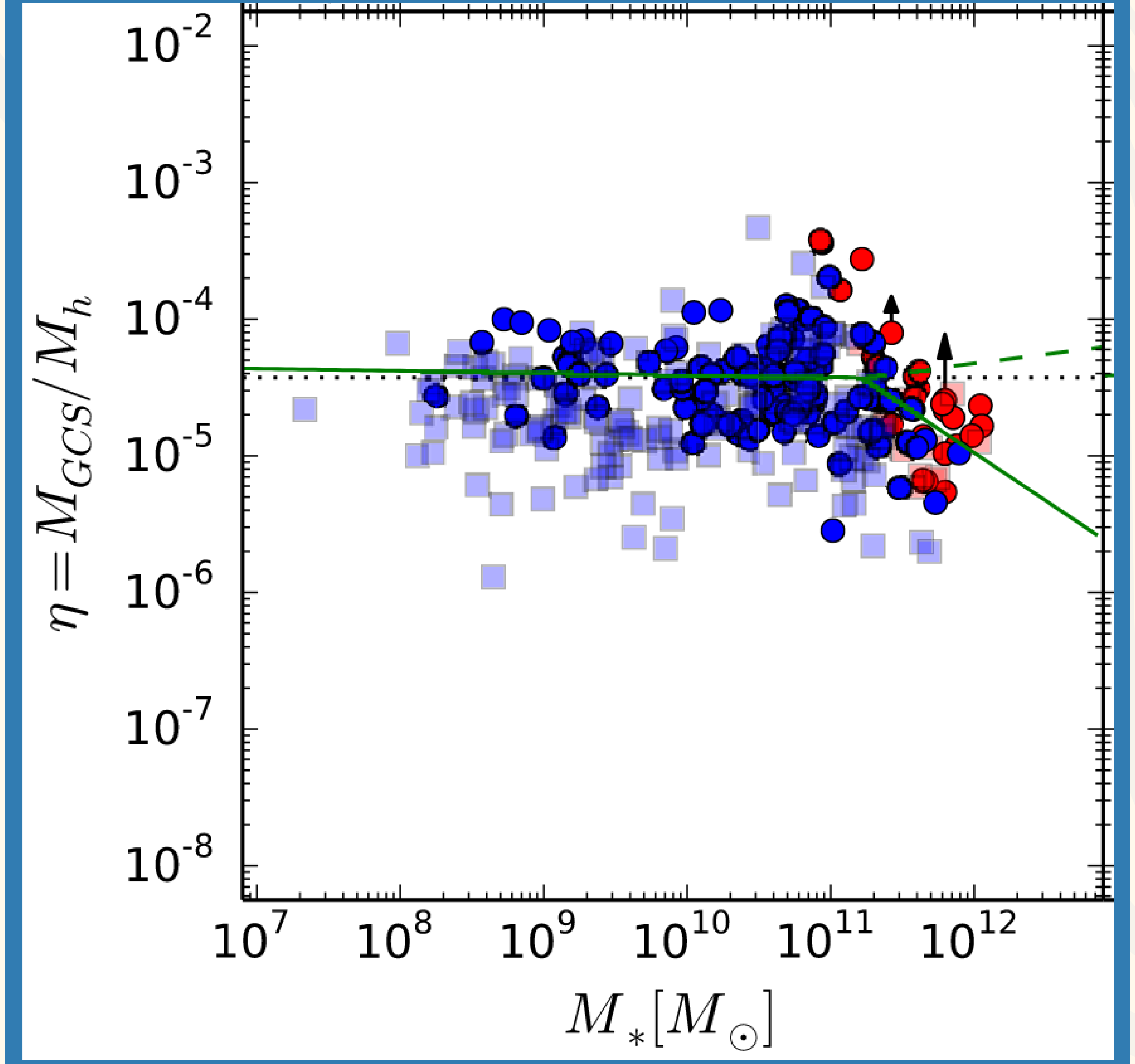
- In the Milky Way, the distribution of metallicity is typically referenced as bimodal;

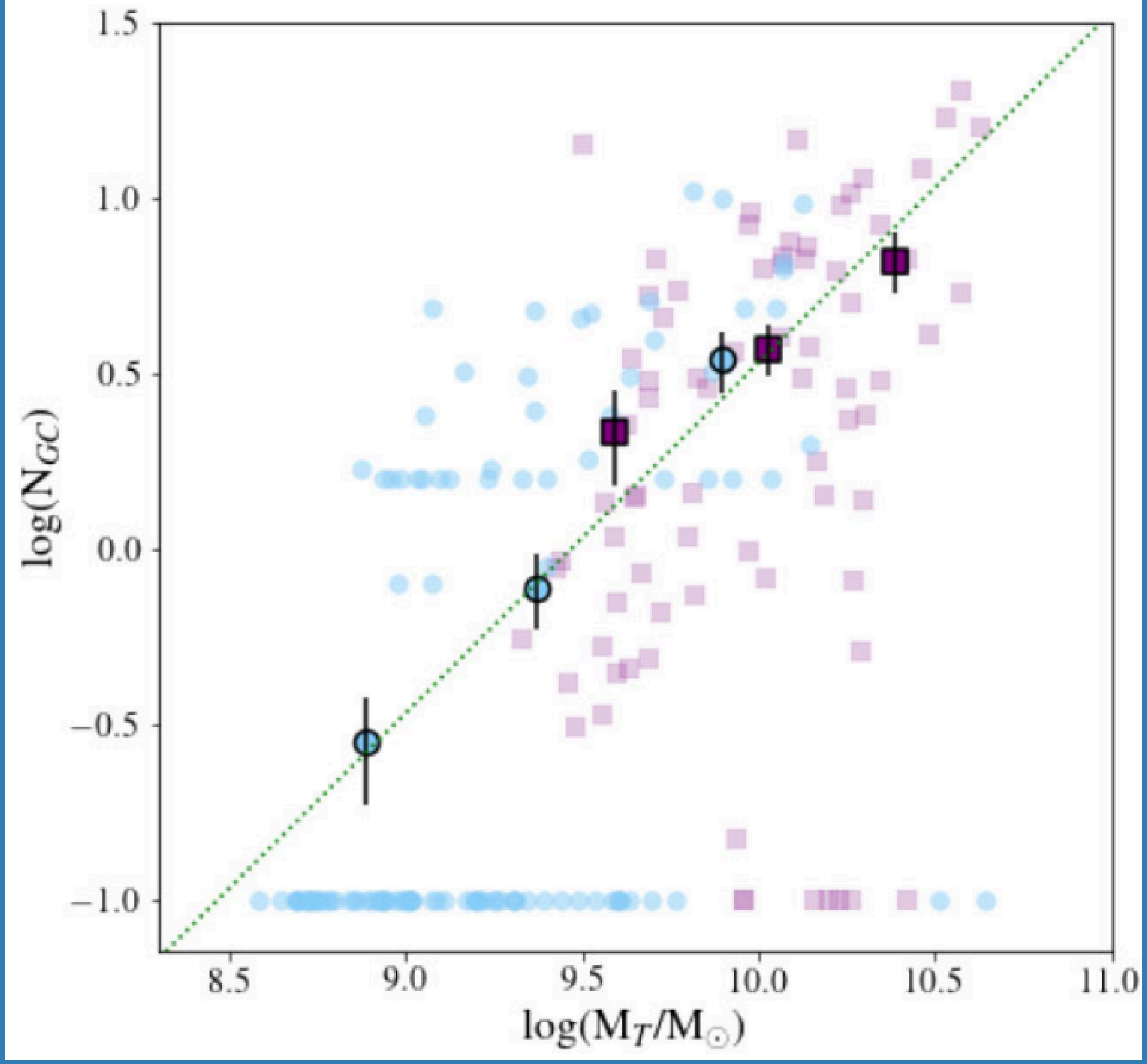
- GCs are valuable tracers of the assembly history of their host galaxy.



see: Brodie and Strader, 2006; Rossi et al., 2015; Usher et al., 2019, 2024; Bica et al., 2024

Laboratories for Stellar Evolution and Tracers of Dark Matter.

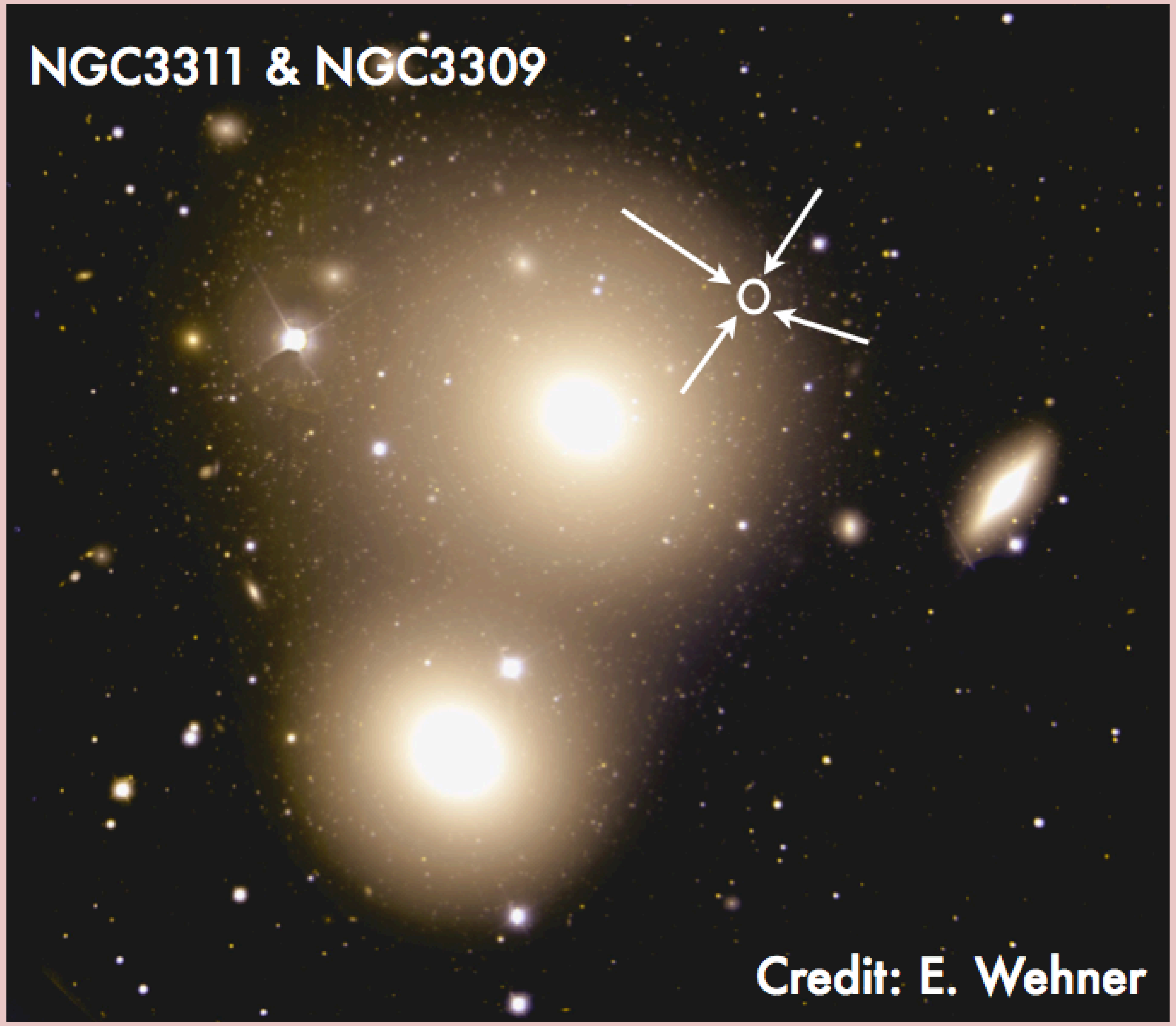




see: Hudson, Harris, & Harris (2014), Zaritsky (2022) and Krumholz et al. (2019)

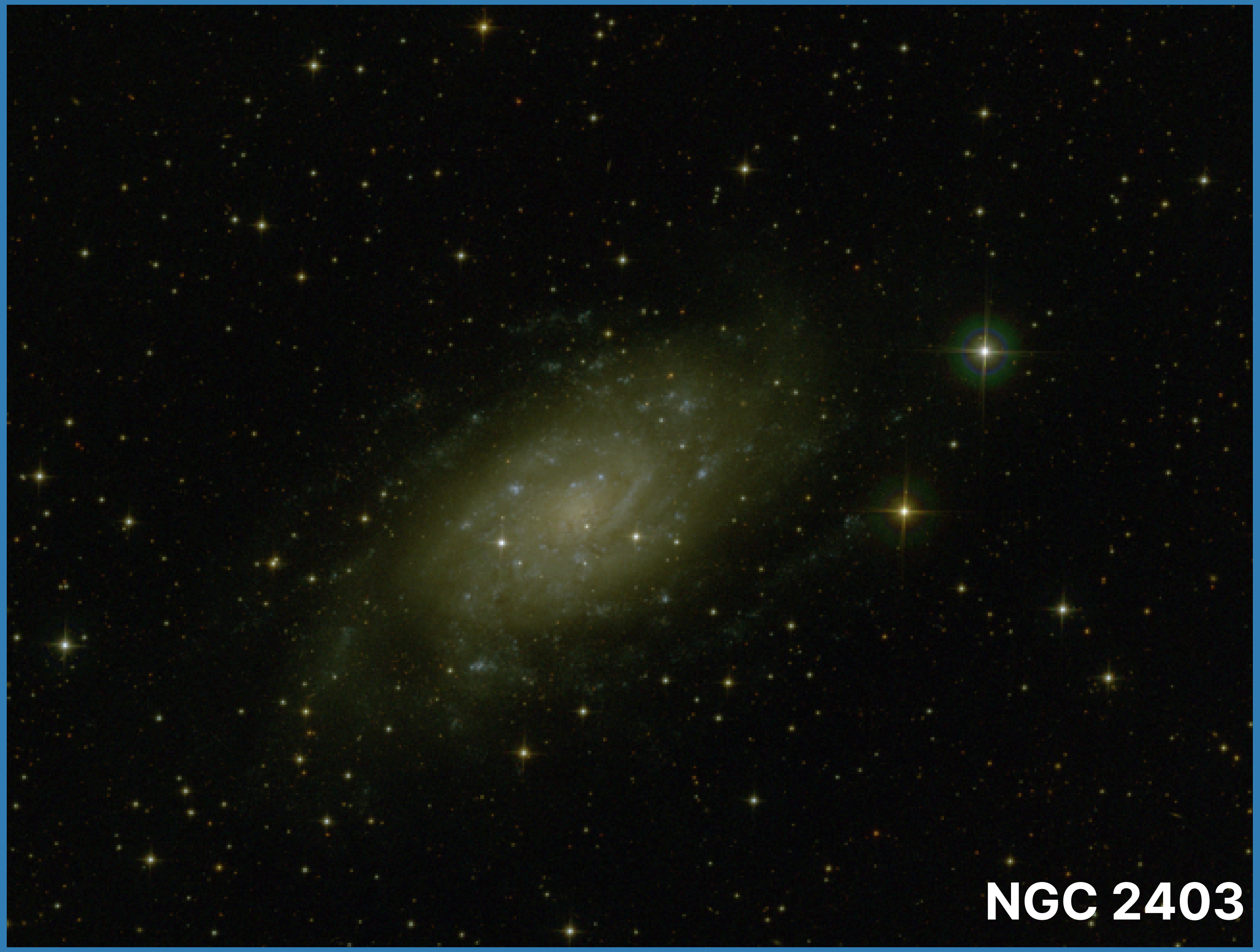
Early-Type

- Smoother light profile;
- Predominantly old, low-mass, and cool stars;
- Mass range: 10^8 - $10^{14} M_\odot$



Late-Type

- Disk, arm structures, and central bulge;
- Active star formation;
- Mass range: 10^9 - 10^{12} $M_\odot$



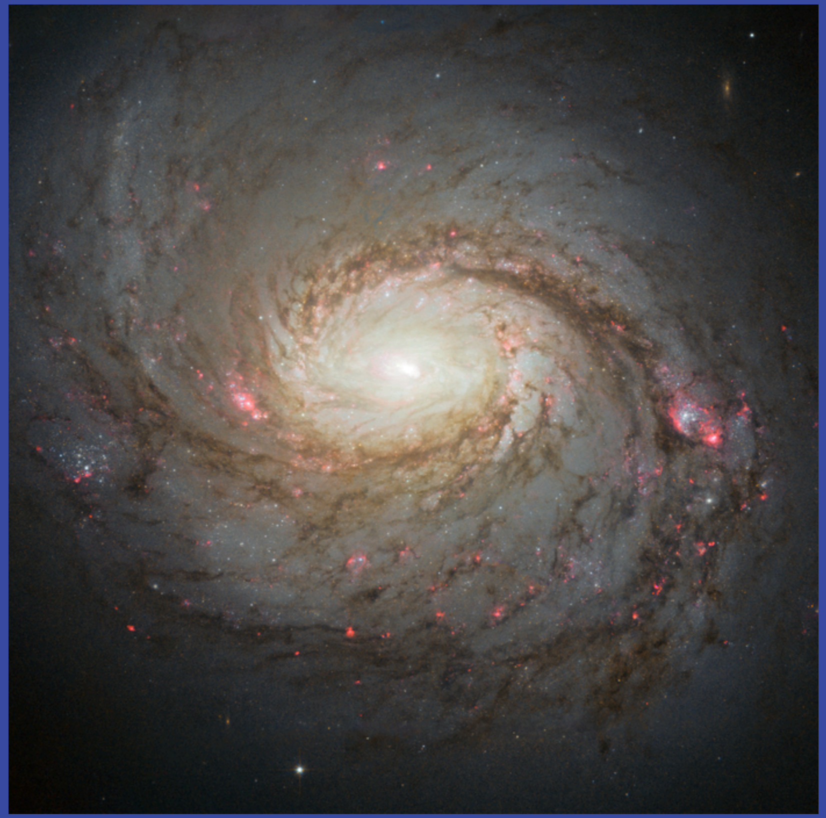
Images: E. Wehner, Gemini Observatory and J-PLUS

How many clusters does each type galaxy have?

