New Herbig Ae/Be and Classical Be candidates using Gaia and Machine Learning

Miguel Vioque
University of Leeds

R. D. Oudmaijer (University of Leeds, UK), M. Schreiner (Desupervised, Denmark), D. Baines (ESAC, Spain), and R. Pérez-Martínez (Isdefe, Spain)

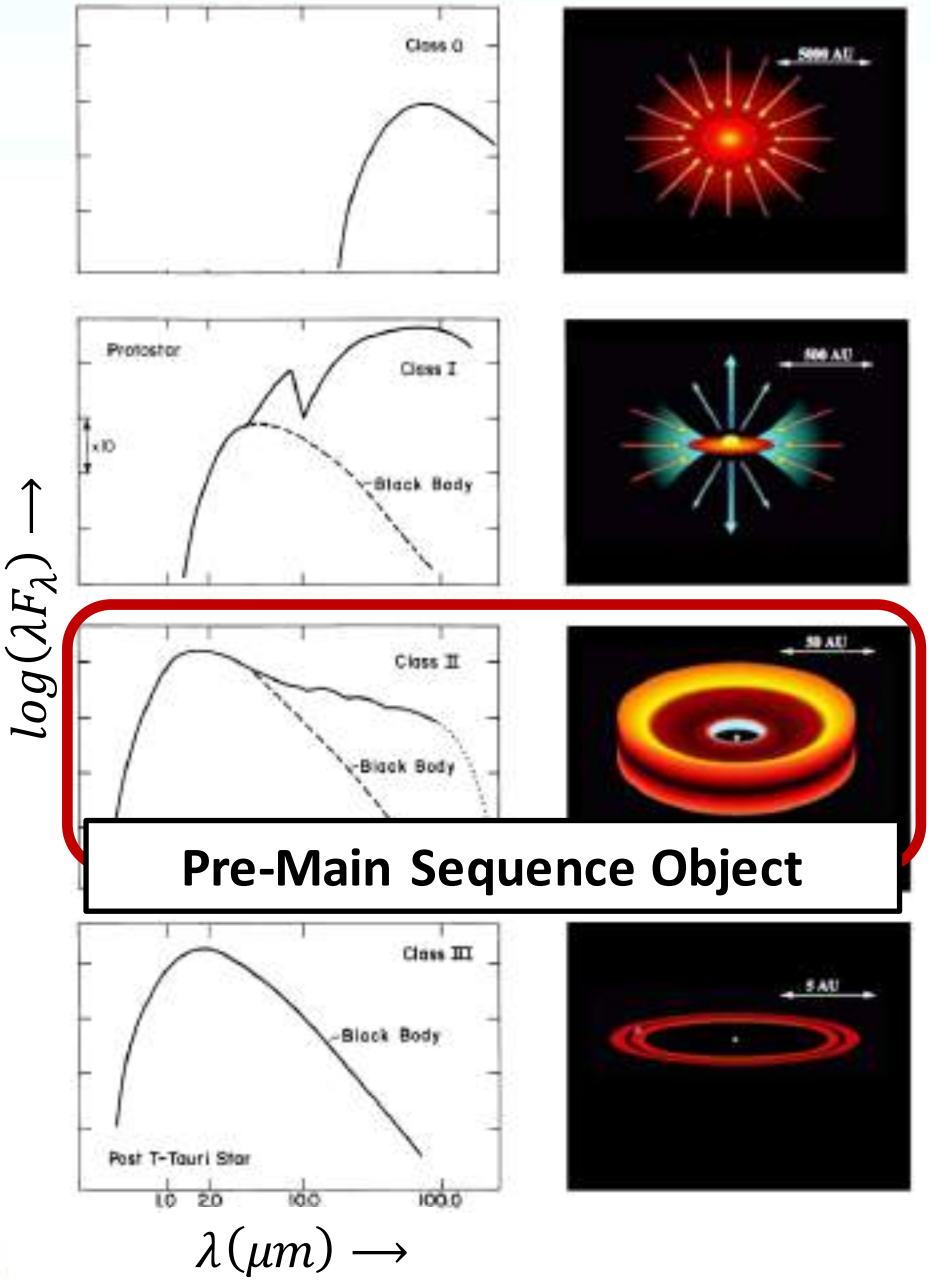
STARRY
STARS THAT 'R' YOUNG

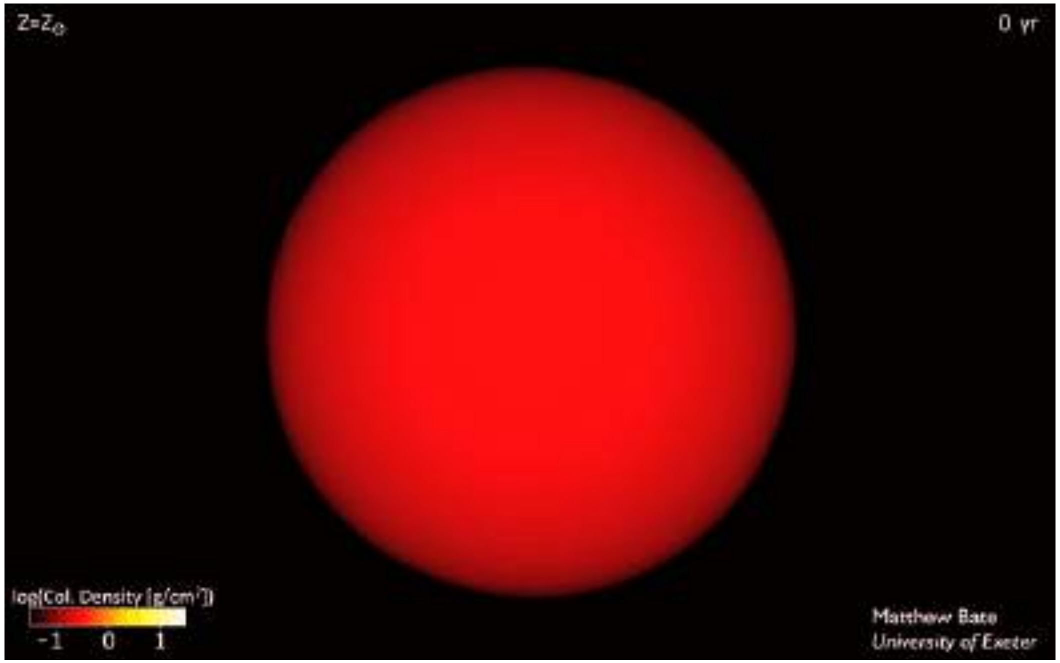
MARIE CURIE ACTIONS

Isdefe

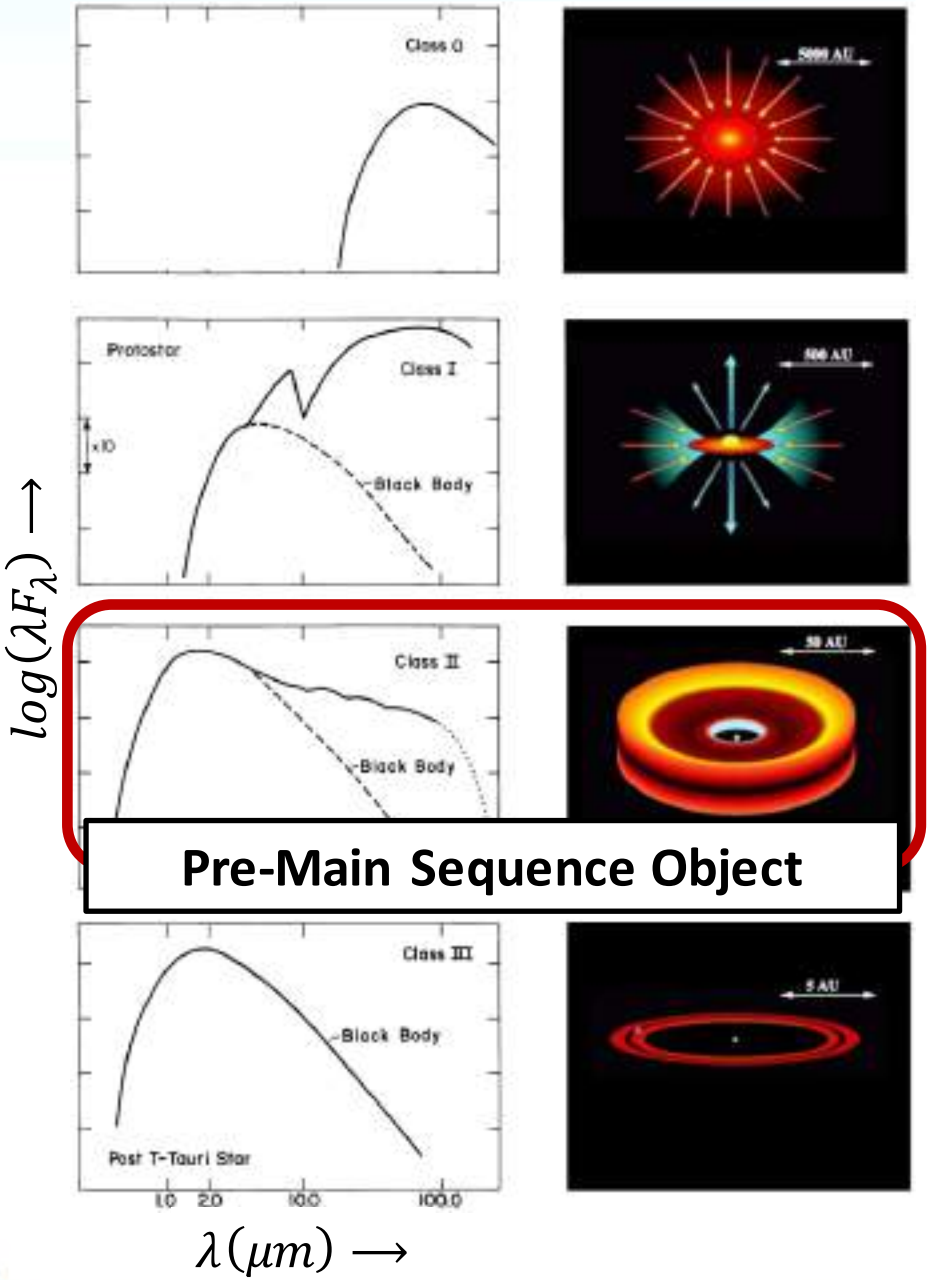
Seminar, ESAC, 12th of September 2019

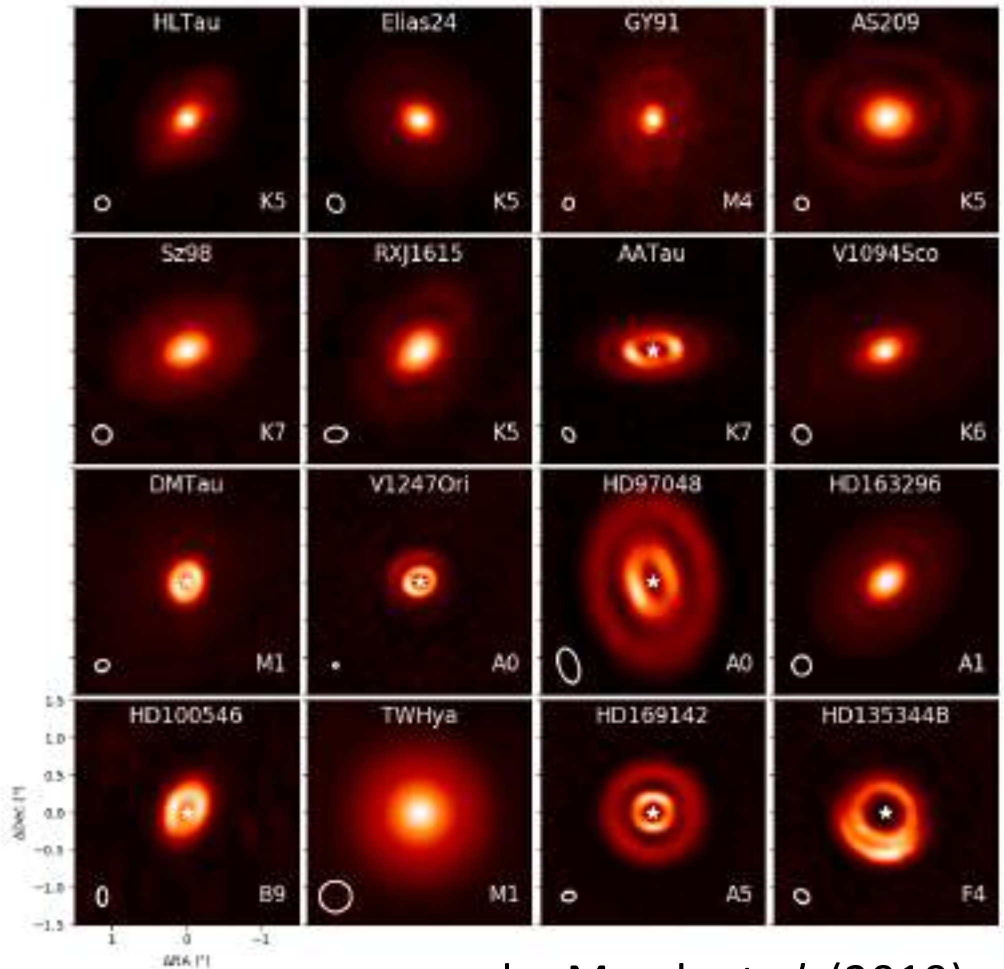
Star Formation in a nutshell





Star Formation in a nutshell

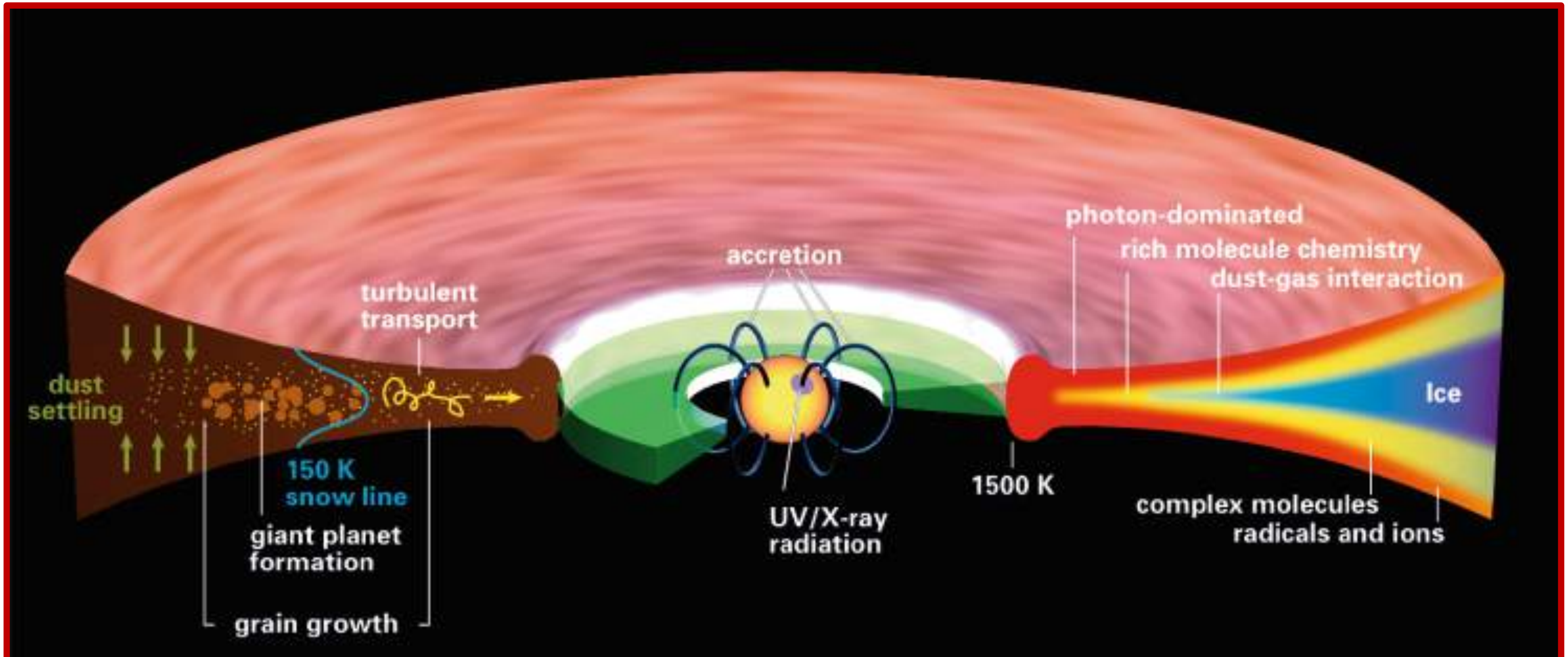




van der Marel, *et al.* (2019)

Star Formation in a nutshell

Low mass Pre-Main Sequence accretion

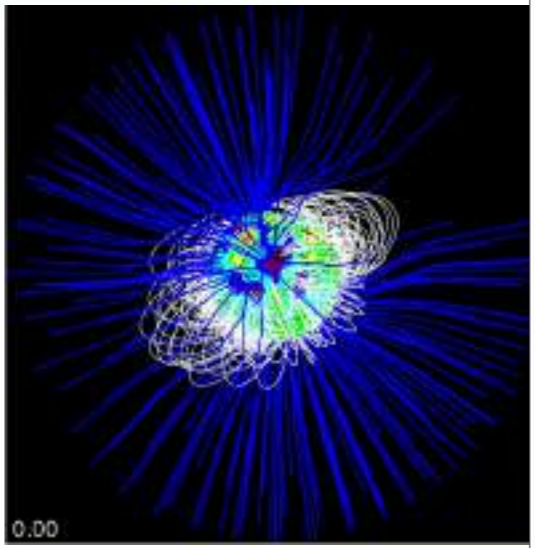


de Wit W., *et al.* (2014)

