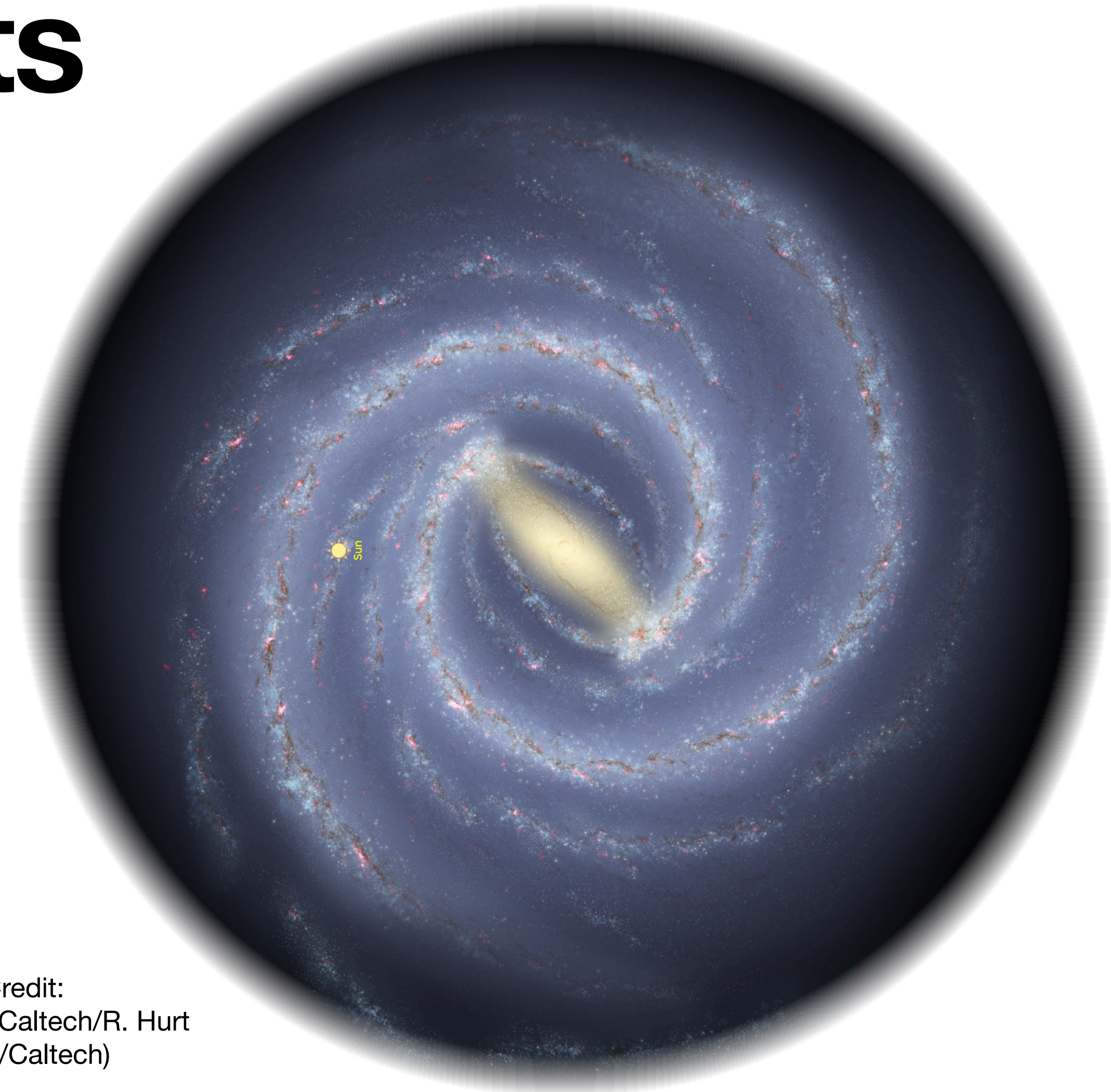# Understanding Astronomical Catalogues: Gaia and its selection function

Alfred Castro Ginard

Credit:
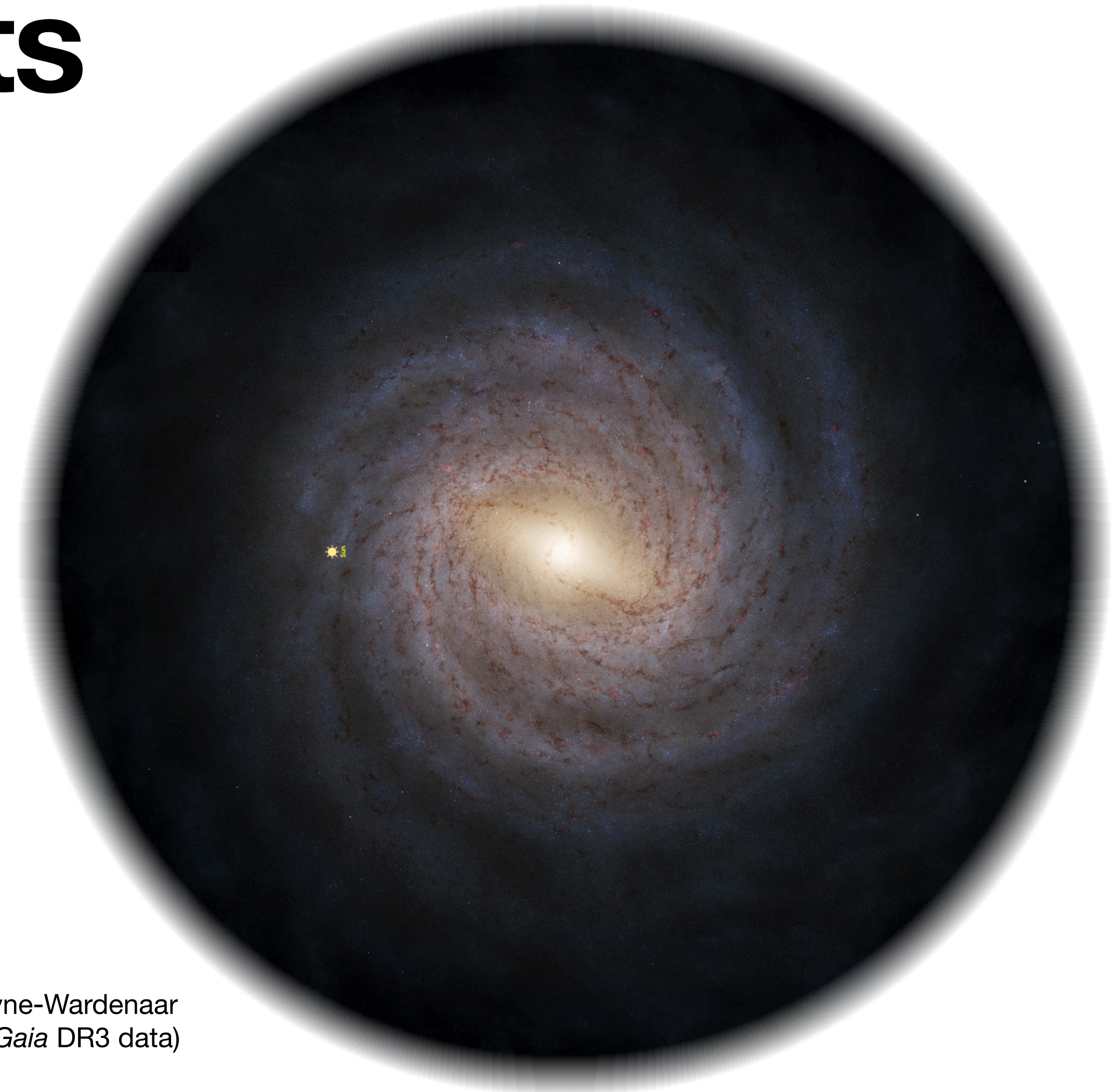NASA/JPL-Caltech/R. Hurt
(SSC/Caltech)

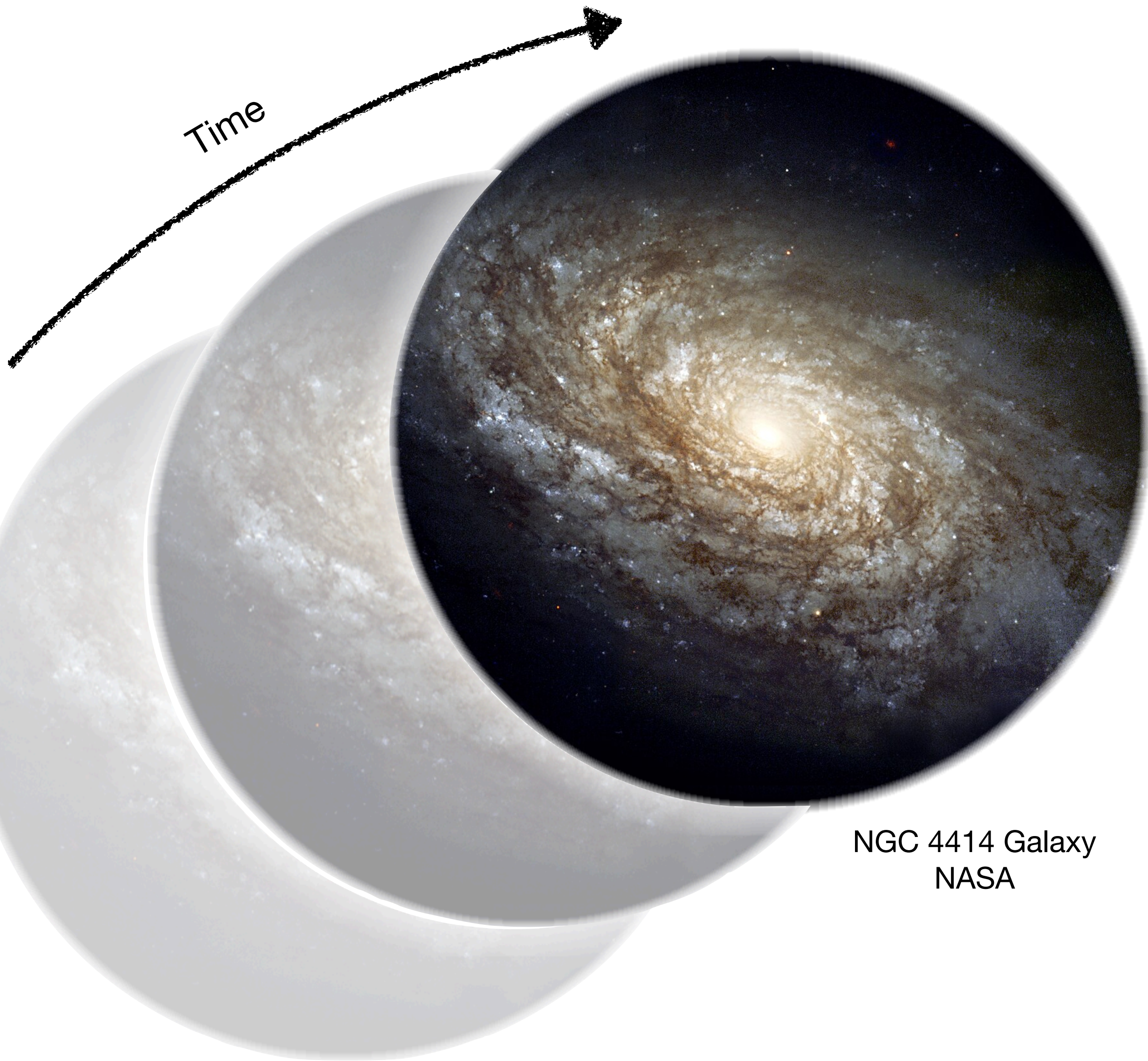# Understanding Astronomical Catalogues: Gaia and its selection function

Alfred Castro Ginard

Stefan Payne-Wardenaar
(including *Gaia* DR3 data)

What type of galaxy is the Milky Way?

Time

NGC 4414 Galaxy
NASA

Whirlpool Galaxy
NASA

Where is the Milky Way?

We don't know our Galaxy's true number of spiral arms, their location or how and why they formed.
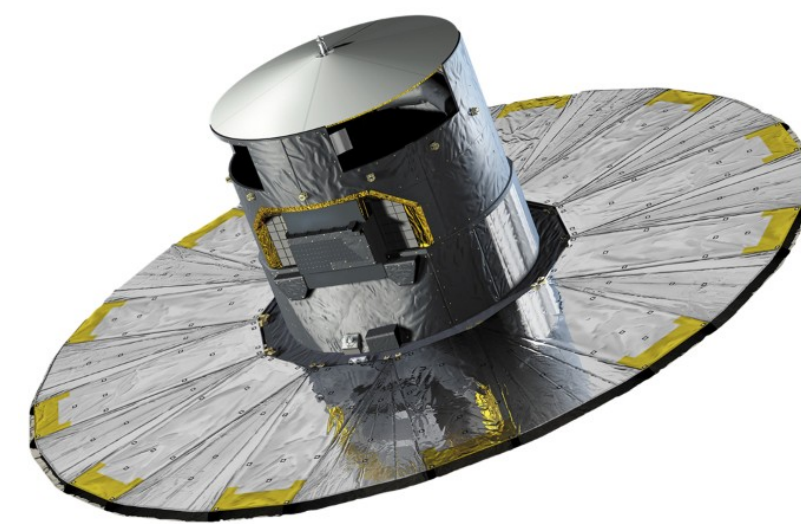
The *Gaia* revolution

Gaia is being transformative to astronomy by measuring the positions and motions of **over a billion stars in our Galaxy**.
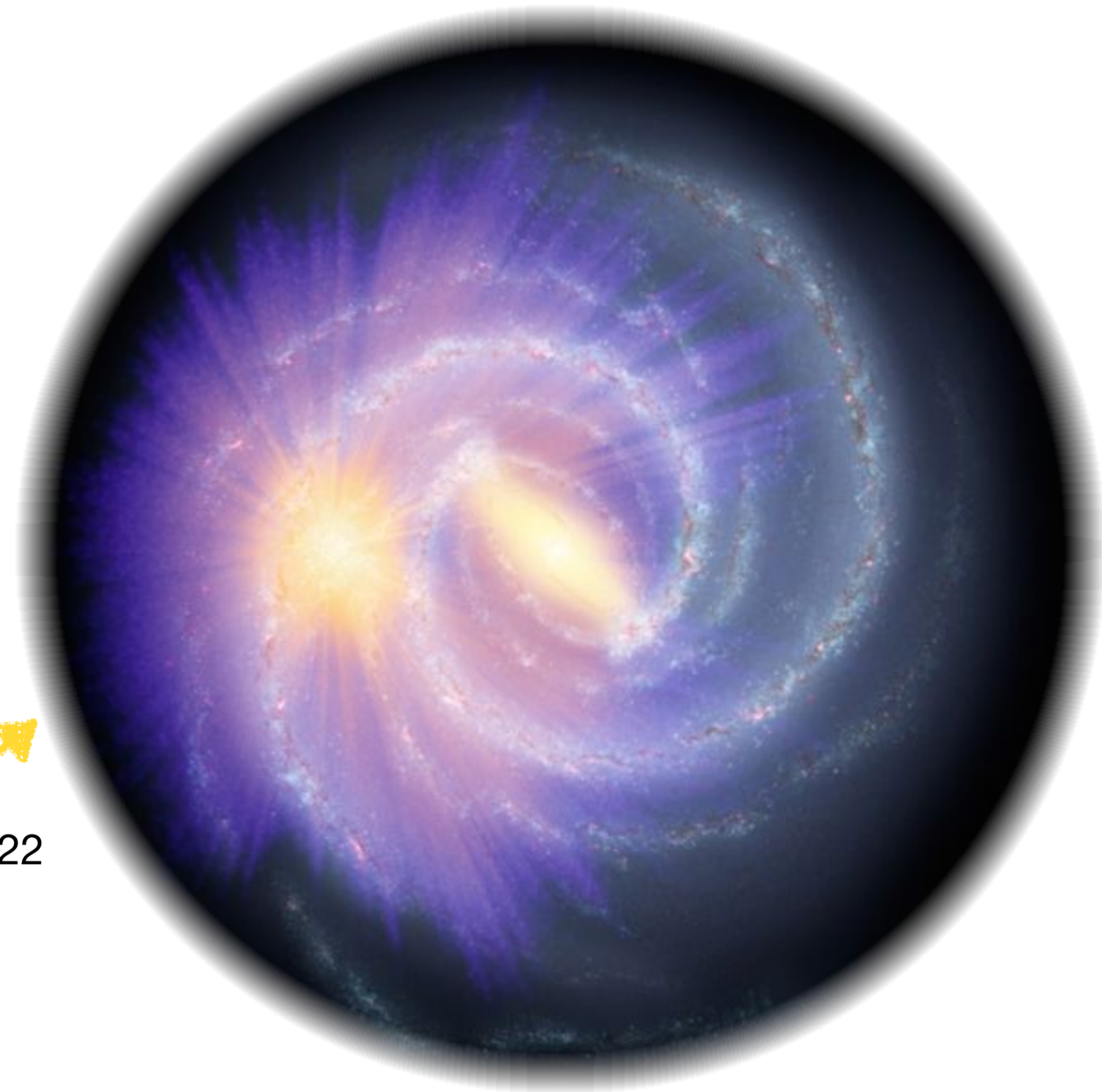
We see the Milky Way from within, so we need the aid machine-learning to analyse these data and understand what does it mean for our Galaxy.
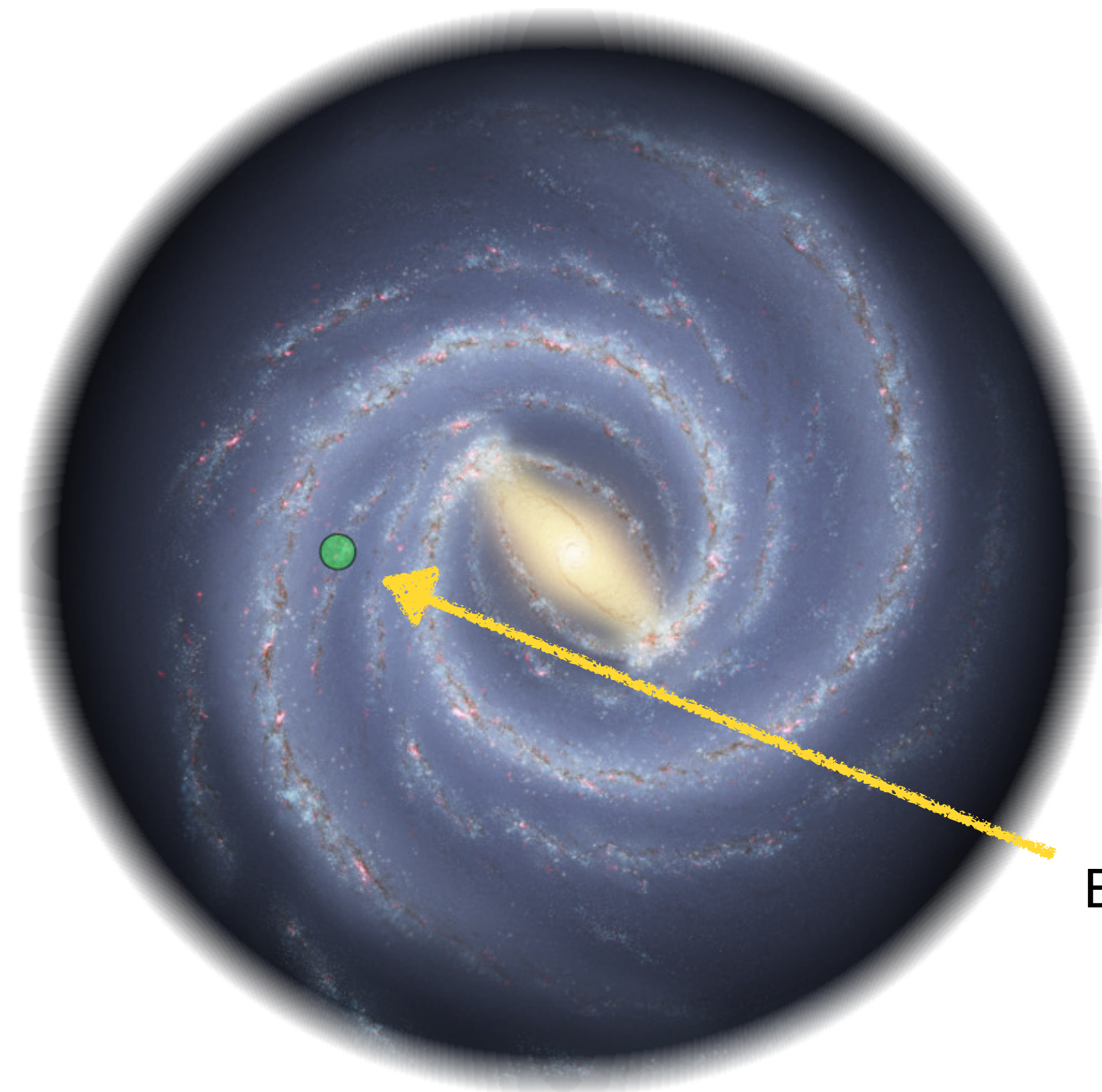
Before *Gaia*

Gaia DR3, 2022

The *Gaia* revolution (but…)

Gaia's view of the sky, i.e. how we see our Galaxy from the inside.

Need reliable tracers of the Galaxy's structure to make sense of the huge amount of data provided by astronomical surveys.

The *Gaia* revolution (but…)

Gaia's view of the sky, i.e. how we see our Galaxy from the inside.

Need reliable tracers of the Galaxy's structure to make sense of the huge amount of data provided by astronomical surveys.