GCs have a strict relation to the galaxy's mass

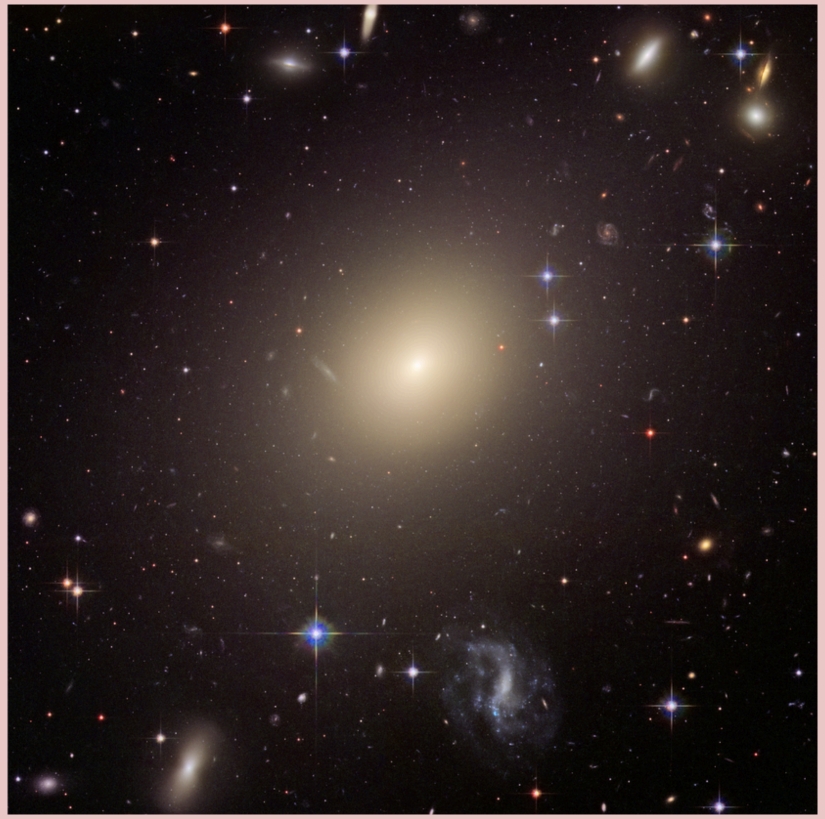
DWARF

0 - 10



SPIRAL

10 - 100



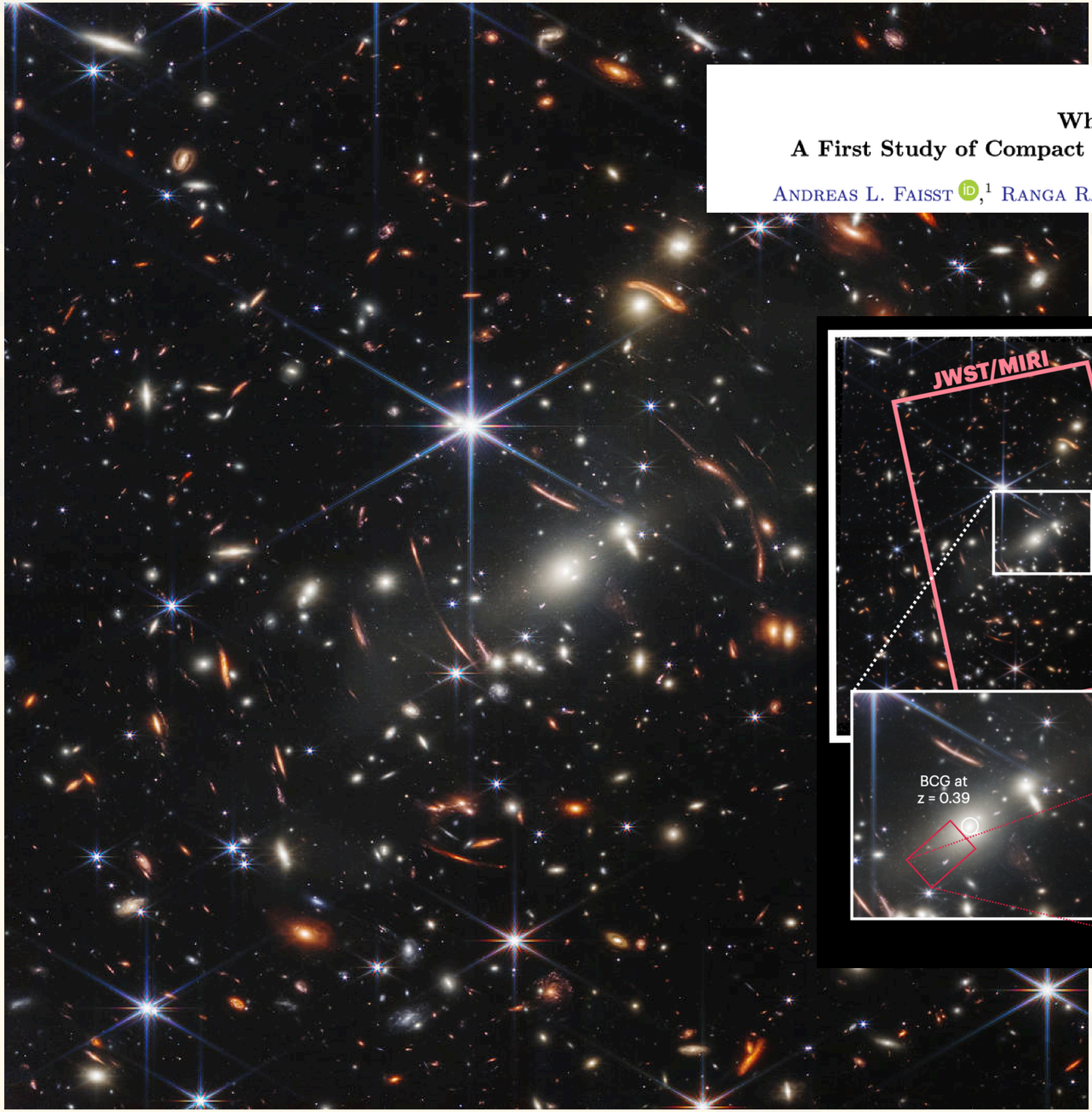
ELLIPTICAL

100 - 10000

Images: ESO and NASA

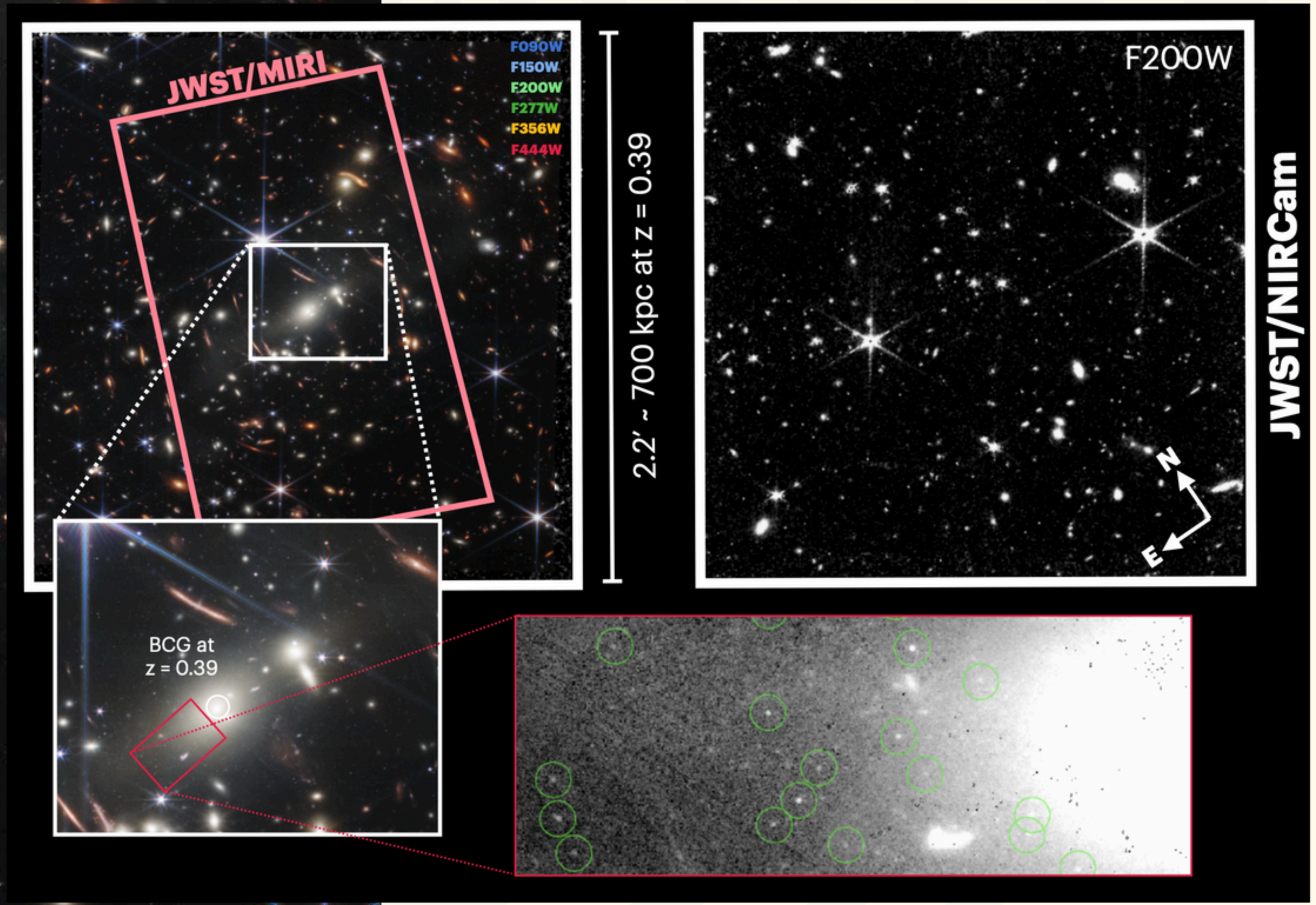
INTRACLUSTER

Image: JWST



What Are Those Tiny Things?
A First Study of Compact Star Clusters in the SMACS0723 Field with JWST

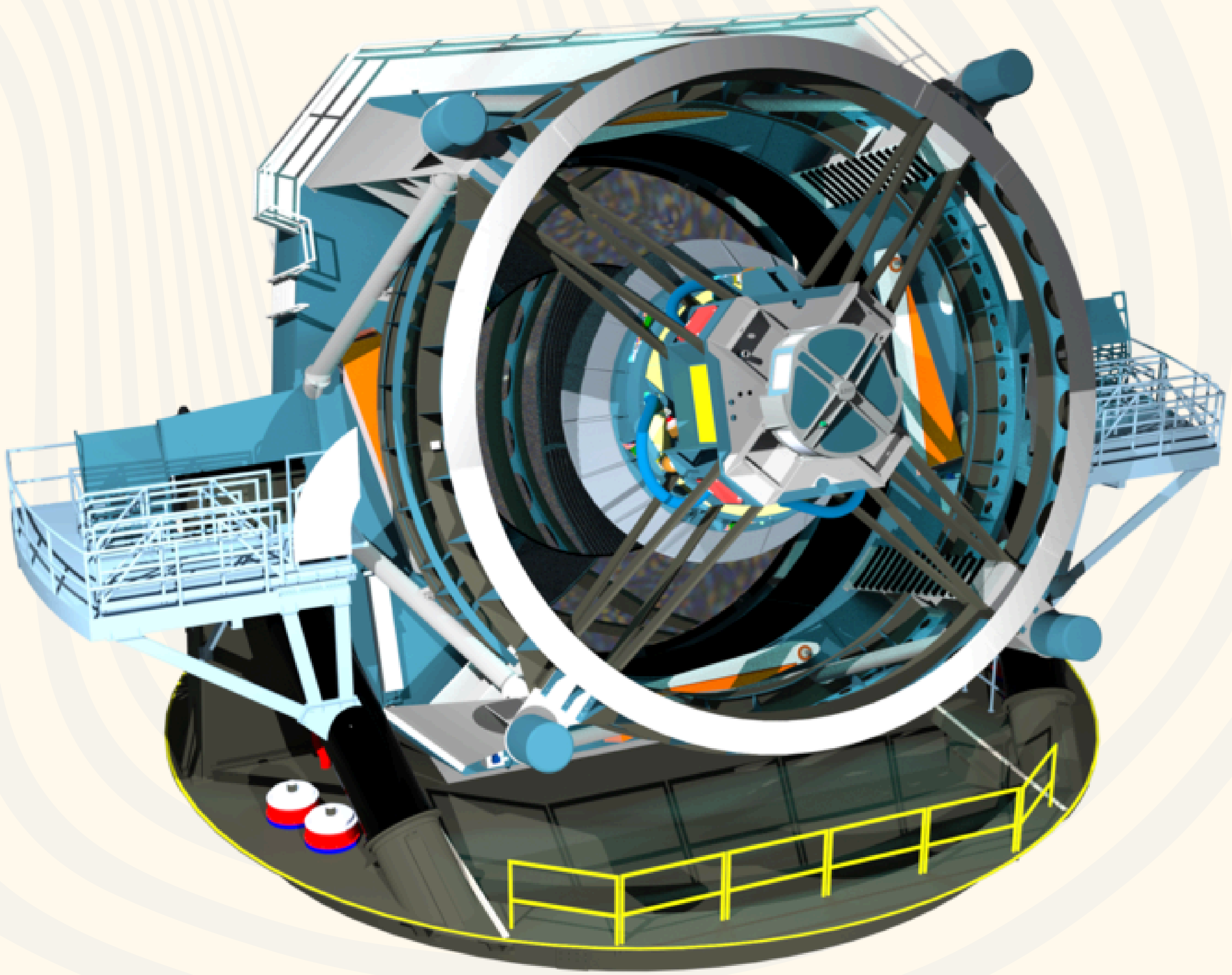
Andreas L. Faisst,[1] Ranga Ram Chary,[1] Gabriel Brammer,[2,3] and Sune Toft[2,3]