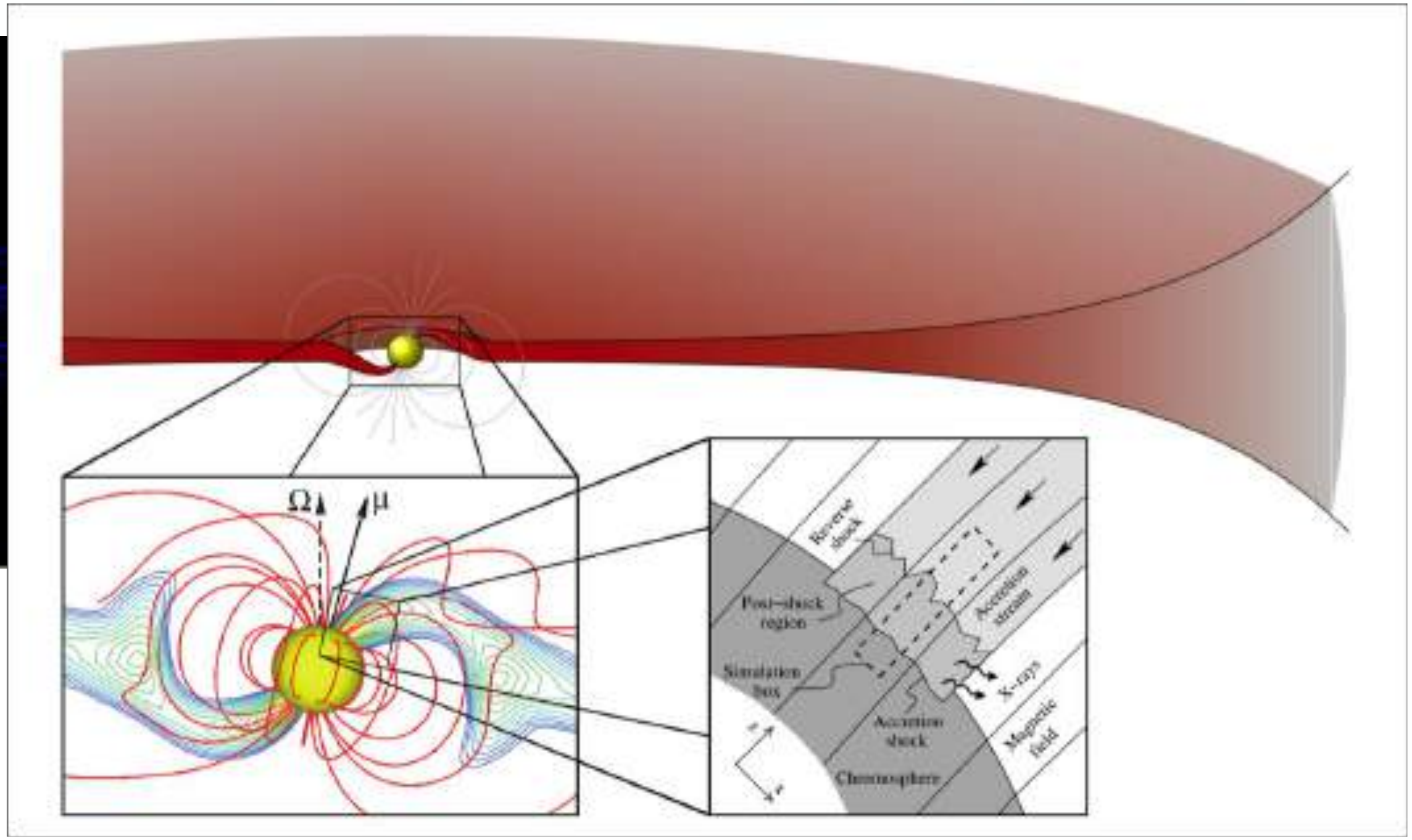
However, high mass Pre-Main Sequence accretion is still largely not understood

Star Formation in a nutshell

Low mass Pre-Main Sequence accretion



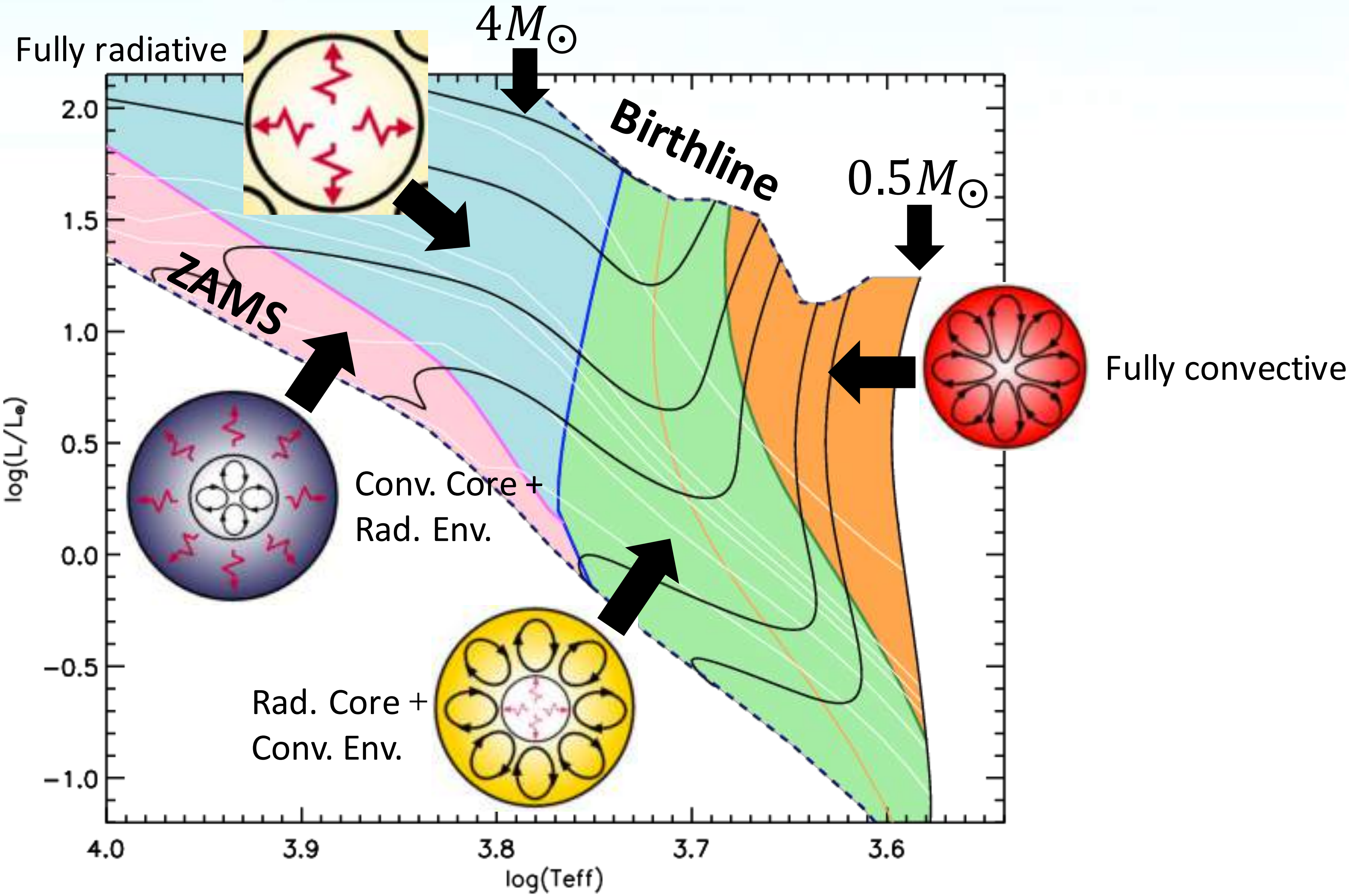
Hill, *et al.* (2019)



de Sá, *et al.* (2014)

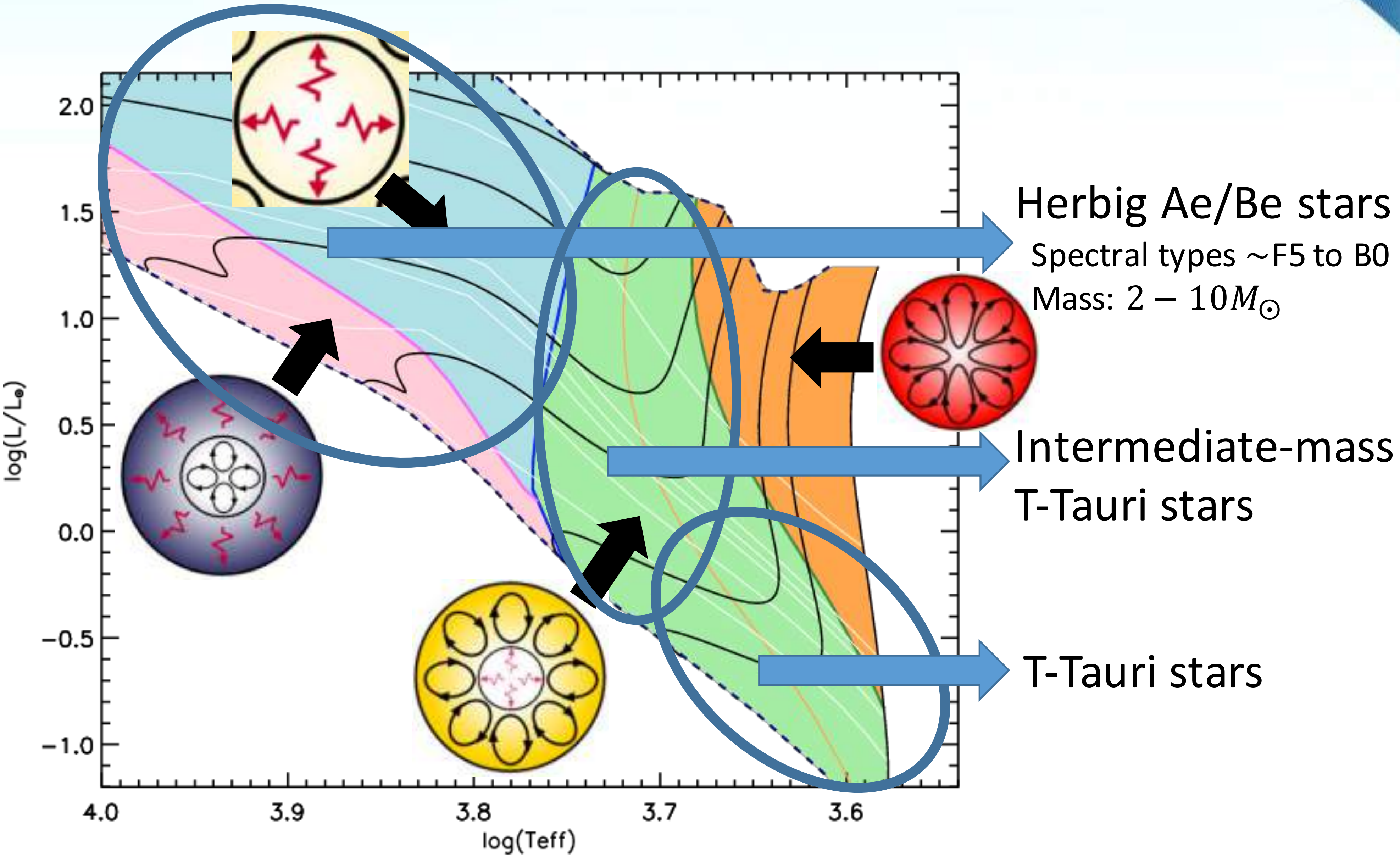
However, high mass Pre-Main Sequence accretion is still largely not understood

Star Formation in a nutshell



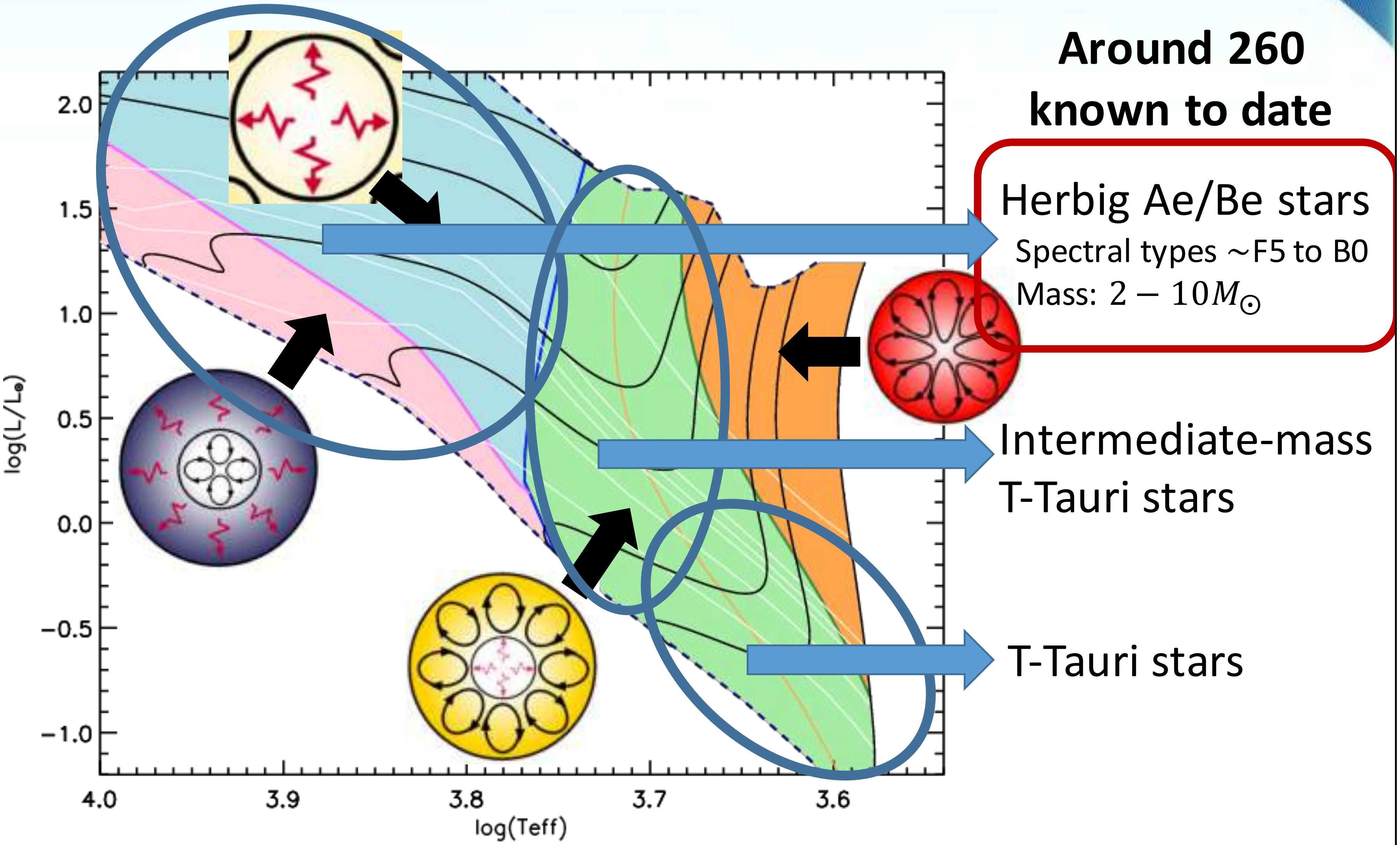
Alecian, *et al.* (2013), Villebrun, *et al.* (2019)

Star Formation in a nutshell



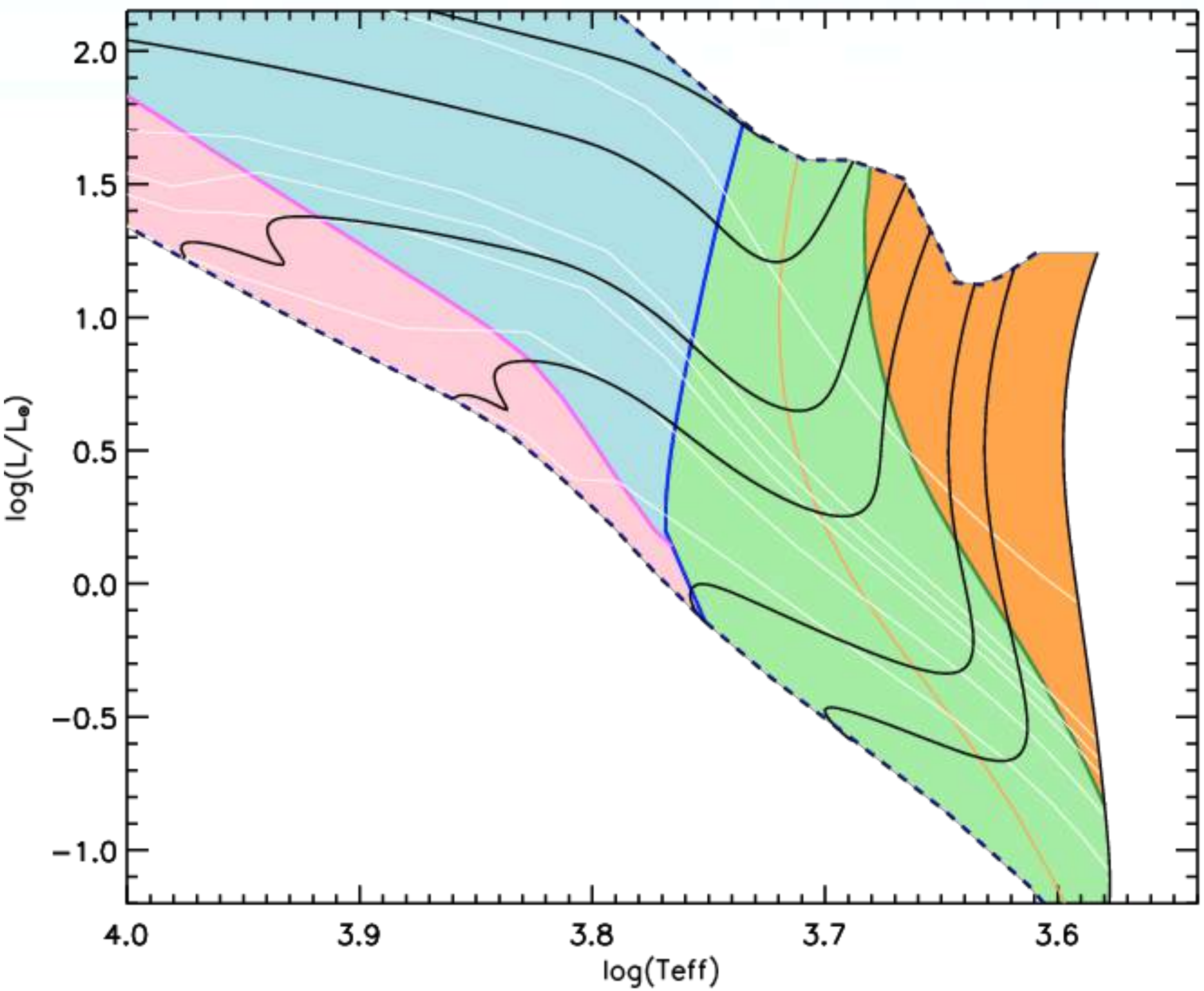
Alecian, *et al.* (2013), Villebrun, *et al.* (2019)

Star Formation in a nutshell



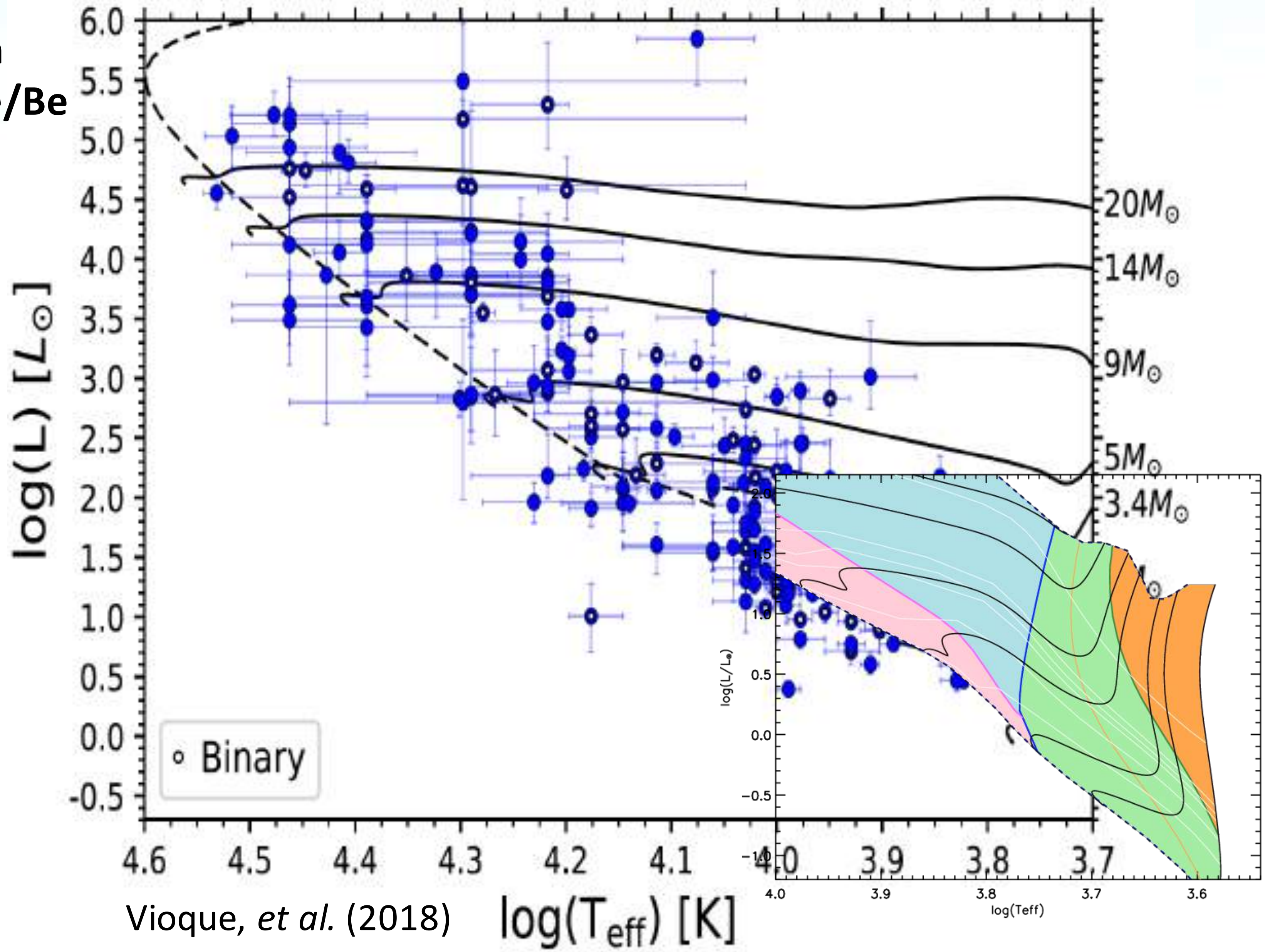
Alecian, *et al.* (2013), Villebrun, *et al.* (2019)