The picture includes the three closest stars to Earth:

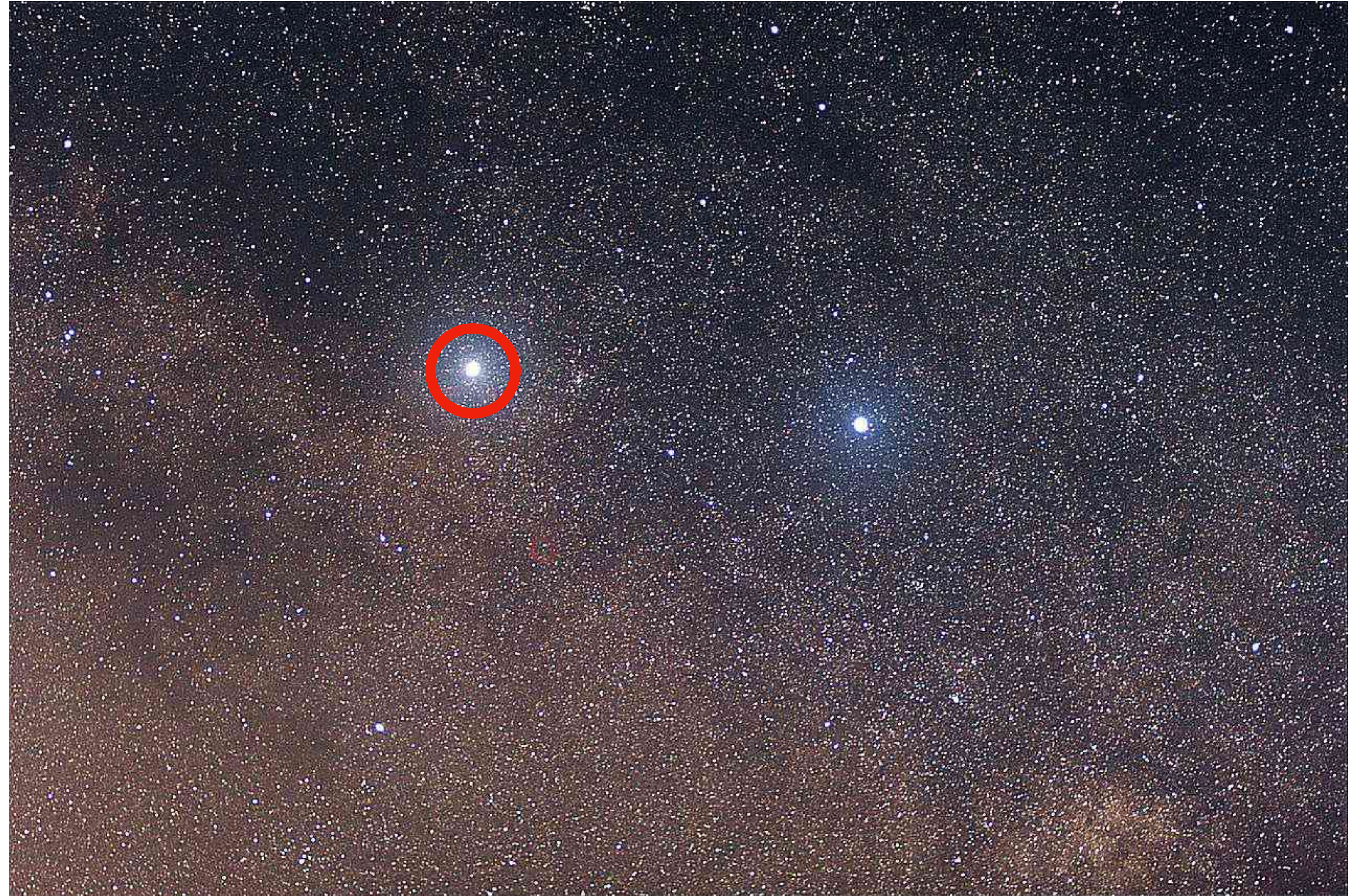- Proxima Centauri
- Alpha Centauri A
- Alpha Centauri B

The *Gaia* revolution (but...)

Gaia's view of the sky, i.e. how we see our Galaxy from the inside.

Need reliable tracers of the Galaxy's structure to make sense of the huge amount of data provided by astronomical surveys.

The picture includes the three closest stars to Earth:

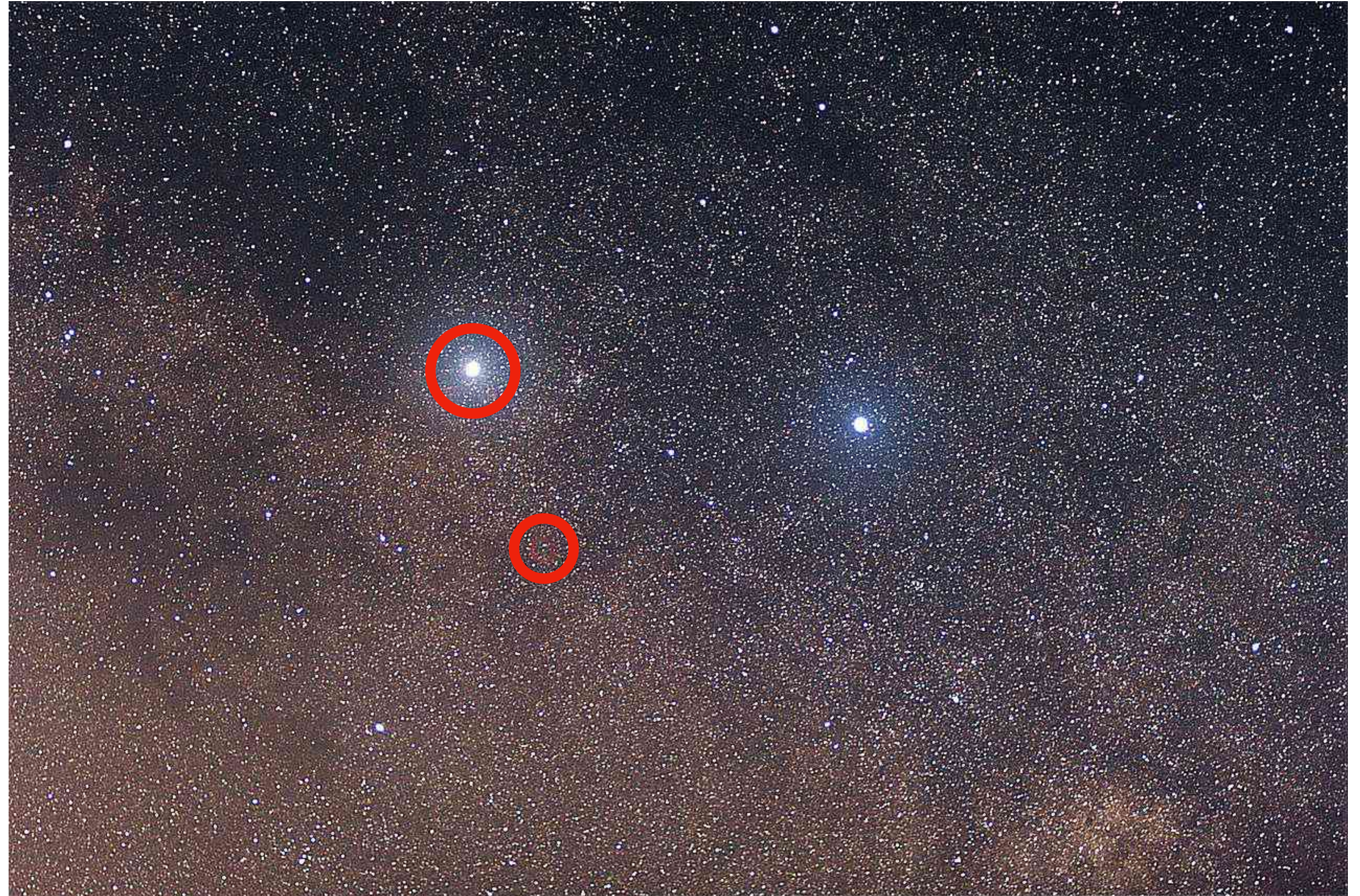- Proxima Centauri
- Alpha Centauri A
- Alpha Centauri B

The *Gaia* revolution (but…)

Gaia's view of the sky, i.e. how we see our Galaxy from the inside.

Need reliable tracers of the Galaxy's structure to make sense of the huge amount of data provided by astronomical surveys.

The picture includes the three closest stars to Earth:

- Proxima Centauri
- Alpha Centauri A
- Alpha Centauri B

How can we make it better? Use OC as main tracers for Galactic structure

Open clusters are gravitationally bound stellar groupings, born at the same event from the same molecular gas cloud

Stars in these typically young structures ($\lesssim$ 1 Gyr) share common location, velocity, age, initial chemical compostition, etc.

In terms of *Gaia* observables they can be described in an 9 dimensional space:

- Stellar overdensities in $(l, b, \varpi, \mu_{\alpha^*}, \mu_\delta, V_{rad})$

- Follow an isochrone pattern in a CMD, (G, G_{BP}, G_{RP})

Smithsonian Magazine

Detection of OCs in Gaia

PI development of OCfinder. Set the foundation for later methods for OC search in *Gaia*
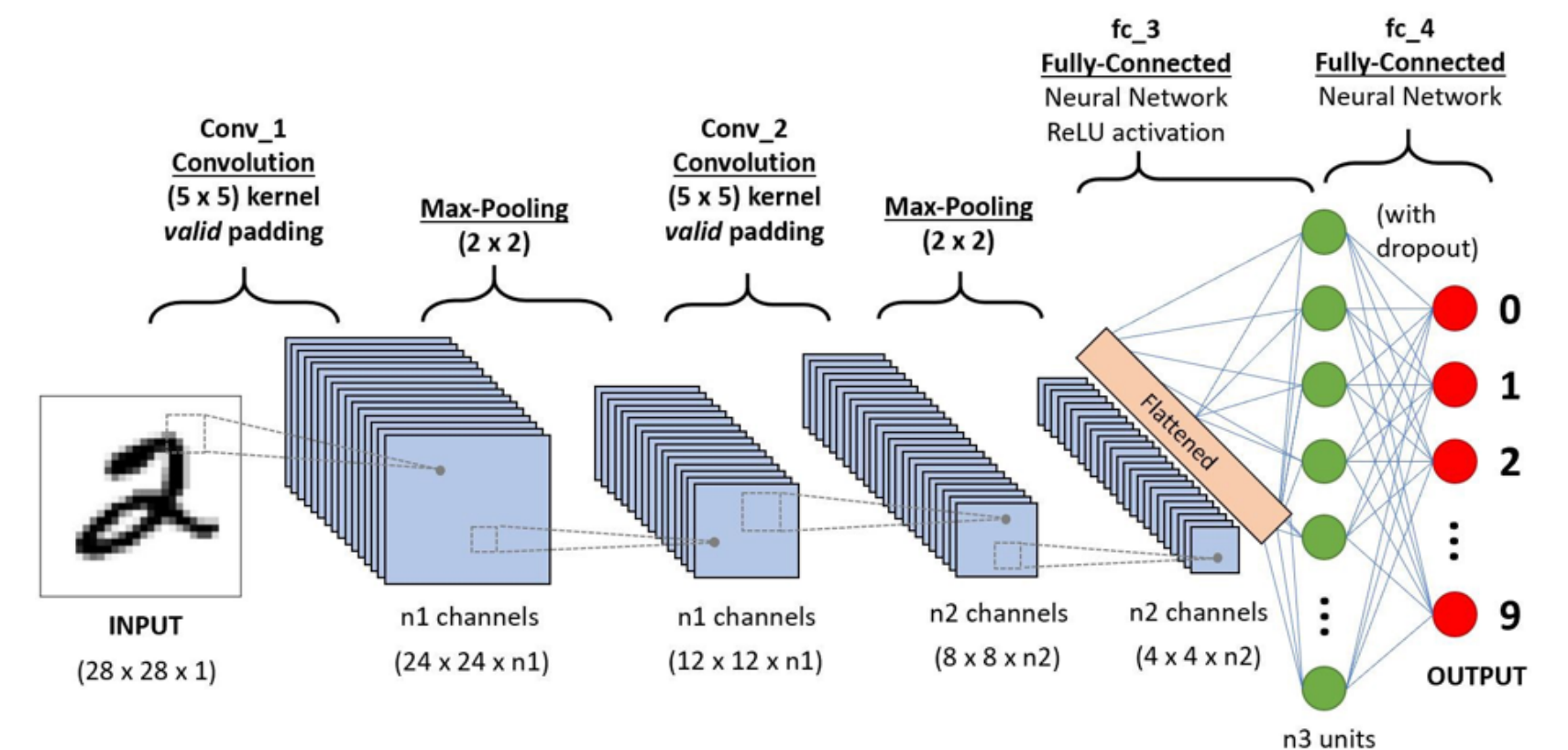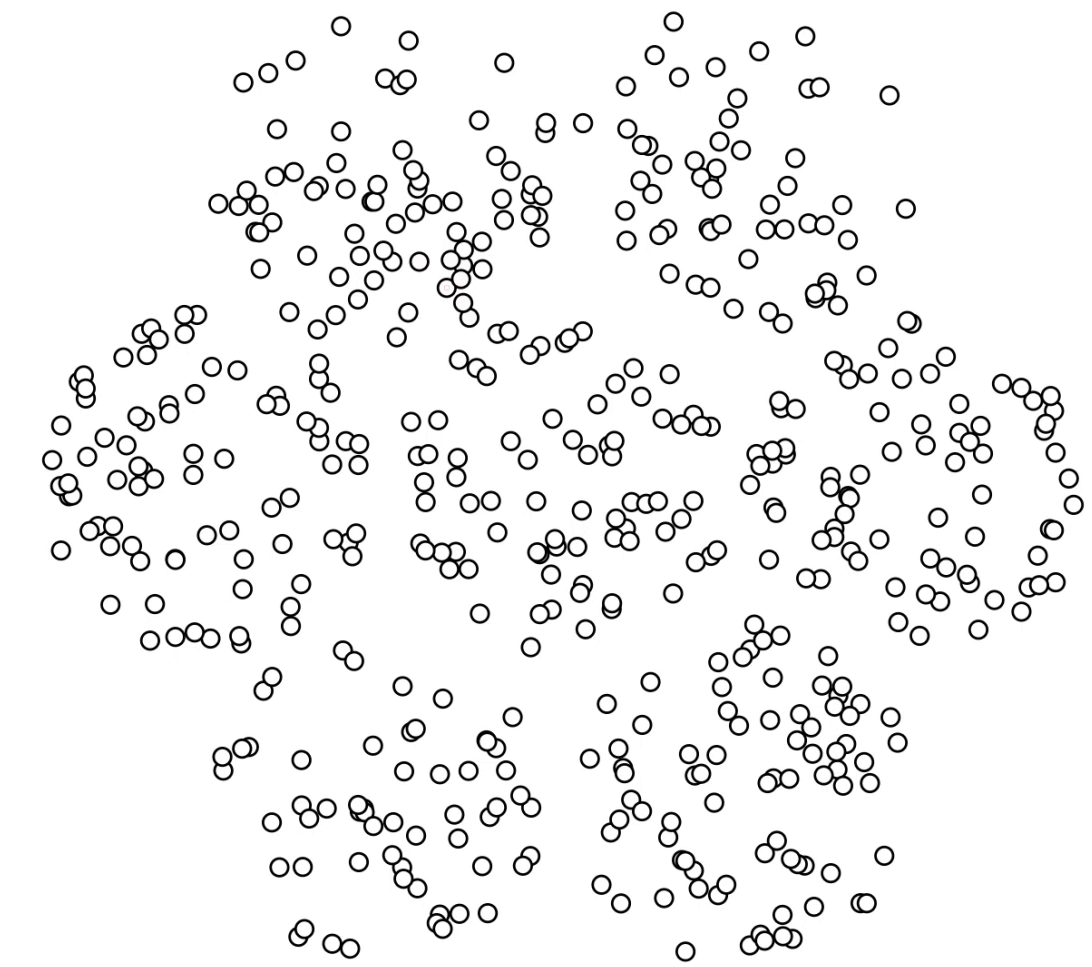

- Clustering on 5D *Gaia* data $\mathcal{O}(10^8)$ stars

MareNostrum at BSC

- Pattern recognition on photometry

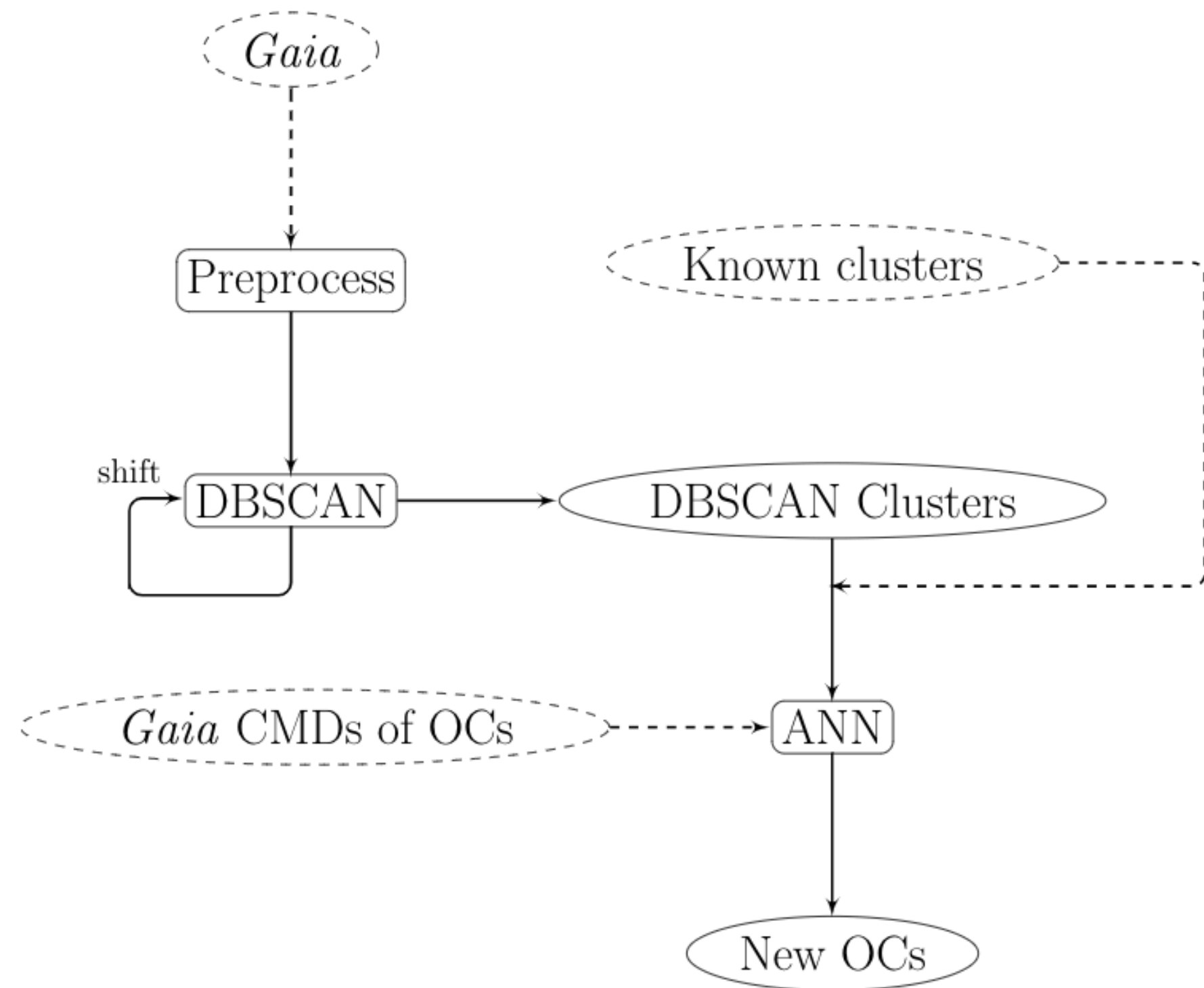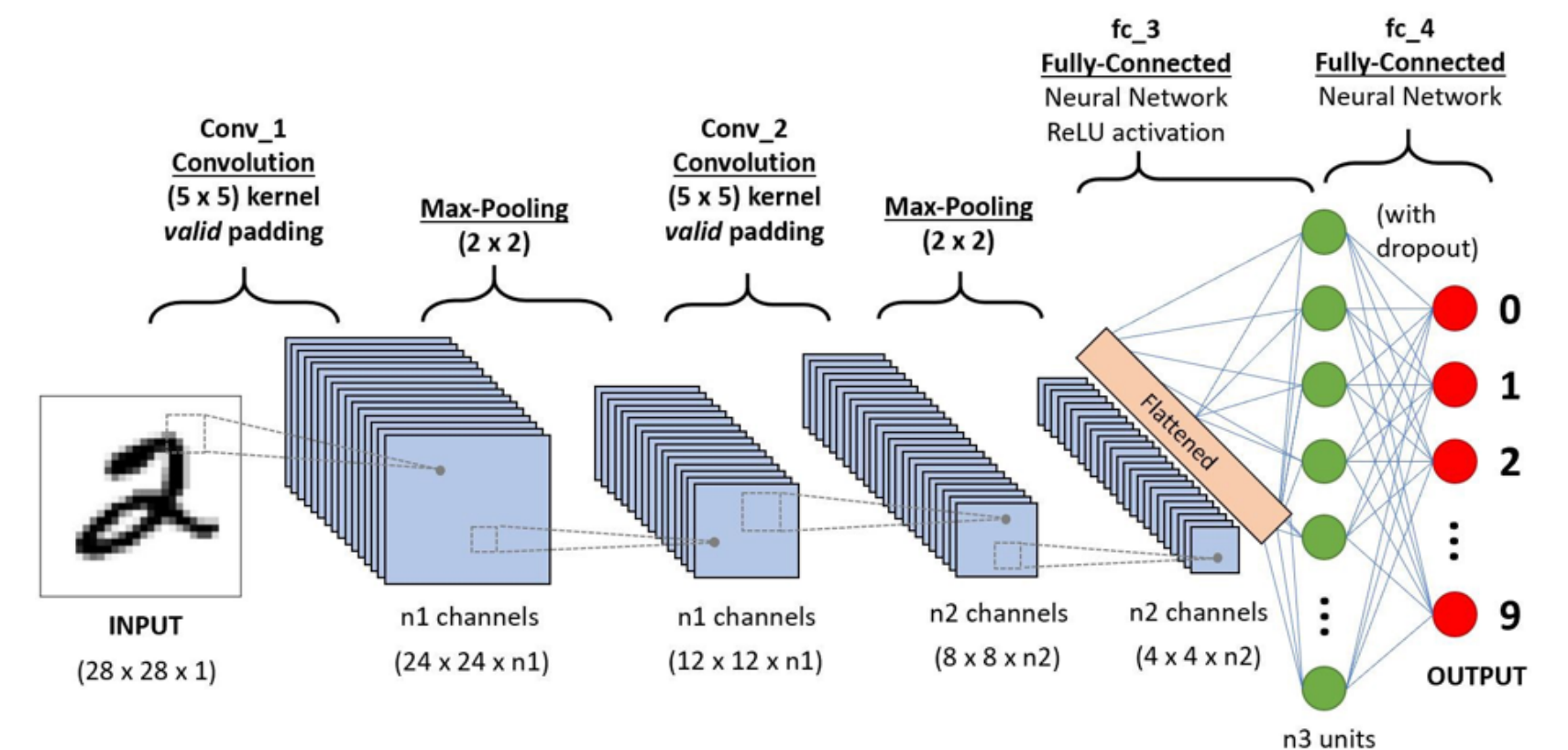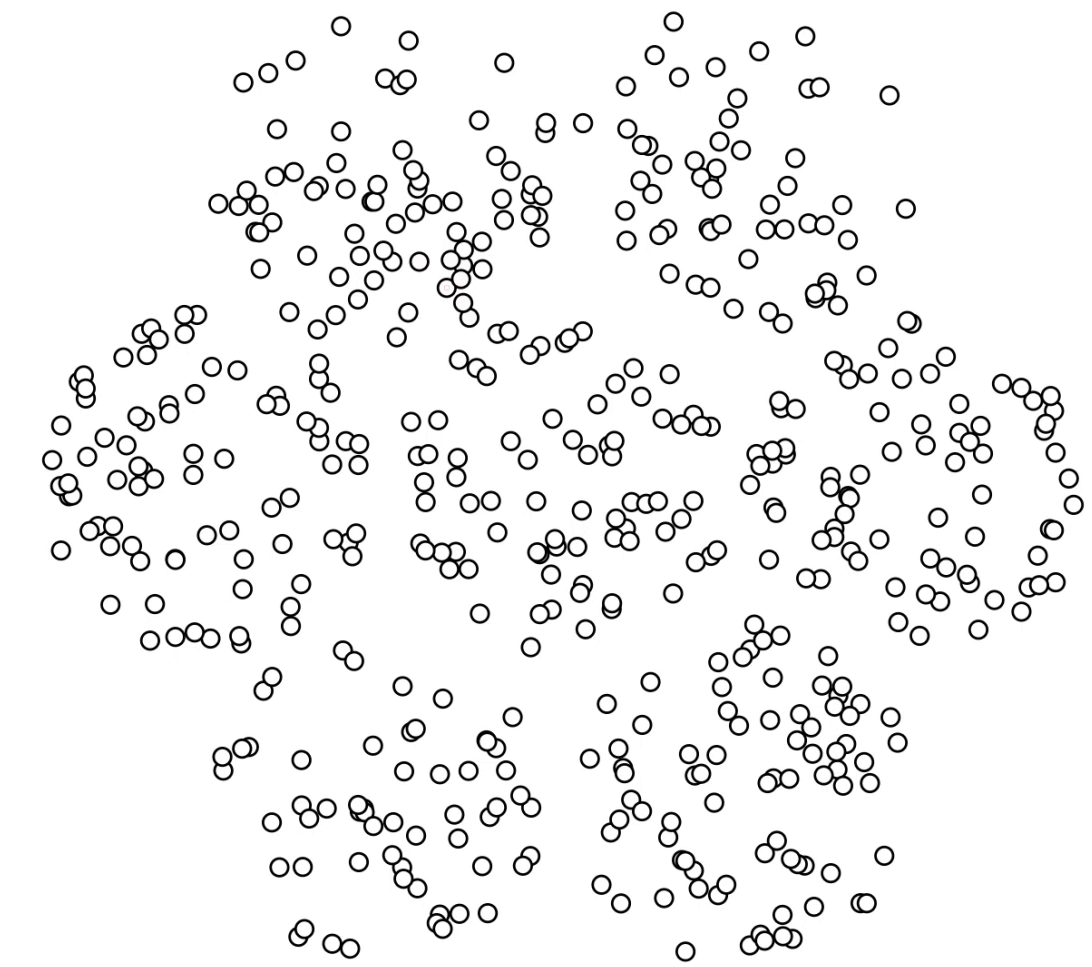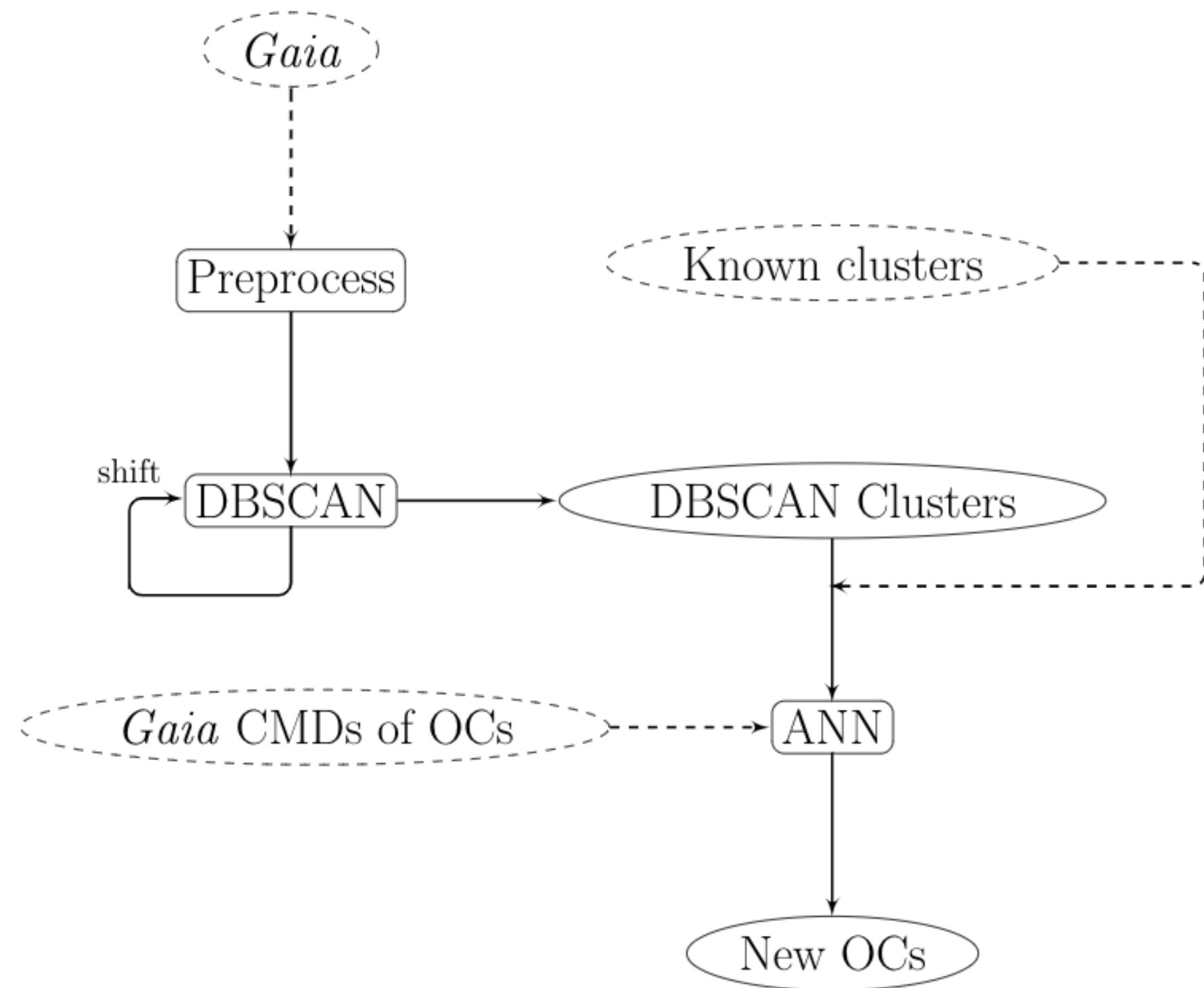[Castro-Ginard+18,19,20,22]