SMACS 0723 at z=0.39

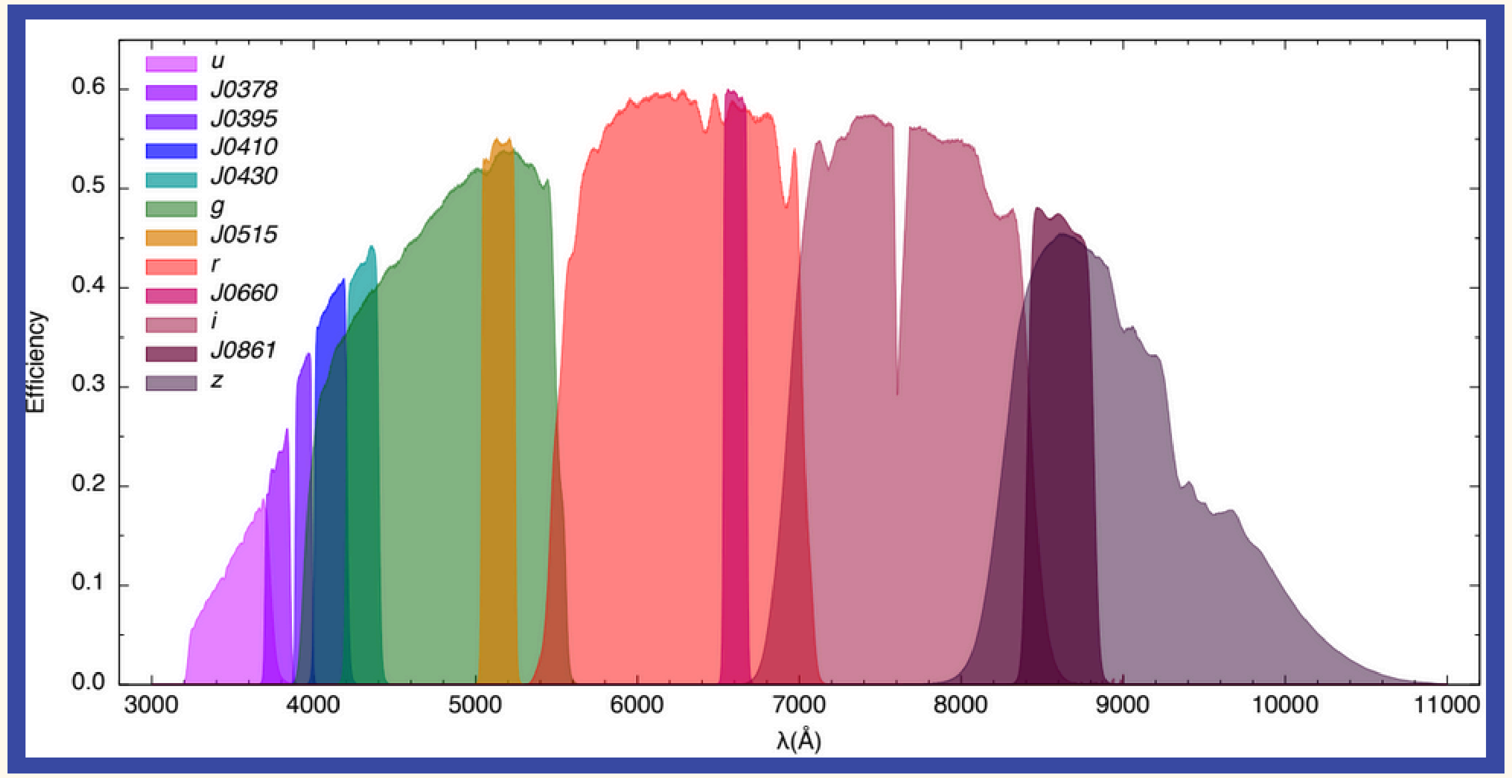
Preparation for LSST

(a) Band	(b) m Visit	(c) m 10 yr	(d) N_{GC}
u	23.60	25.45	4×10^5
g	24.41	26.54	1×10^7
r	23.98	26.71	3×10^7
i	23.39	26.20	2×10^7
z	22.79	25.51	9×10^6
y	22.00	24.73	3×10^6



Usher et al. (2023)

J-PLUS + Euclid

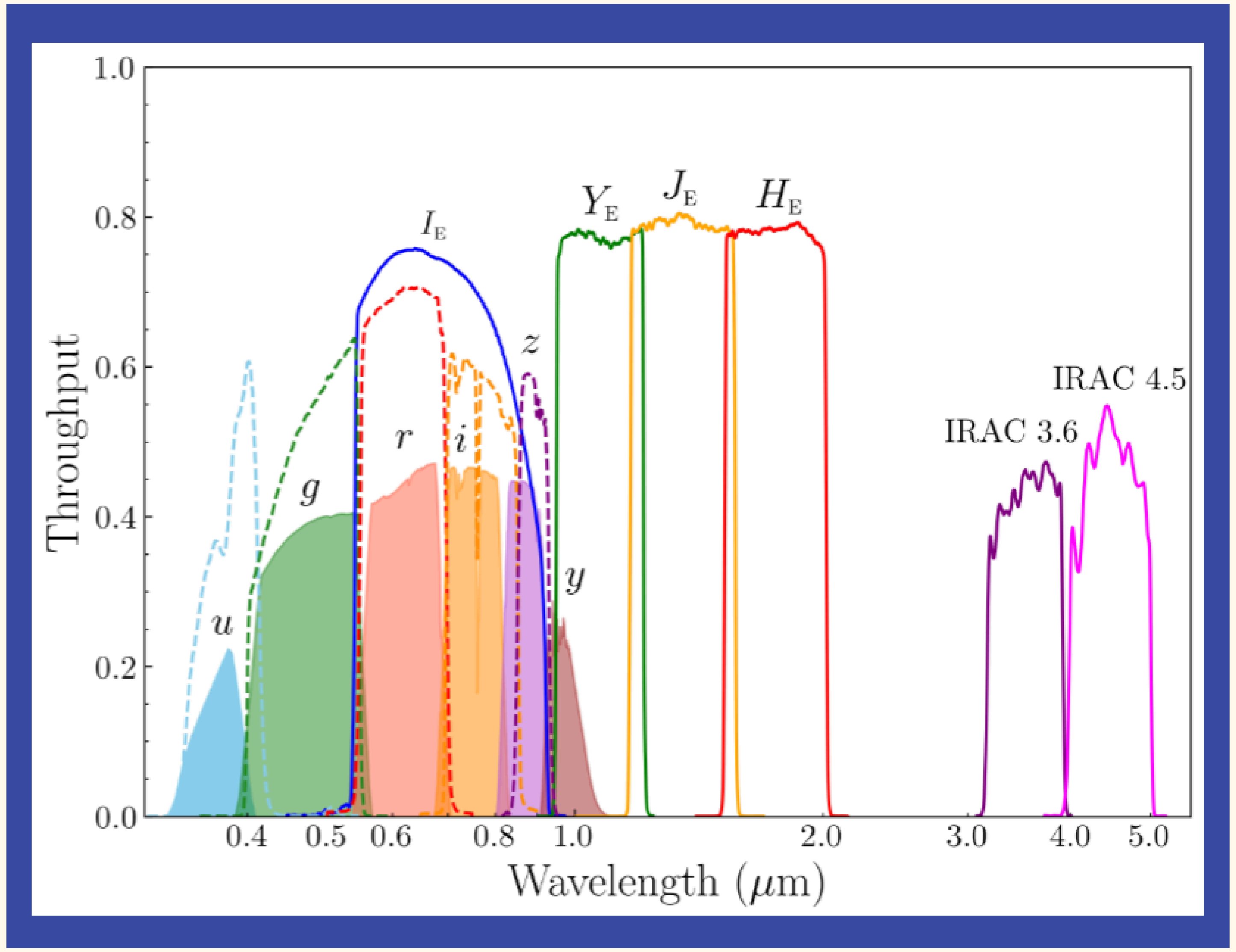


12 Filters

J-PLUS maps the northern hemisphere sky using 12 bands, 5 broad and 7 narrow. Ranging from near-infrared to UV-A light.

4 Filters

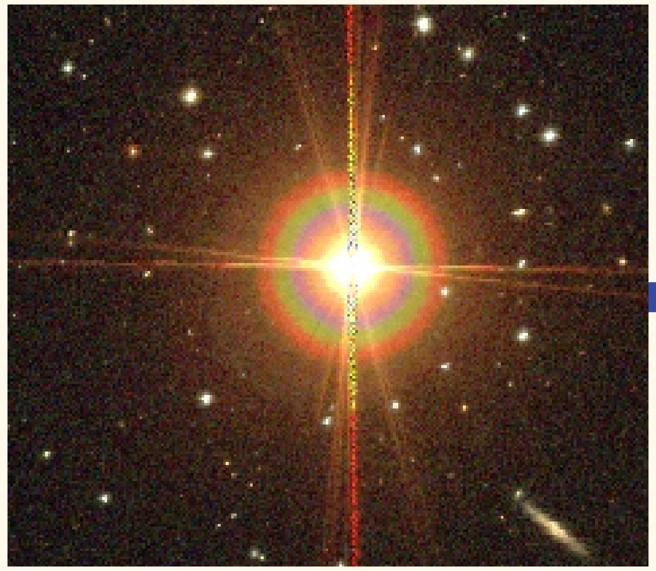
Euclid is a space telescope that observes the sky in 4 filters, 1 for visible light and 3 for infra-red light.



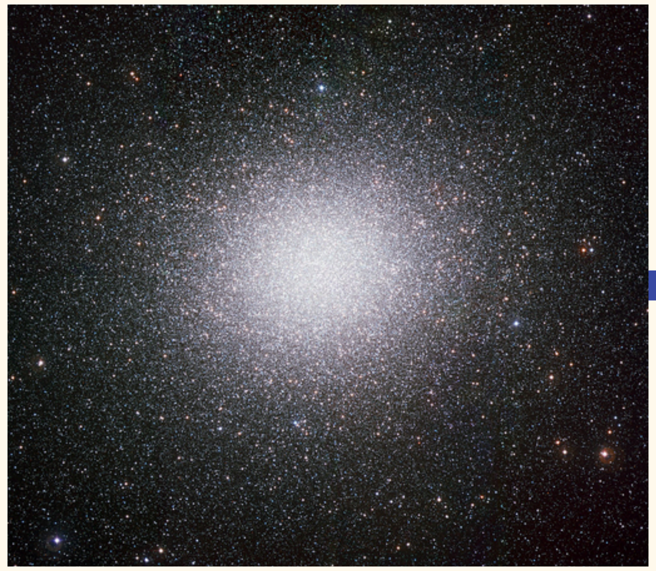
Images: Cenarro et al. (2019)