Star Formation in a nutshell



Star Formation in a nutshell

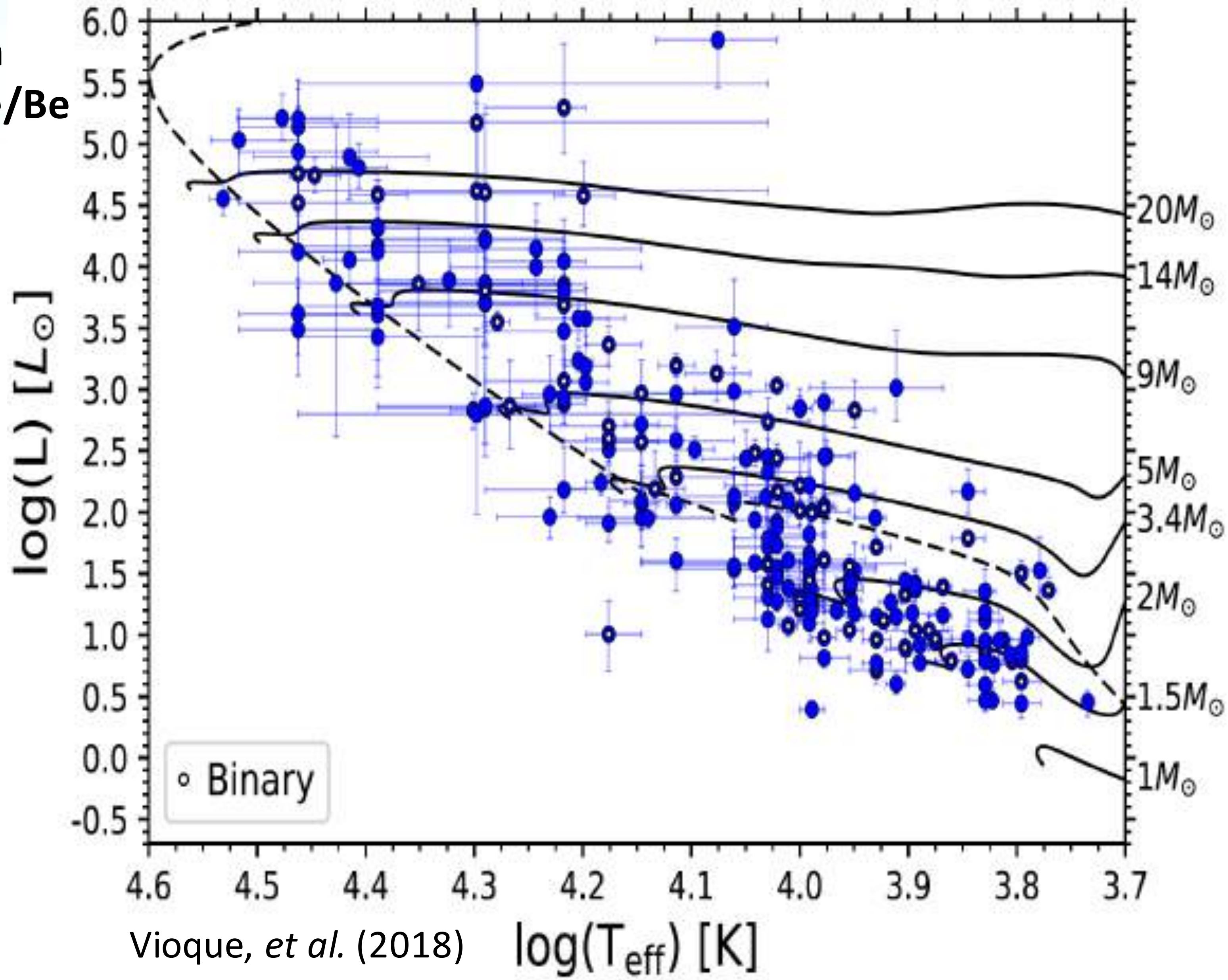
All known Herbig Ae/Be stars



Vioque, *et al.* (2018)

Star Formation in a nutshell

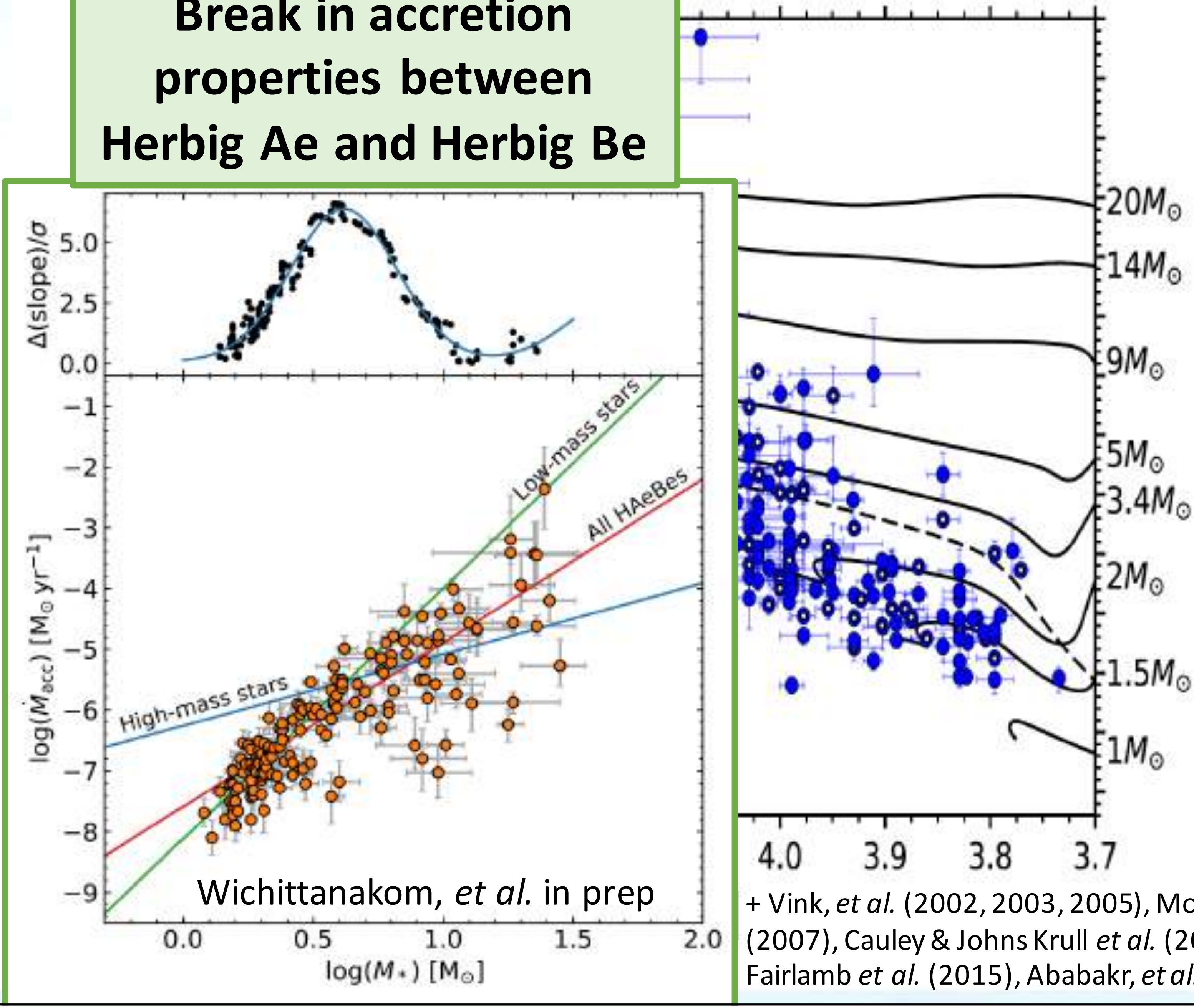
All known Herbig Ae/Be stars



Vioque, *et al.* (2018)

Star Formation in a nutshell

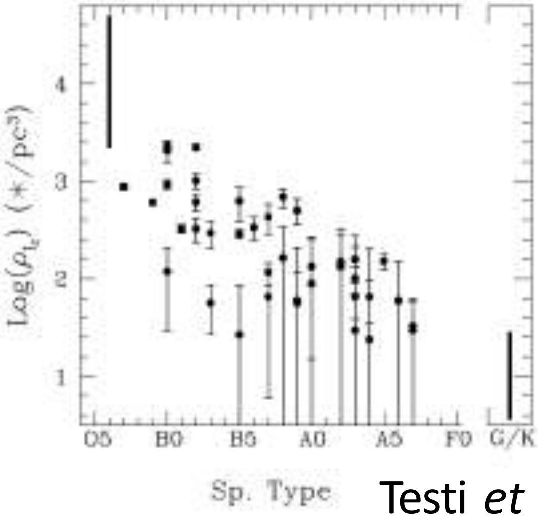
Break in accretion properties between Herbig Ae and Herbig Be



+ Vink, *et al.* (2002, 2003, 2005), Mottram, *et al.* (2007), Cauley & Johns Krull *et al.* (2014, 2015), Fairlamb *et al.* (2015), Ababakr, *et al.* (2017).

Some other open questions

Clustering

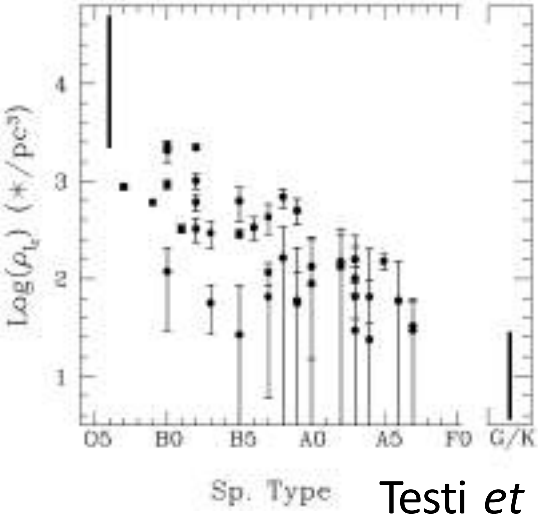


Testi *et al.* (1999)

Do all massive stars form in clusters?

Some other open questions

Clustering



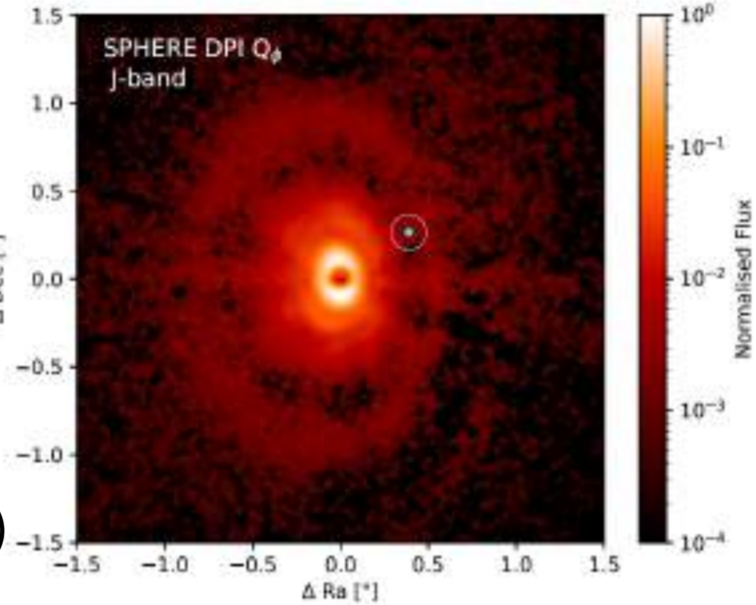
Testi *et al.* (1999)

Do all massive stars form in clusters?

Planet Formation

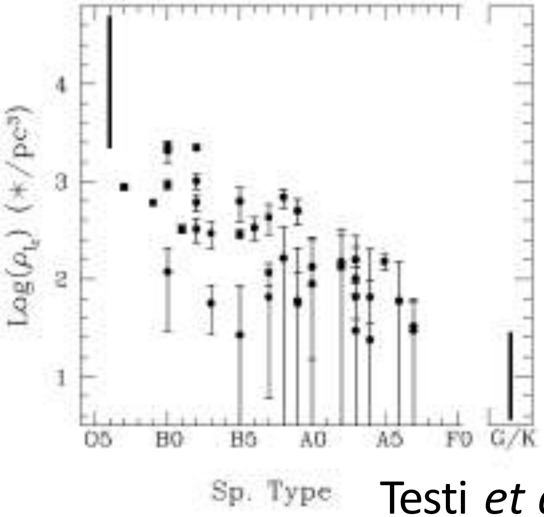
Planet forming or planet hosting disks - when do we start seeing signatures of planet? Class 1? Class 0?

Pinte *et al.* (2019)



Some other open questions

Clustering



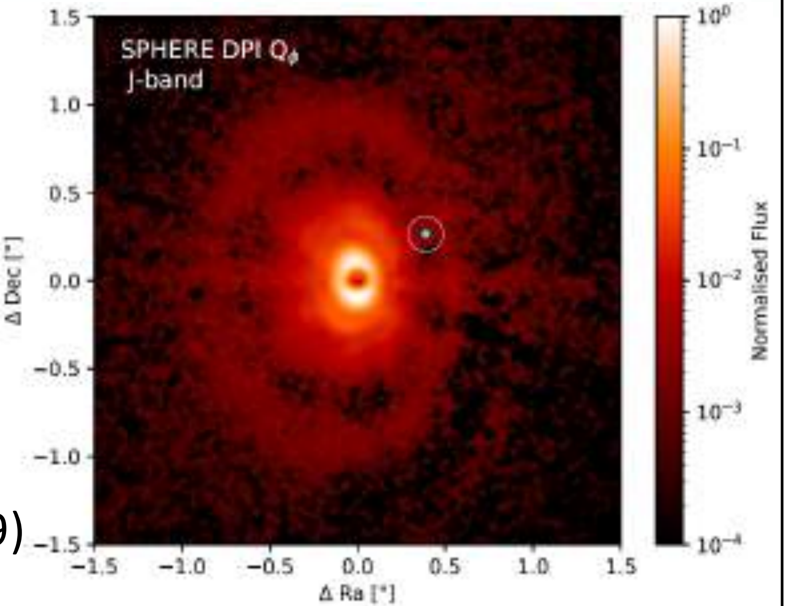
Testi *et al.* (1999)

Do all massive stars form in clusters?

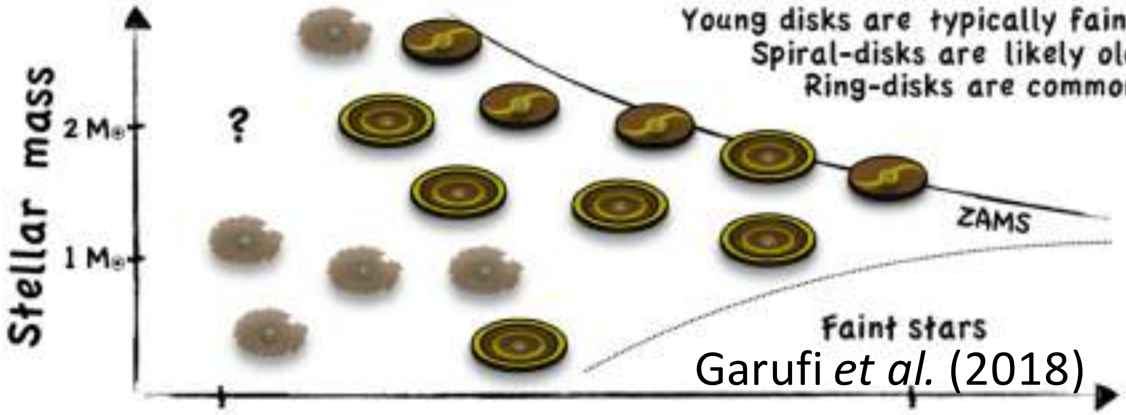
Planet Formation

Planet forming or planet hosting disks - when do we start seeing signatures of planet? Class 1? Class 0?

Pinte *et al.* (2019)



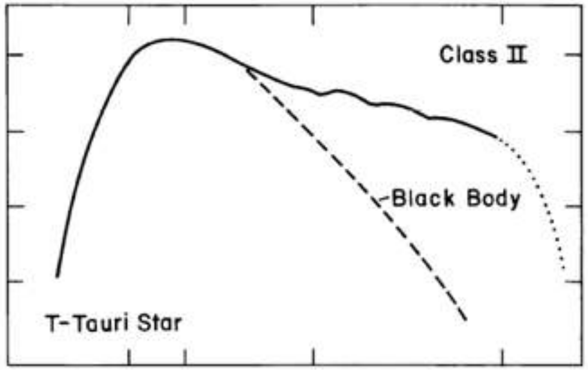
Disk Structure and Evolution

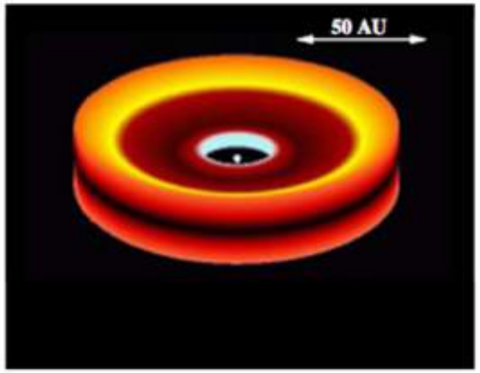


Garufi *et al.* (2018)

- *Is the spiral structure caused by planets or some other effect?*
- *Why do disks around Herbig Ae/Be stars dissipate faster but some seem to persist for very long times?*

Looking for new Pre-Main Sequence (PMS) objects in Gaia!