Detection of OCs in Gaia

PI development of OCfinder. Set the foundation for later methods for OC search in *Gaia*

- Clustering on 5D *Gaia* data $\mathcal{O}(10^8)$ stars

MareNostrum at BSC

- Pattern recognition on photometry

[Castro-Ginard+18,19,20,22]

Detection of OCs in Gaia

PI development of OCfinder. Set the foundation for later methods for OC search in *Gaia*

[Castro-Ginard+18,19,20,22]

• About 650 new UBC clusters (DR2) **[Castro-Ginard+18,19,20]**

    • 23 OC found in DR1 and confirmed in DR2.

    • 53 OCs in *Gaia* DR2 in the direction of the Galactic anticentre.

    • 582 new OCs in the whole Galactic disc (Big Data search)

• About 650 more new UBC clusters (EDR3) **[Castro-Ginard+22]**

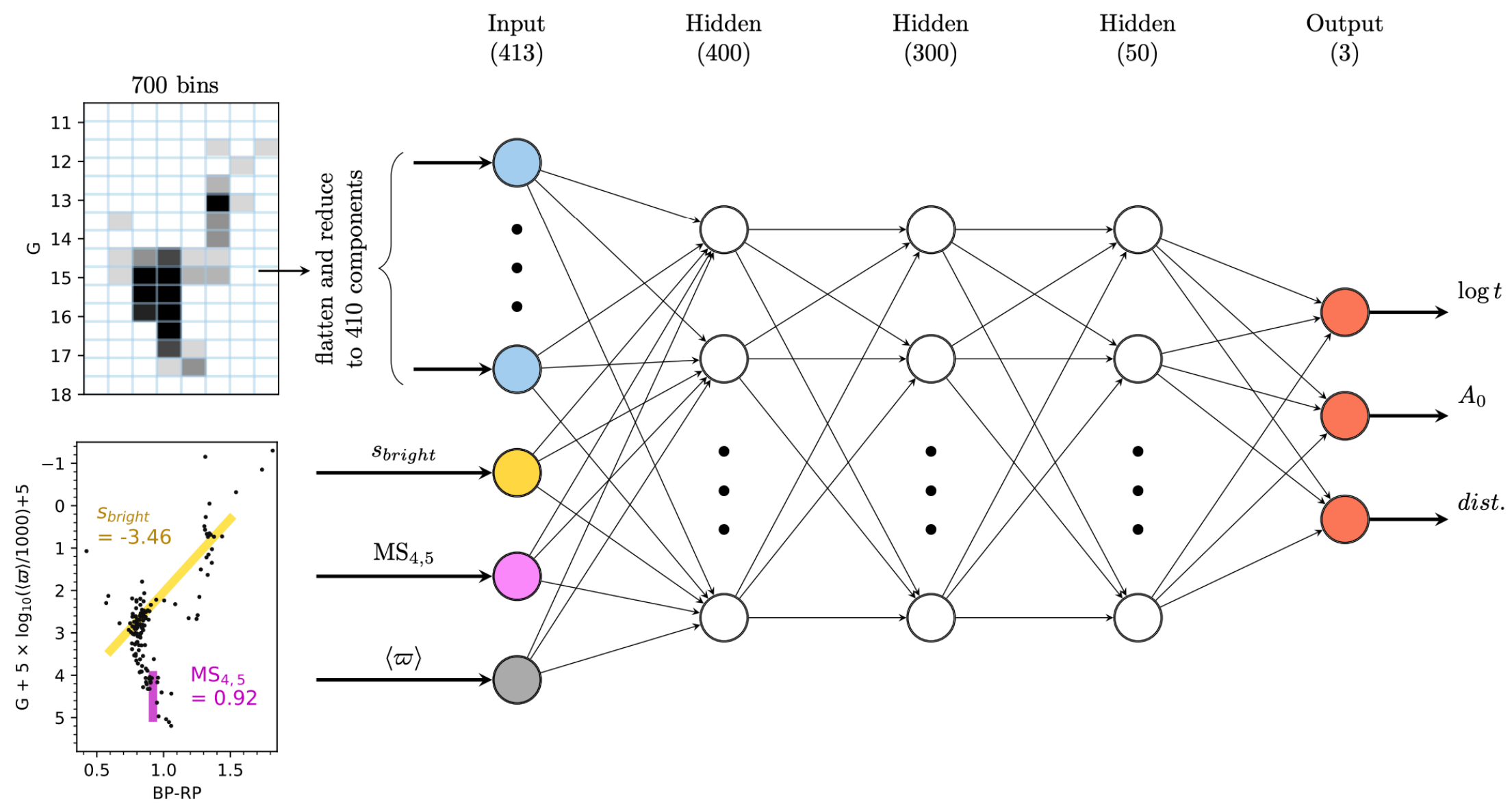## Of all the OC in the Milky Way…
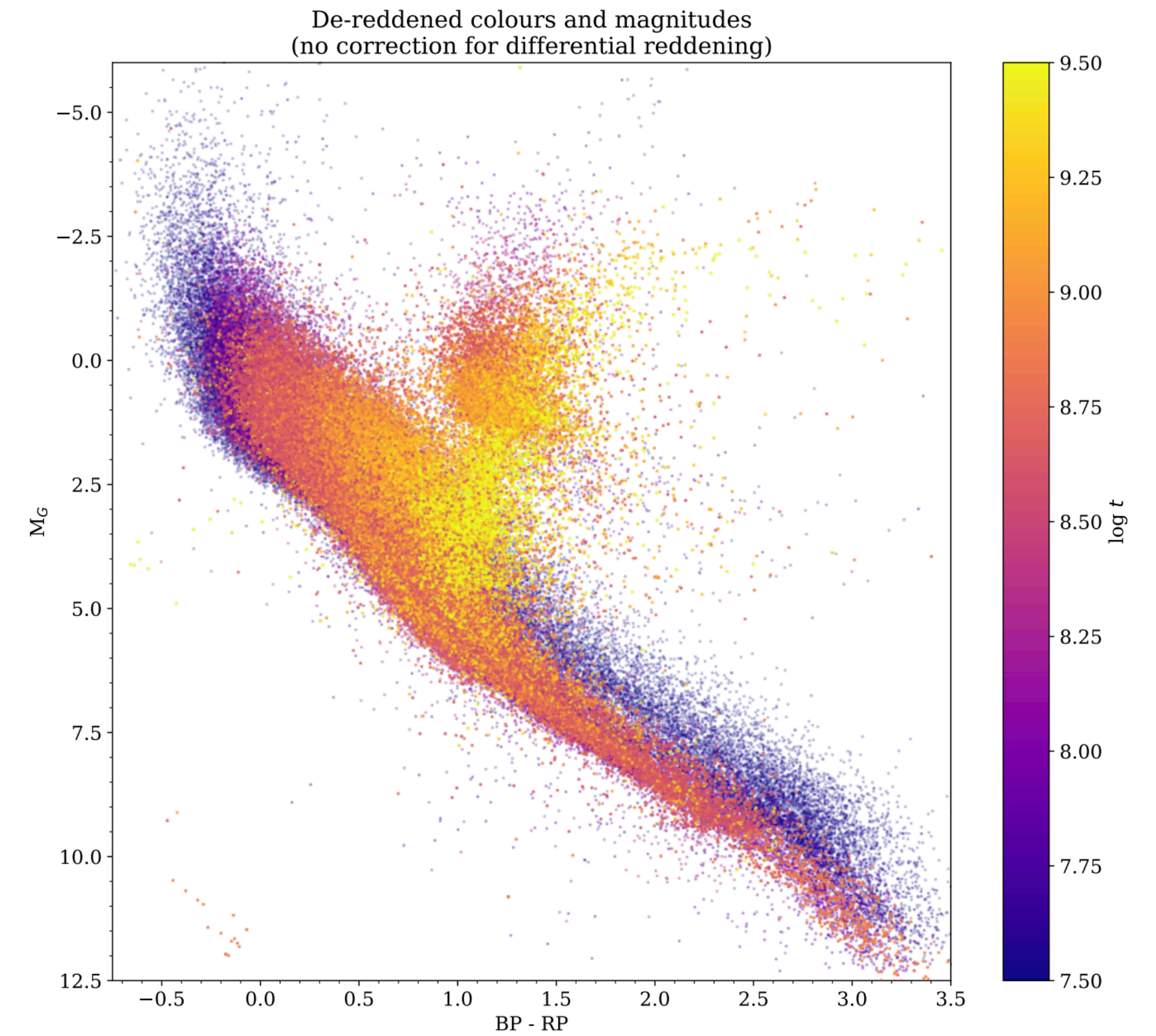
$\lesssim$ 40% are UBC clusters!!

OC Astrophysical characterisation

Astrophysical characterisation of OC parameters

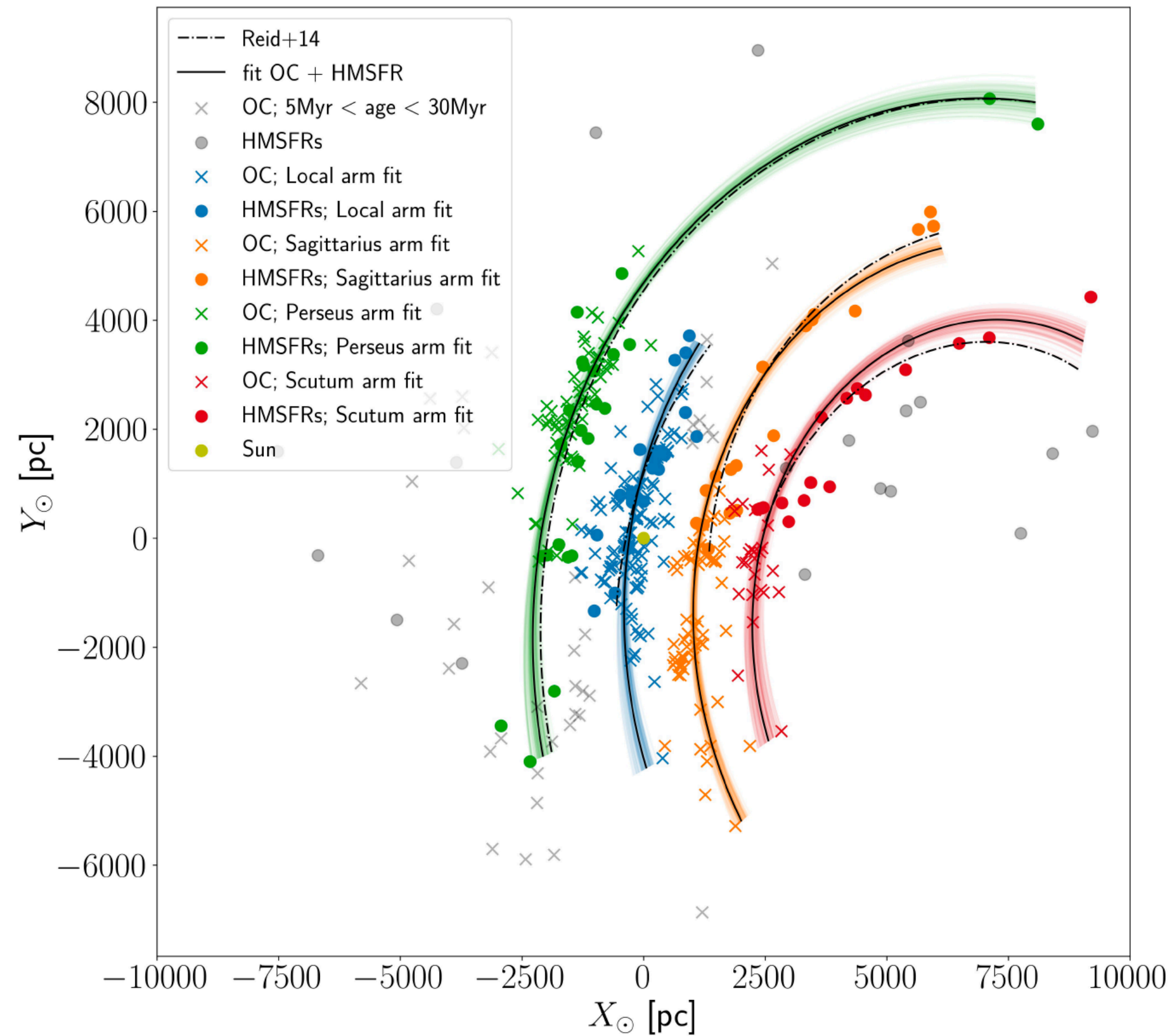Homogeneous estimation of ages, distances, and line-of-sight extinctions for around 2000 OC (in DR2).

[Cantat-Gaudin…**ACG**+20]

De-reddened colours and magnitudes
(no correction for differential reddening)

OCs tracing Galactic structure

Open clusters as tracers of Galactic structures [Castro-Ginard+21]

We traced the evolution of these spiral structures to understand their formation mechanism.

For the first time using OCs, we found a transient behaviour of the spiral arms, disfavouring classical density waves.

Still a highly debated topic in the literature.
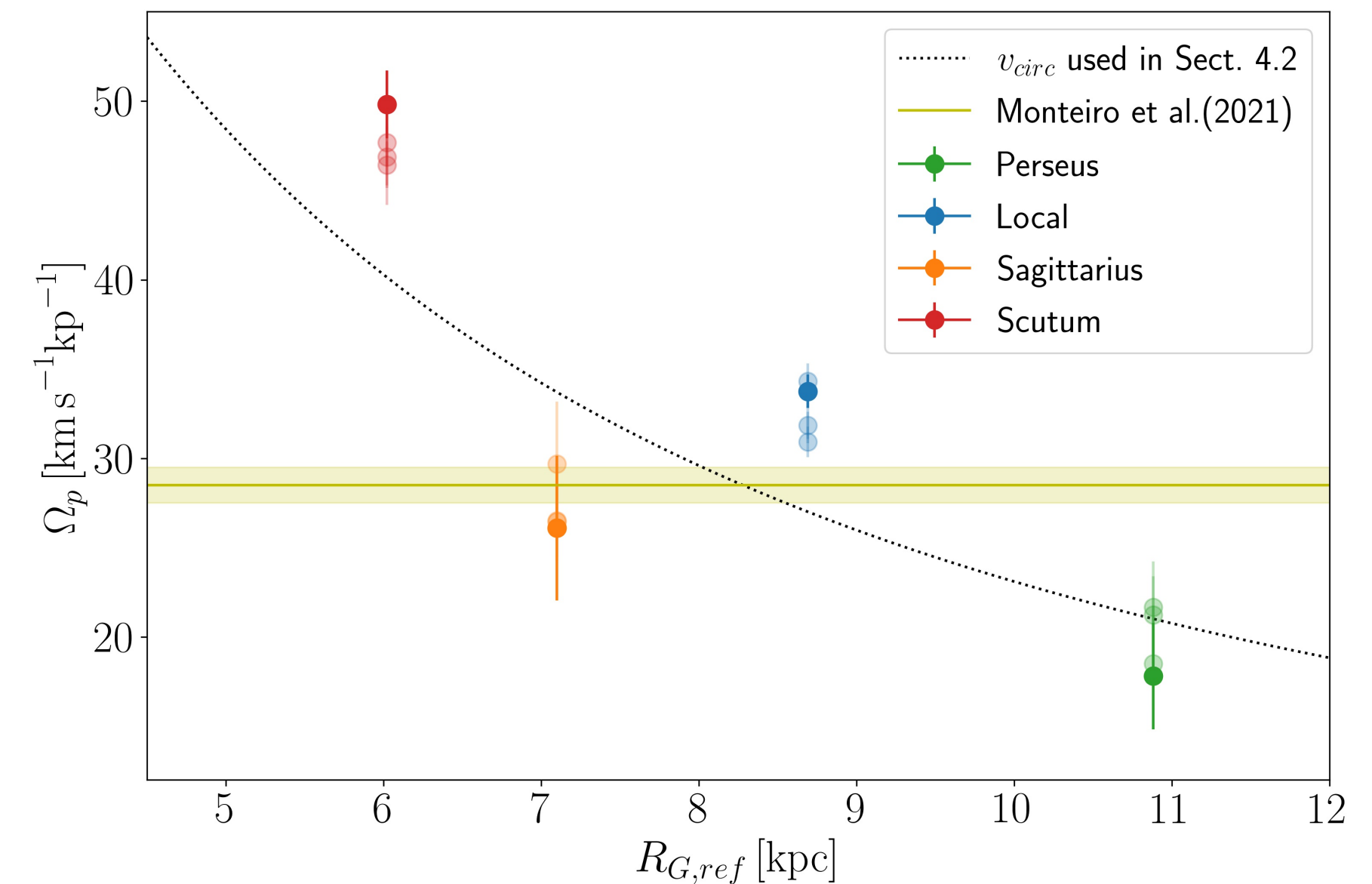
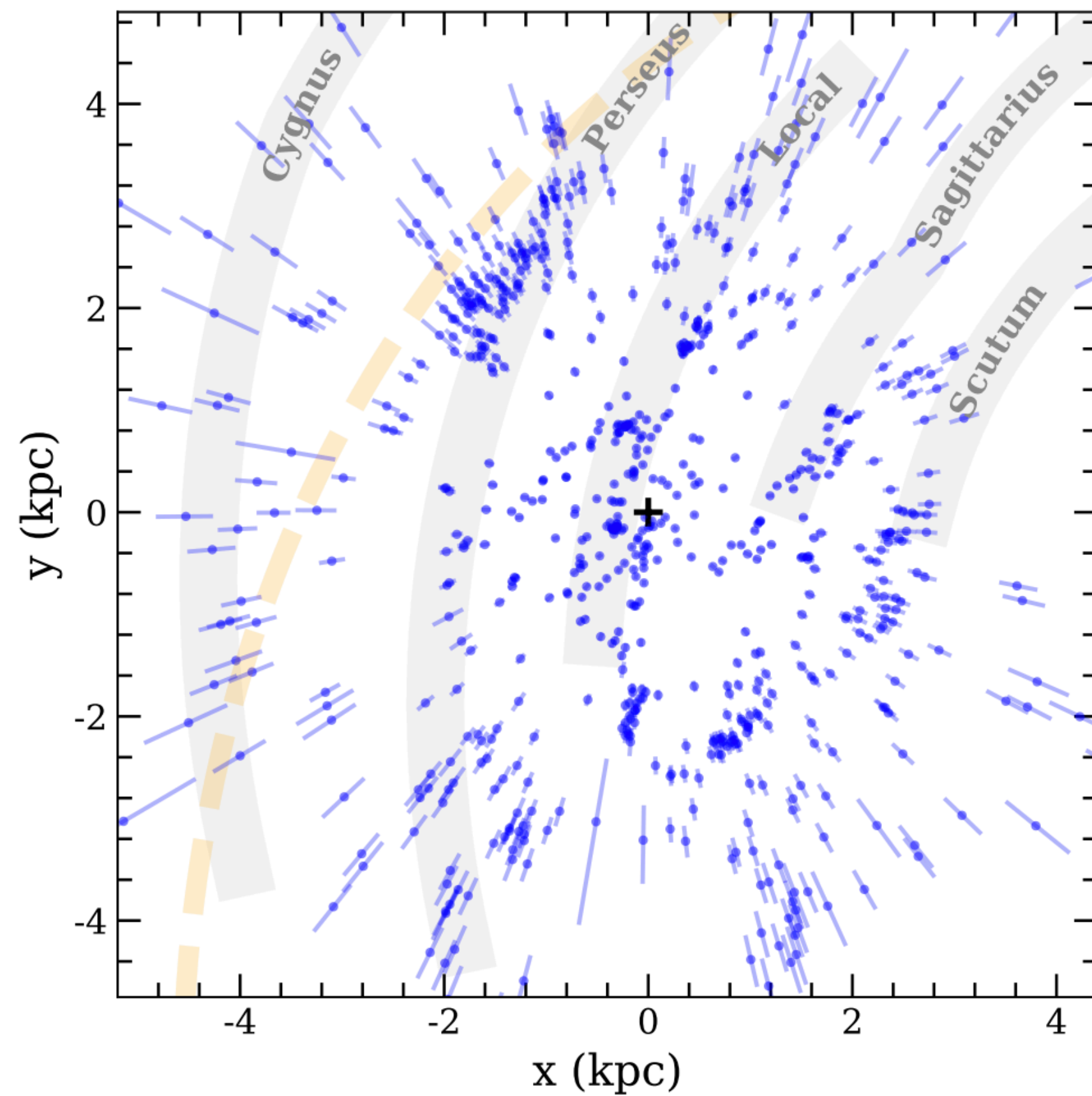# Comparison with other tracers