CHALLENGES FOR SELECTION

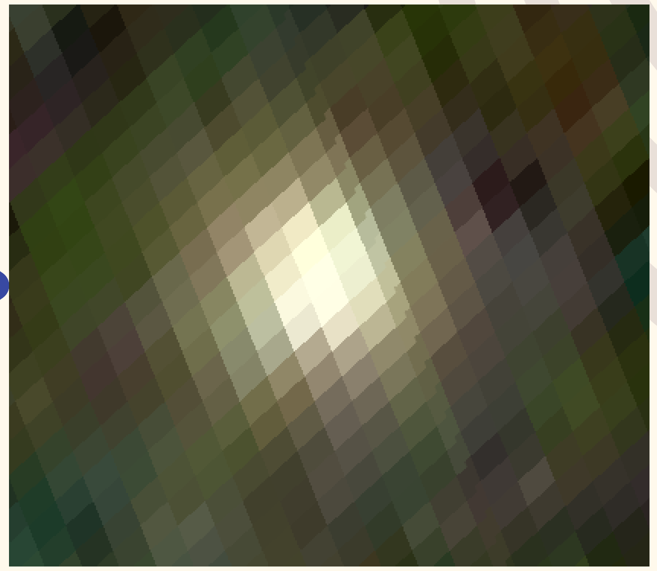
Foreground Star



Globular Cluster

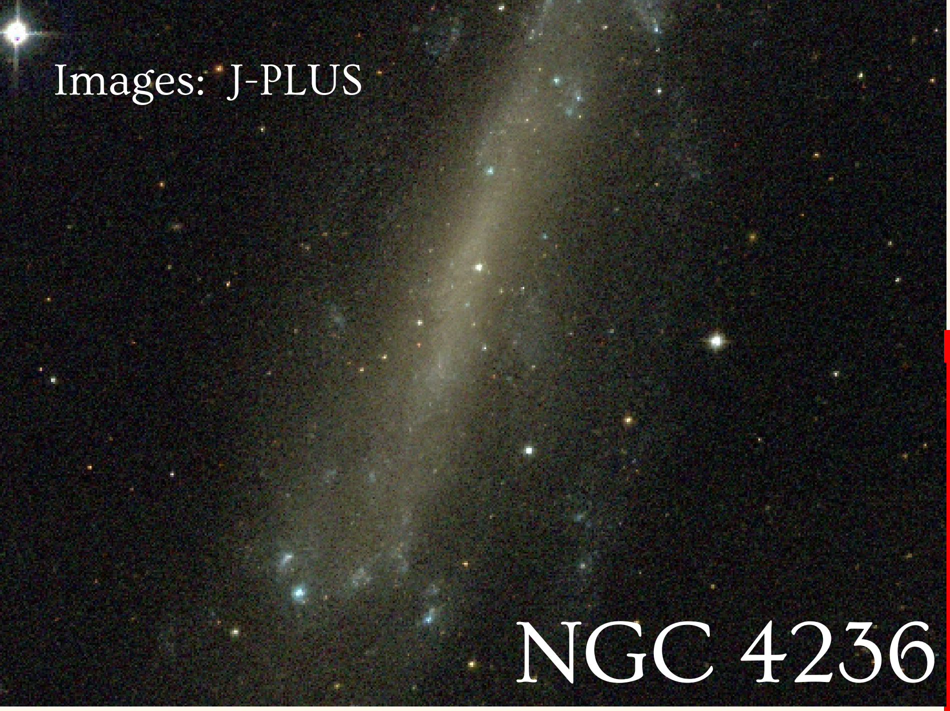


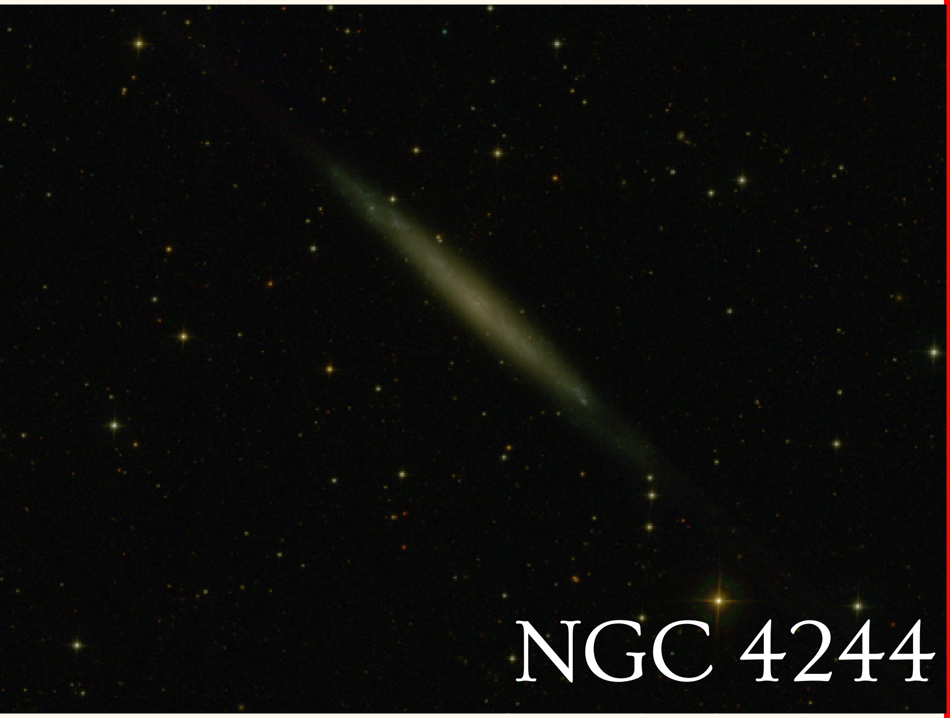
Background Galaxy

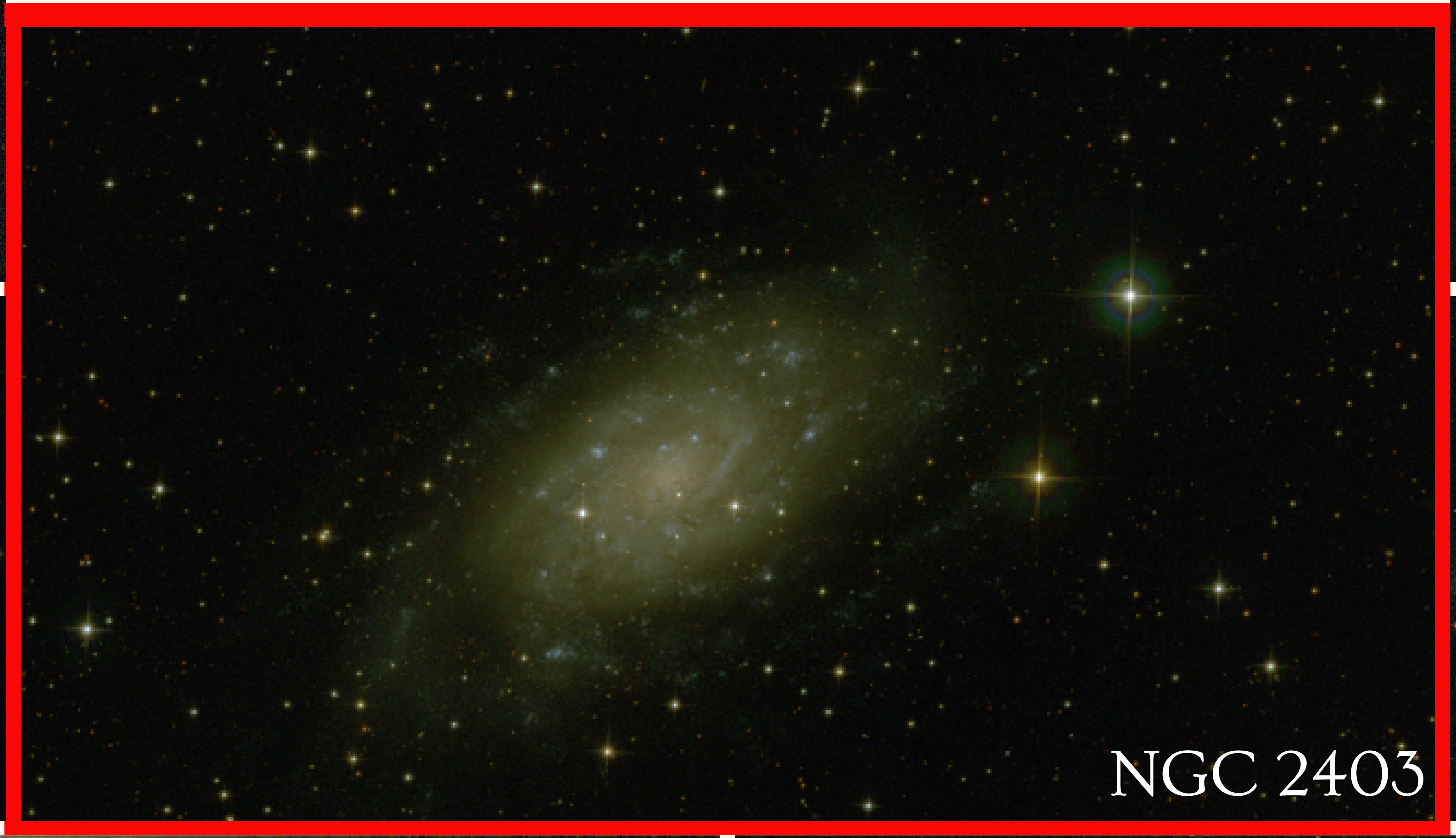


Images: ESO, NASA and J-PLUS

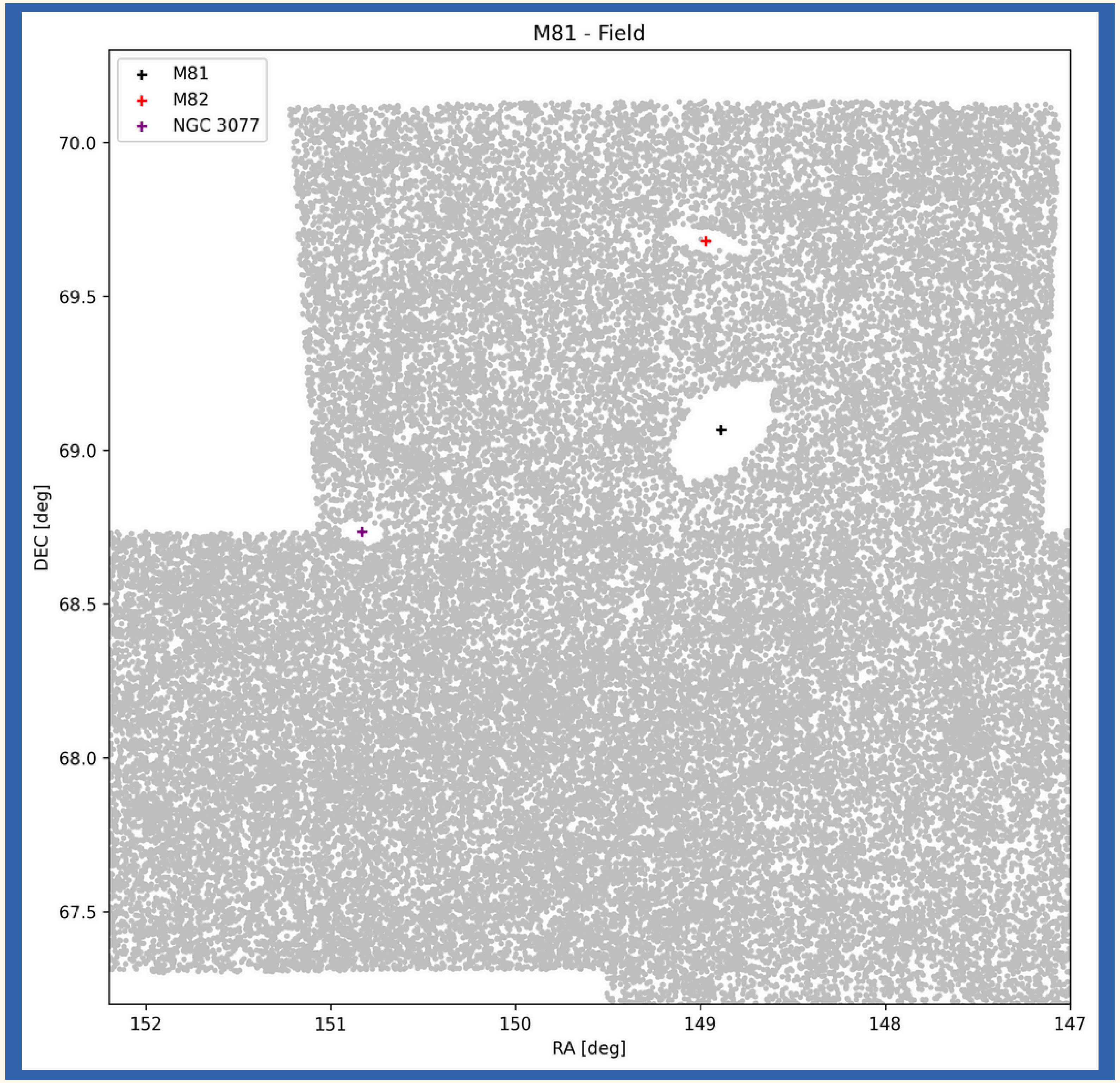
Galaxy Sample







Data Preparation

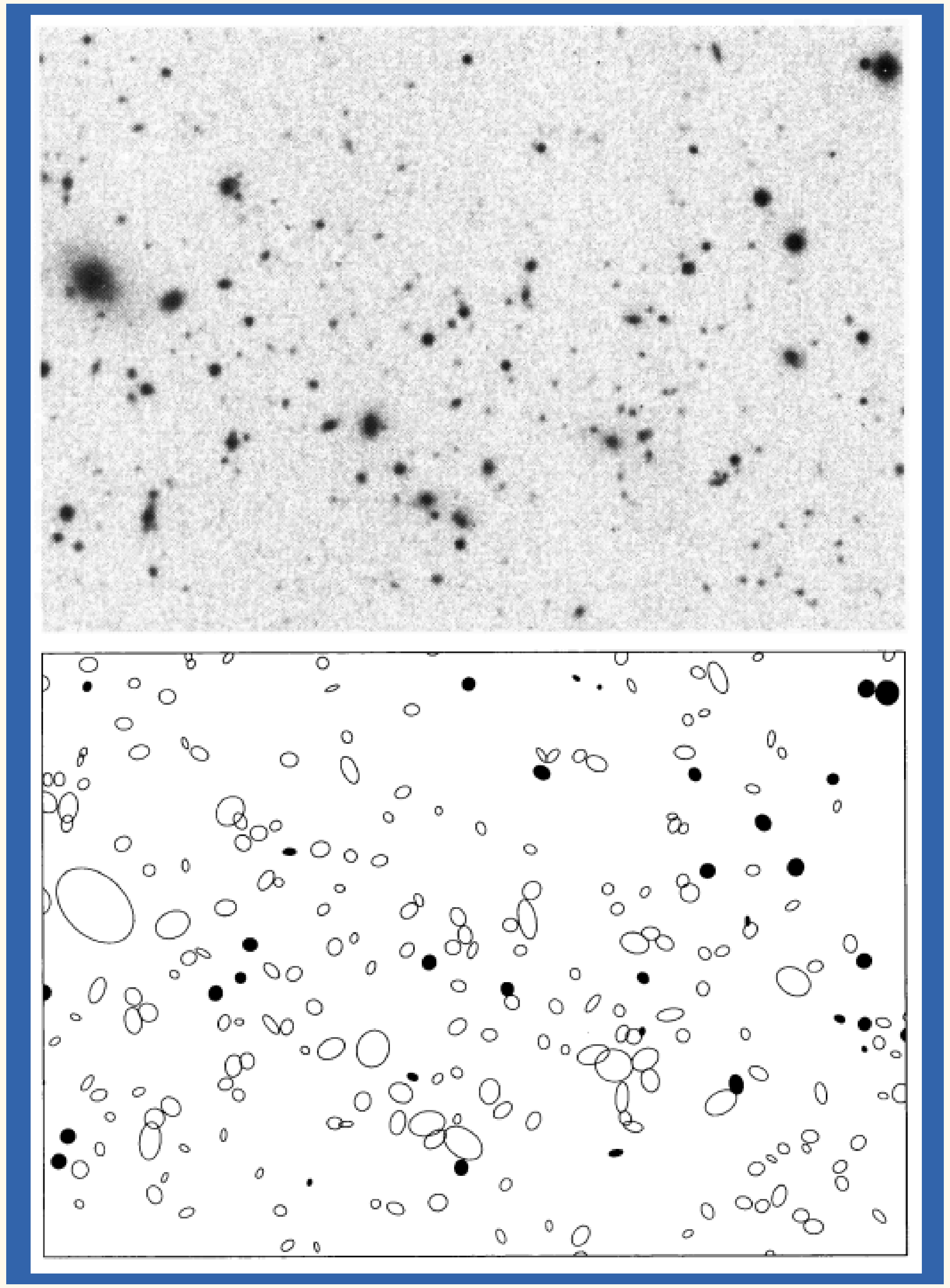


Rubin Early Science – Data Release Scenario

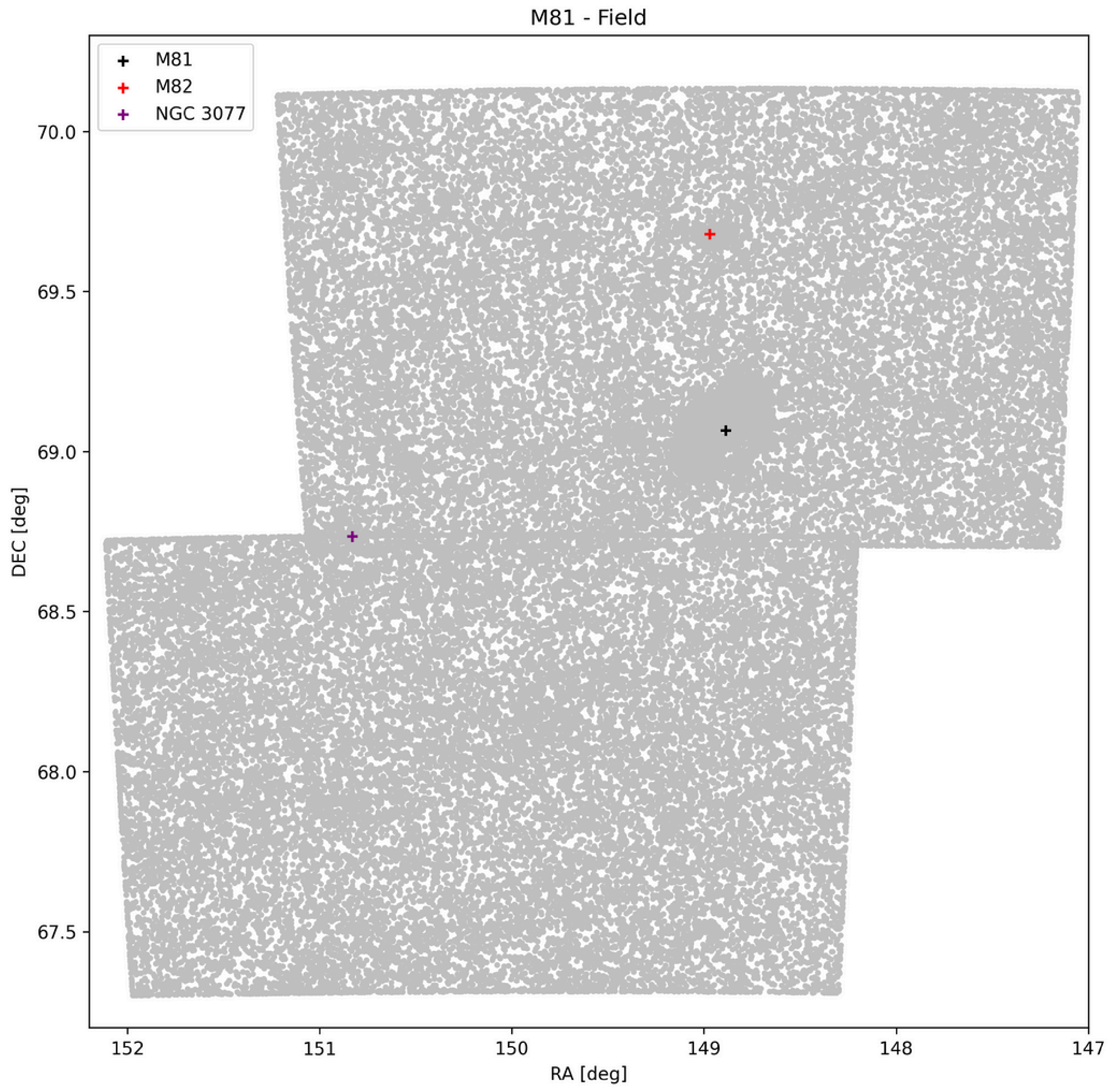
Data Product	Jun 2021	Jun 2022	Jun 2023	Jun 2025 – Jul 2025	Mar 2026 – May 2026	Sep 2026 – Jan 2027	Sep 2027 – Jan 2028	Sep 2028 – Nov 2028
	DP0.1	**DP0.2**	**DP0.3**	**DP1**	**DP2**	**DR1**	**DR2**	**DR3**
	DC2 Simulated Sky Survey	**Reprocessed DC2 Survey**	**Solar System PPDB Simulation**	**ComCam Data**	**LSSTCam Science Validation Data**	**LSST First 6 Months Data**	**LSST Year 1 Data**	**LSST Year 2 Data**
Raw Images	●	●	–	●	●	●	●	●
DRP Processed Visit Images and Source Catalogs	●	●	–	●	●	●	●	●
DRP Coadded Images and Object Catalogs	●	●	–	●	●	●	●	●
DRP Cell-based Coadded Images and ShearObject Catalog	–	–	–	–	○	●	●	●
DRP ForcedSource Catalogs	●	●	–	○	●	●	●	●
DRP Difference Images and DIA Catalogs	–	●	–	○	●	●	●	●
DRP SSP Catalogs	–	–	●	–	○	●	●	●

At first, we work with the automatic pipeline catalogues

Data Preparation by SExtractor



Bertin (1996)



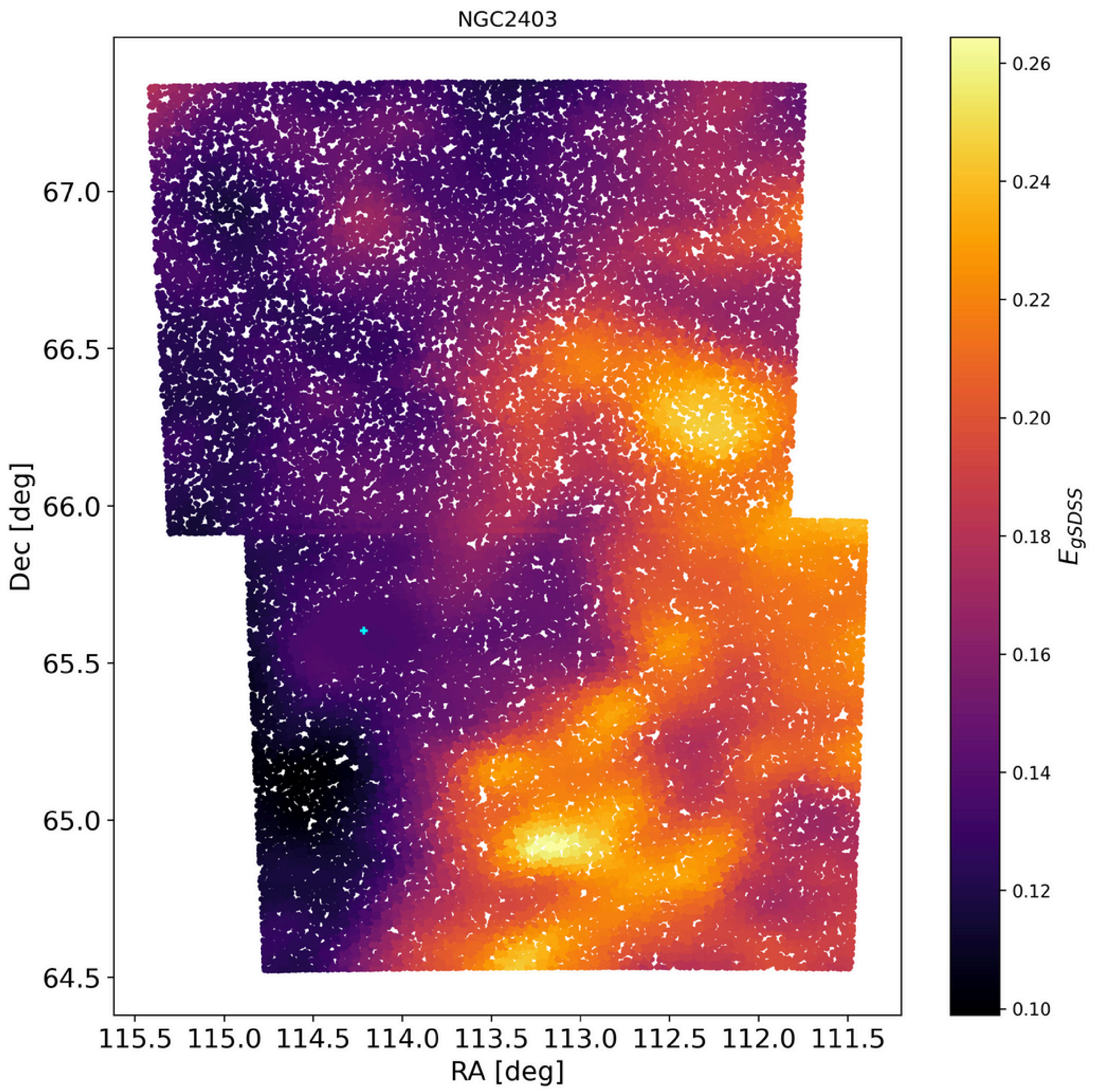
When necessary, we work with images and run photometry ourselves