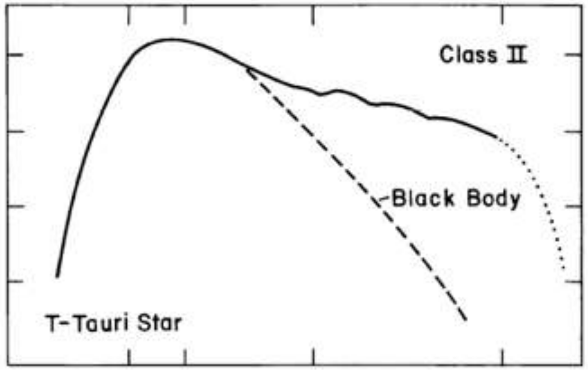


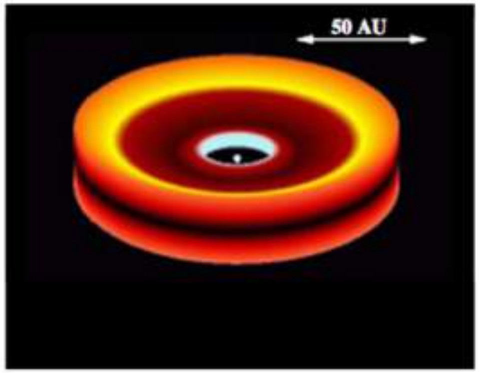


Main characteristics of PMS objects:

- Infrared excesses
- Hα emission
- Photometric variability

Looking for new Pre-Main Sequence (PMS) objects in Gaia!



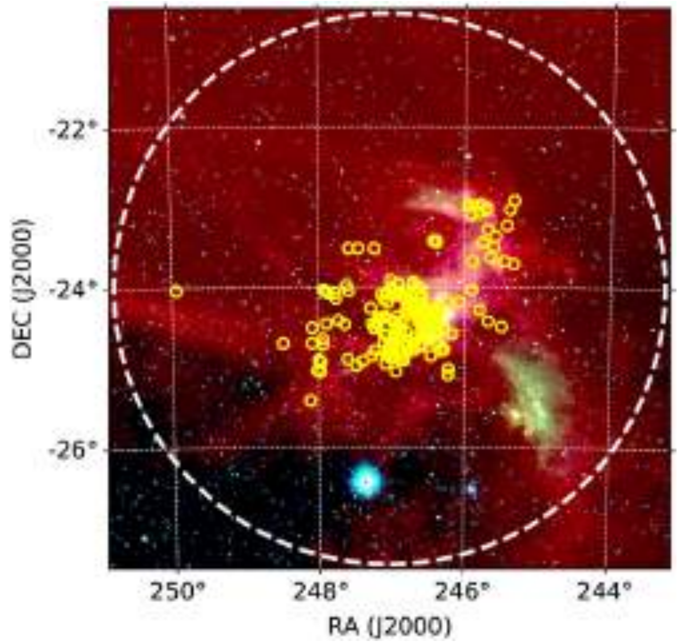


Main characteristics of PMS objects:

- Infrared excesses
- Hα emission
- Photometric variability

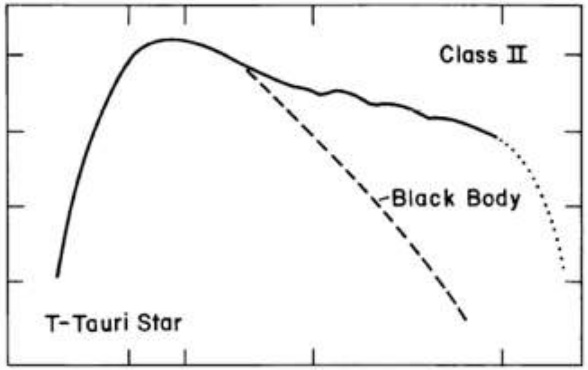
"Easy" to do for low-mass objects:

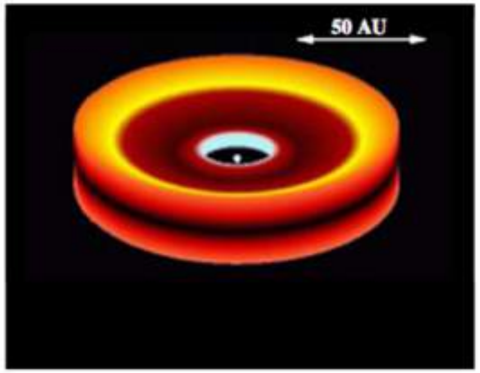
eg., Ksoll et al. (2018), Marton et al. (2019)



Cánovas et al. (2019)

Looking for new Pre-Main Sequence (PMS) objects in Gaia!

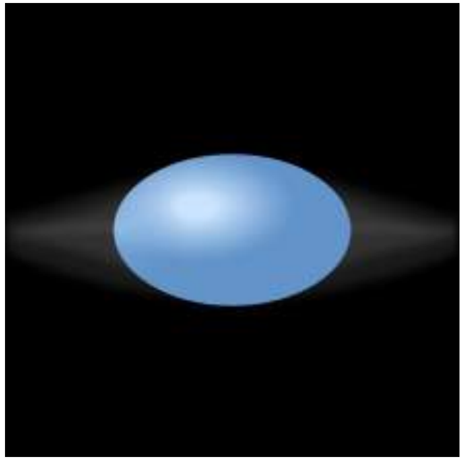




Main characteristics of PMS objects:

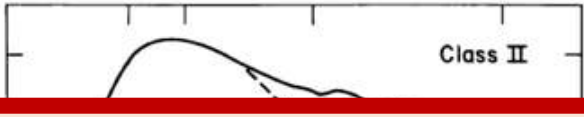
- Infrared excesses
- Hα emission
- Photometric variability

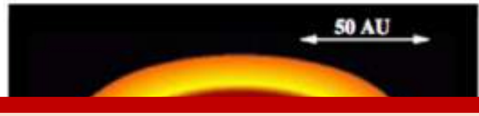
High mass PMS objects (Herbig Be stars) are very similar to **Classical Be stars**



... and supergiants, B[e] stars, ...

Looking for new Pre-Main Sequence (PMS) objects in Gaia!





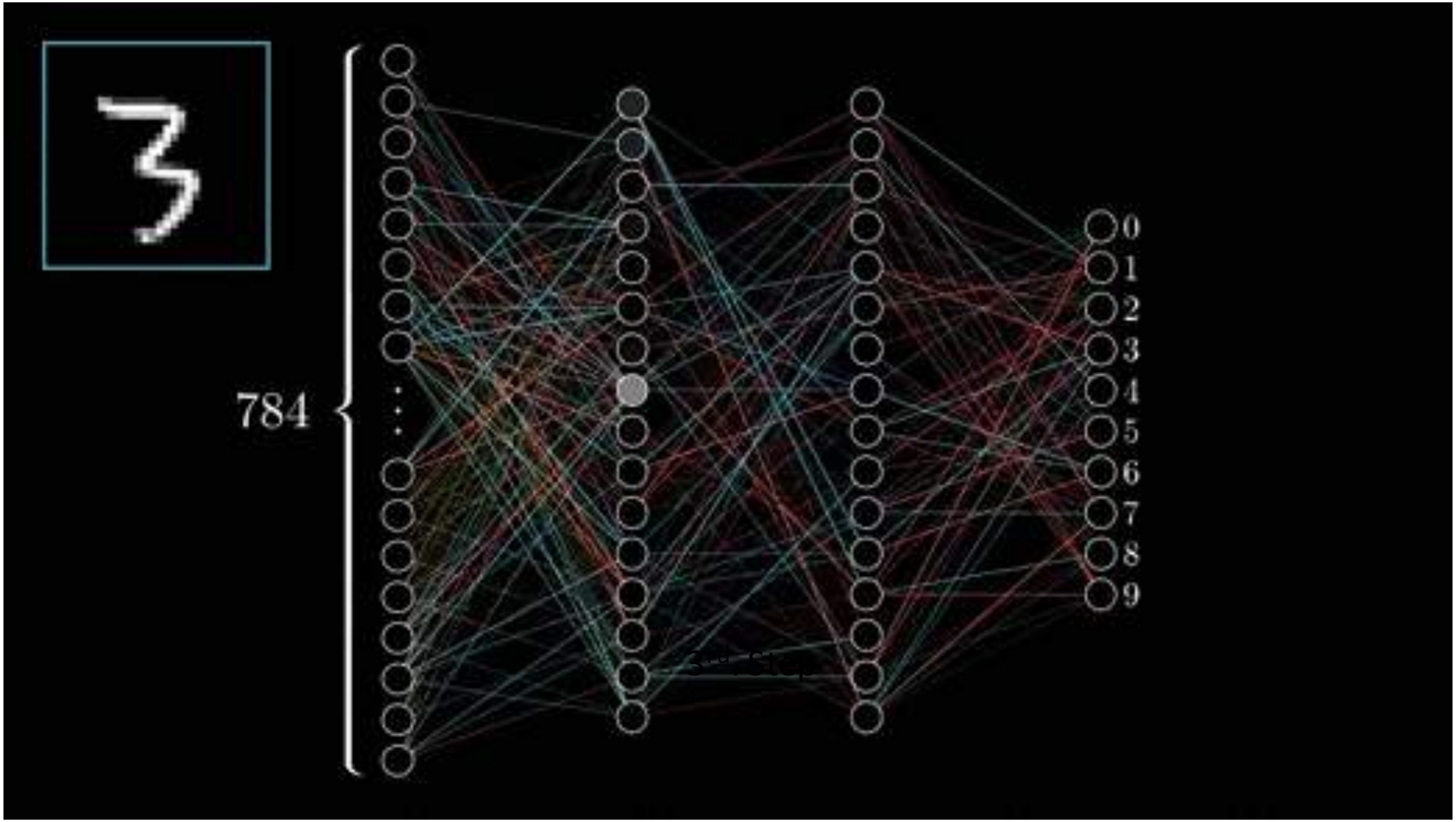
Perform an homogeneous selection, distance and position independent!

High mass PMS objects (Herbig Be stars) are very similar to **Classical Be stars**

... and supergiants, B[e] stars, ...

Neural Network

1st Step: The algorithm learns from labeled data (**Training Set**) through minimizing error in successive iterations



Credit: 3Blue1Brown

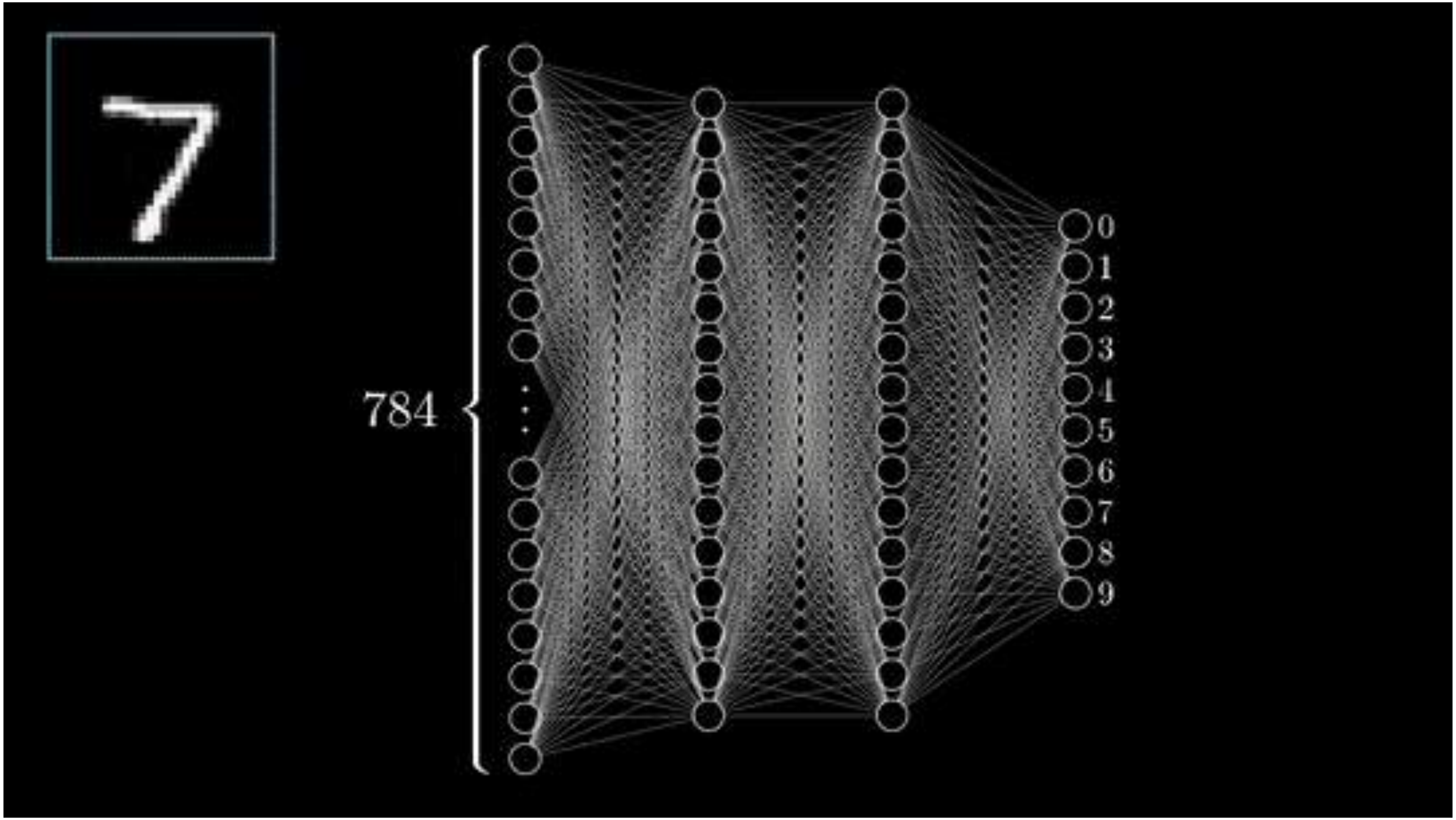
Neural Network

2nd Step: The algorithm improves itself by finding non-trivial patterns hidden within the data

Credit: 3Blue1Brown

Neural Network

3rd Step: Trained algorithm can be applied to data it hasn't seen before (and hopefully it is good at classifying it)

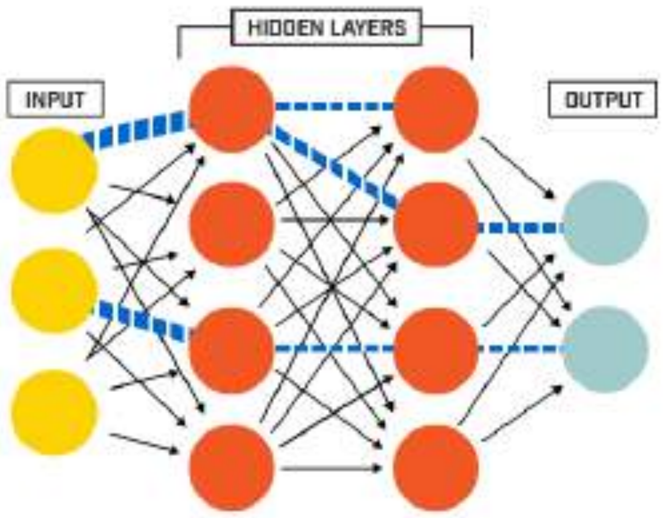


Credit: 3Blue1Brown

Neural Network

Algorithm is trained with known labeled data

Before training:
- Training Set
- Set of characteristics
- Set of categories



After generalizing:
- Each category gets a probability
- Efficiency of the algorithm

The best architecture is selected

Neural Network

Selection of the **characteristics**:

Neural Network

Selection of the **characteristics**:

- Infrared excesses

Neural Network

Selection of the **characteristics**:

AllWISE (WISE+2MASS)

- Infrared excesses

$W1, W2, W3, W4$

J, H, K_s

Neural Network

Selection of the **characteristics**:

AllWISE (WISE+2MASS)

$\boldsymbol{W1, W2, W3, W4}$

$\boldsymbol{J, H, K_s}$

- Infrared excesses

- Hα emission

Neural Network

Selection of the **characteristics**:

AllWISE (WISE+2MASS)

- Infrared excesses

$W1, W2, W3, W4$

J, H, K_s

IPHAS

VPHAS+

- Hα emission

$r - H_\alpha$

Neural Network

Selection of the **characteristics**:

AllWISE (WISE+2MASS)

$W1, W2, W3, W4$

J, H, K_s

- Infrared excesses

IPHAS

VPHAS+

$r - H_\alpha$

- Hα emission

- Photometric variability

Neural Network

Selection of the **characteristics**:

AllWISE (WISE+2MASS)

- Infrared excesses

$W1, W2, W3, W4$

J, H, K_s

IPHAS

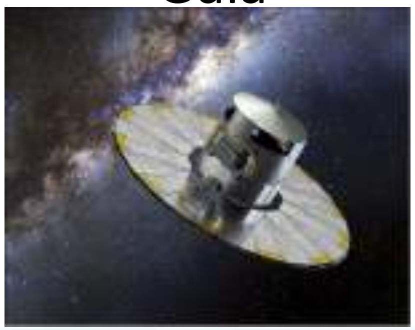
VPHAS+

- Hα emission

$r - H_\alpha$

Gaia

- Photometric variability



2 variability indicators

B_p, G, R_p

Neural Network

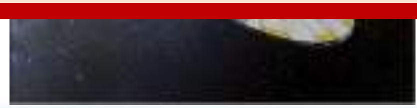
Selection of the **characteristics**:

Create **all possible colours**

Distance and position independent!