OCs younger than 63 Myr

OB stars

Comparison with other tracers

OCs younger than 63 Myr

No OC detected here

[Cantat-Gaudin+19]

[Castro-Ginard+19]

OB stars
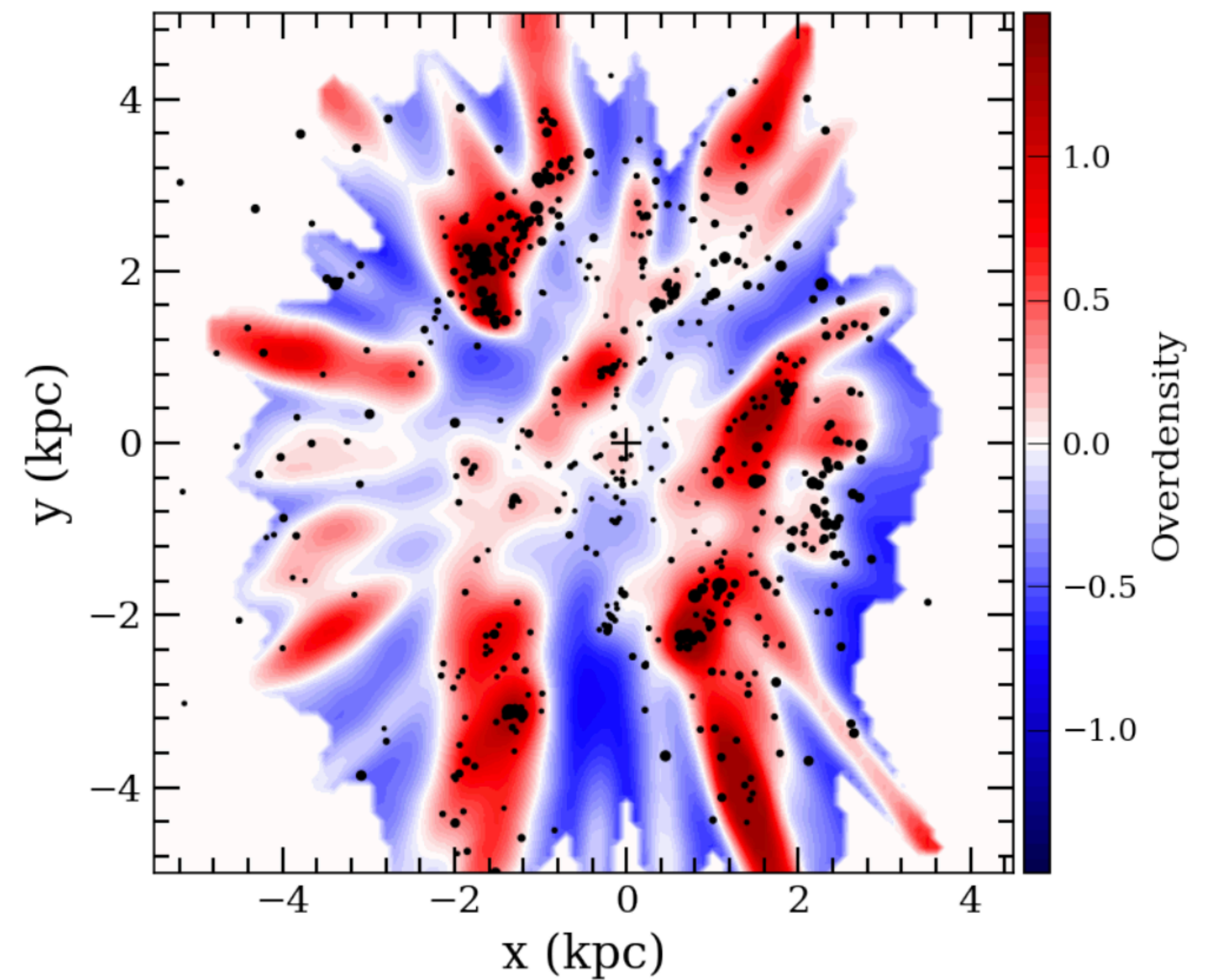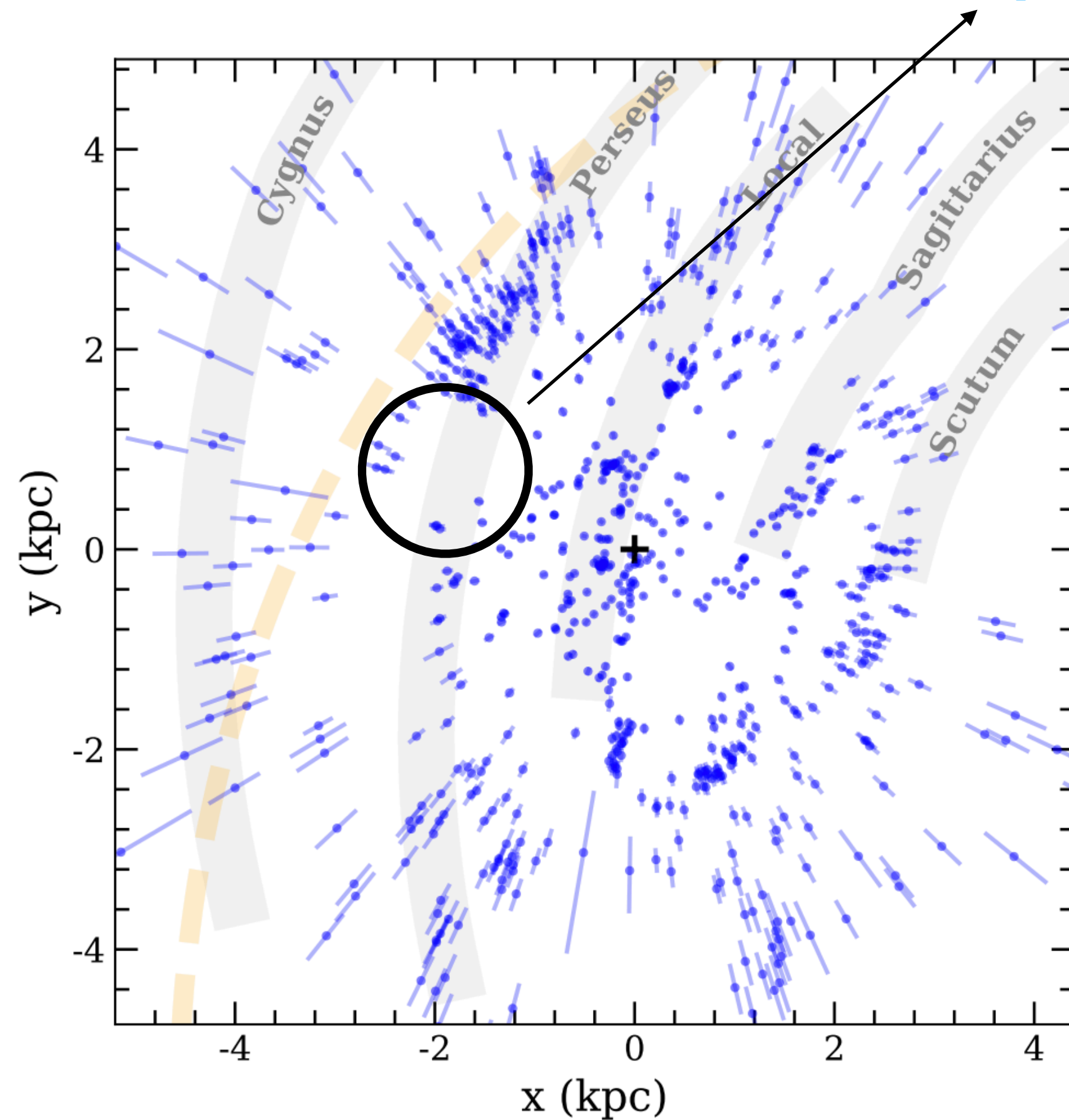
Fig. 14 in GaiaCollaboration,Drimmel+22

Fig.13 in GaiaCollaboration,Drimmel+22
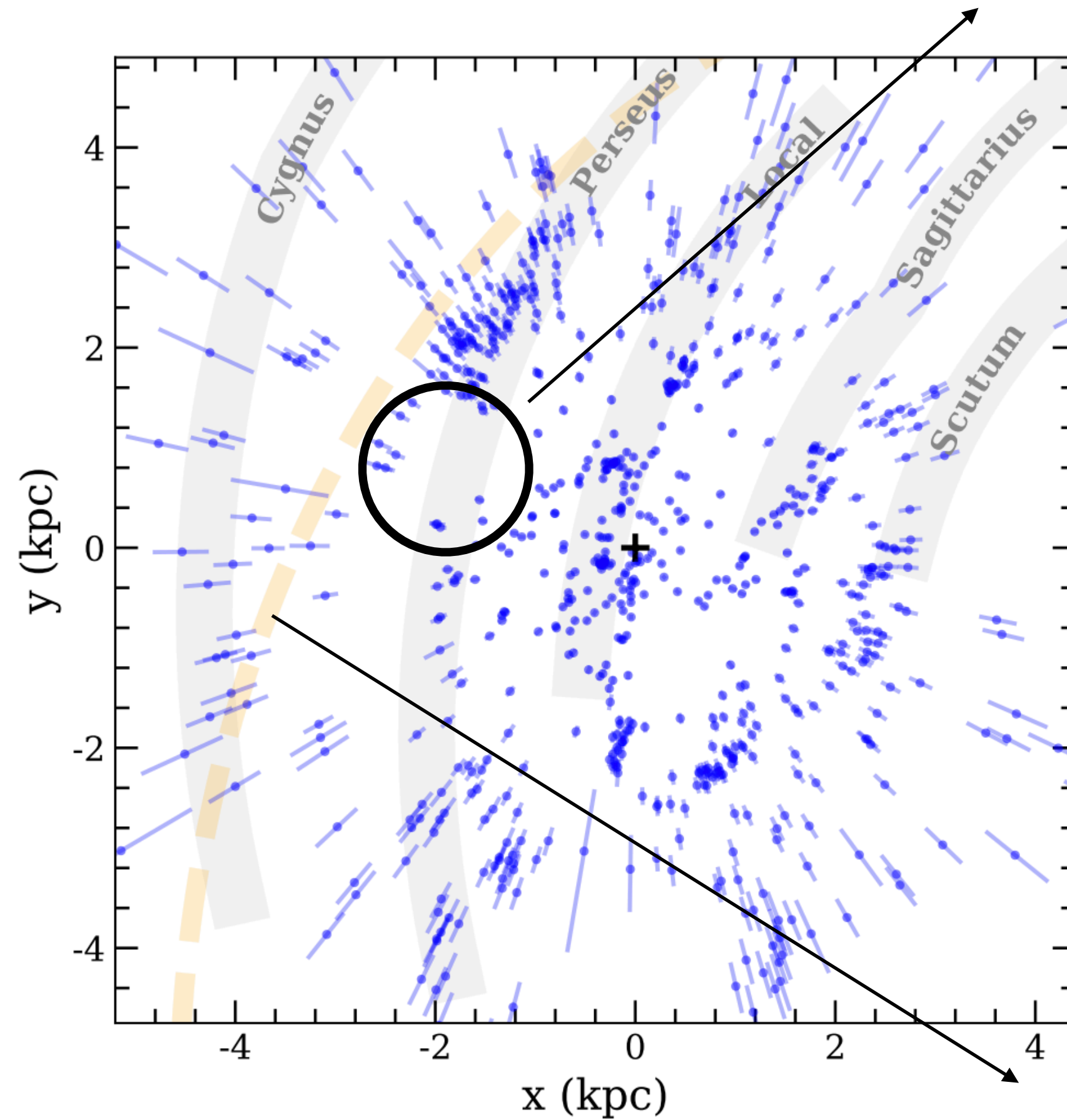
Comparison with other tracers

OCs younger than 63 Myr

No OC detected here

[Cantat-Gaudin+19]

[Castro-Ginard+19]

OB stars

[Levine+06] arm model

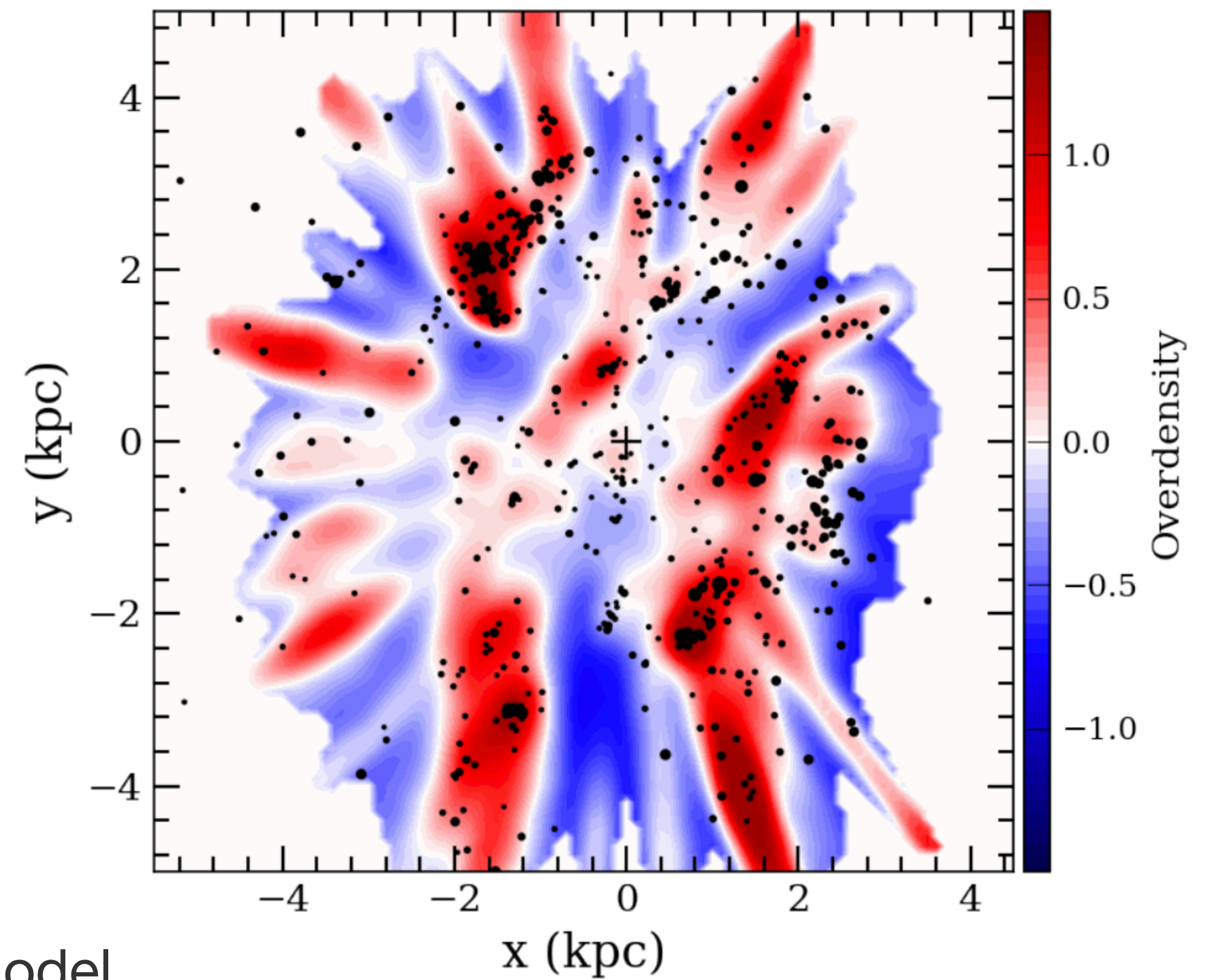Fig. 14 in GaiaCollaboration,Drimmel+22
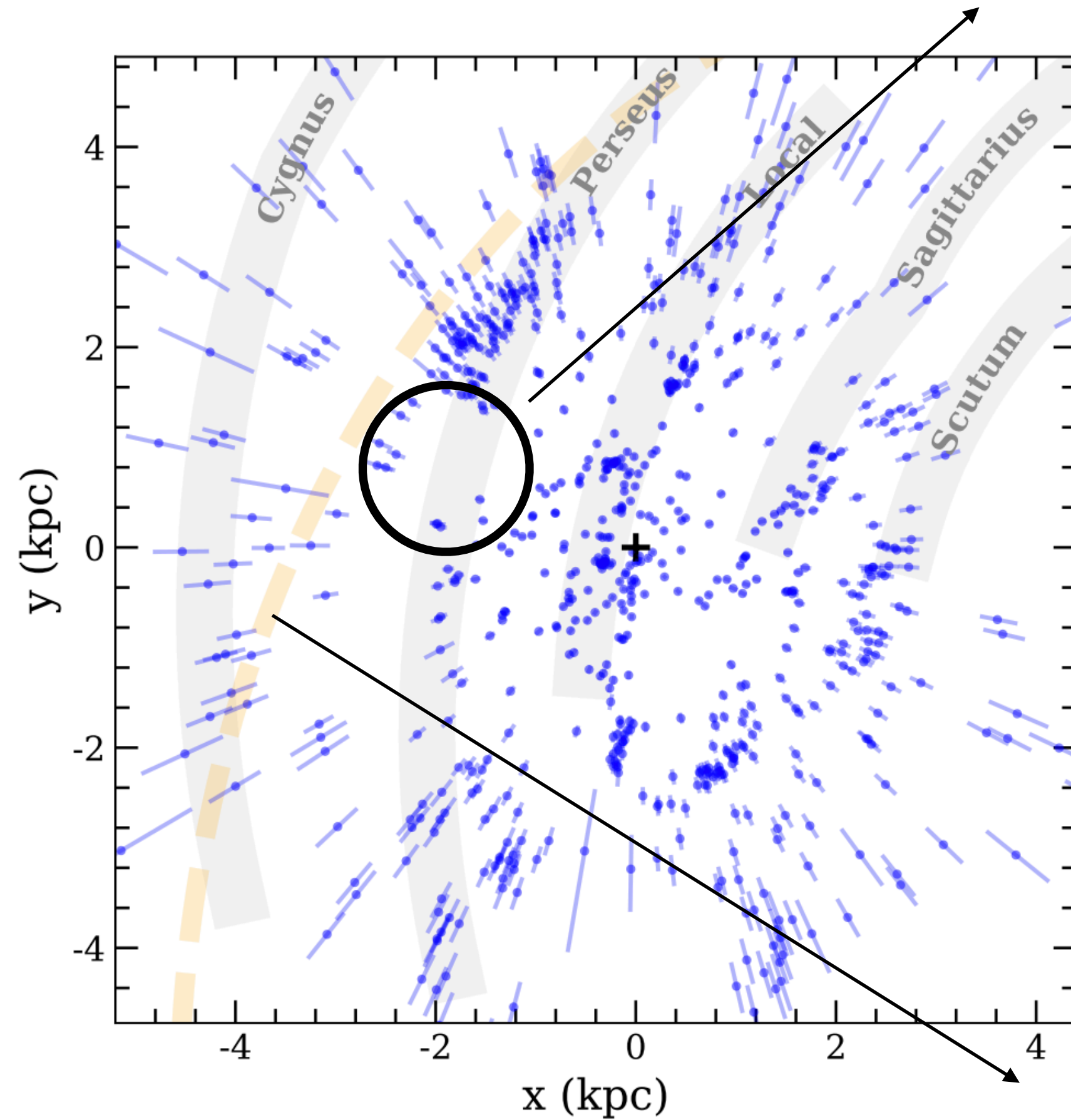
Fig.13 in GaiaCollaboration,Drimmel+22

Comparison with other tracers

No OC detected here

[Cantat-Gaudin+19]

[Castro-Ginard+19]