Photometric Correction

Band	Zero Point	Aperture
rSDSS	0.026501	0.256944
gSDSS	0.031947	0.288879
iSDSS	0.030163	0.255187
zSDSS	0.029848	0.261464
uJAVA	0.040926	0.309303
J0378	0.043606	0.322511
J0395	0.038029	0.317831
J0410	0.026175	0.299767
J0430	0.035034	0.292709
J0515	0.023473	0.268894
J0660	0.033441	0.258988
J0861	0.028965	0.280862

NGC 2403: Main Field

$$\mathrm{Mag_{corr}} = \mathrm{Mag_{orig}} + \text{Zero Point} - \text{Reddening} - \text{Aperture}.$$

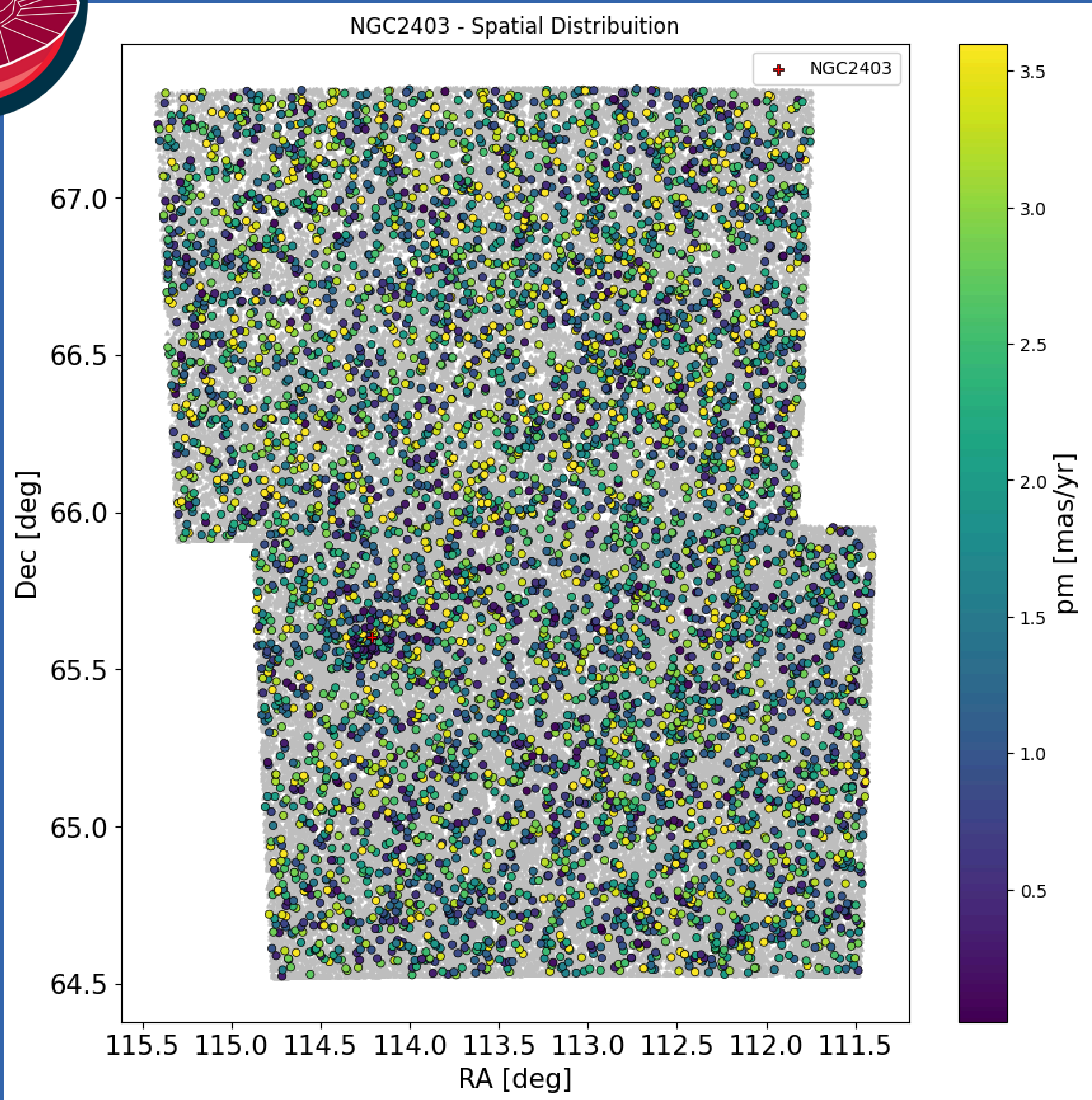


NGC 2403: Extinction Map, gSDSS-filter

Schlegel, Finkbeiner & Davis (1998)

External Data

<u>GAIA EDR3</u>: Astrometric data is used in the identification of non-cluster contaminants and enhancement of candidate selection.

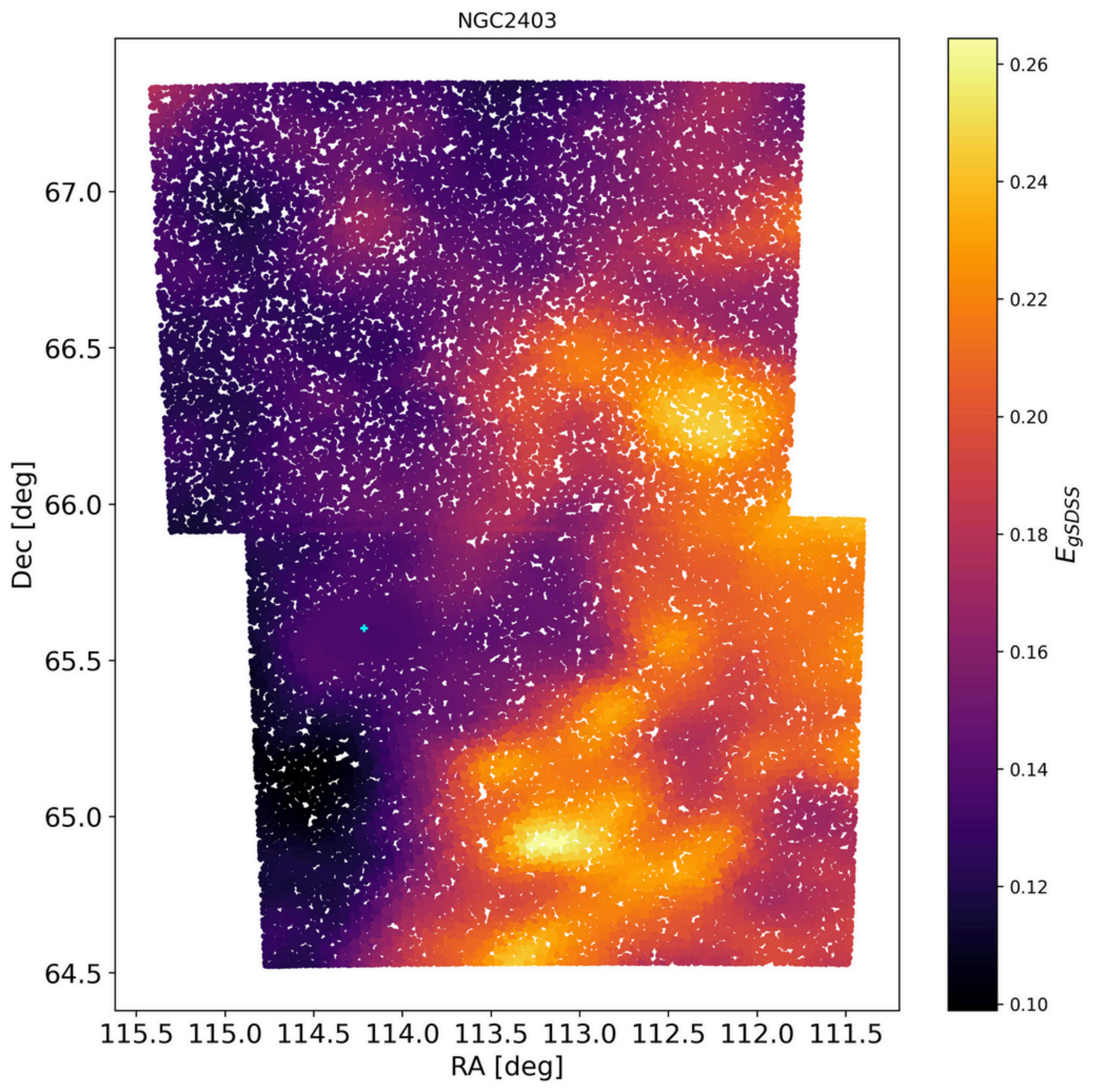


NGC 2403: Candidates - Proper Motion

WHAT DO WE NEED TO WORK

We need Catalogues, and in the following DRs, we need Images to create more complete catalogues and to have visual inspection

Nancy Grace Roman Space Telescope



Extinction Map

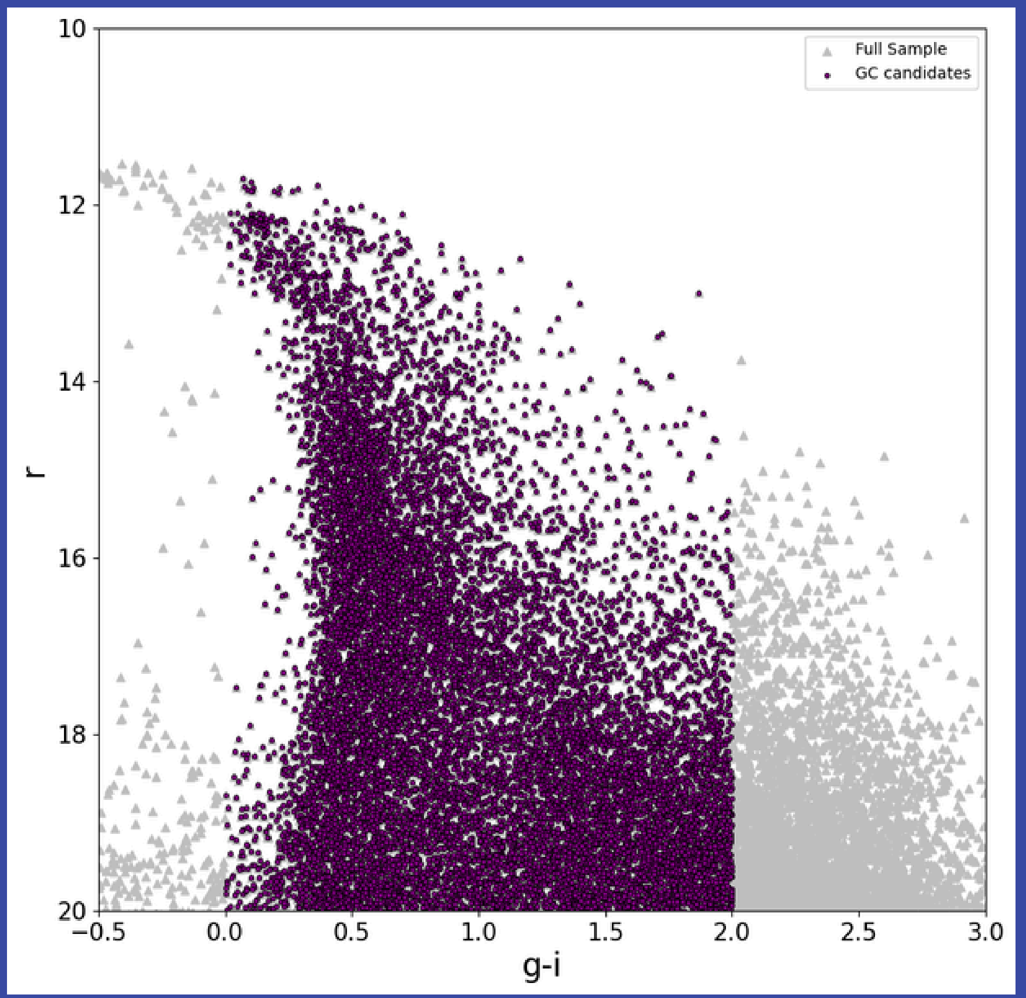
CANDIDATE SELECTION

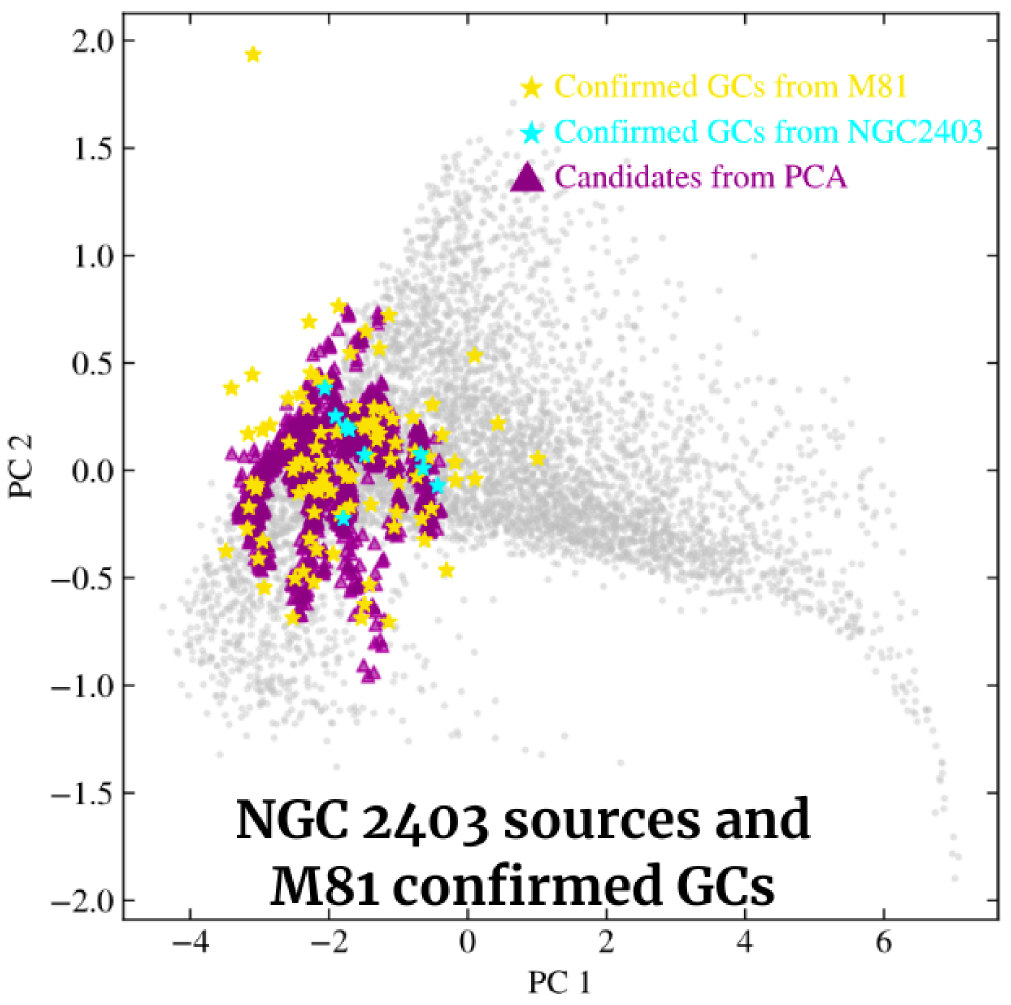
Simple Cuts

The candidates are selected by using singular magnitude and color cuts based on SSP models or spectroscopically confirmed GCs.