Remove all linear dependency **(PCA)**

12 dimensions (**characteristics**)



$W1, W2, W3, W4$

J, H, K_s

$r - H_\alpha$

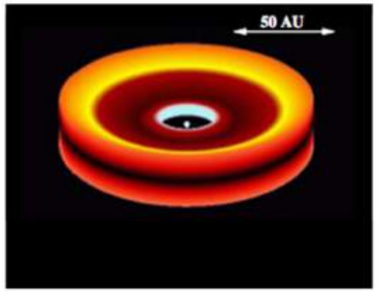
2 variability indicators

B_p, G, R_p

Selection of the **categories**:

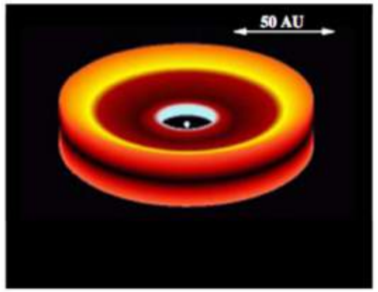
Selection of the **categories**:

PMS category

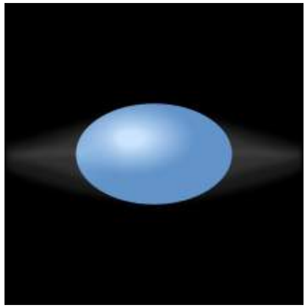


Selection of the **categories**:

PMS category

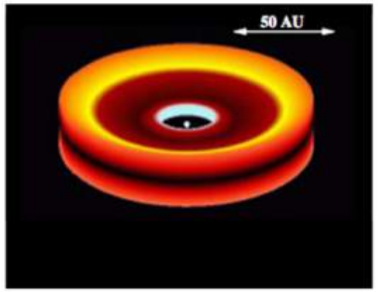


Classical Be category

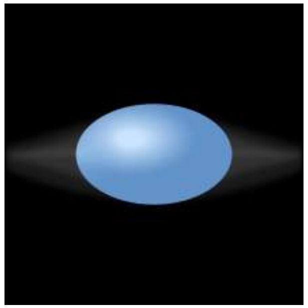


Selection of the **categories**:

PMS category



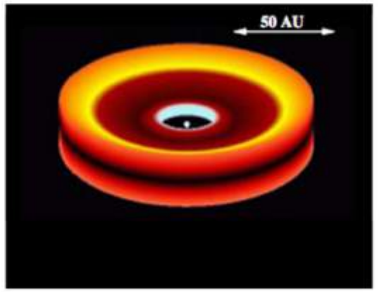
Classical Be category



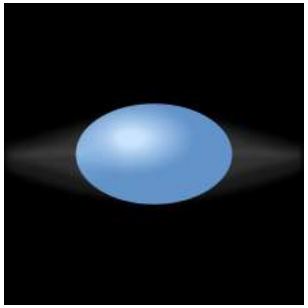
Other sources

Selection of the **Training Set**:

PMS category



Classical Be category



Other sources

Selection of the **Training Set**:

AllWISE

\+

IPHAS VPHAS+

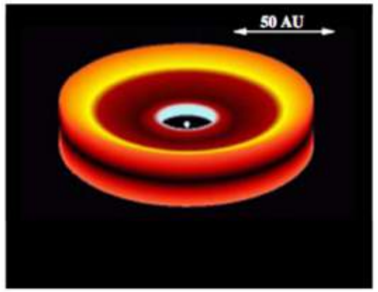
\+

Gaia

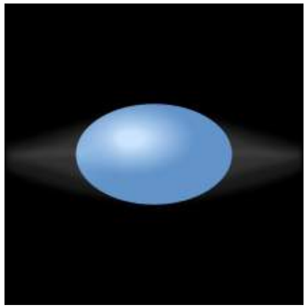
=

4,151,538 sources

PMS category



Classical Be category



Other sources

Selection of the **Training Set**:

AllWISE

\+

IPHAS

VPHAS+

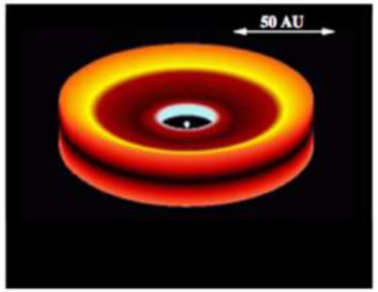
\+

Gaia

=

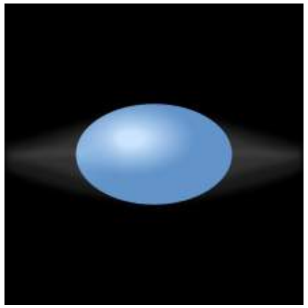
4,151,538 sources

PMS category



- **848** Pre-Main Sequence objects (**163** Herbig Ae/Be)

Classical Be category



- **775** Classical Be stars

Other sources

- **471,111** random sources

Selection of the **Training Set**:

AllWISE

\+

IPHAS

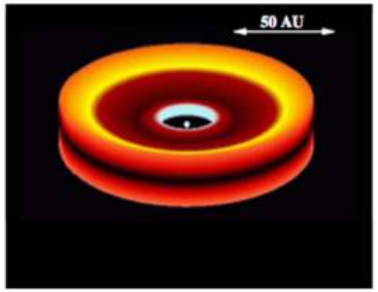
VPHAS+

\+

Gaia

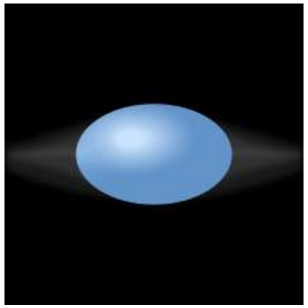
= **4,151,538 sources**

PMS category



- **848** Pre-Main quence objects (**163** H big Ae/Be)

Classical Be category



- **775** Classical stars

Other sources

- **471,111** random sources

Training the Neural Network

Trained Neural Network

AllWISE

+

IPHAS

VPHAS+

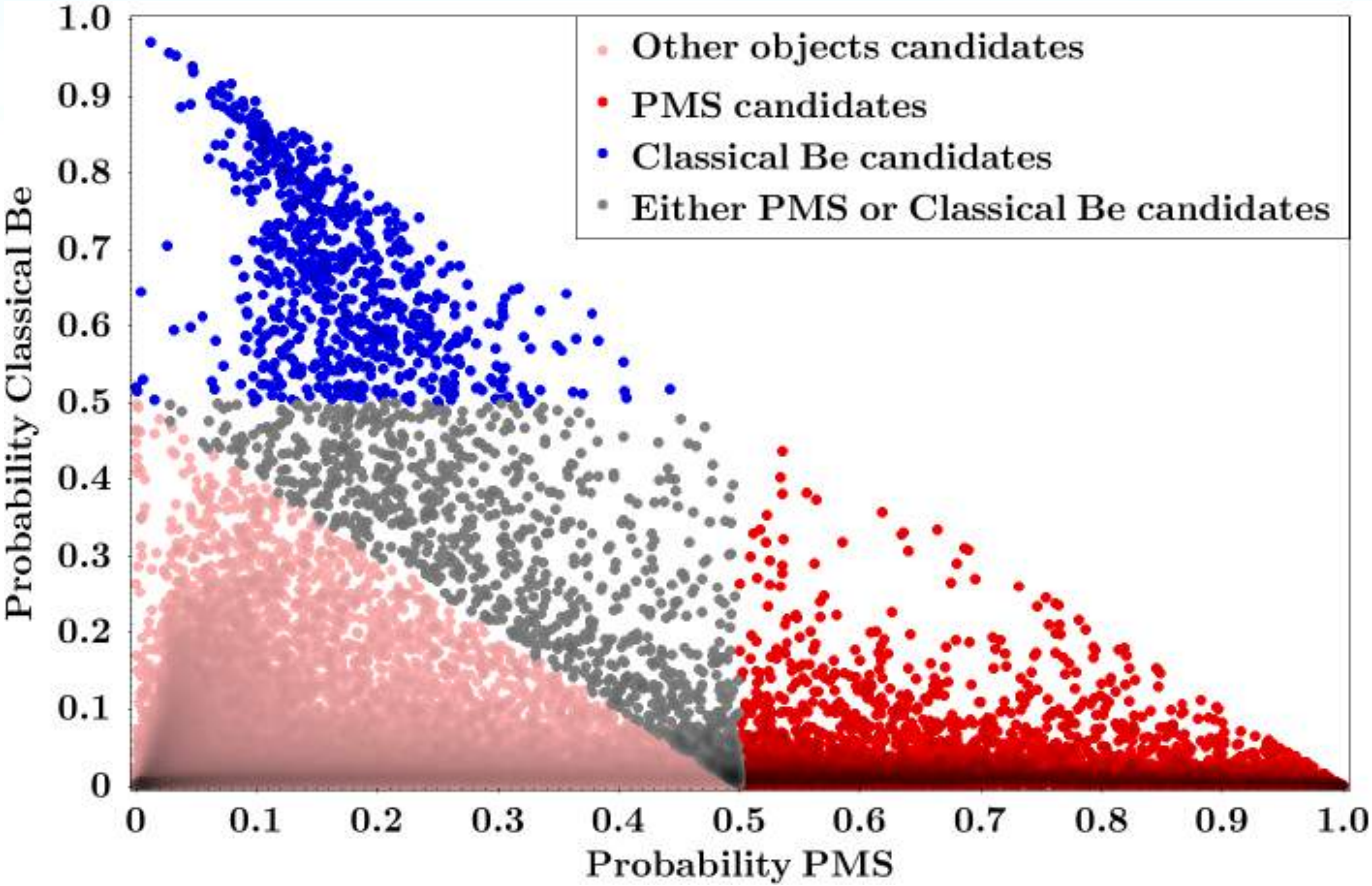
+

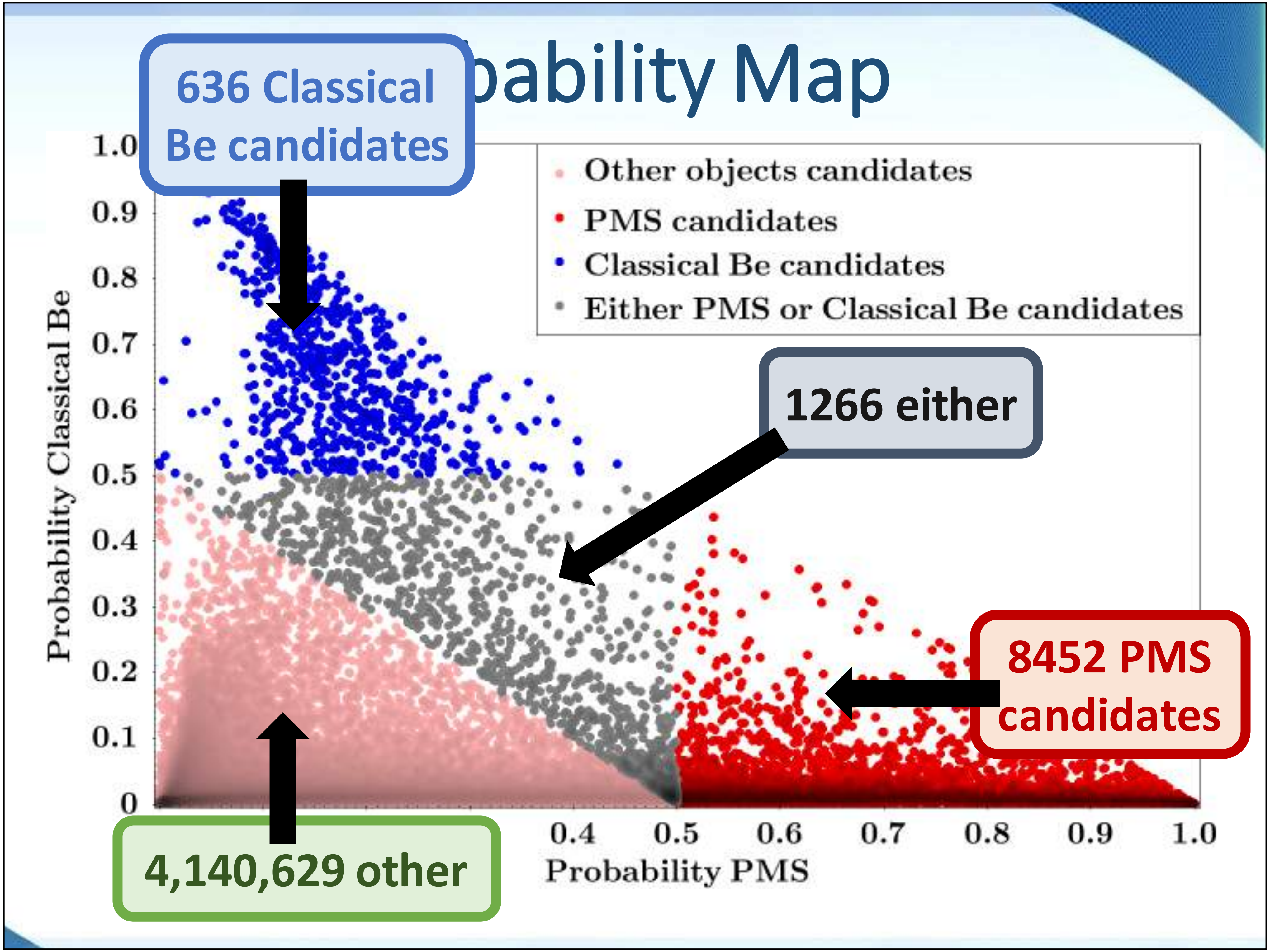
Gaia

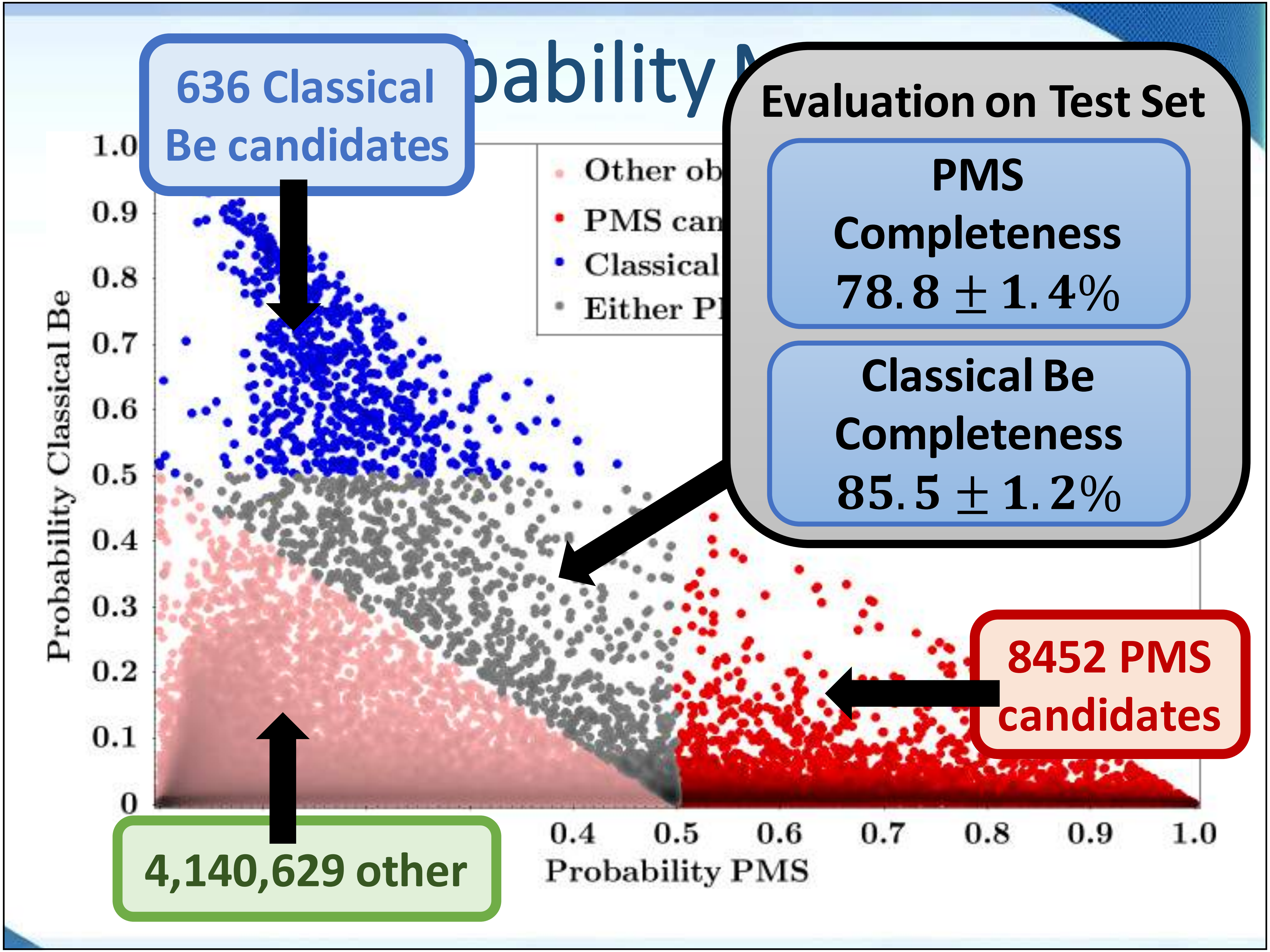
=

4,151,538 sources

Probability Map




bability Map
636 Classical Be candidates
1266 either
8452 PMS candidates
4,140,629 other
Other objects candidates
PMS candidates
Classical Be candidates
Either PMS or Classical Be candidates
Probability Classical Be
Probability PMS
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.4
0.5
0.6
0.7
0.8
0.9
1.0



636 Classical Be candidates
Evaluation on Test Set
PMS Completeness
$78.8 \pm 1.4\%$
Classical Be Completeness
$85.5 \pm 1.2\%$
8452 PMS candidates
4,140,629 other
Probability Classical Be
Probability PMS
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Selection of the **Training Set**:

AllWISE

+

IPHAS

VPHAS+

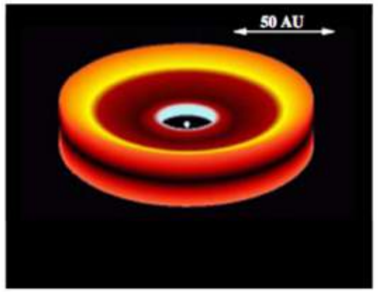
+

Gaia

=

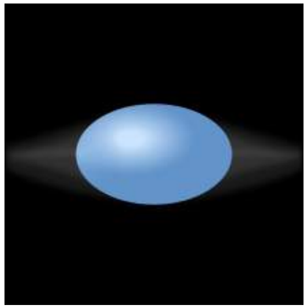
4,151,538 sources

PMS category



- **848** Pre-Main Sequence objects (**163** Herbig Ae/Be)

Classical Be category



- **775** Classical Be stars

Other sources

- **471,111** random sources

Selection of the **Training Set**:

AllWISE

\+

IPHAS

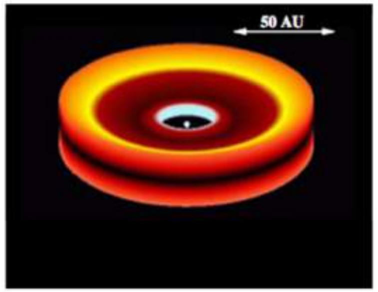
VPHAS+

\+

Gaia

= **4,151,538 sources**

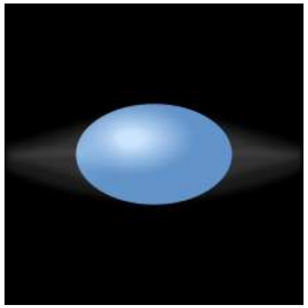
PMS category



- **848** Pr
  object
  Ae/Be

There is a large contamination between categories!

Classical Be category

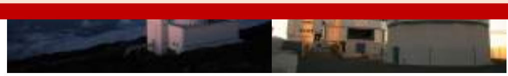


- **775** classical Be stars

Other sources

- **471,111** random sources

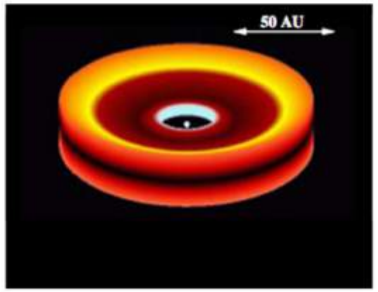
This algorithm cannot assess itself, we need a **totally independent** analysis



sources

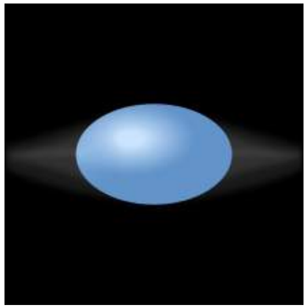
There is a large contamination between categories!