OCs younger than 63 Myr

Cepheids younger than 200 Myr

[Levine+06] arm model

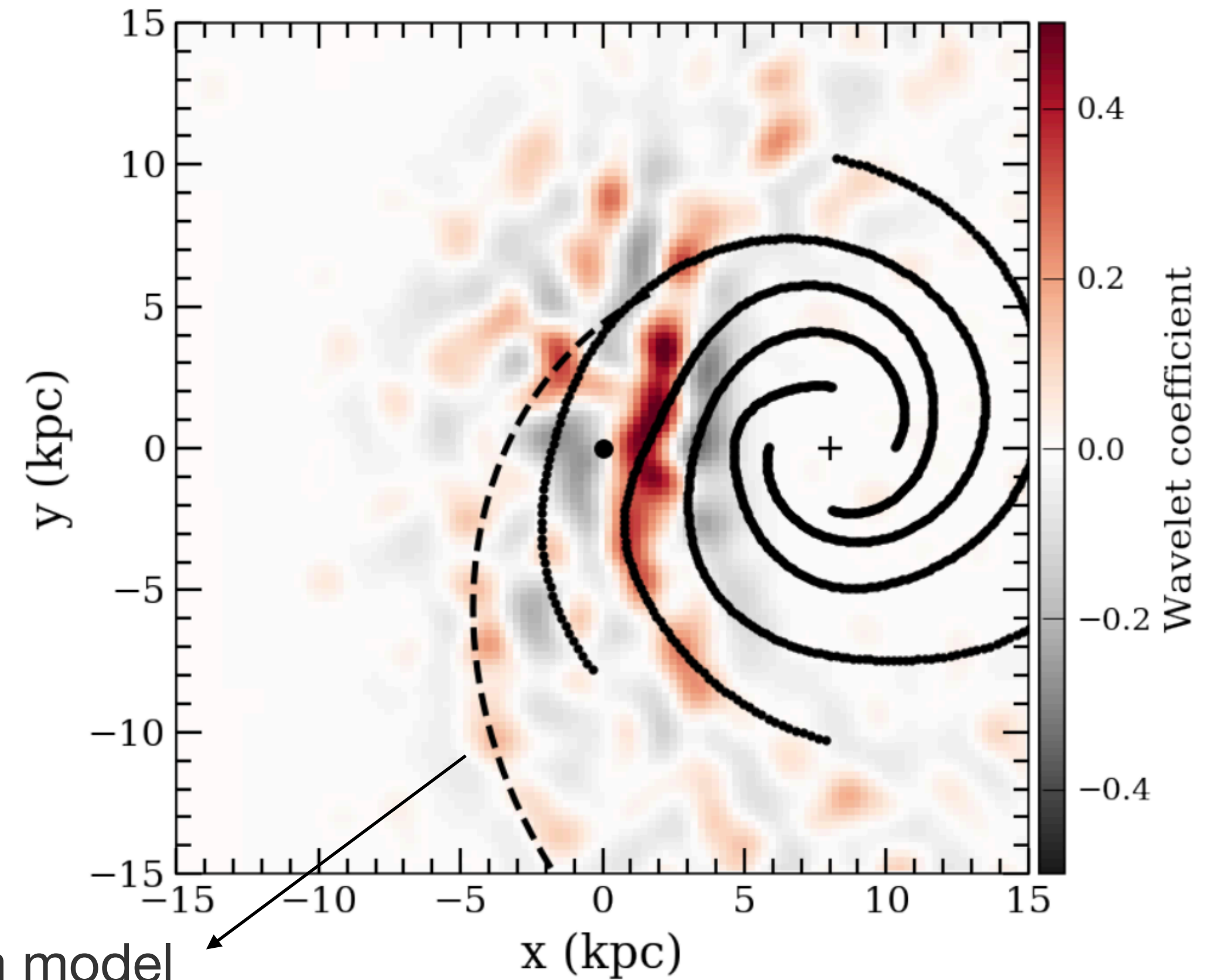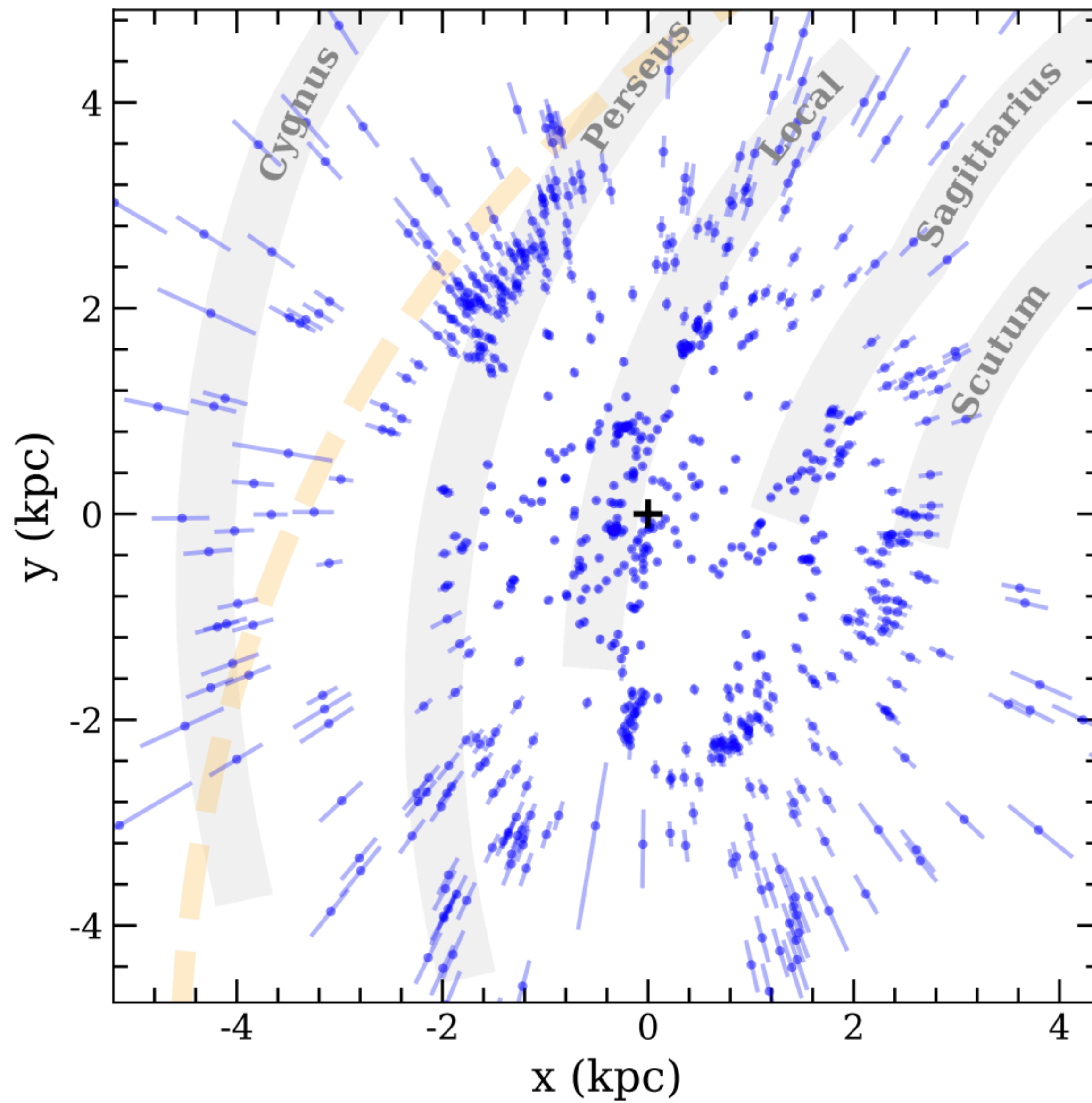Fig. 14 in GaiaCollaboration,Drimmel+22

Fig.15 in GaiaCollaboration,Drimmel+22
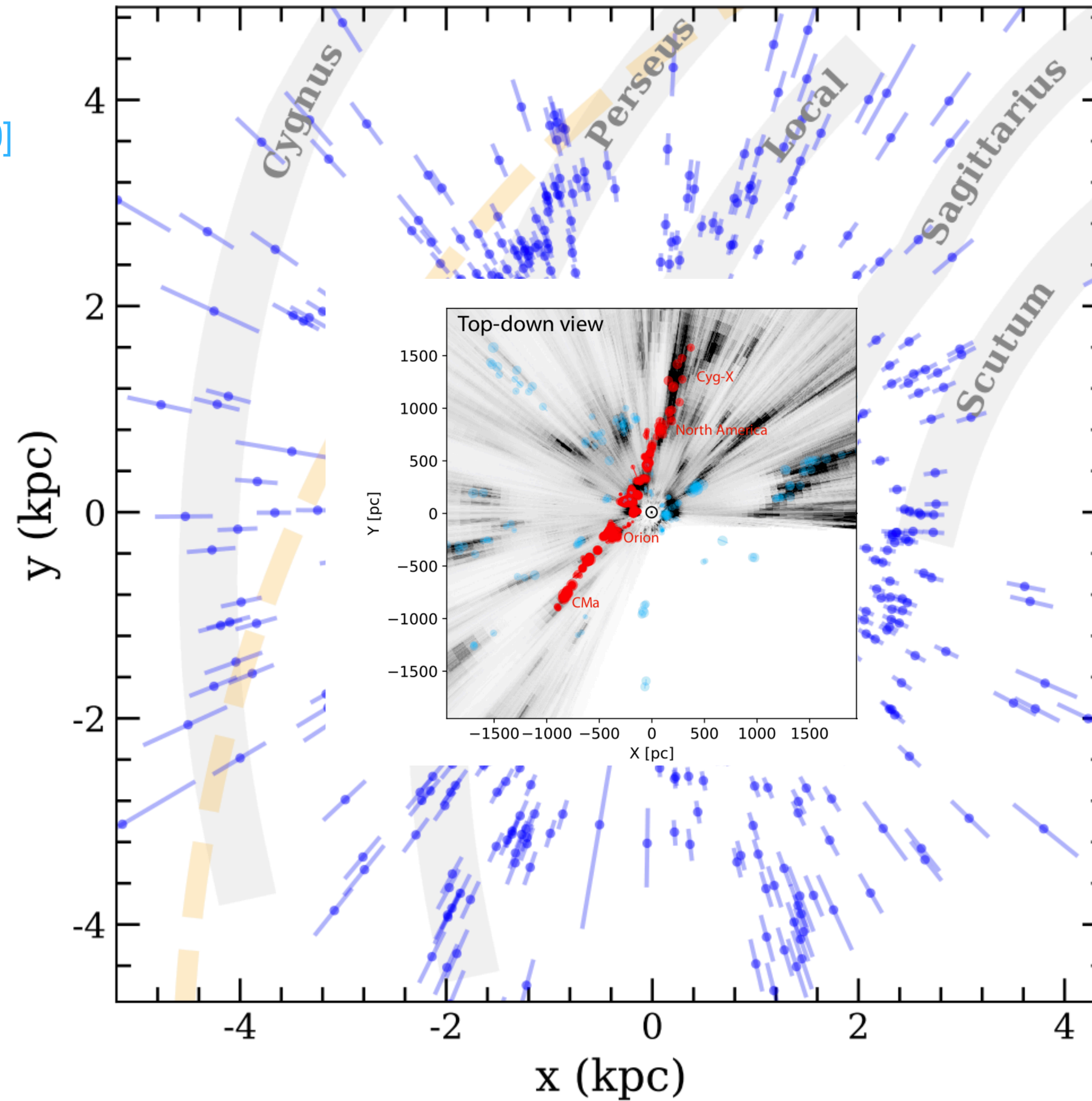
Deviation from log-periodic spiral arms

Deviation from log-periodic spiral arms
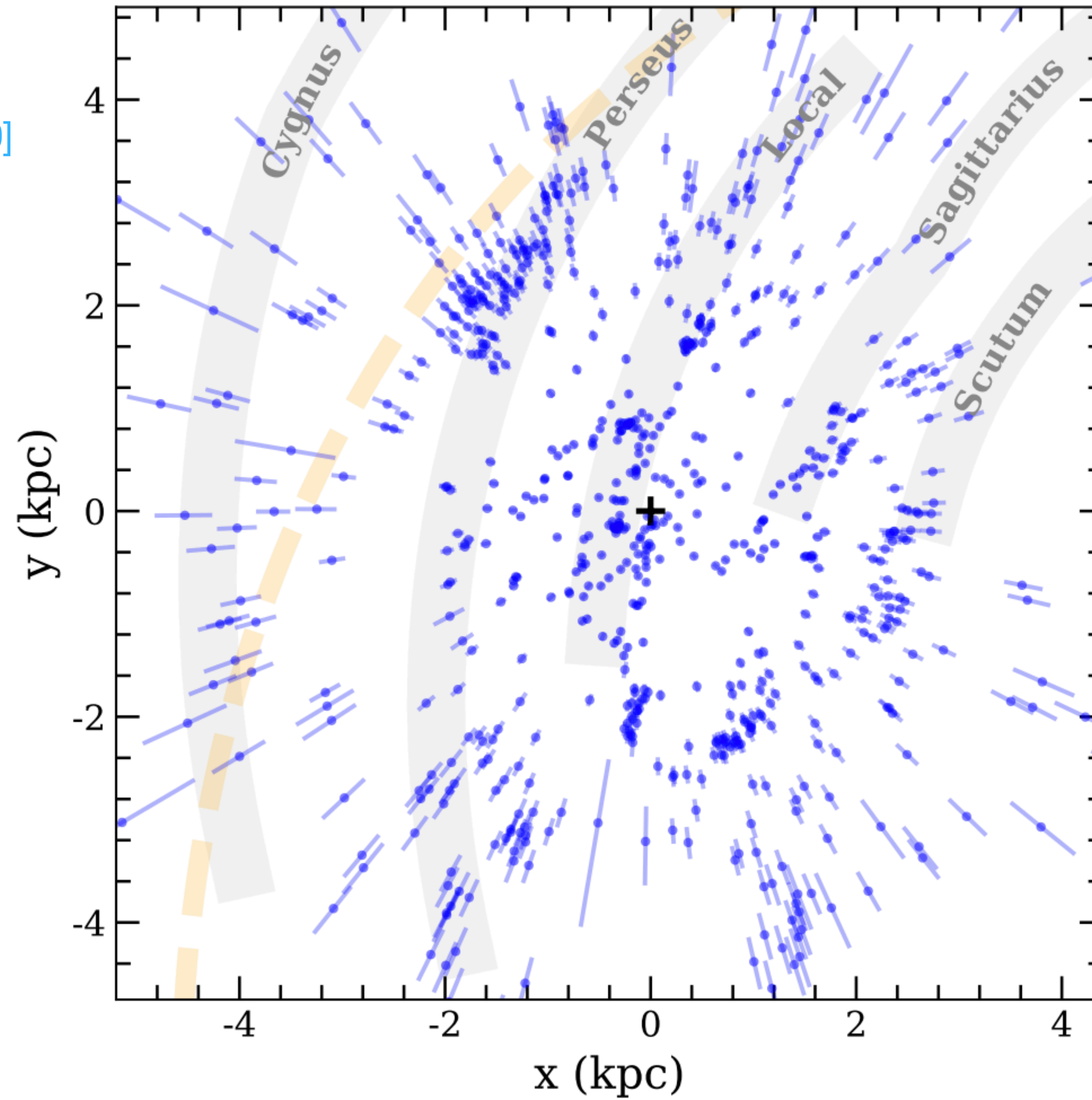
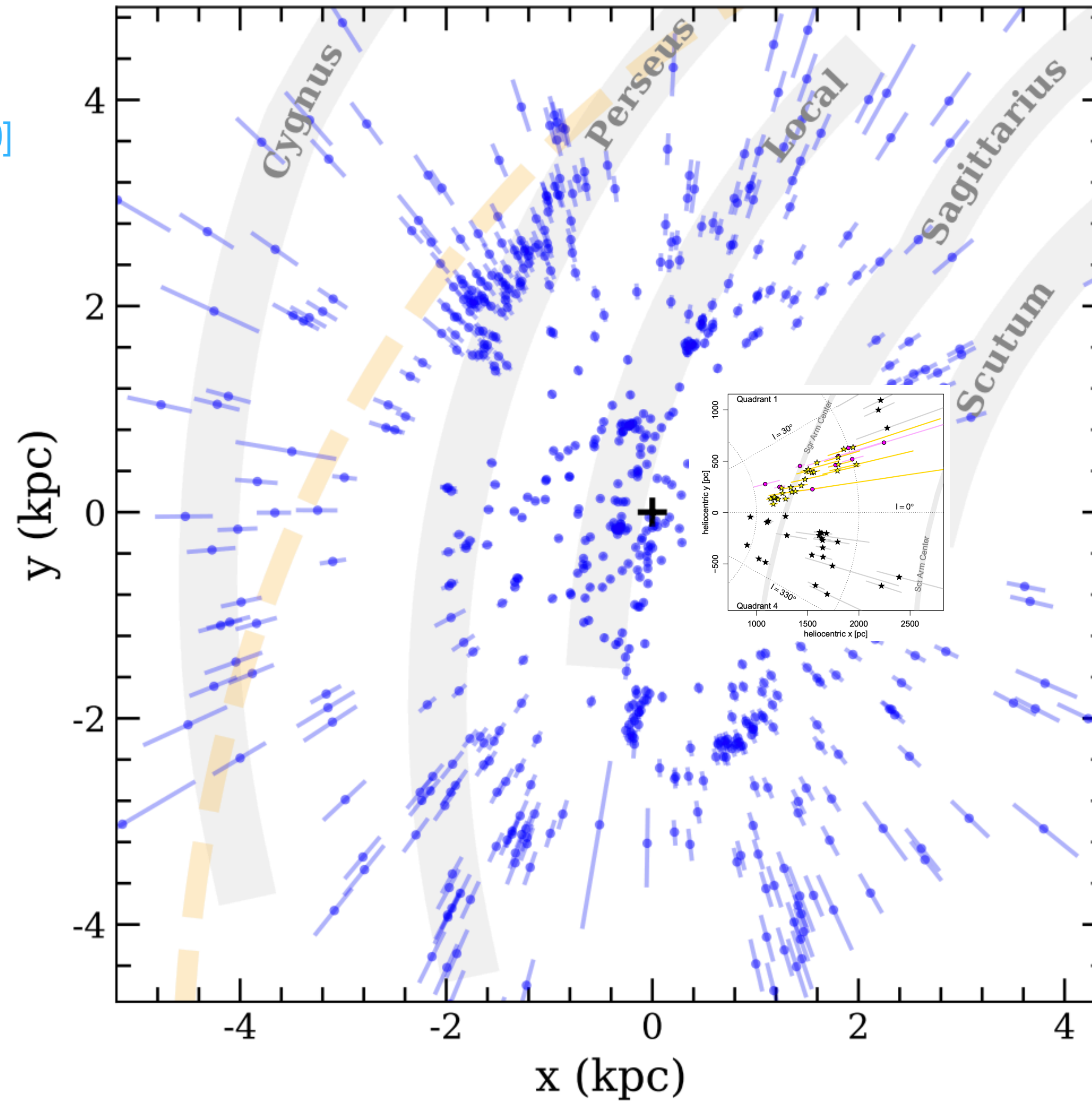- Radcliffe wave [Alves+20]

Deviation from log-periodic spiral arms

- Radcliffe wave [Alves+20]

Deviation from log-periodic spiral arms

- Radcliffe wave [Alves+20]

- Sag. spur [Kuhn+21]

What about the stars that we are missing?

Gaia catalogue selection function

Two approximations to estimate the *Gaia* Selection Function:

- Forward modelling (based on [Castro-Ginard+*in prep*])

    - From *Gaia* data only: Scanning law, ability to convert observations to detections, constraints in the used detections to get measurements, …

- Empirical model [Cantat-Gaudin+22]