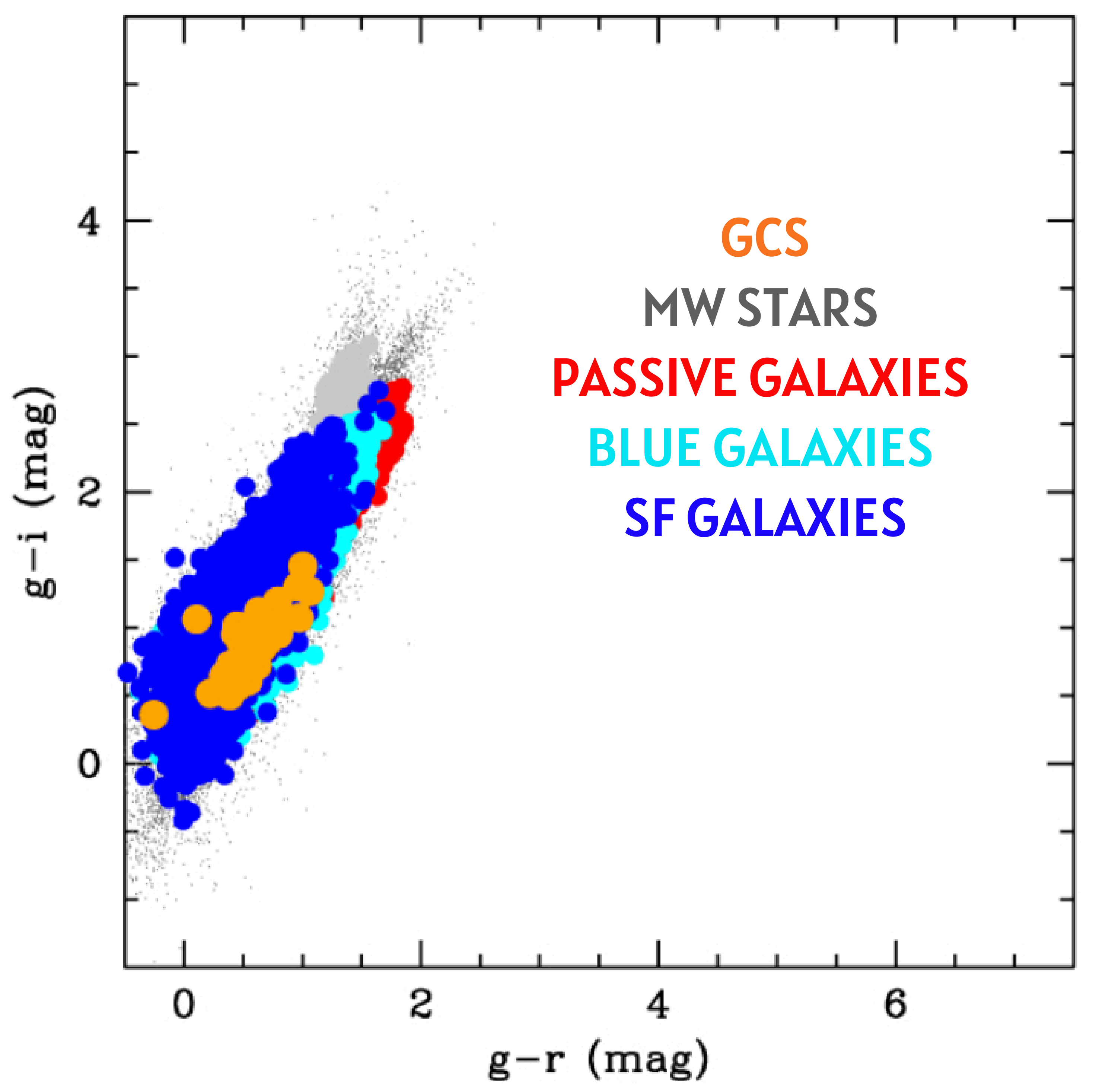
Multi-Dimensional Analysis

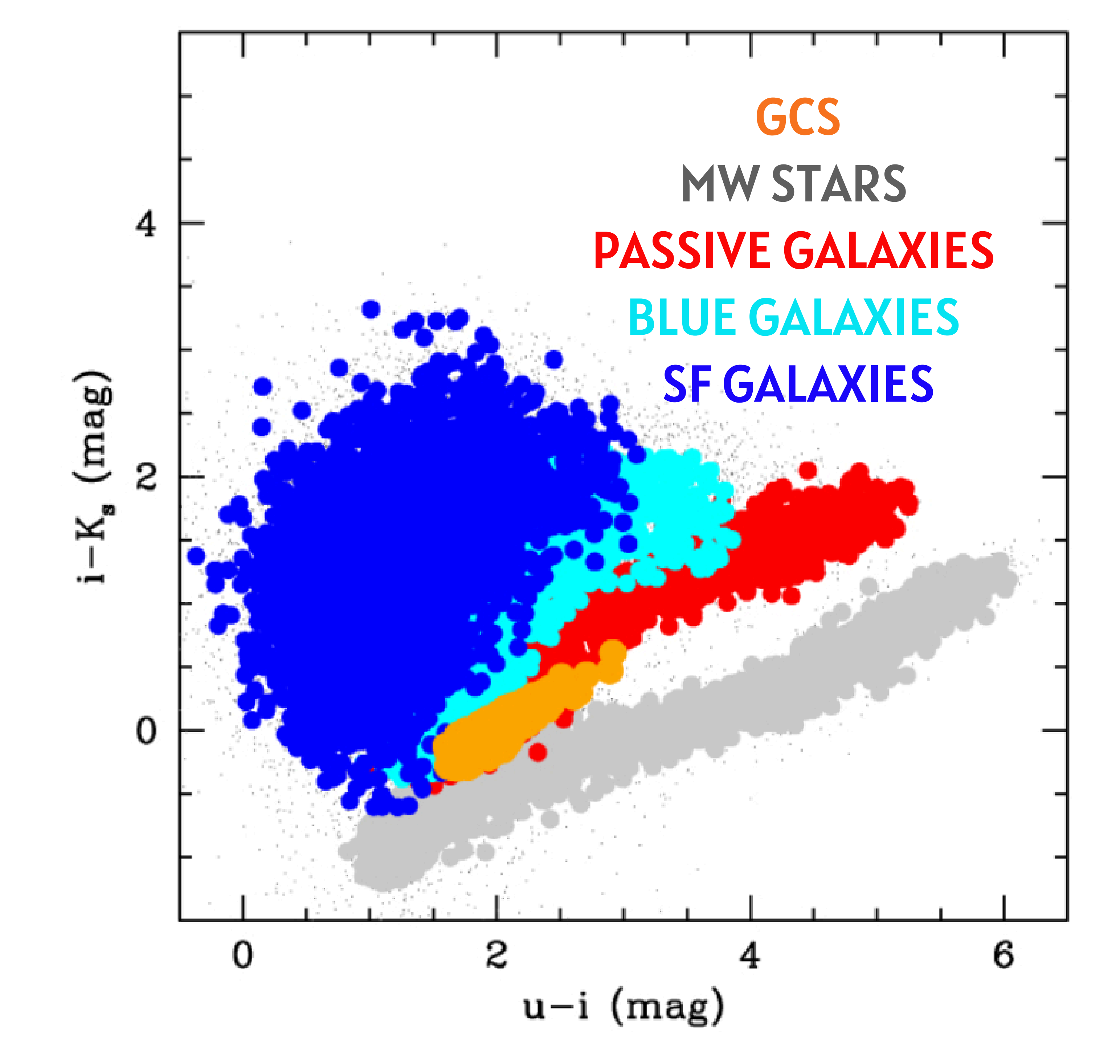
Is there a more clever way to prepare and/or transform the available data to better separate promising GC candidates?





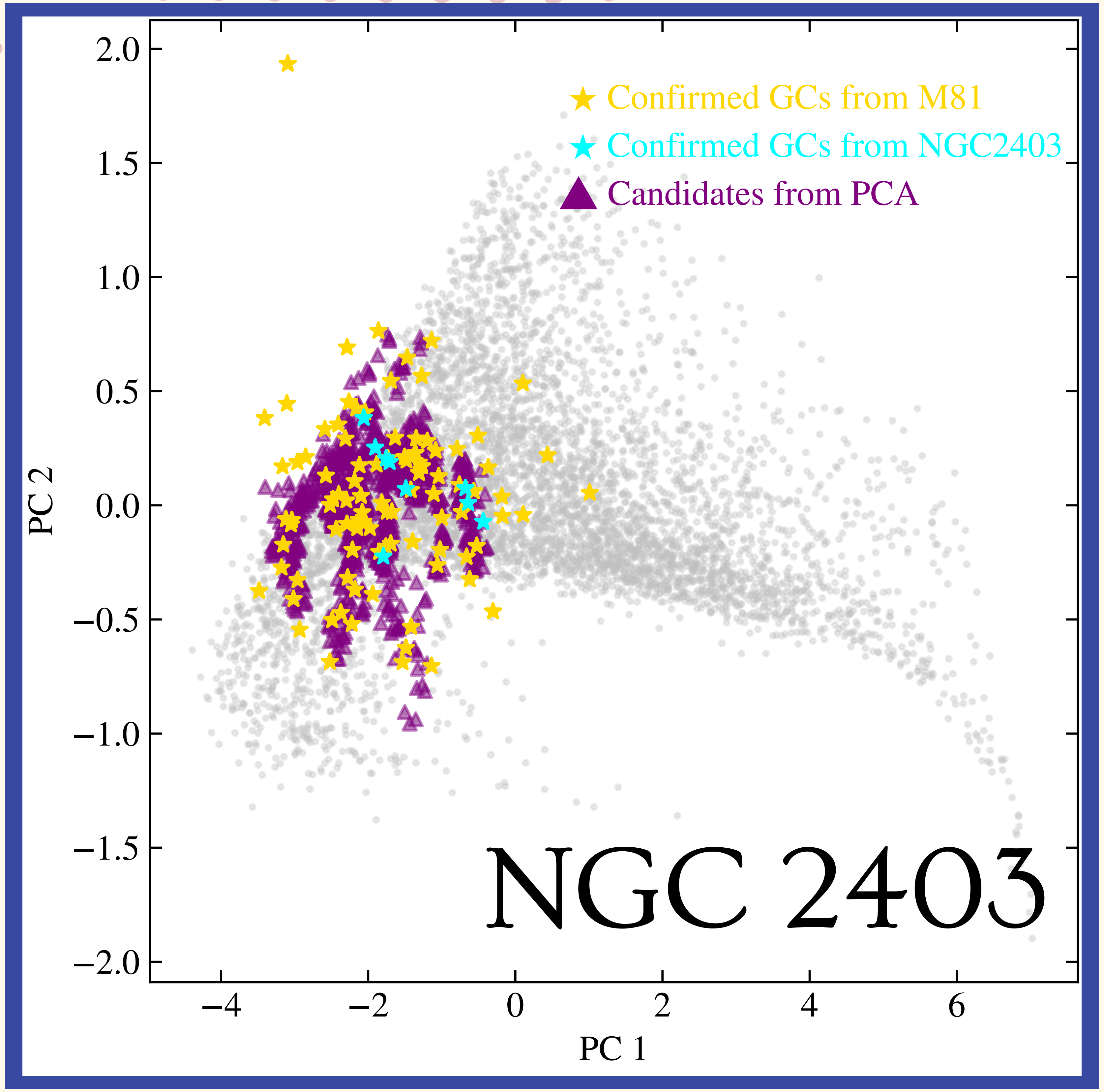
Simple Cuts





Cantiello et al. 2017

Multi-Dimensional Analysis

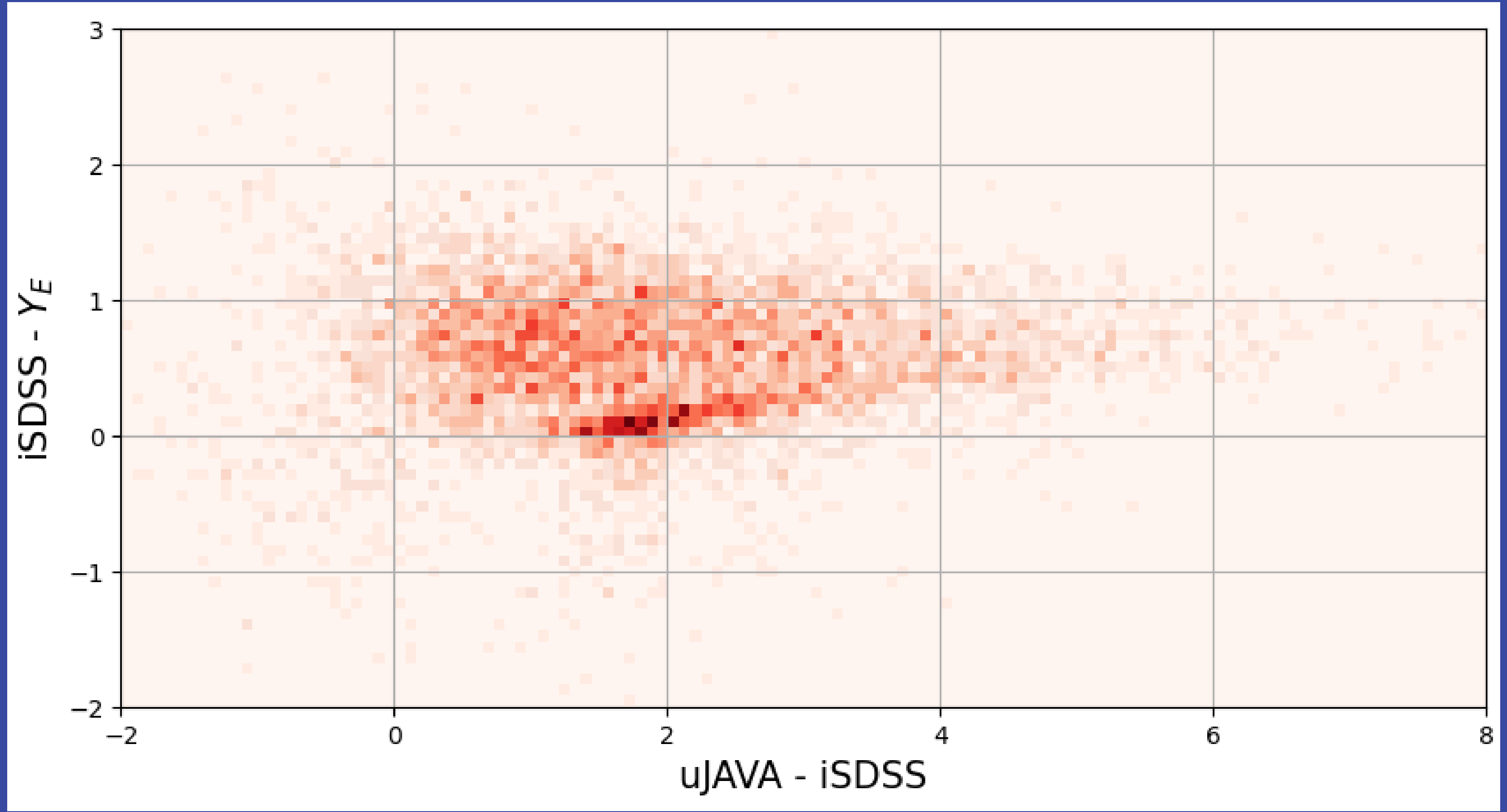


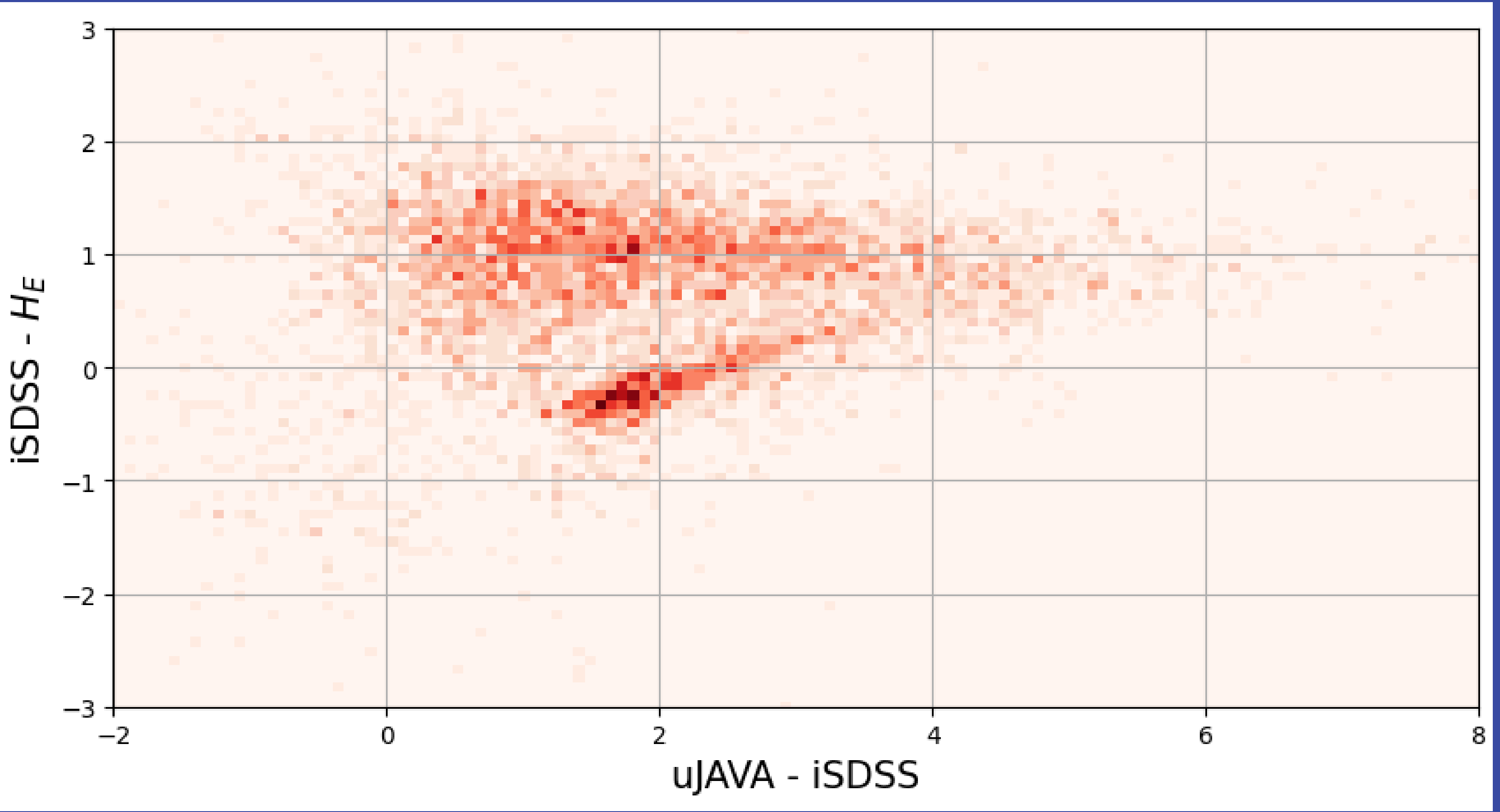
BEYOND COLOR-COLOR DIAGRAMS

- Dimensionality reduction with PCA
- Statistical matching
- Combine labelled data from different galaxies and surveys