PMS category



- **848** Pr
  object
  Ae/Be

Classical Be category



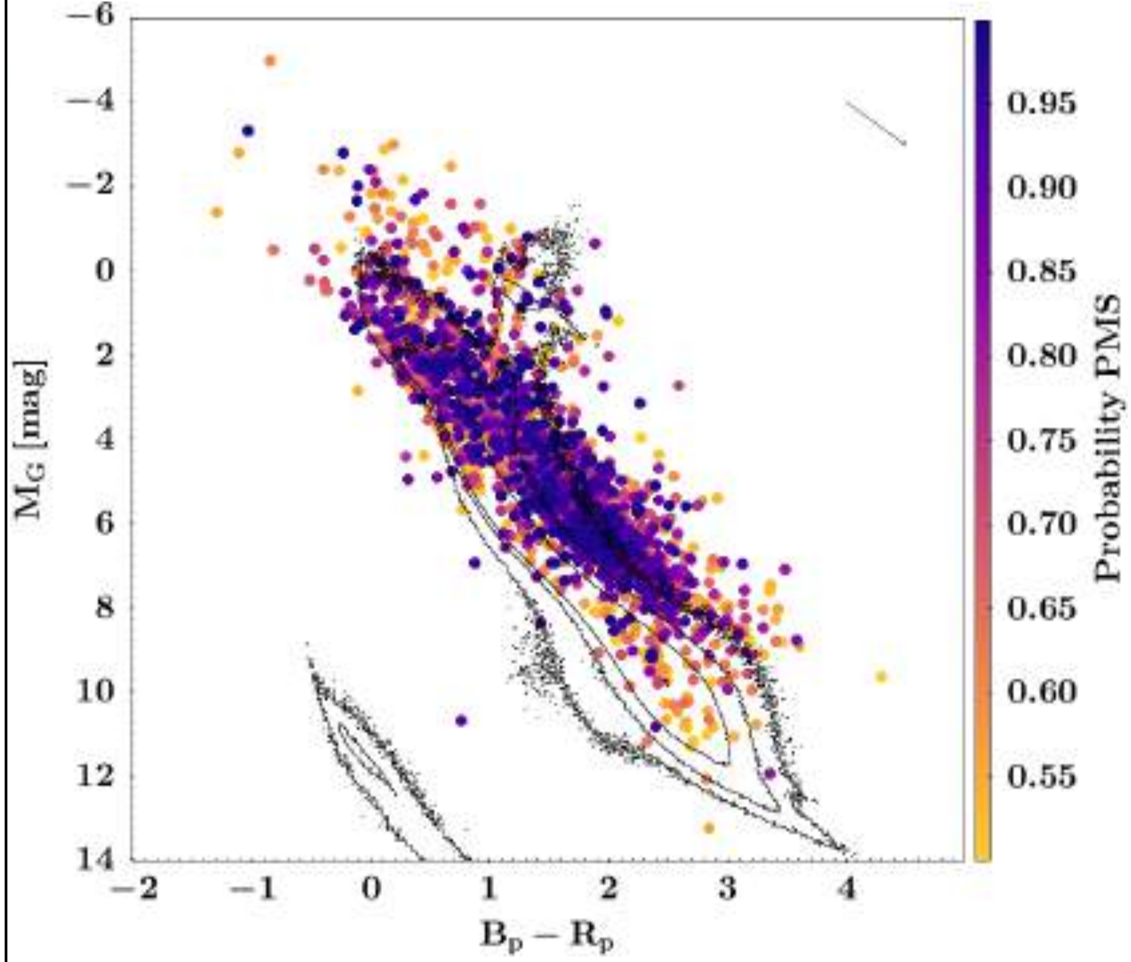
- **775** lassical Be stars

Other sources

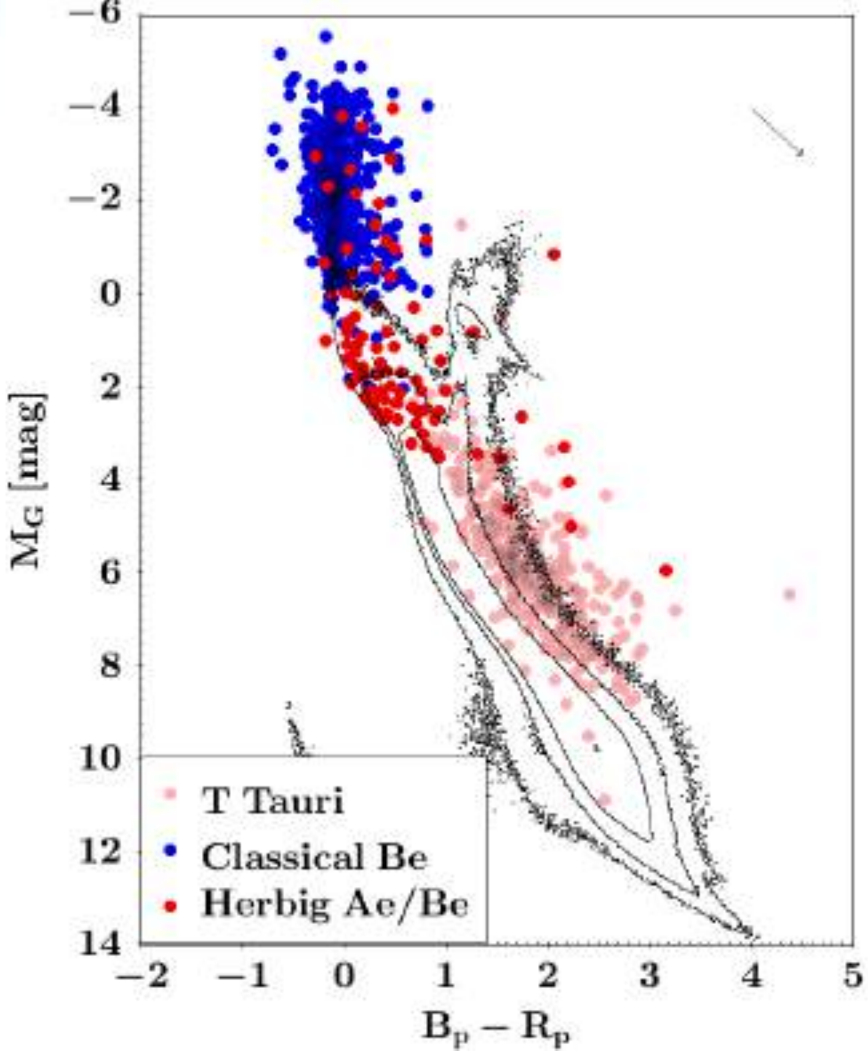
- **471,111** random sources

Gaia HR diagram

PMS candidates

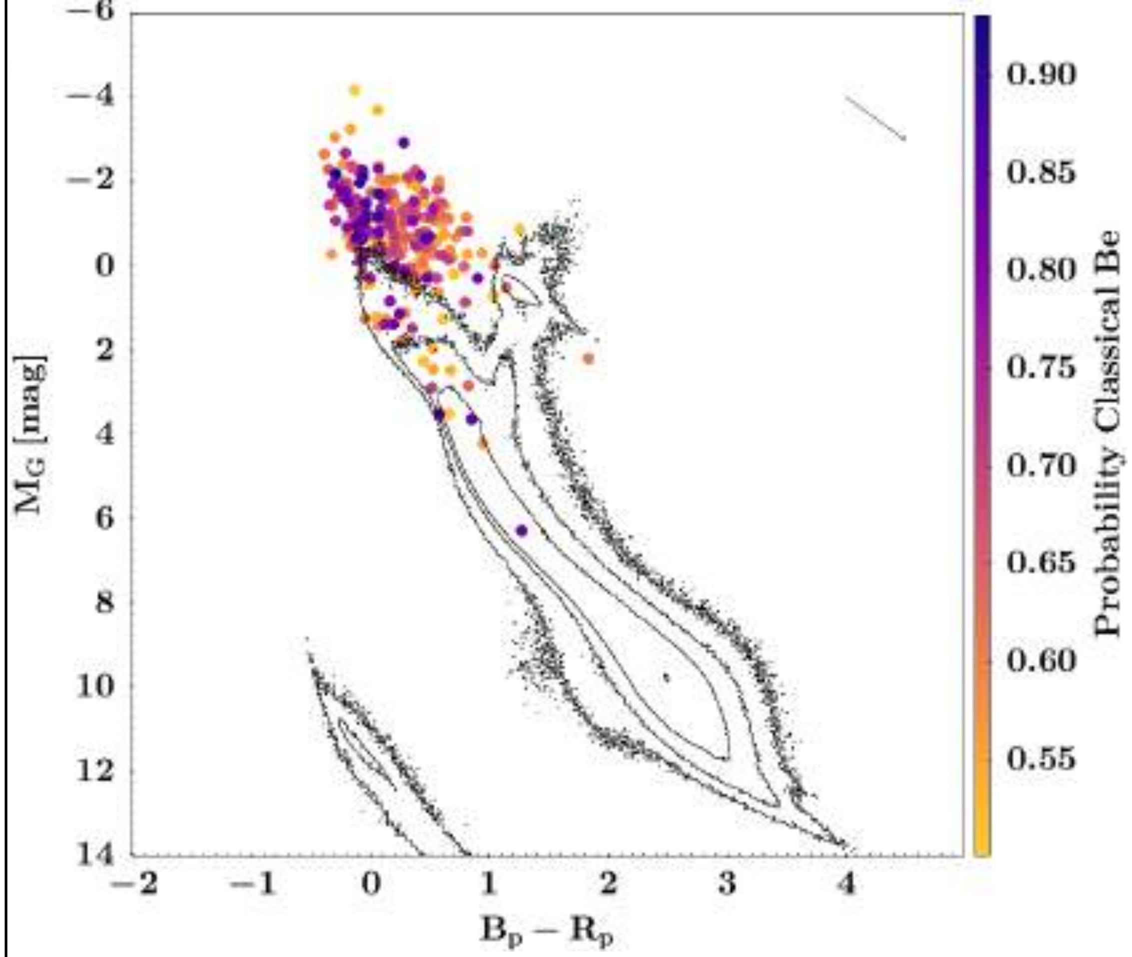


Training Set

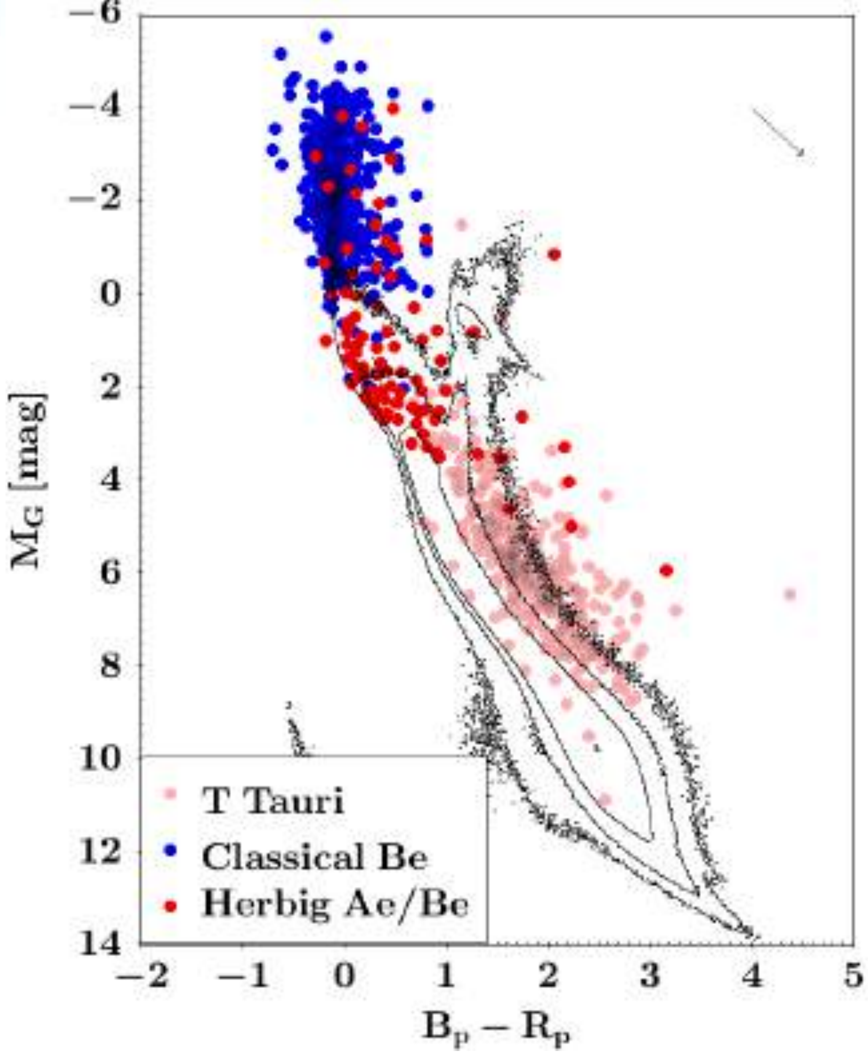


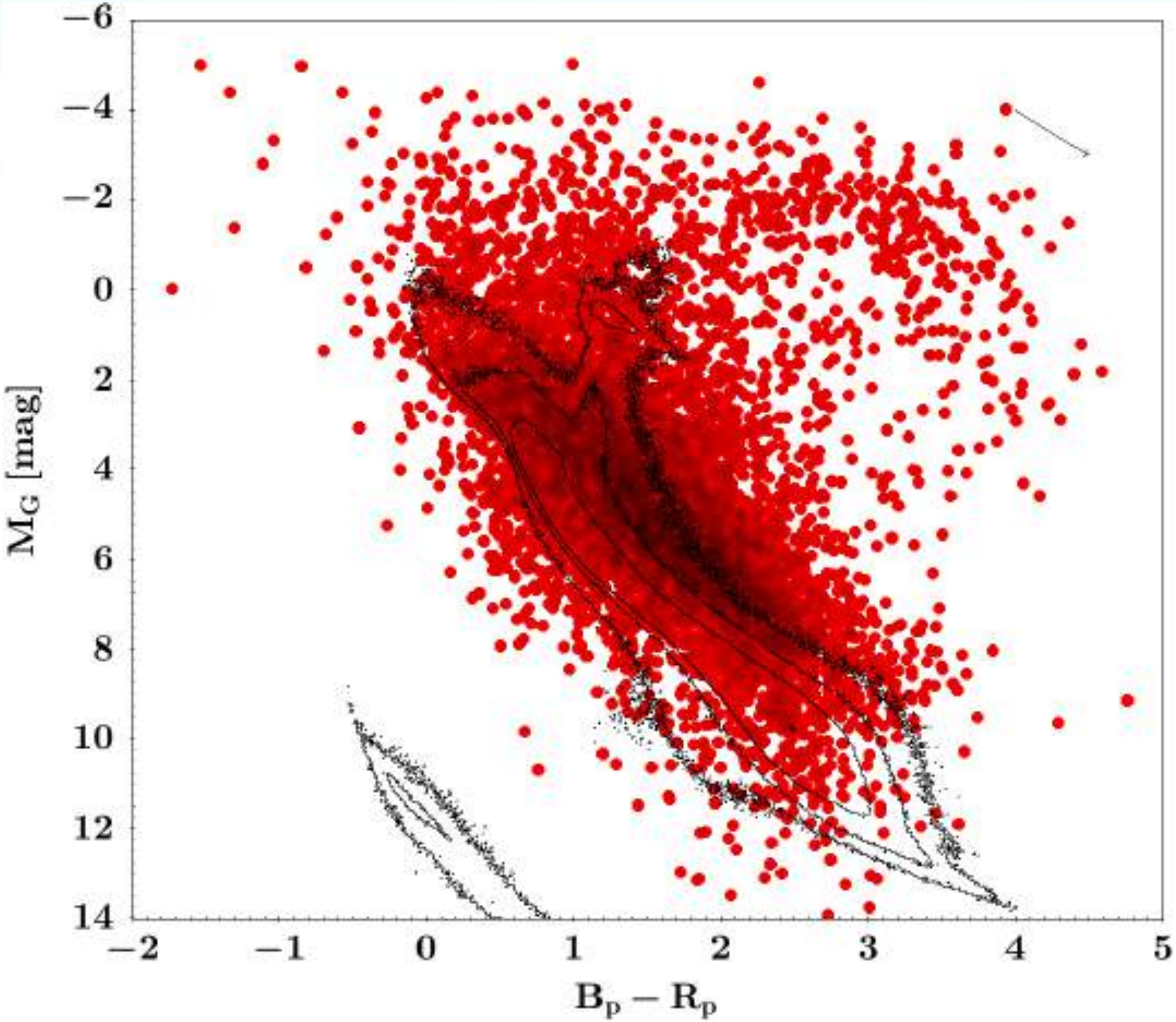
Gaia HR diagram

Classical Be candidates



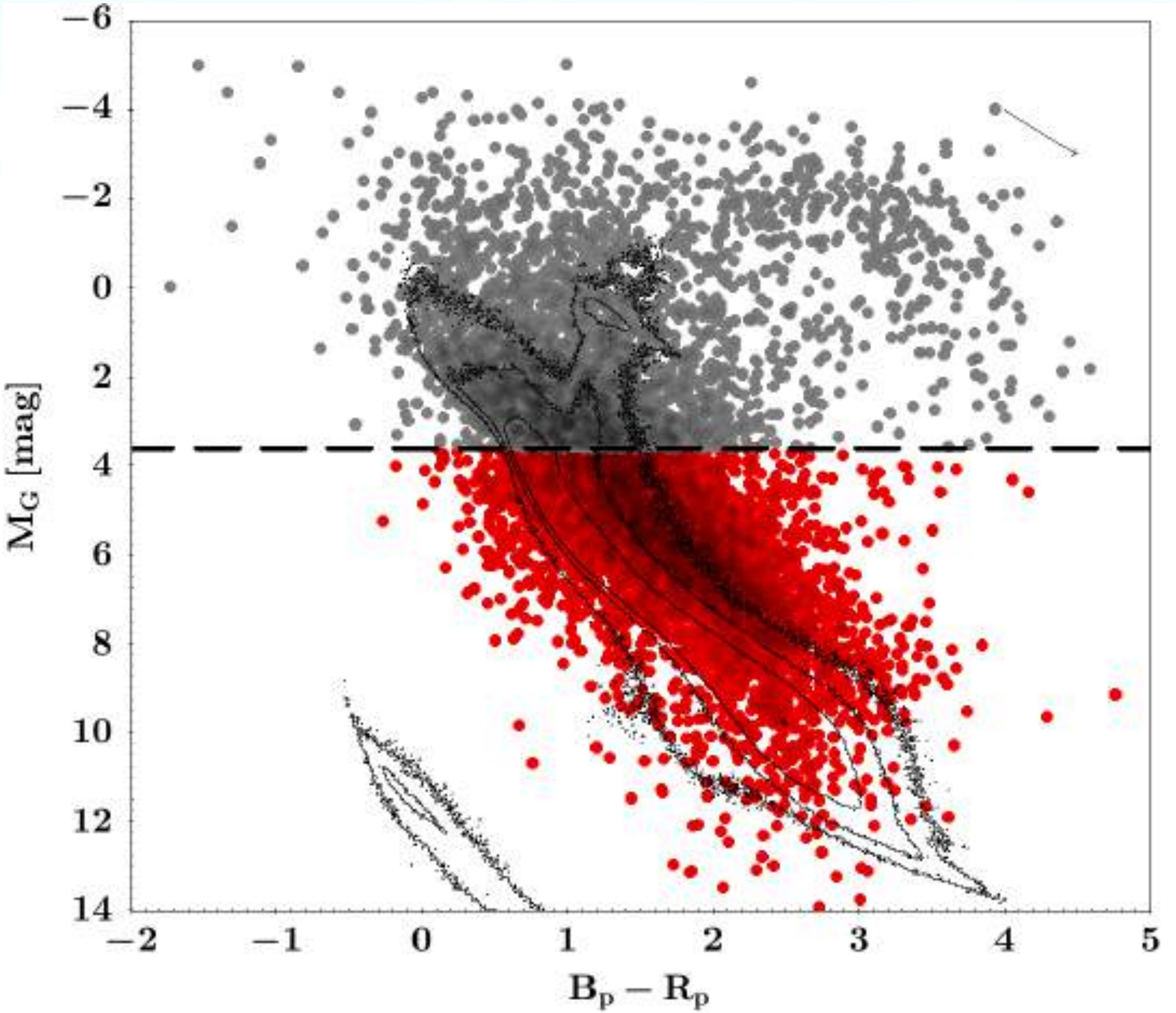
Training Set