    - Comparing to deeper photometric catalogues (here, DECaLS), considered as 'ground truth'
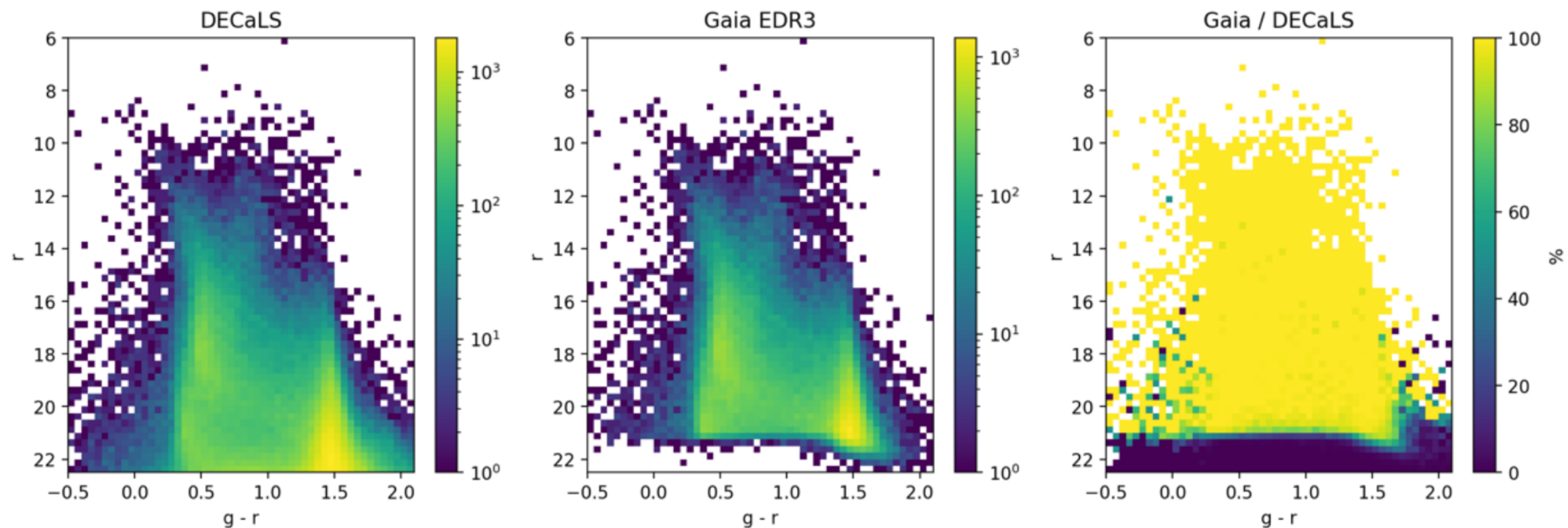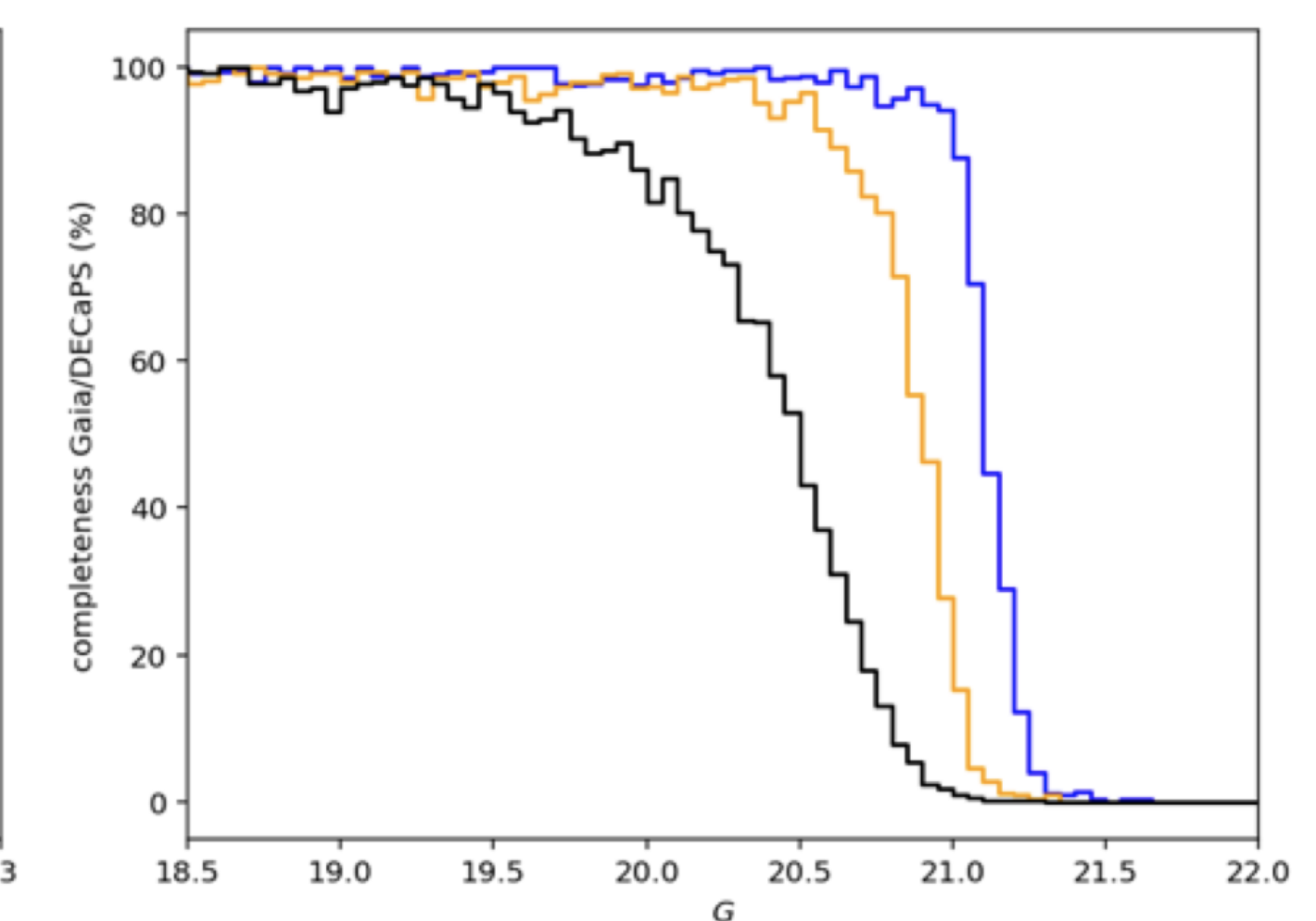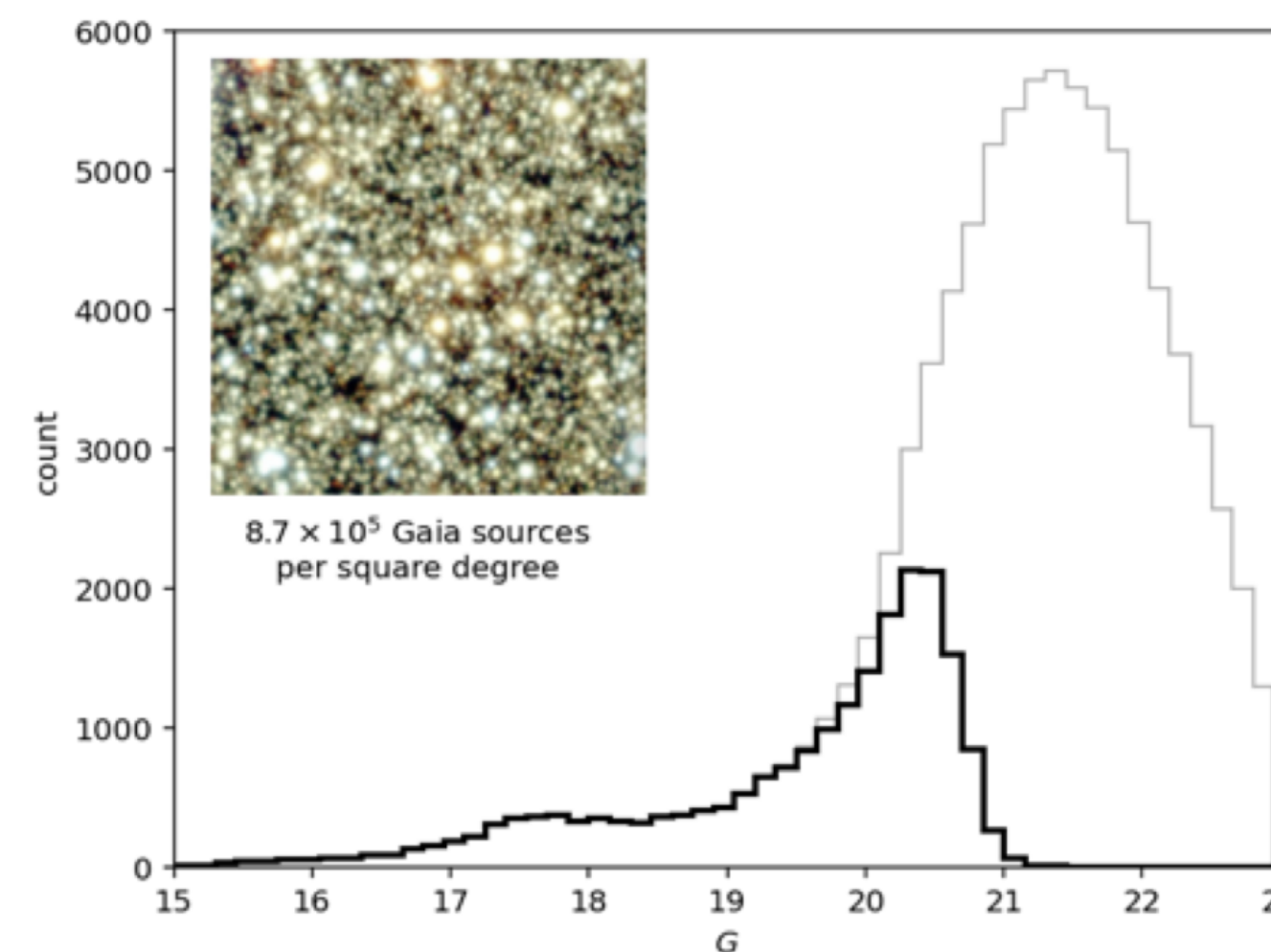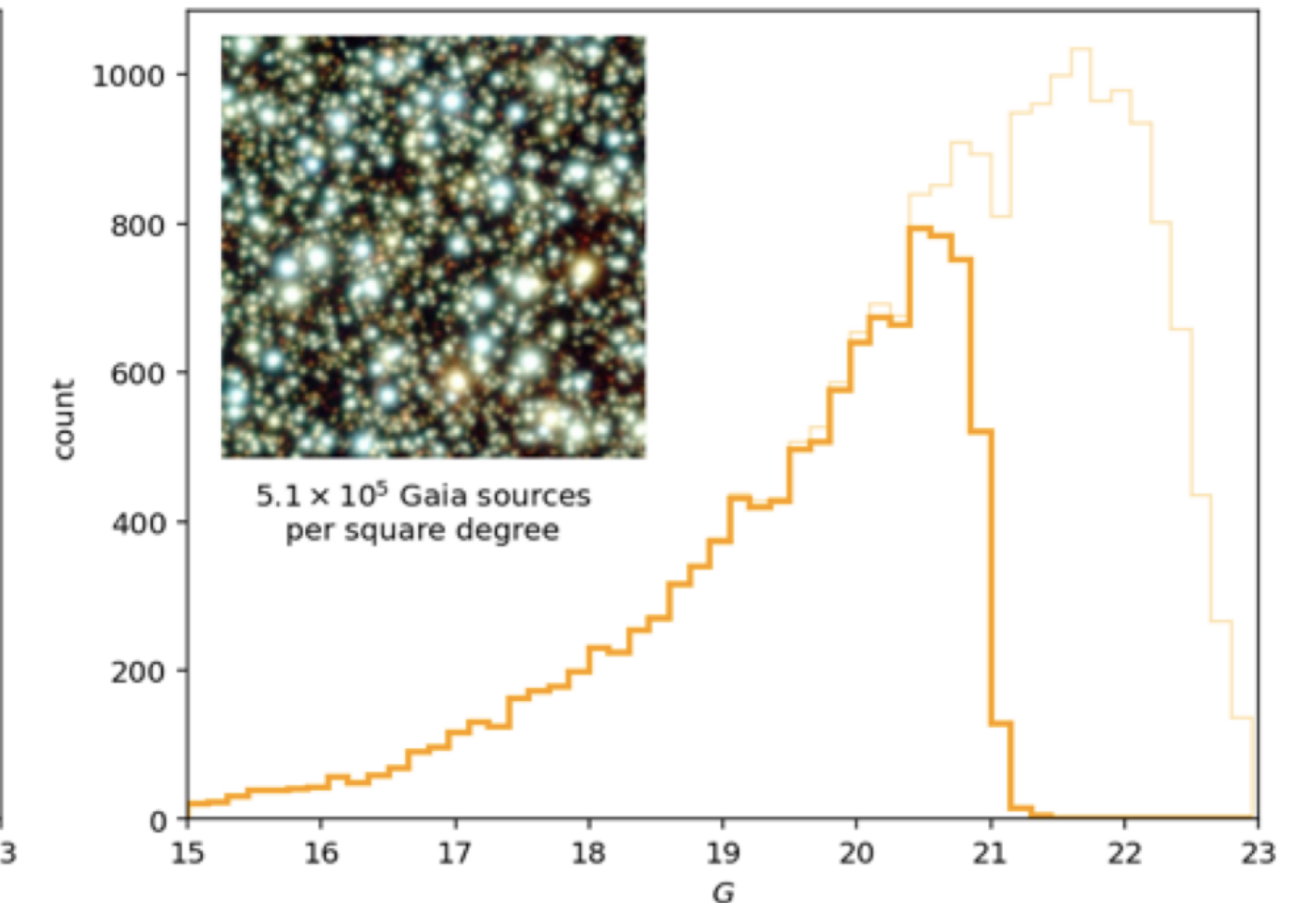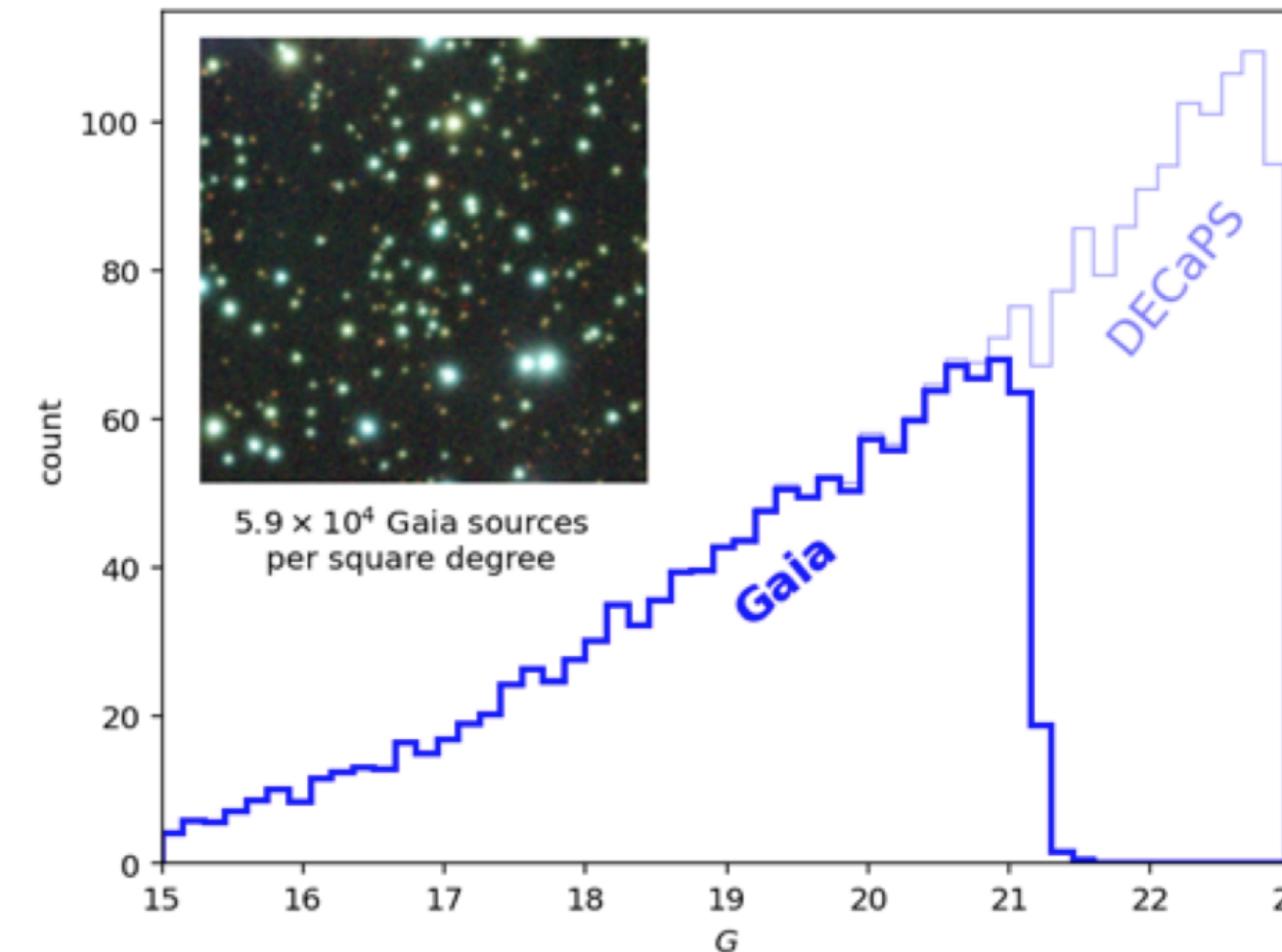
Gaia selection function: empirical model

The "true" source density of DECaPS is a good predictor of completeness as a function of G

- DECaPS data are not available for the entire sky
- We need to find a proxy that can be computed from *Gaia* data alone

Gaia selection function: forward model

Based on [Boubert&Everall20]: probability of observation as coin
toss exercise

- $$p(k|n, \theta) = \binom{n}{k} \theta^k (1 - \theta)^{(n-k)}$$

- n: number of potential observations (scanning law)
- k: number of detections used to produce the catalogue: astrometric
  matched transits
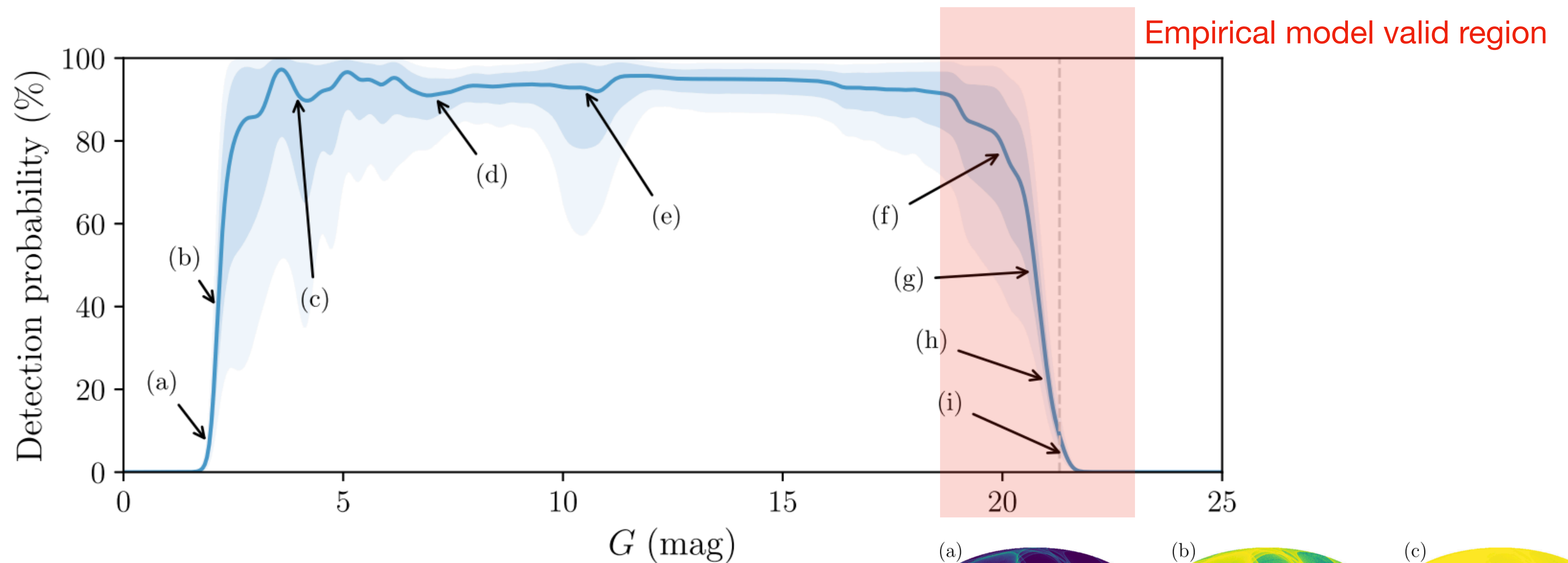- θ: estimated from Gaia data using Bayesian model
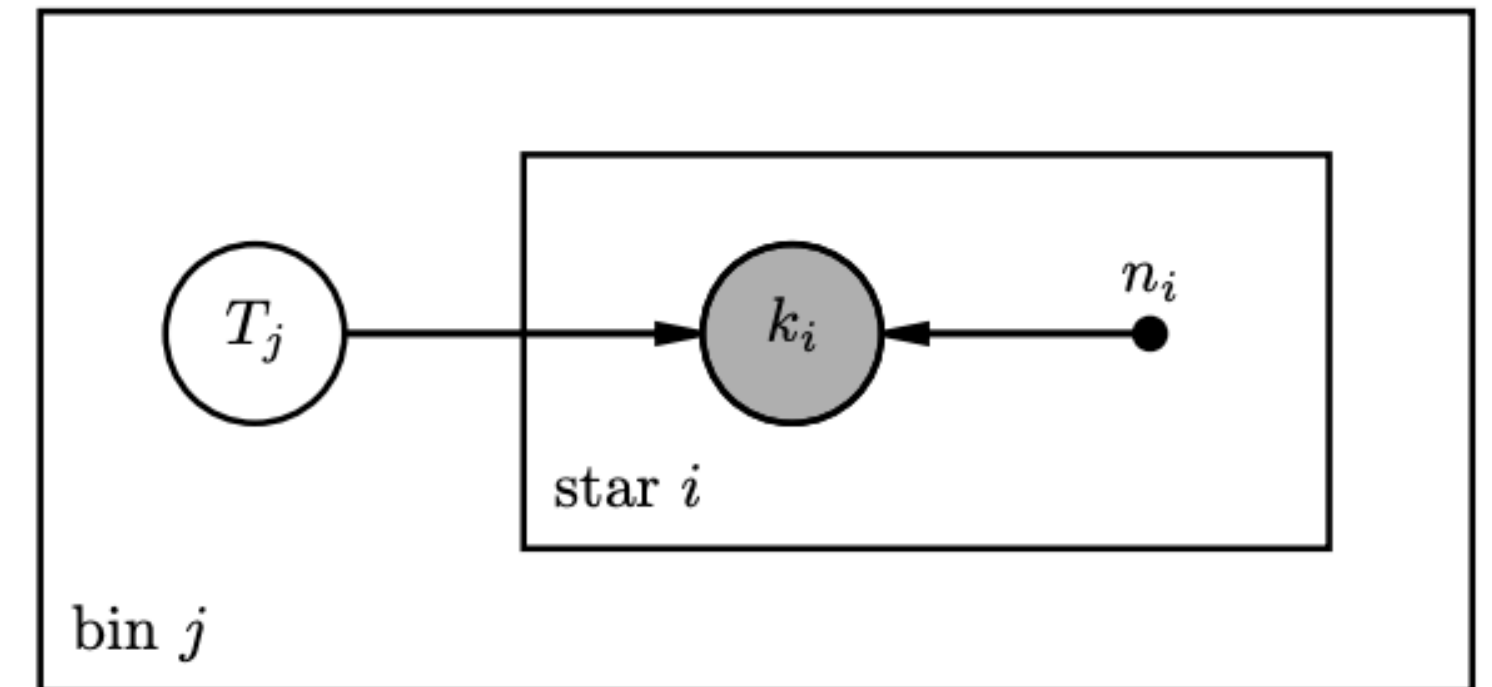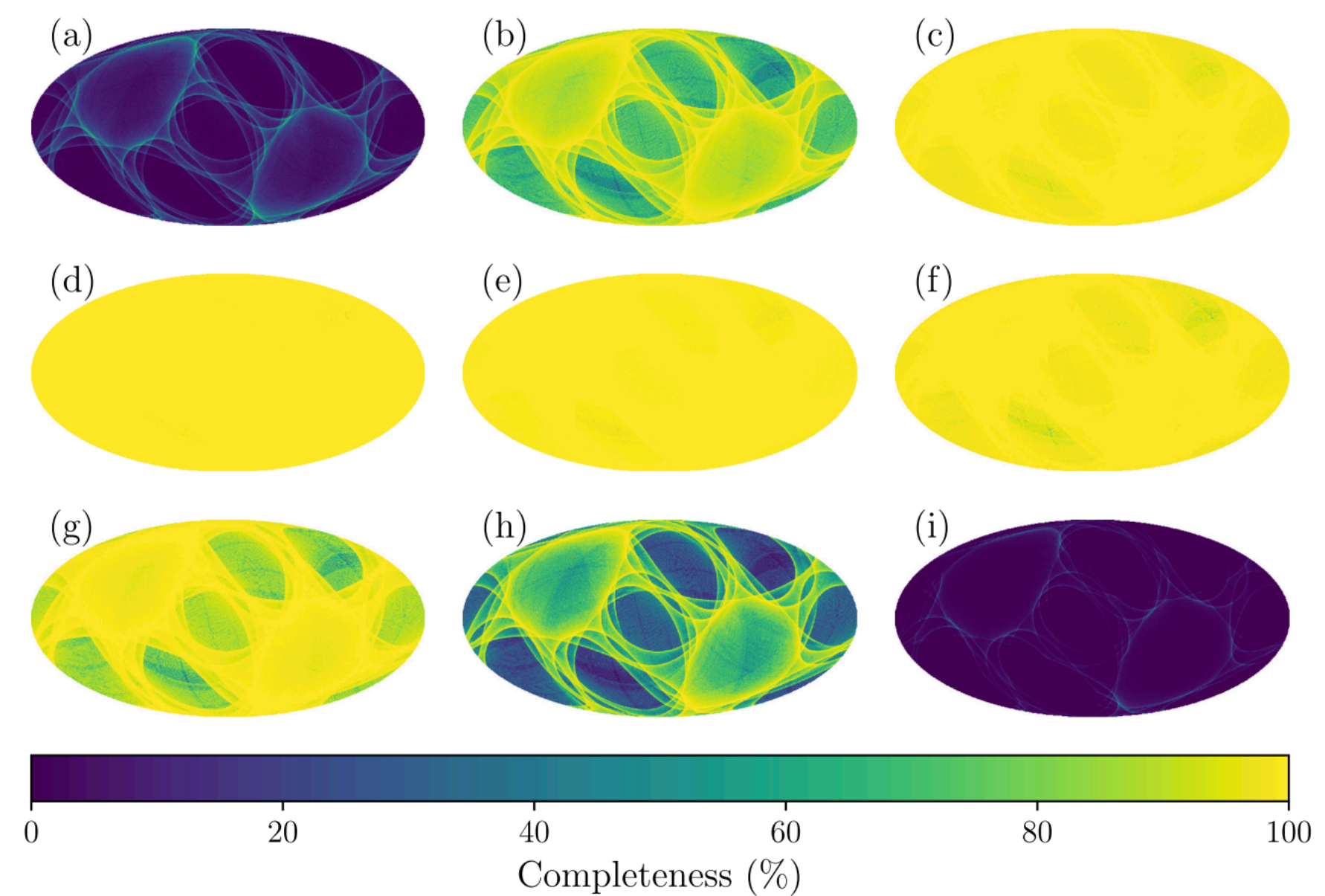