J-PLUS + Euclid

NGC 2403



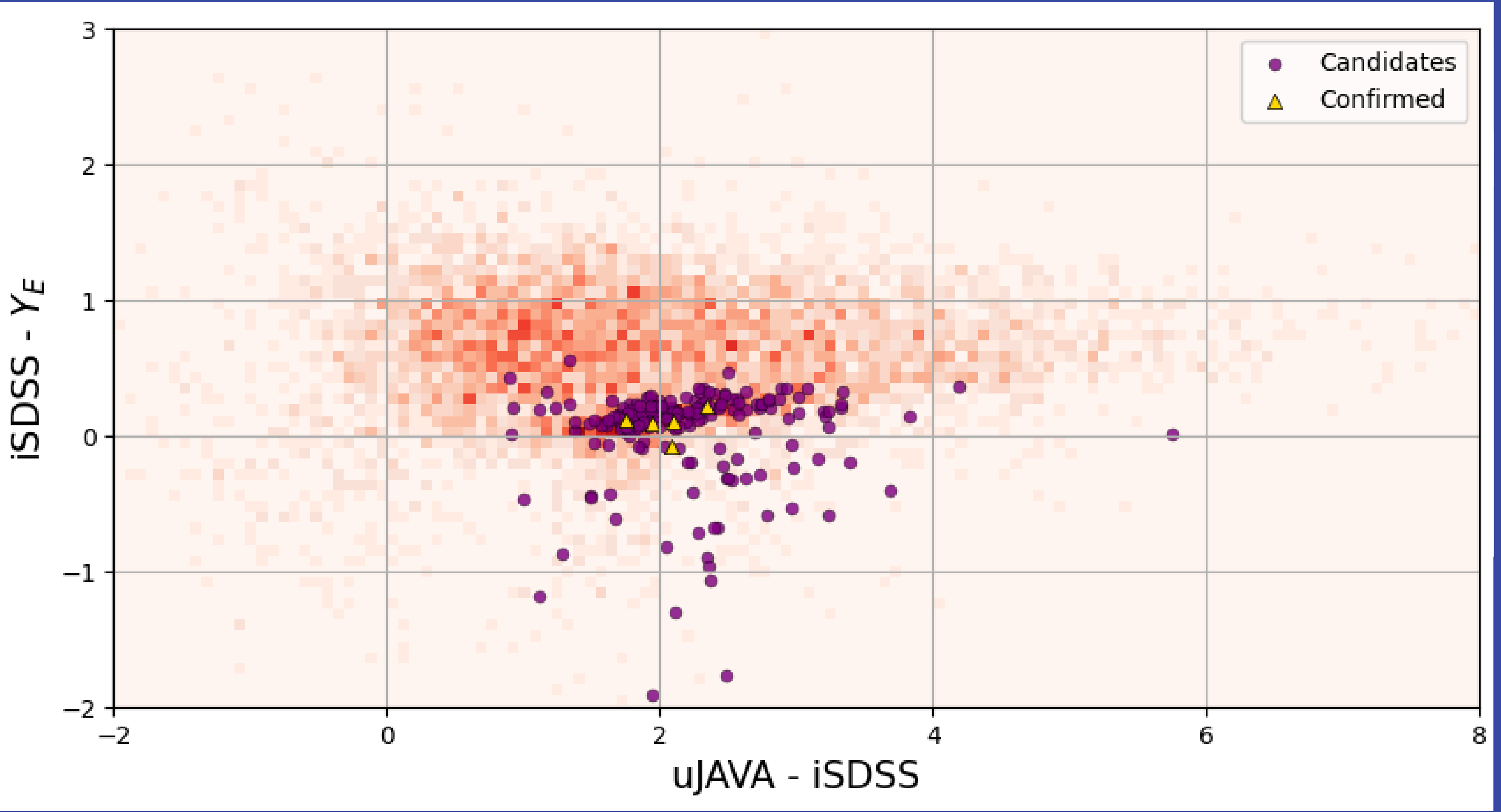


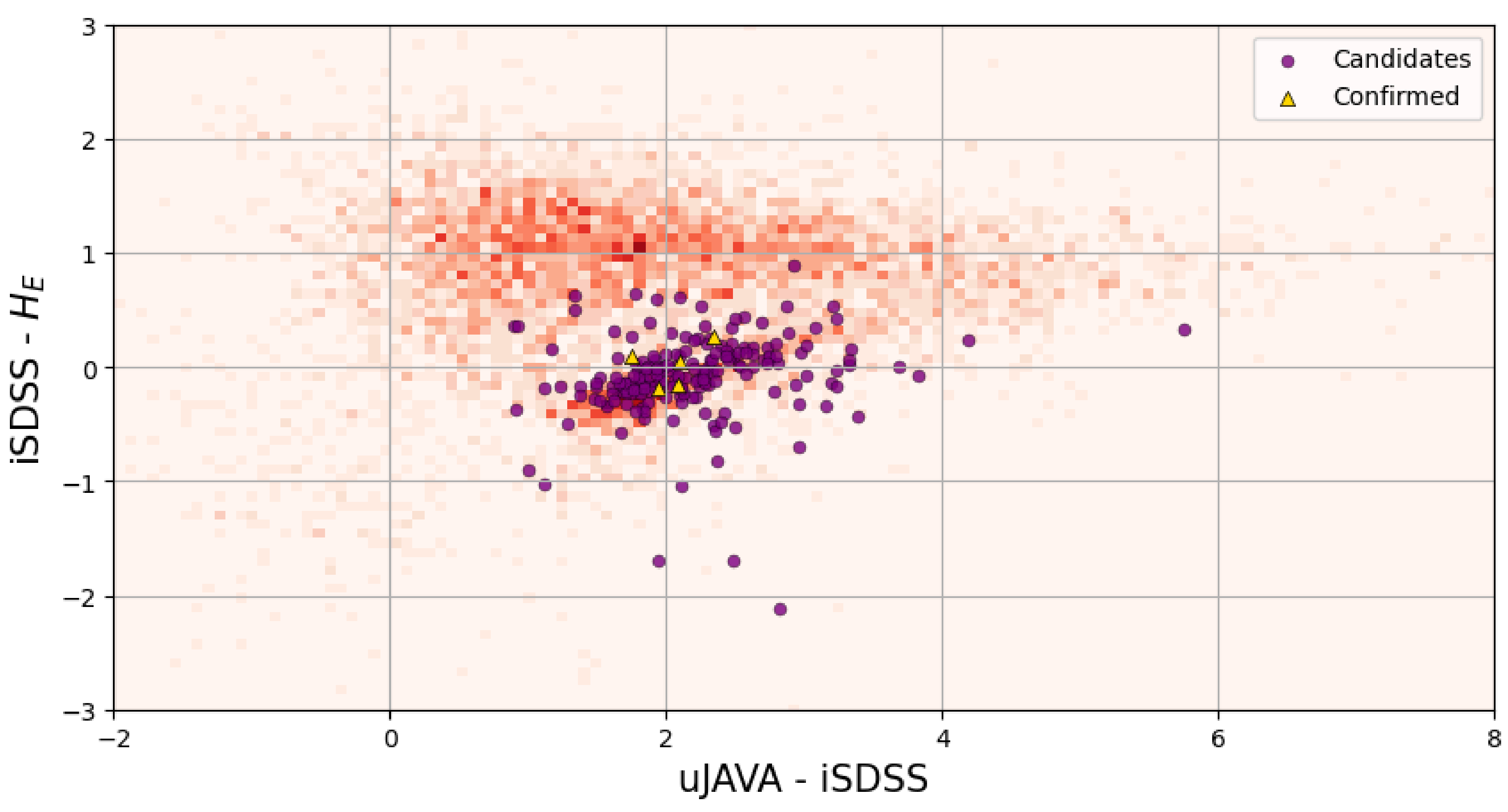
Euclid ERO data for NGC 2403

L. K. Hunt et al. (2024)

J-PLUS + Euclid

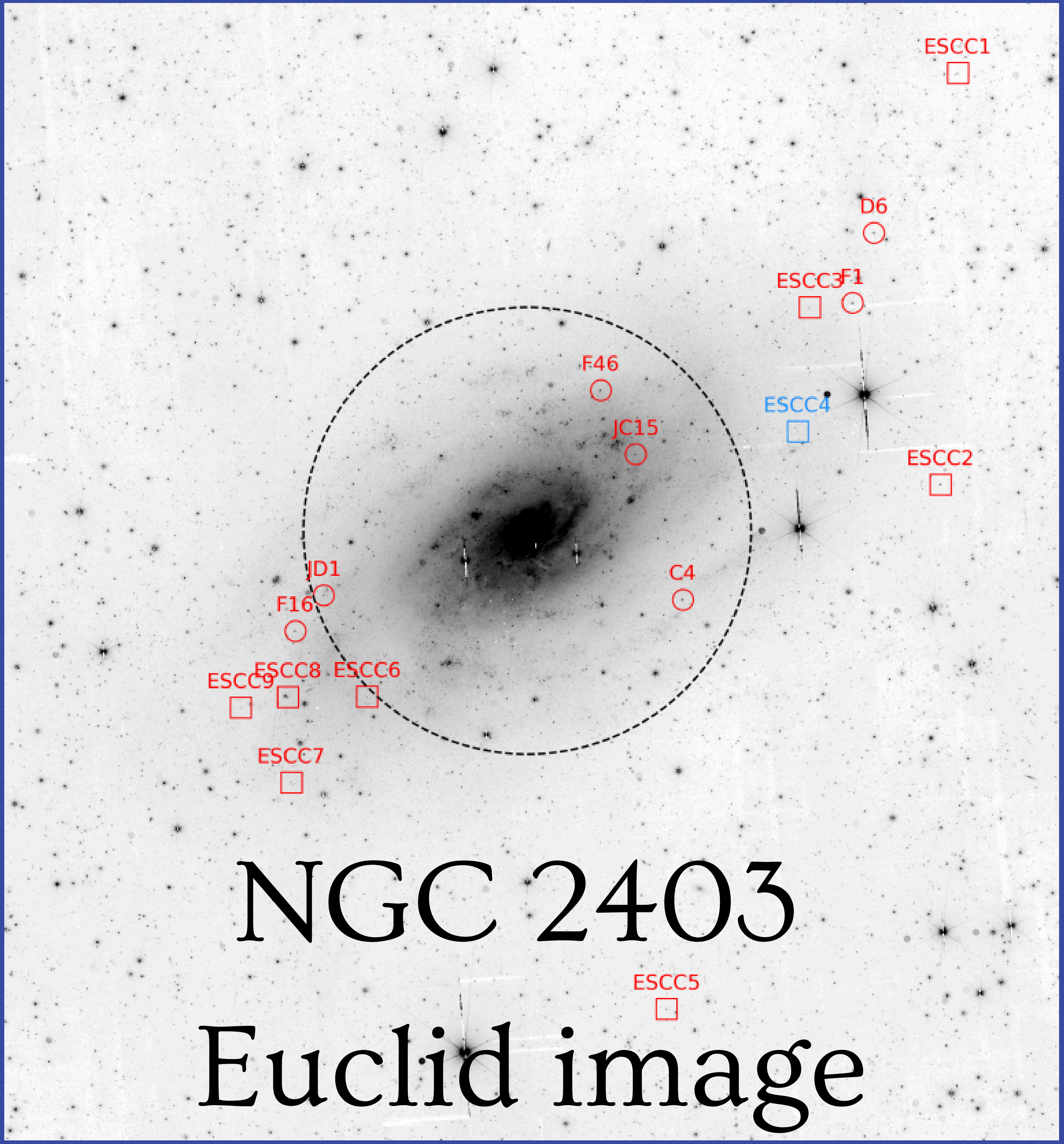
NGC 2403





<u>The candidates (selected without Y or H bands) coincide with an overdensity known to host GCs</u>

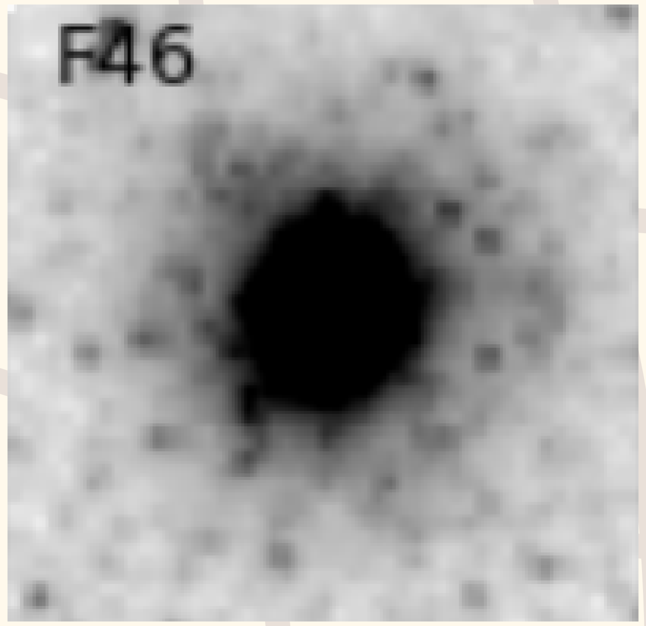
L. K. Hunt et al. (2024)

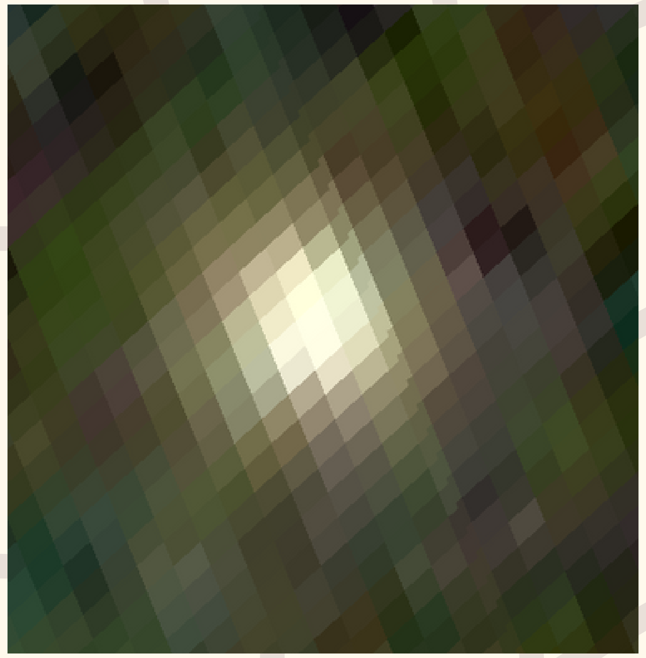


L. K. Hunt et al. (2024)

Visual Inspection

- 72 stars
- 40 galaxies
- 39 potential GCs
- 9 undefined





Euclid vs J-PLUS

Image: J-PLUS, Euclid

LSST-LIKE PHOTOMETRY OF THE FORNAX CLUSTER

- Fornax Deep Survey (FDS): ugri
- Dark Energy Survey (DES): grizY
- LSST: ugrizy
- ~ 1000 spect. conf. GCs