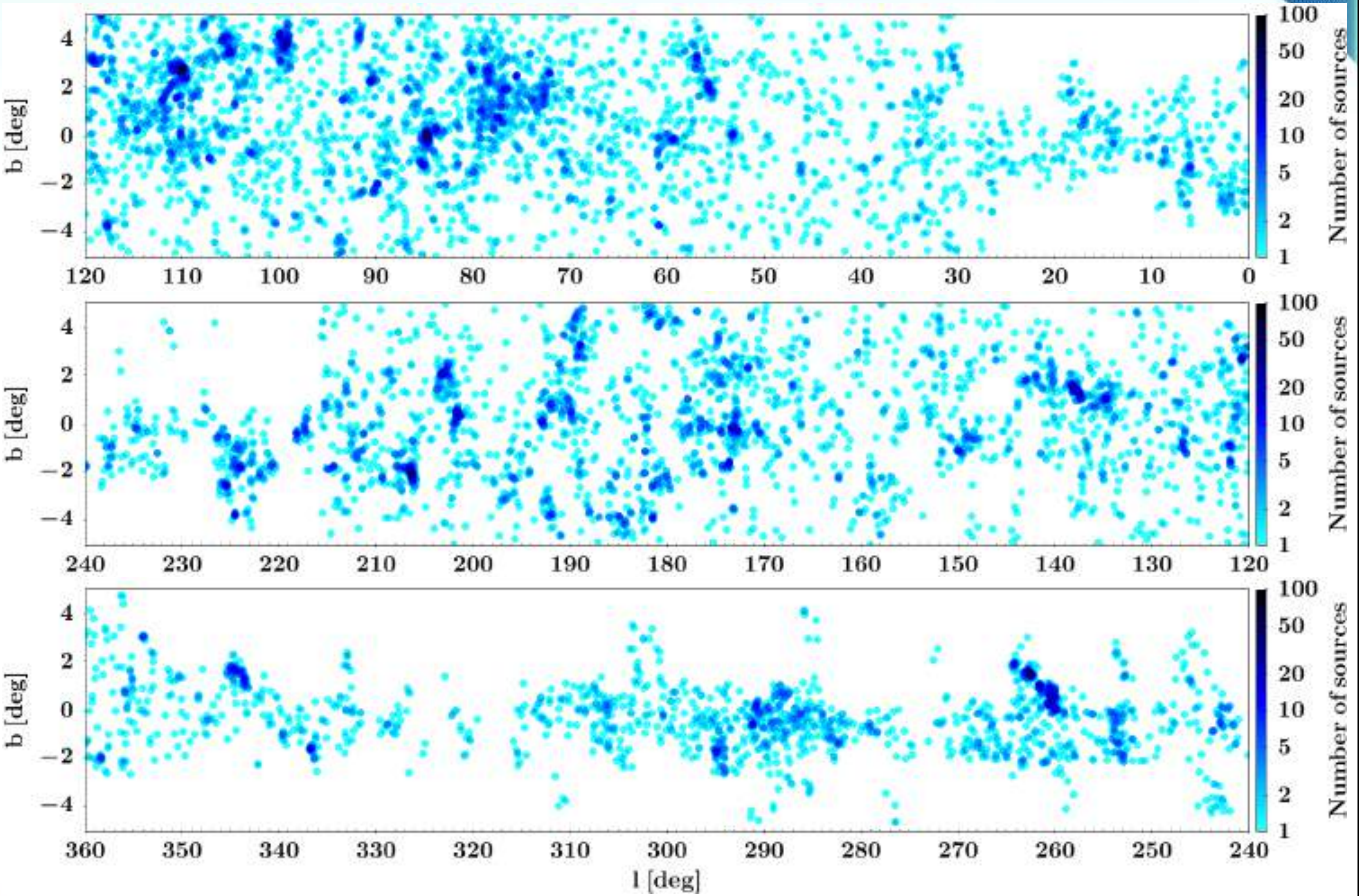


−6
−4
−2
0
2
4
6
8
10
12
14
M_G [mag]
−2
−1
0
1
2
3
4
5
$B_p - R_p$

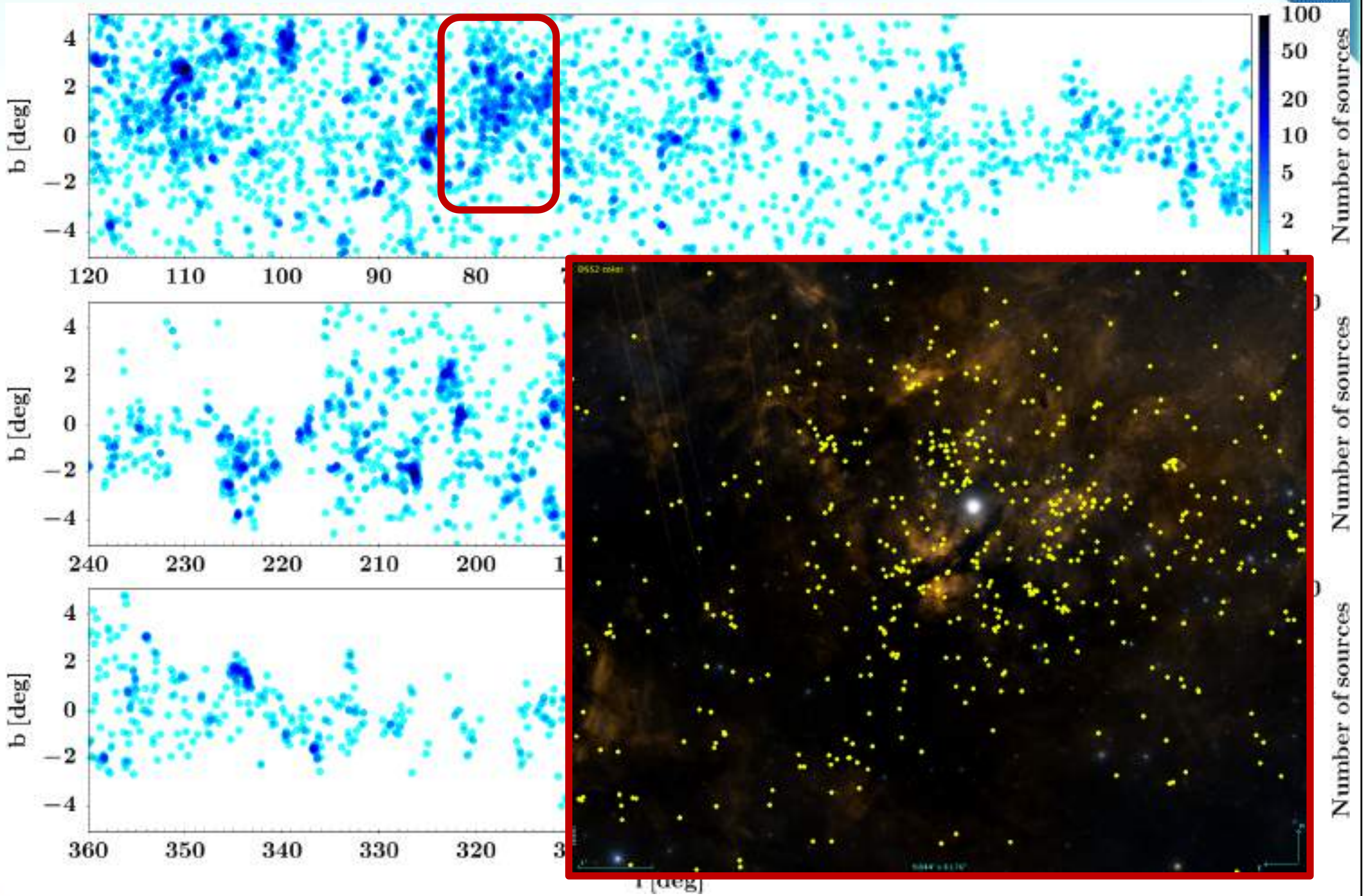
3131 potential high mass (**682** with good Gaia solution)



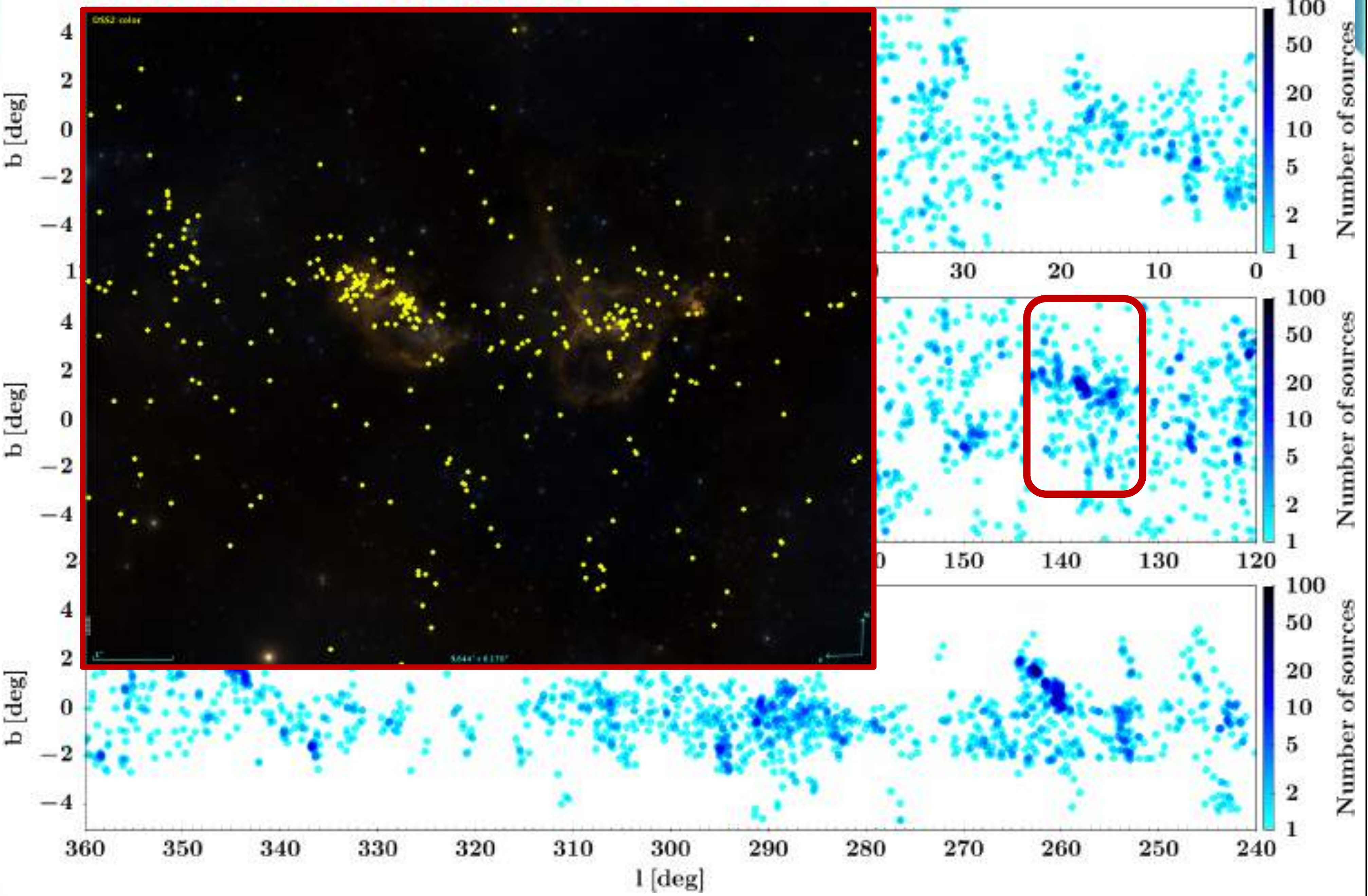
Coordinates



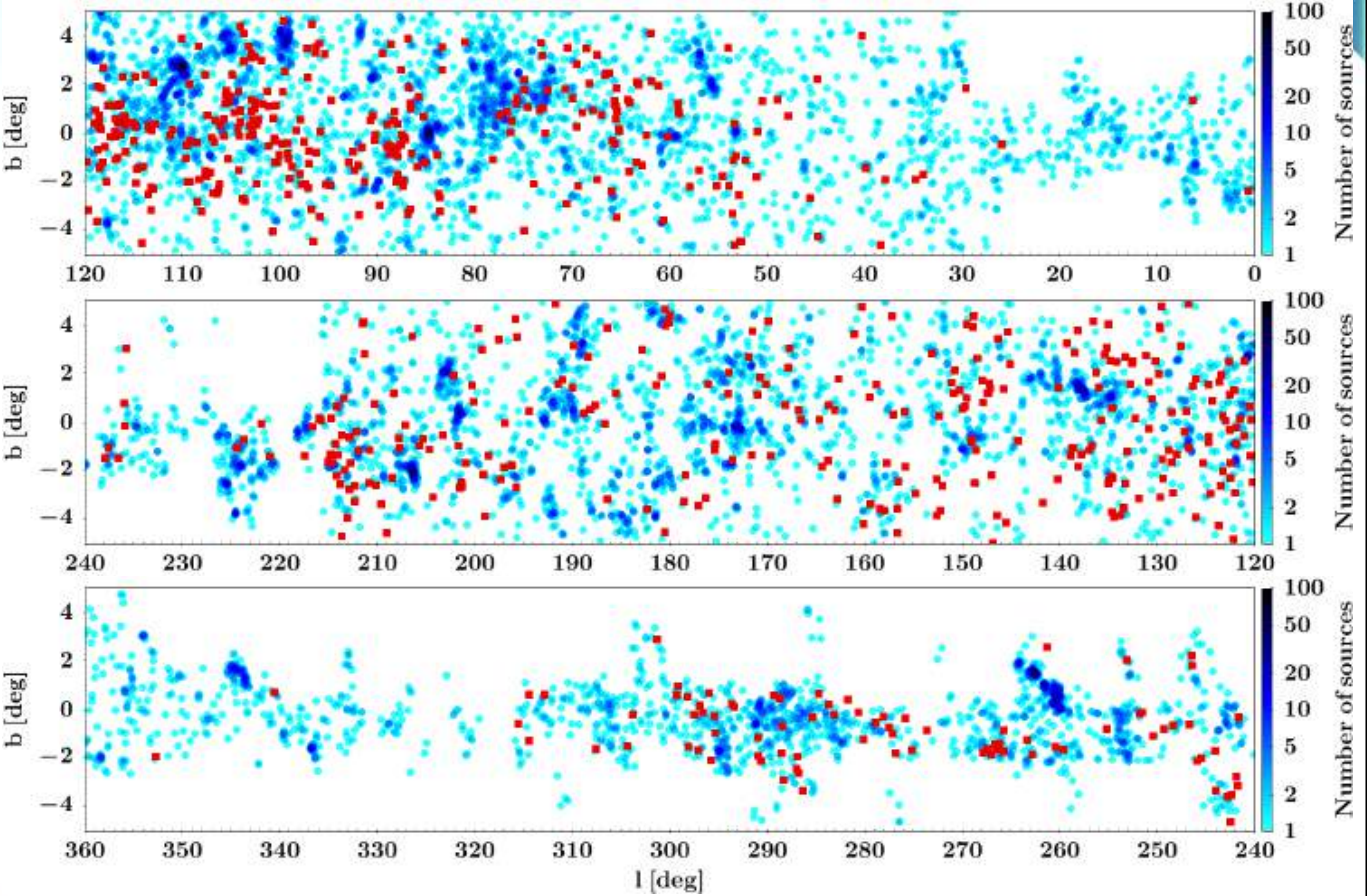
Coordinates



Coordinates



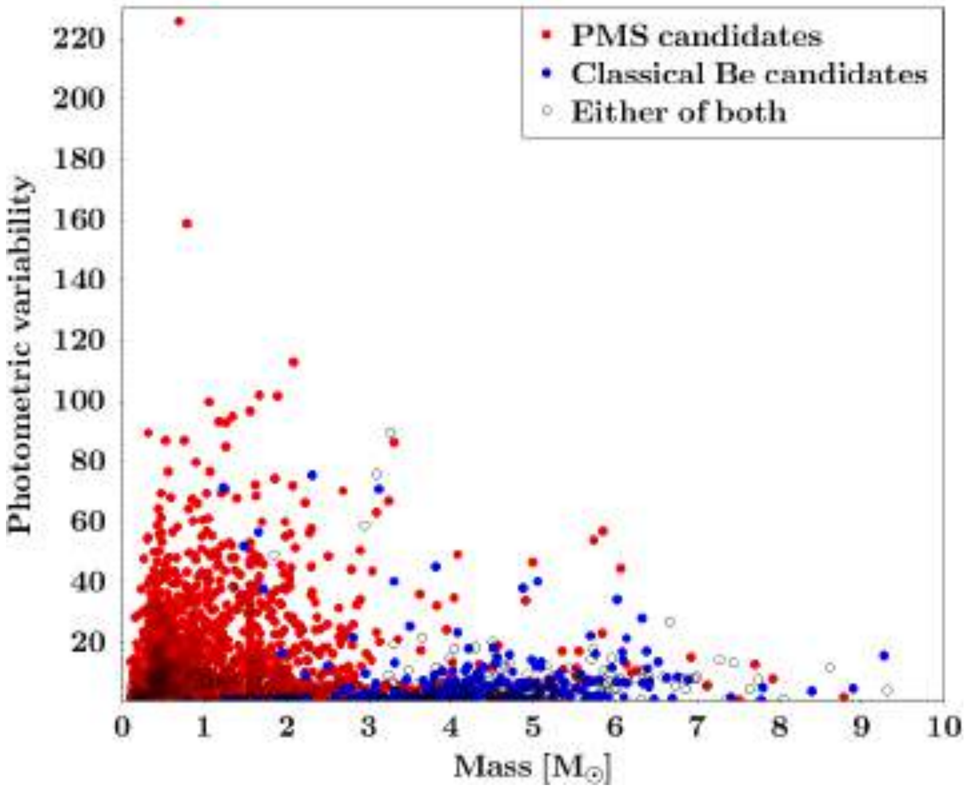
Coordinates



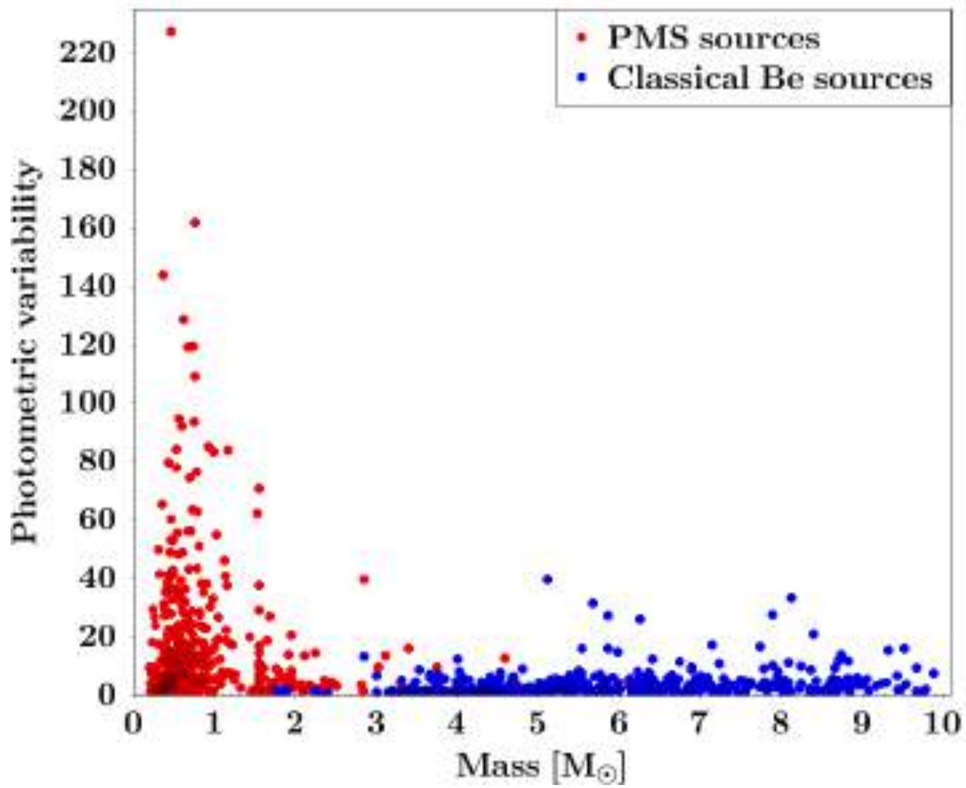
Physical behaviour

Variability vs. Masses (lower limits)