(a) Model T
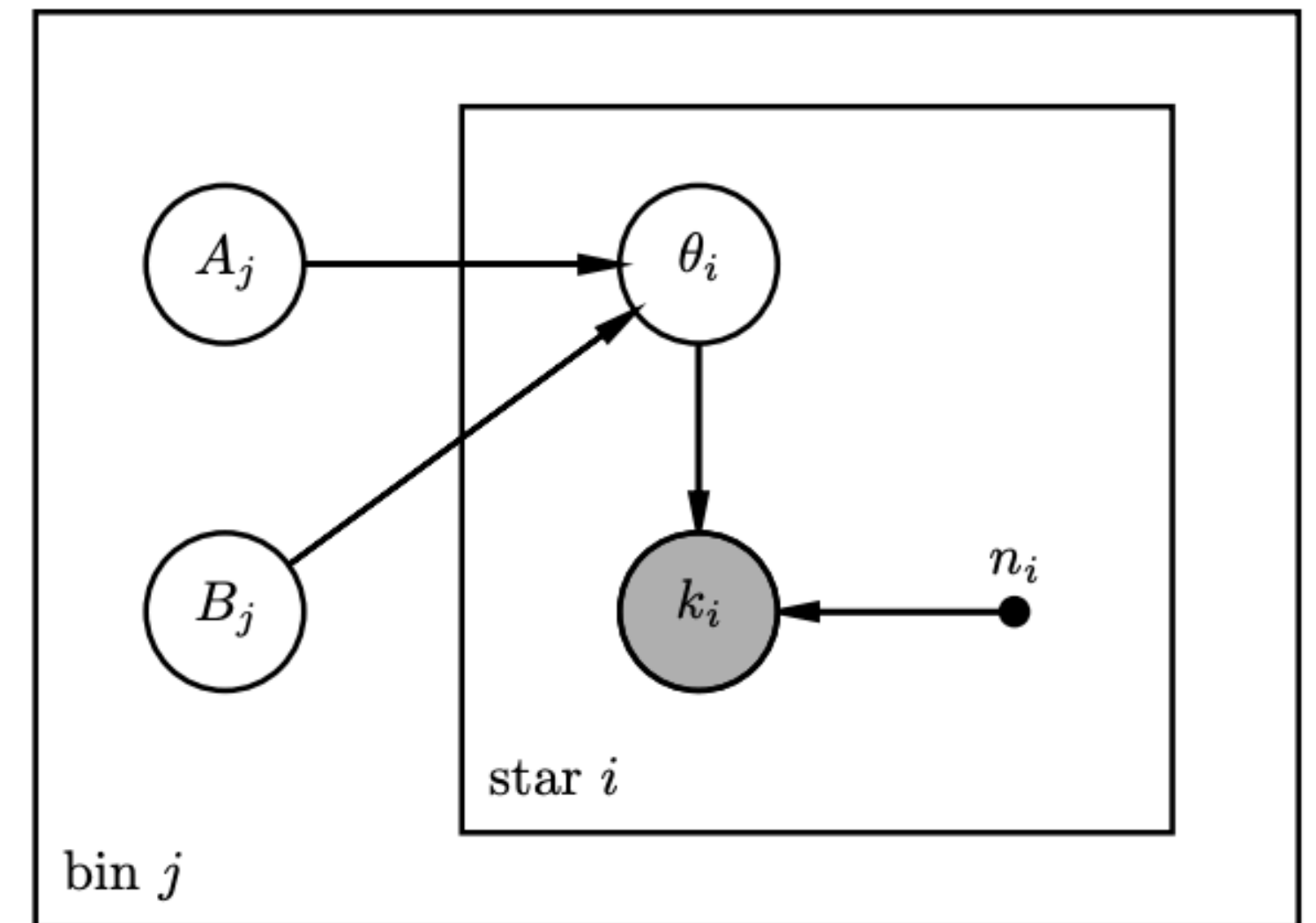
(b) Model AB

Gaia selection function

Empirical model valid region
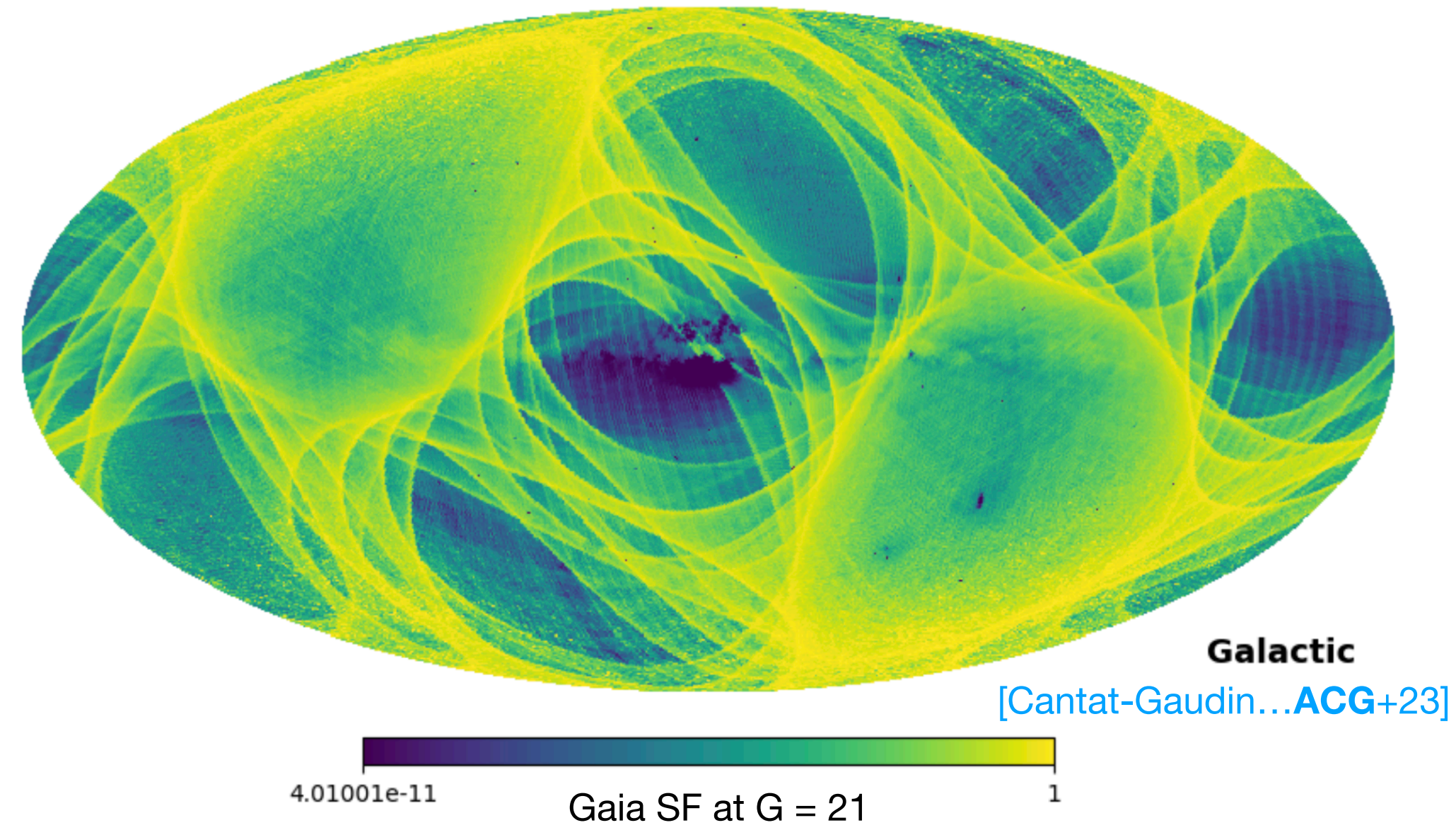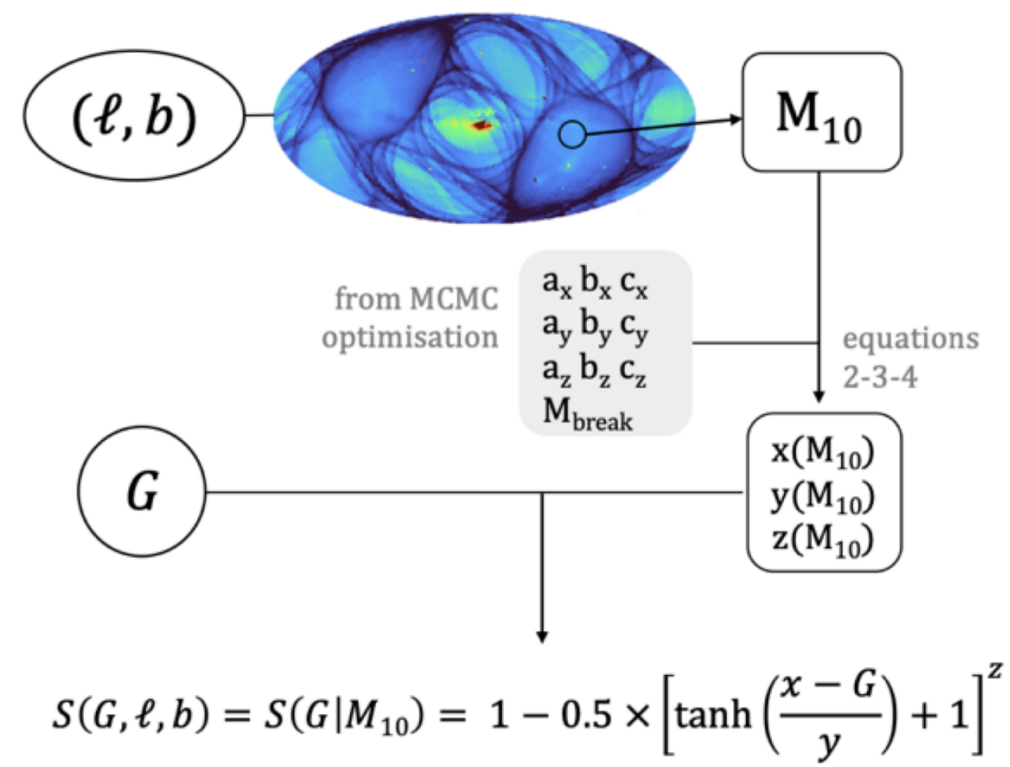
Need both models to:

- Estimate Gaia completeness across all G (forward model)

- Validate the model/results on the faint end (empirical model)

Some other Gaia selection functions

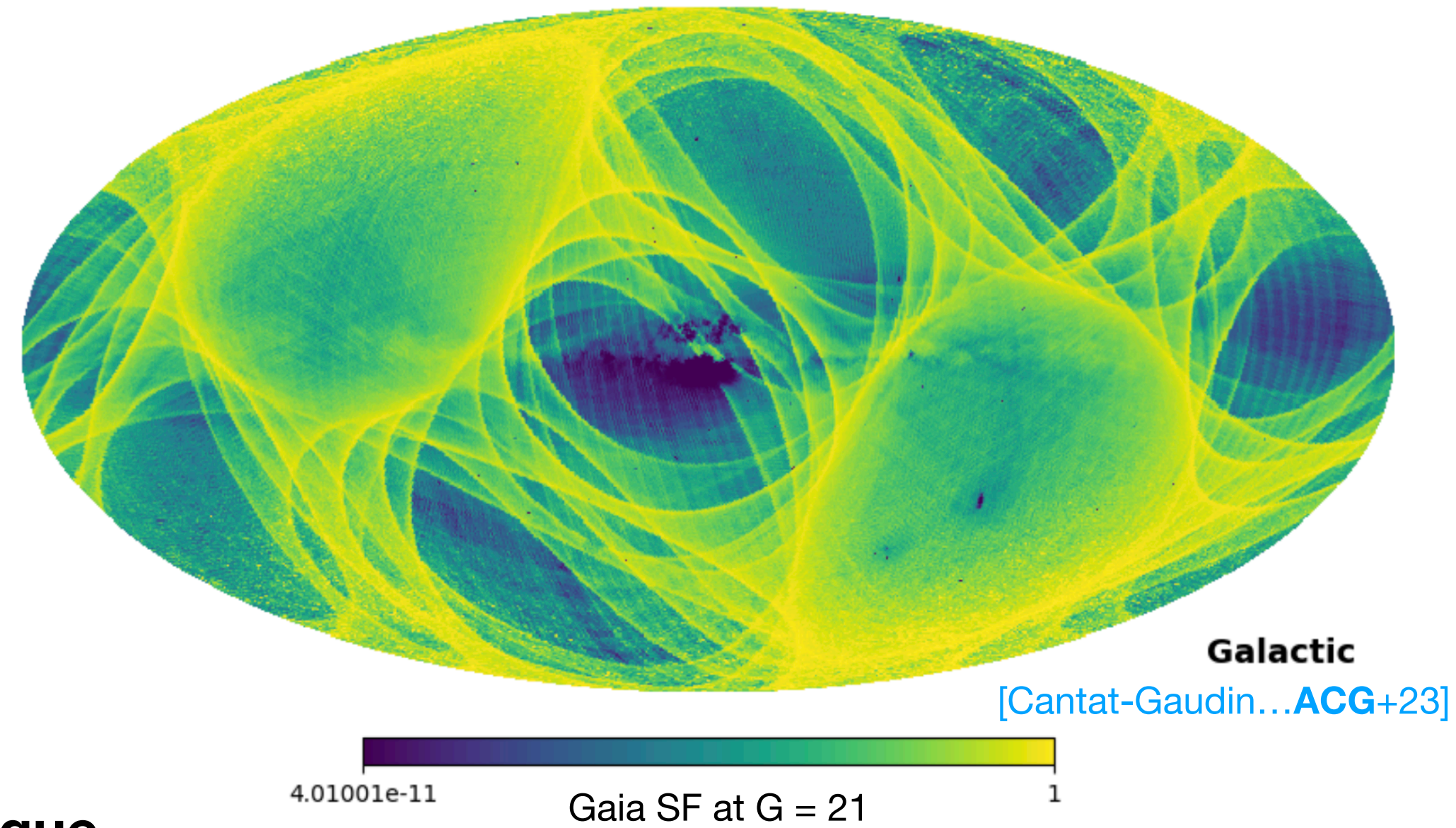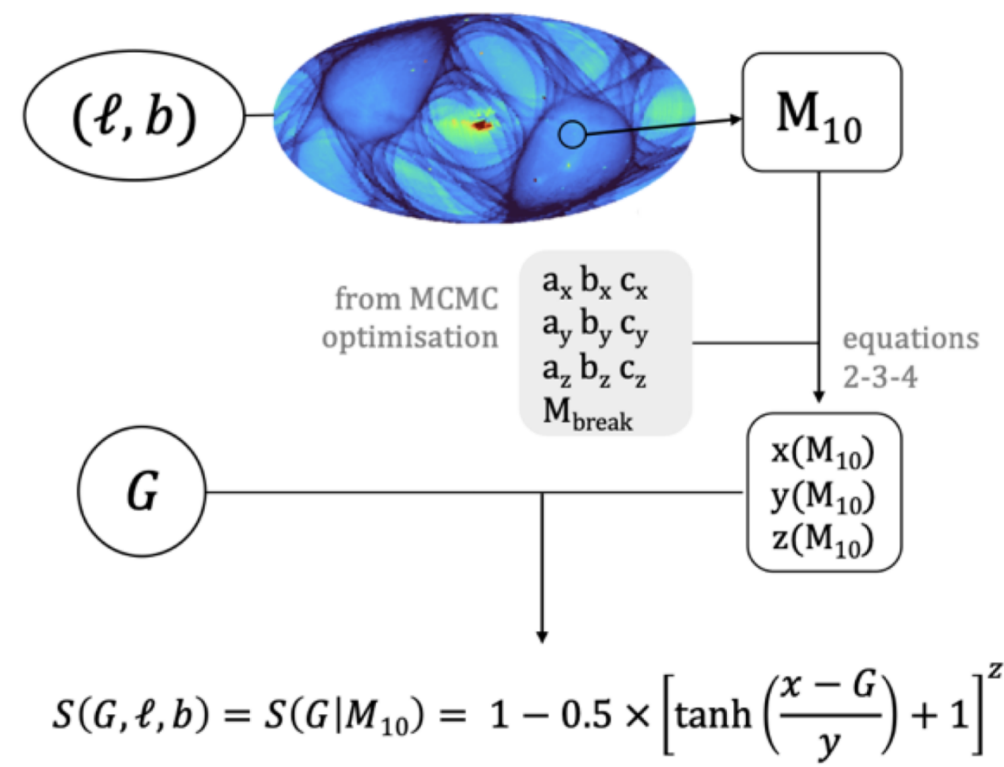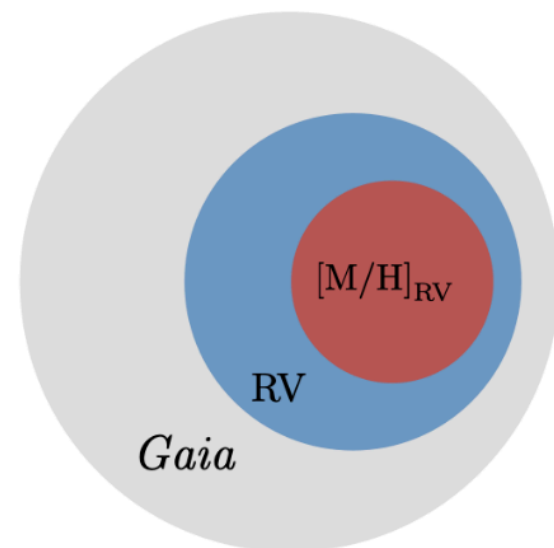**Selection function of the *Gaia* catalogue, …**

$$S(G, \ell, b) = S(G|M_{10}) = 1 - 0.5 \times \left[\tanh\left(\frac{x - G}{y}\right) + 1 \right]^{z}$$

Galactic

[Cantat-Gaudin…**ACG**+23]

4.01001e-11   Gaia SF at G = 21   1

Some other Gaia selection functions

Selection function of the *Gaia* catalogue, …

$$S(G, \ell, b) = S(G|M_{10}) = 1 - 0.5 \times \left[\tanh\left(\frac{x-G}{y}\right) + 1\right]^z$$

Galactic

[Cantat-Gaudin…**ACG**+23]

4.01001e-11   Gaia SF at G = 21   1

… subsamples of the catalogue …

$$\mathcal{S}([\text{M/H}]_{\text{RV}}) = \mathcal{S}(Gaia) \cdot \mathcal{S}(\text{RV}|Gaia) \cdot \mathcal{S}([\text{M/H}]_{\text{RV}}|\text{RV}, Gaia)$$
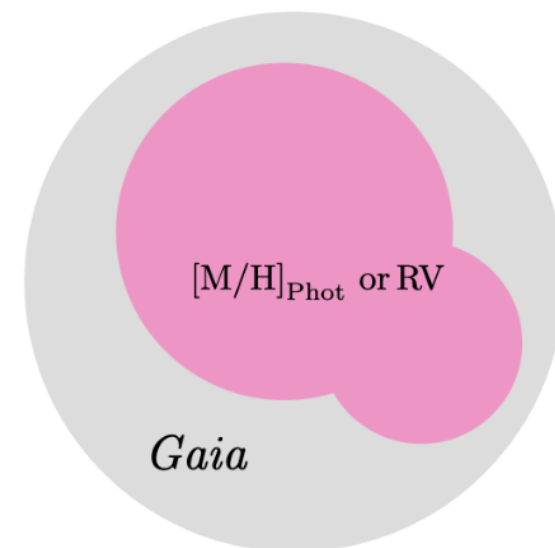
$$\mathcal{S}([\text{M/H}]_{\text{Phot}} \text{ or RV}) = \mathcal{S}(Gaia) \cdot \mathcal{S}([\text{M/H}]_{\text{Phot}} \text{ or RV}|Gaia)$$

[Castro-Ginard+23]