Abbott et al. (2021)
Cantiello et al. (2020)
Pota et al. (2018)
Schuberth et al. (2010)

Multi-Dimensional Analysis

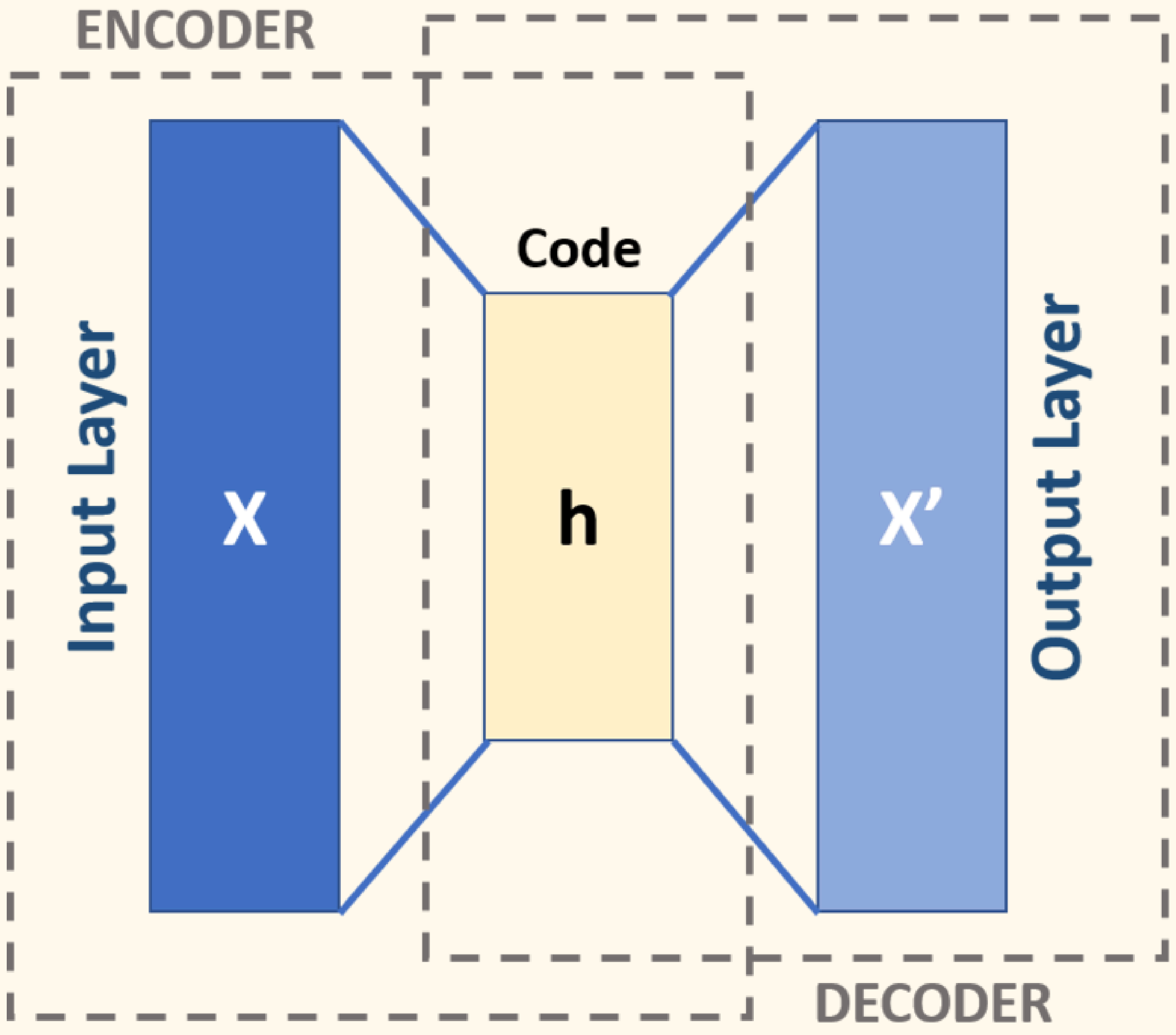
PROSPECTS

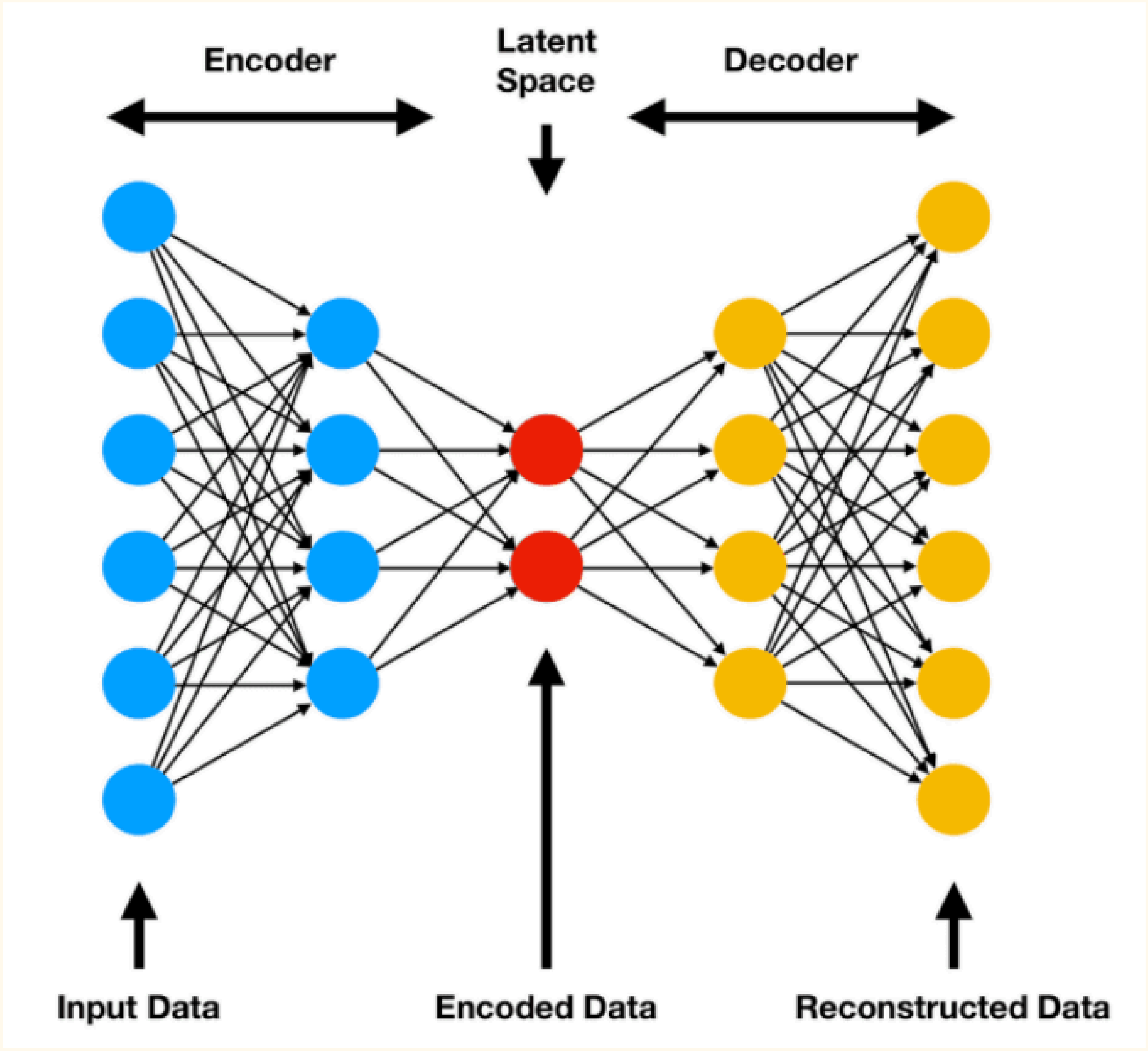
- We are convicted that exploring more sophisticated methods, creating machine learning models, etc. is the way to go.

Multi-Dimensional Analysis

PROSPECTS

- We are convicted that exploring more sophisticated methods, creating machine learning models, etc. is the way to go.





We are making tests with autoencoders!

Schweder-Souza et al. (in prep.)

LSST Discovery Alliance

LINCC Frameworks Incubators

- We wrote a proposal that's being analyzed: we're on the 2nd phase.
- Make our autoencoder model compatible with LSDB and the HATS format in order to select GC candidates over very large amounts of data.

An Automated Tool for Selecting Extragalactic Globular Cluster Candidates for the Rubin/LSST Community

Uncovering and understanding the population of extragalactic globular clusters (GCs) is important for a wide array of research areas in astronomy, from galaxy formation and evolution to dark matter halos to high energy and multi-messenger astrophysics. Globular clusters –densely-packed clusters of 10^5 or more stars– are known to birth unusual high energy and transient astrophysical sources thanks to the dynamical evolution that takes place within their crowded interiors. Unlike the Milky Way, which apparently hosts fewer than 200 globular clusters, galaxies in the nearby Universe can host many thousands of globular clusters; the spatial distribution and properties of the clusters can directly inform on the properties of the dark matter halos and the assembly history of the host galaxy and its environment (e.g. Beasley et al. 2020).

Many extragalactic GC system studies select cluster candidates based on photometric measurements; globular clusters fall into a specific parameter space in terms of magnitude and color. These "color cuts" can be straightforward to implement, but they lead to imperfect selection. Depending on how the selection is carried out and which filters are used, there are substantial rates of contamination from other sources – namely, foreground Galactic stars and compact background galaxies – that have similar magnitudes and colors.

Due to its broad sky coverage, excellent image quality, and accurate photometry in multiple filters across a wide wavelength range, LSST will enable a thorough census of the GC populations of the galaxies that lie in the survey footprint (see e.g. Usher et al. 2023). In particular, LSST's six-filter photometry will help minimize contamination in the GC candidate samples, since color selection in three or more filters allows background galaxies with low to moderate redshifts to be eliminated. LSST's wide wavelength coverage (from ~3200Å to ~1 micron) will also make it possible to more accurately quantify the color distributions and color gradients of the GC populations in the galaxies. The survey will thus enable robust measurements of important quantities like the total number of GCs, the specific frequency of the GC system (i.e., the number of globular clusters normalized by the mass or luminosity of the galaxy), the spatial distribution of the GC population around the galaxy, and the colors of the GC populations (which serve as a useful proxy for metallicity in old stellar populations).

Rubin Observatory will usher in a new era of large-scale survey astrophysics. The challenge of GC identification on a large scale with LSST data is daunting because astronomers interested in star clusters currently lack a general methodology capable of producing very clean extragalactic GC candidate catalogues. To contribute to the community interested in studying star clusters, we envision a selection tool capable of **flagging promising GC candidates within multi-band photometric catalogues with millions of sources**, covering areas as large as an entire galaxy cluster in the nearby universe, utilizing the available data in the most optimal way. **What is needed is a model capable of identifying probable extragalactic GC candidates while maximizing detection completeness and minimizing contamination from non-clusters.** More specifically, our efforts aim at facilitating future investigations about extragalactic GCs **in the context of LSST**, as our potential data products will allow researchers to start their studies with an already filtered list of possible sources of interest. We propose to use AI/machine learning techniques to achieve such a model. As we will handle large amounts of data, we wish to make our model compatible with the LSDB framework: one should be able to input into the model a large catalog written in the HATS format, and/or be able to call our code from within LSDB.

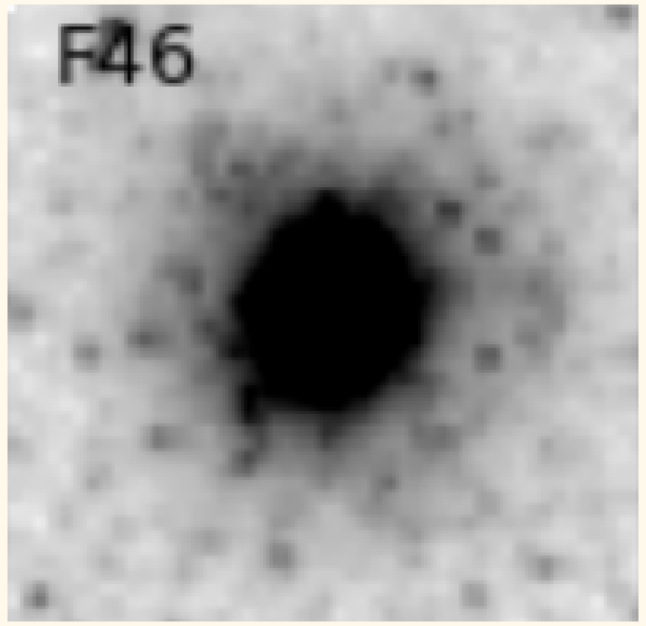
Machine learning model and LSST-like datasets

To achieve our goal, we will construct a model composed of an auto-encoder that will receive multi-band photometric data as input and generate an optimal lower-dimensional embedding (Michelucci 2022), upon which a statistical tool –e.g. propensity score matching (Ho et al. 2007 and Austin 2011)– can be used to select promising GC candidates. We plan to select GC candidates over a combination of data releases already available in the HATS formats, as described in the documentation of LSDB: cross-matching S-PLUS DR4 (dual) and DES DR2 to represent LSST-like photometry (in the bands *ugrizy*) over a very large area of the southern sky. In what regards the training of our model, we will use LSST-like photometry of spectroscopically confirmed GCs in the Fornax Cluster (compilation from Saifollahi et al. 2024 and references therein), which contains galaxies of all types that can be used to test our

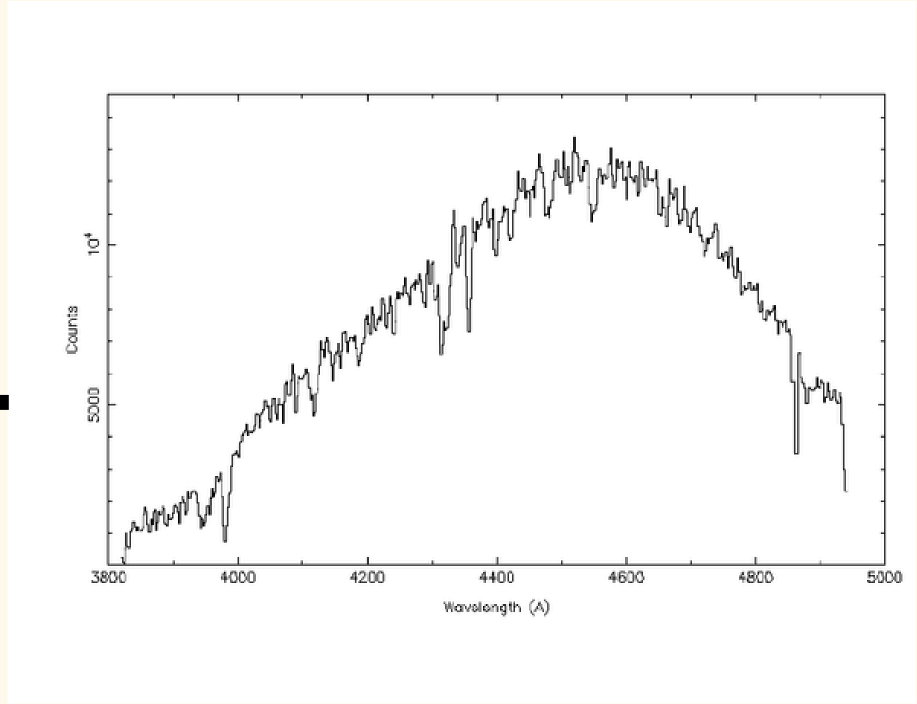
Multi-Dimensional Analysis

PROSPECTS

- Photometry data in the LSST filter-set is inevitably limitated.
  - We wanto to eventually incorporate different types of data into our model to improve it: complementary photometry, images, spectra



+



+

$F275W$	$F336W$	$F475W$	$F814W$
...	...	24.869	25.920
...	26.015	24.718	25.926
22.656	22.578	22.958	23.512
23.235	22.673	22.097	23.865
23.282	22.878	22.416	23.986
...	...	24.827	25.762
23.545	22.882	22.850	24.346
22.766	22.619	22.917	23.720
23.178	22.295	21.837	23.356
...	...	23.712	25.093

COMMUNITY ACCESS

Make our GC candidate finder model available to the community via a web portal.

Hosted by LIneA? Associated with the Brazilian IDAC?

Let's talk!

WHAT DO WE NEED TO WORK

- At first: CATALOGUES
- Then: IMAGES
- Future: Multiple types of data from multiple surveys

jupyter

The Astropy Project

NumPy

pandas

matplotlib

TensorFlow

scikit learn

THANK YOU!

OBRIGADO!

contact:
pedlook@gmail.com (Pedro Lopes)
nicholas9357@gmail.com (Nicholas Schweder)