Some other Gaia selection functions
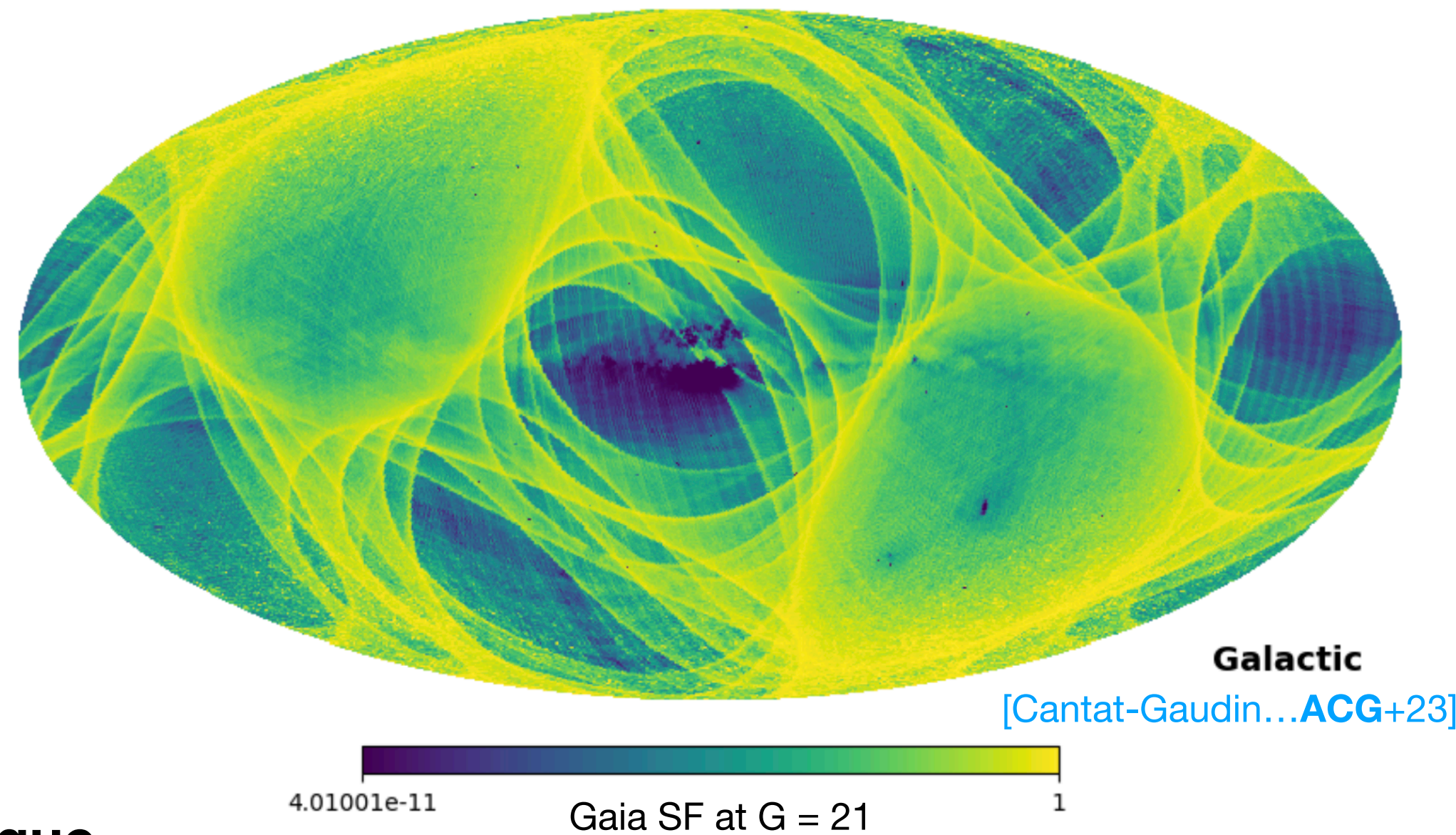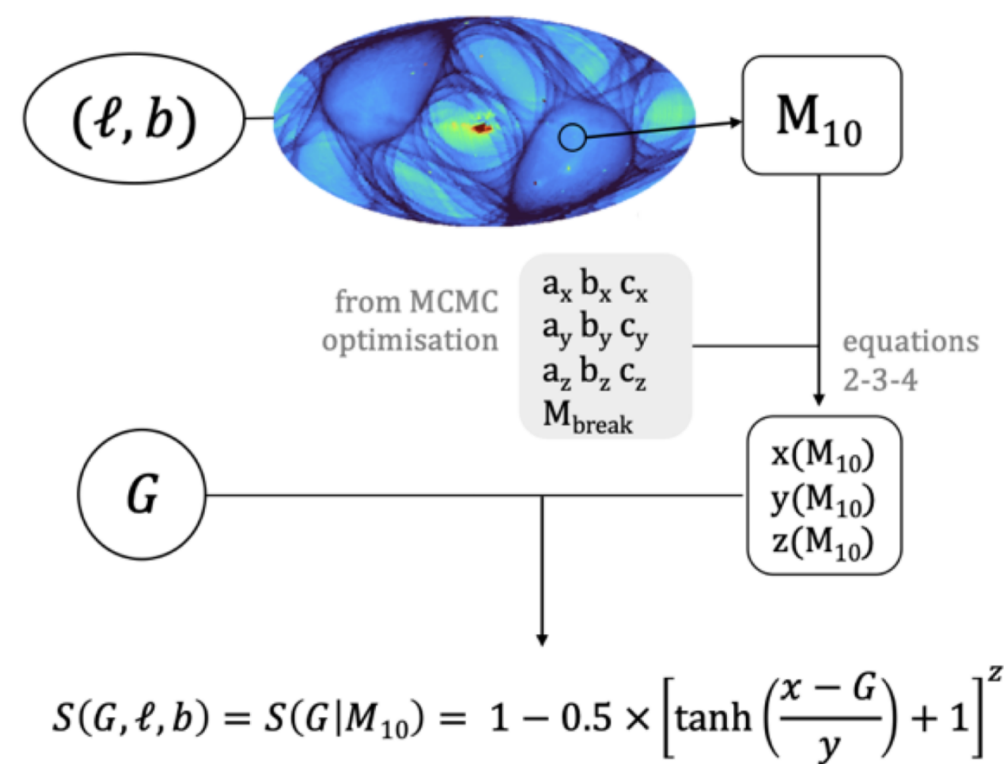
Selection function of the *Gaia* catalogue, …

$$S(G, \ell, b) = S(G|M_{10}) = 1 - 0.5 \times \left[\tanh\left(\frac{x-G}{y}\right) + 1\right]^z$$

Galactic

[Cantat-Gaudin…**ACG**+23]

4.01001e-11   Gaia SF at G = 21   1

… subsamples of the catalogue …
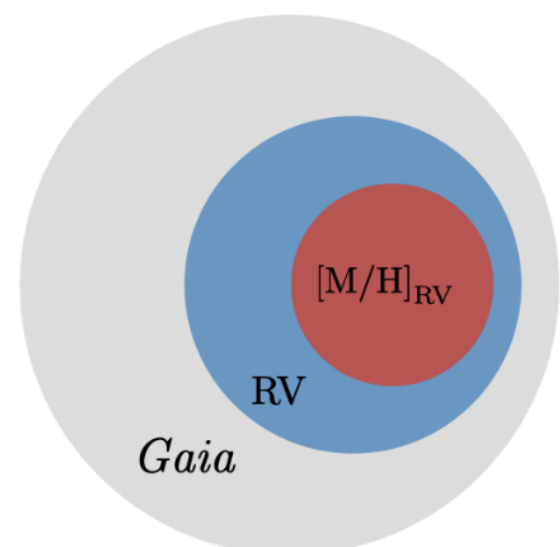
$$\mathcal{S}([\text{M/H}]_{\text{RV}}) = \mathcal{S}(Gaia) \cdot \mathcal{S}(\text{RV}|Gaia) \cdot \mathcal{S}([\text{M/H}]_{\text{RV}}|\text{RV}, Gaia)$$
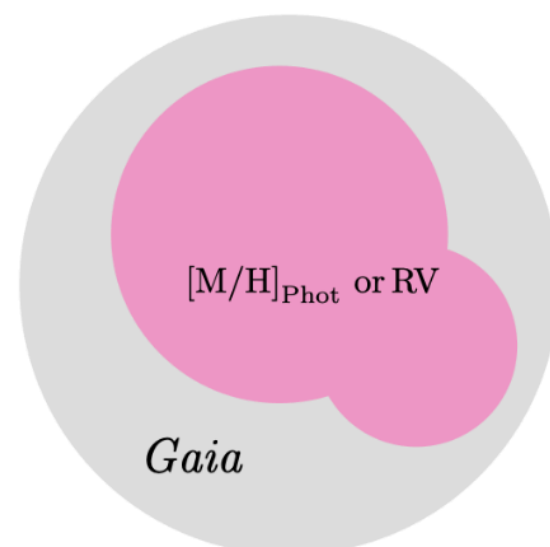
$$\mathcal{S}([\text{M/H}]_{\text{Phot}} \text{ or RV}) = \mathcal{S}(Gaia) \cdot \mathcal{S}([\text{M/H}]_{\text{Phot}} \text{ or RV}|Gaia)$$

[Castro-Ginard+23]

… and binary systems

Stellar Binary   eDR3   DR2

δ Dec [mas]

δ RA* [mas]
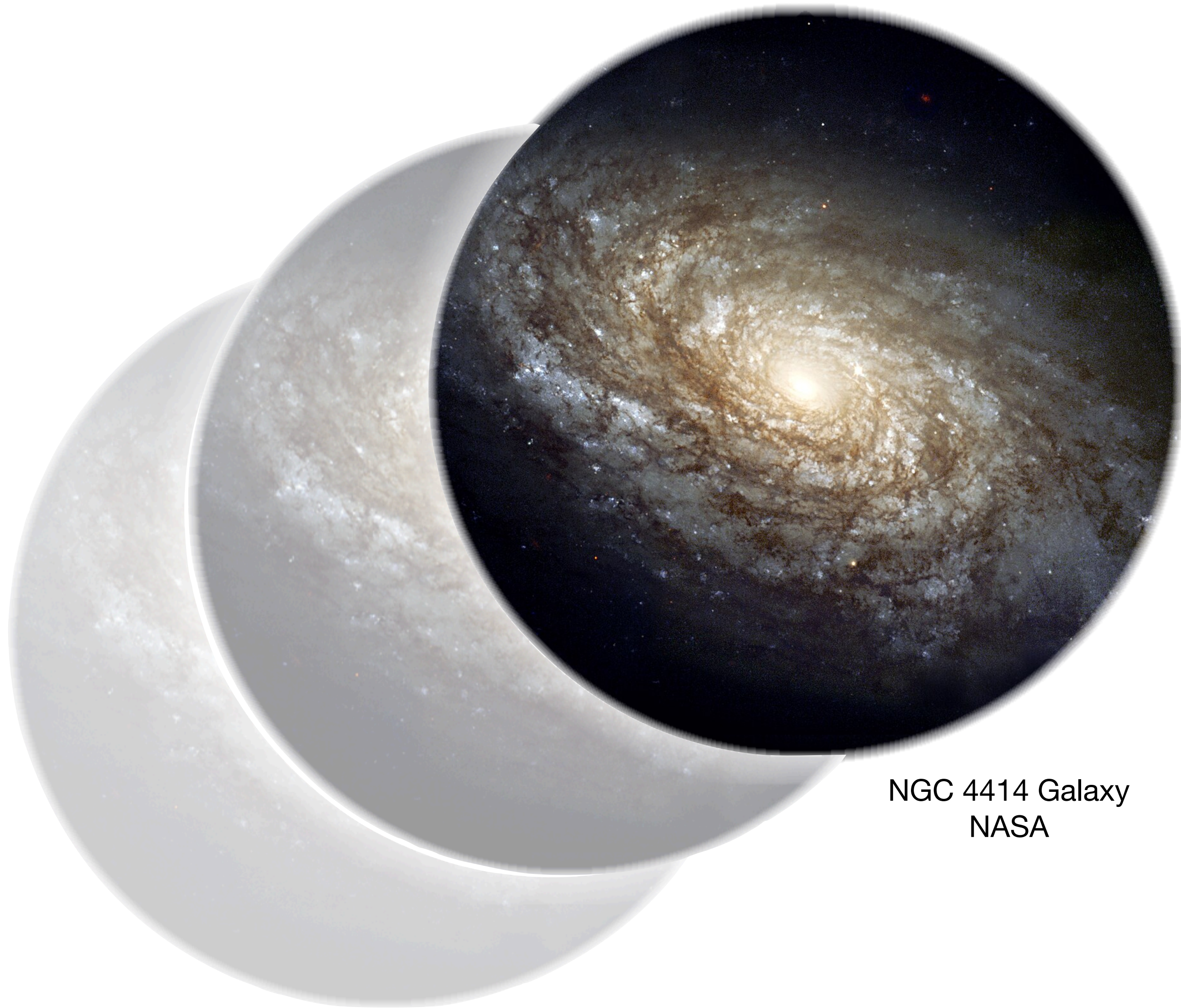
[Penoyre+21]

Computed ruwe threshold

Galactic

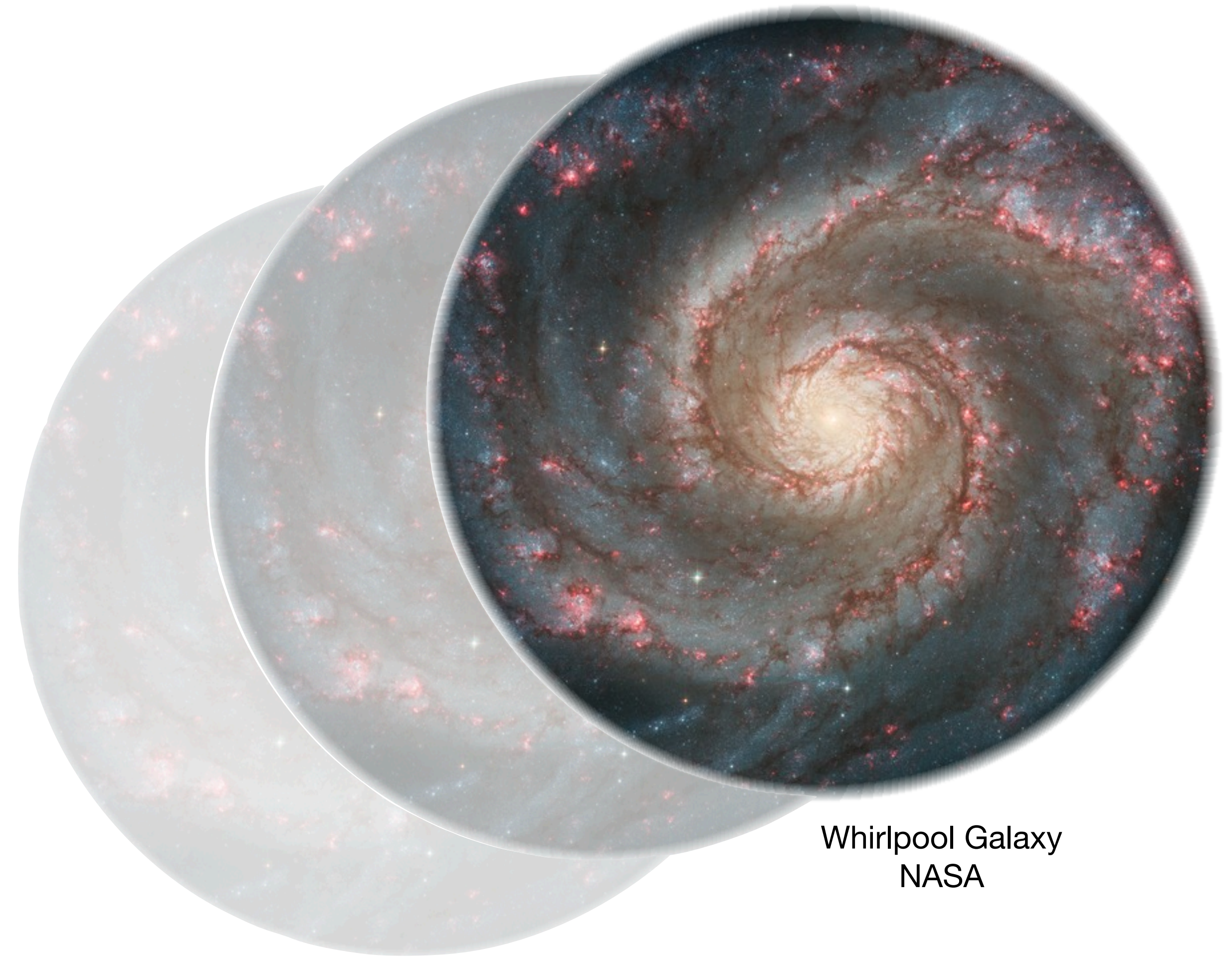[Castro-Ginard+24]

1.15252   1.36345
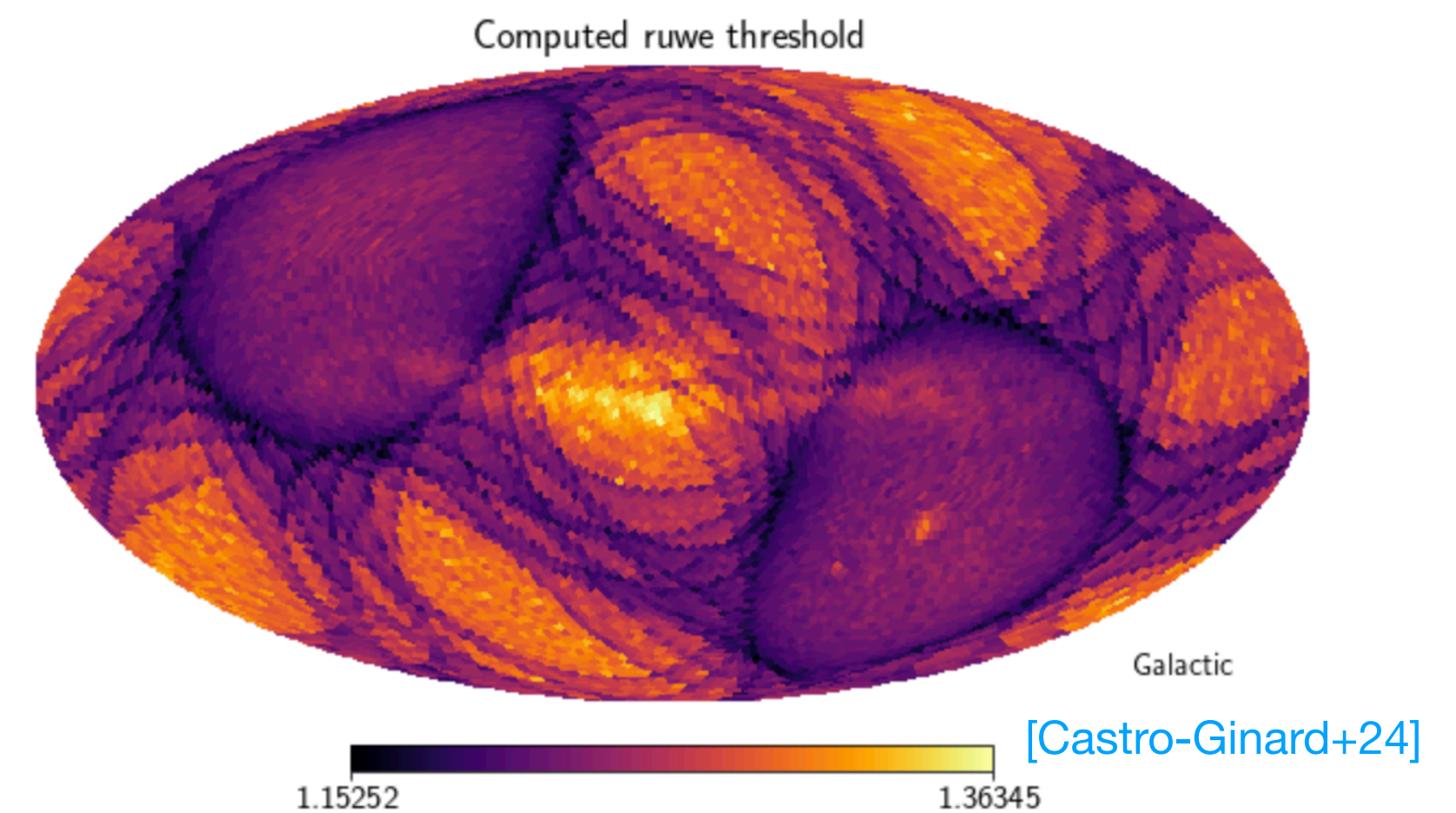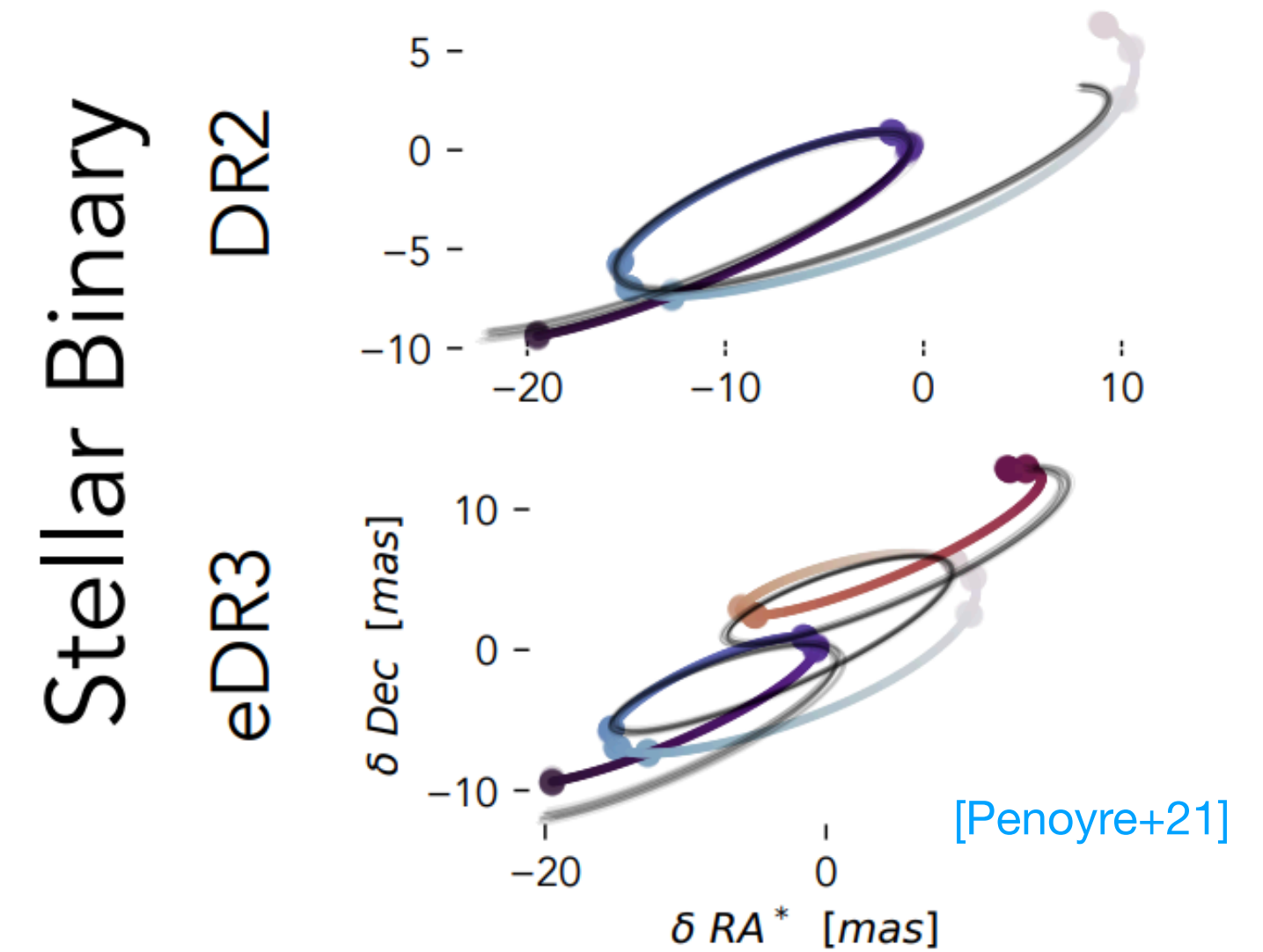
Conclusions

NGC 4414 Galaxy
NASA

Whirlpool Galaxy
NASA

- Need to choose adequate tracers: statistically and astrophysically robust

- Need to know our catalogue caveats and characterise its biases