Candidates



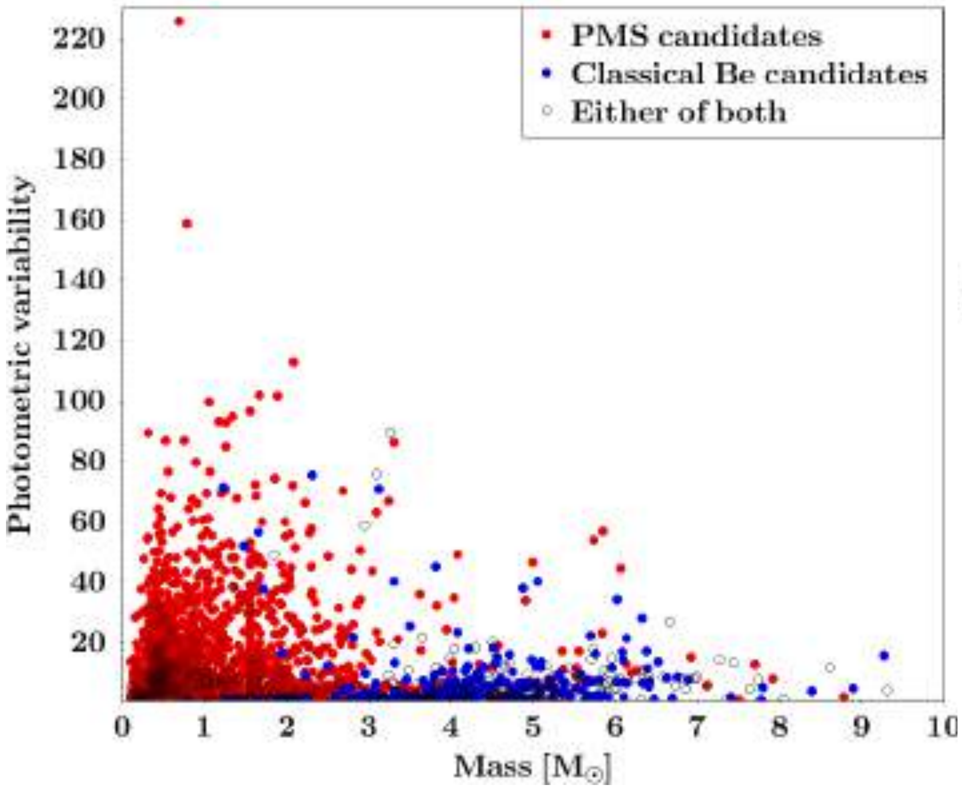
Training Set

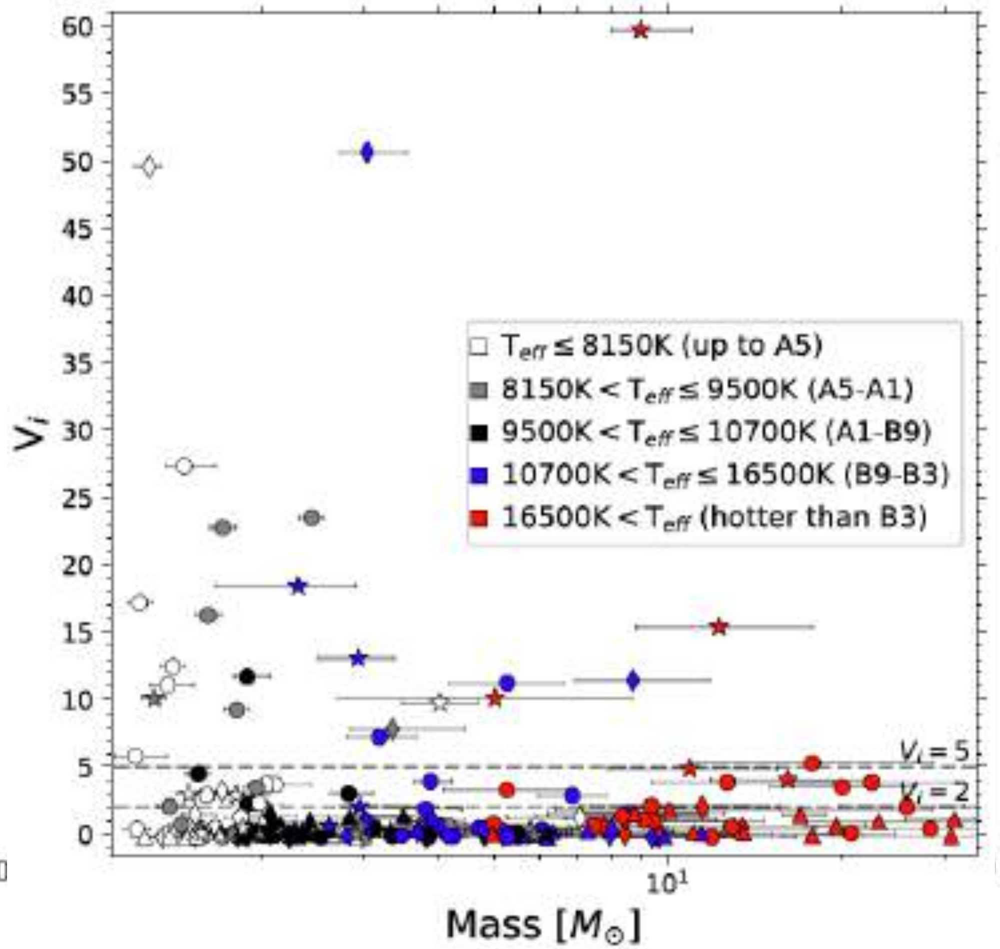


Physical behaviour

Variability vs. Masses (lower limits)

Candidates

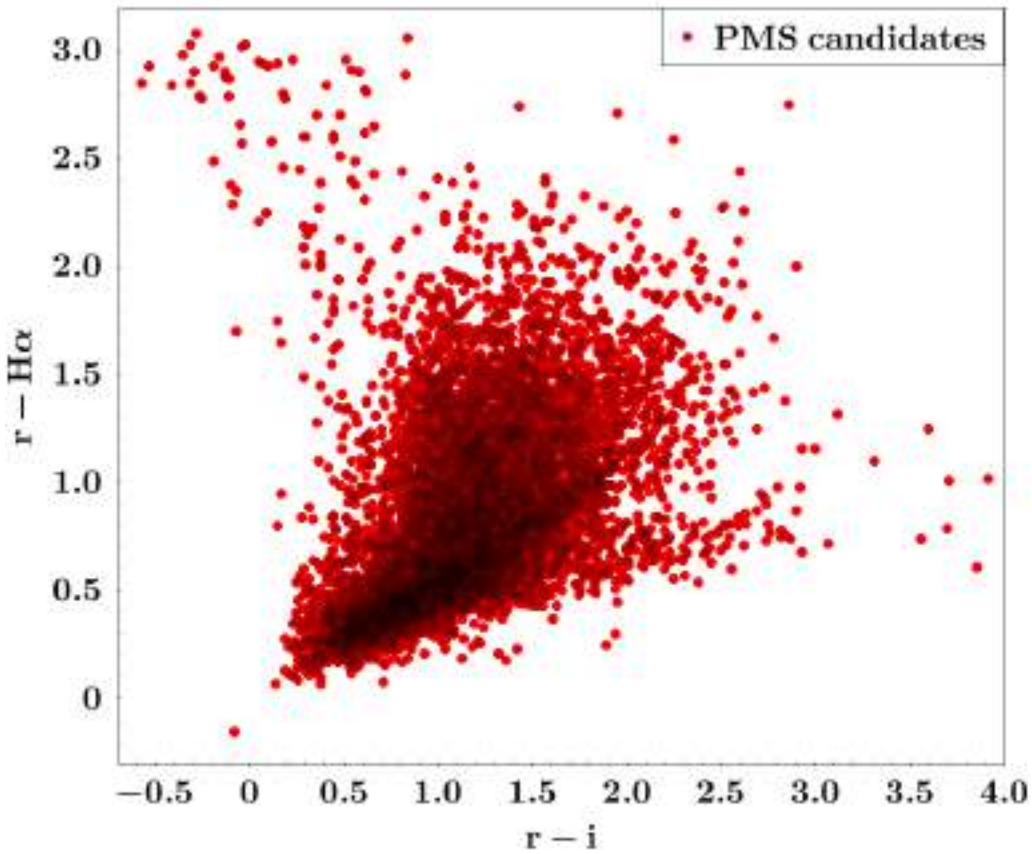


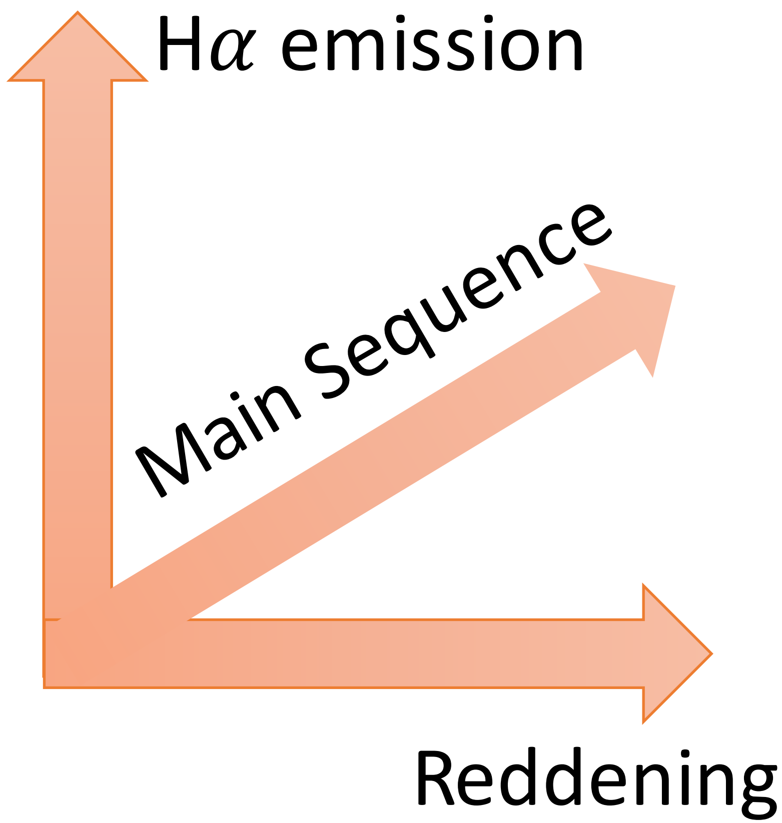


Vioque et al. 2018

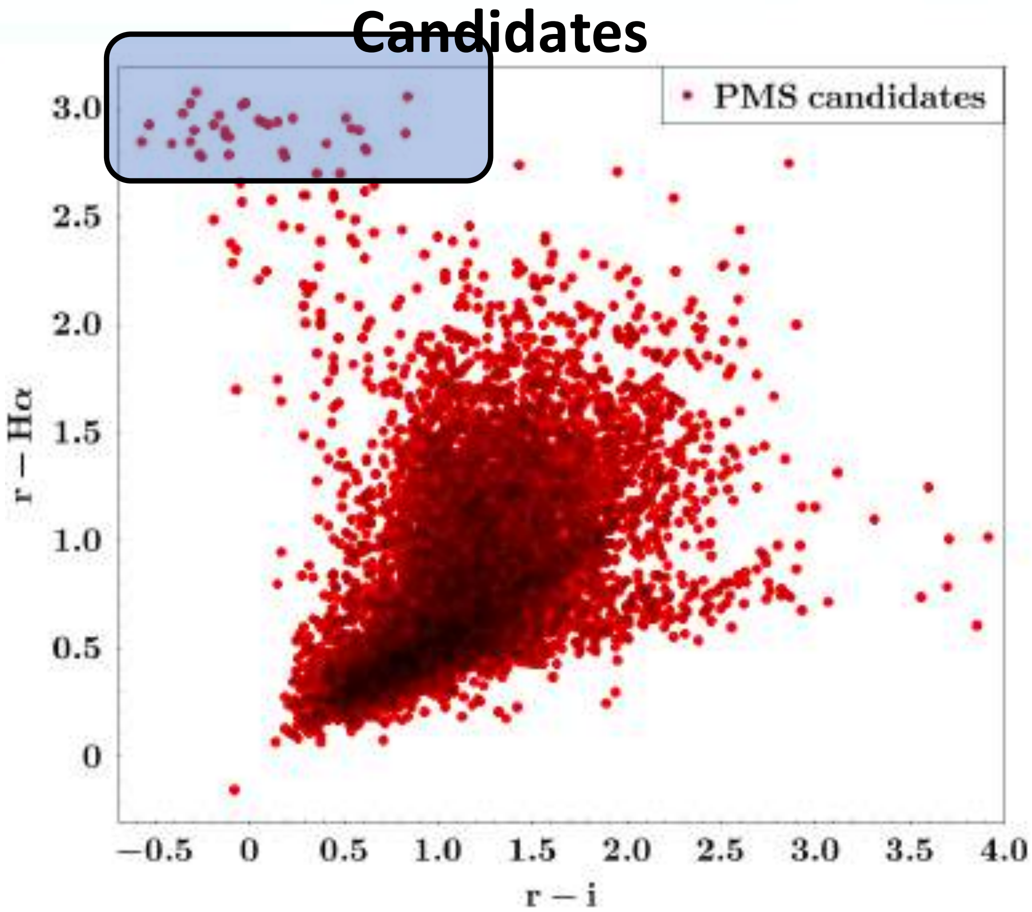
Caveats

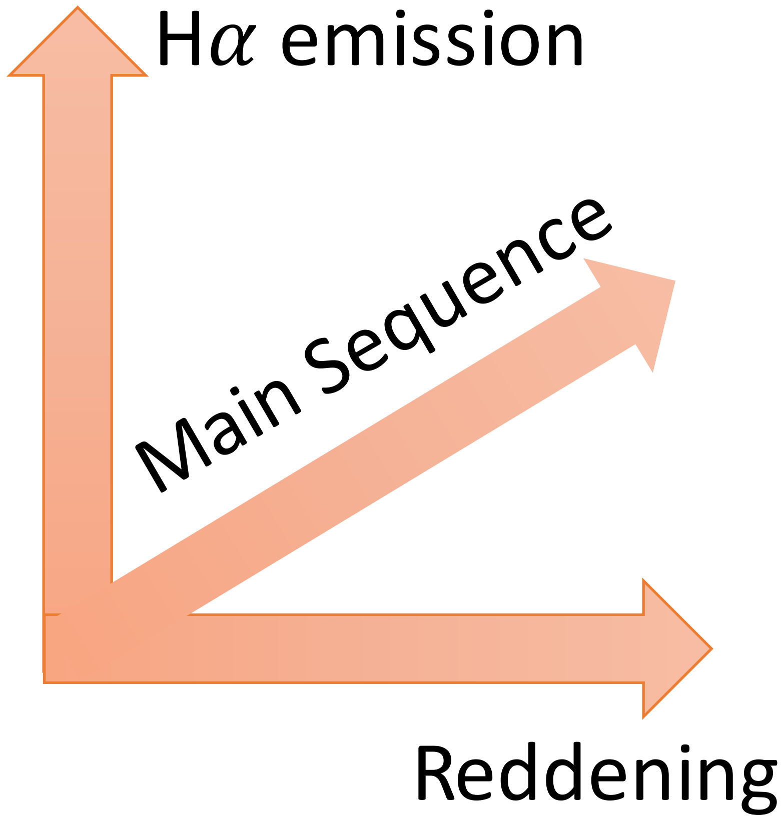
Candidates



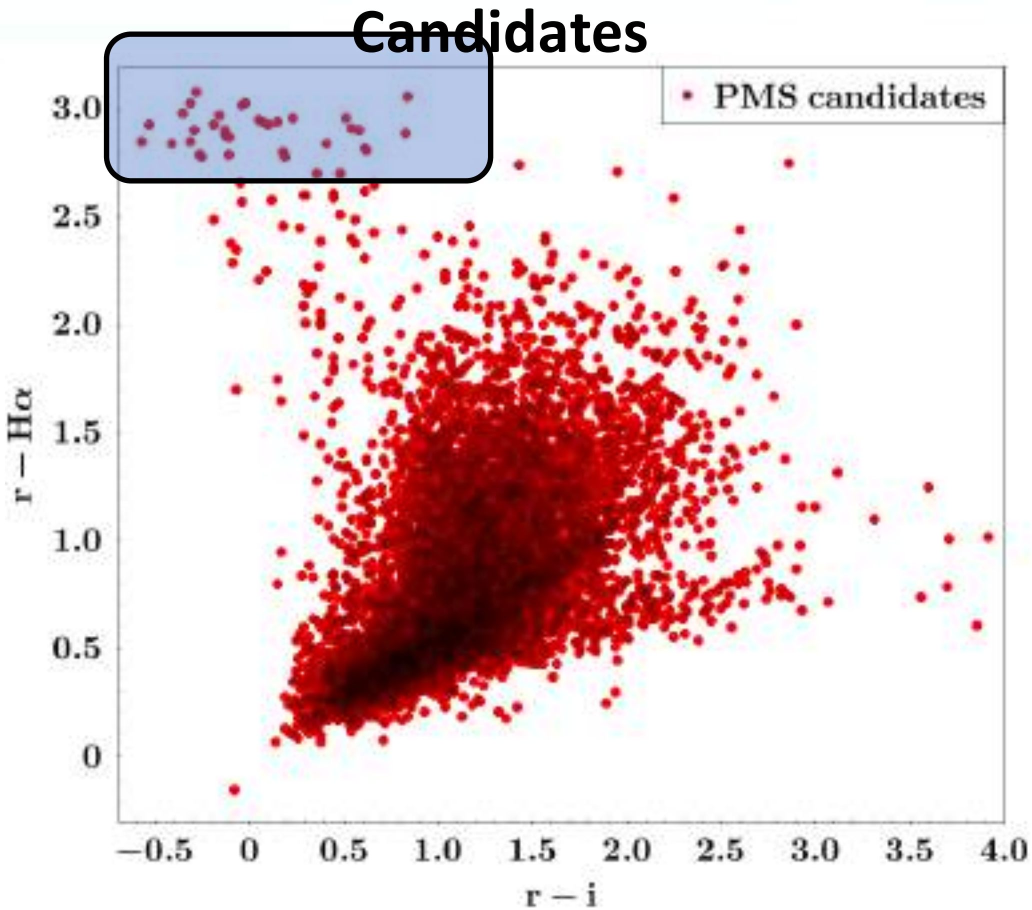


Caveats


Candidates
PMS candidates
3.0
2.5
2.0
1.5
1.0
0.5
0
r − Hα
−0.5
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
r − i


Hα emission
Main Sequence
Reddening

Caveats



Planetary Nebula!

Future work

Past and future observations

INT

2.2m Calar Alto

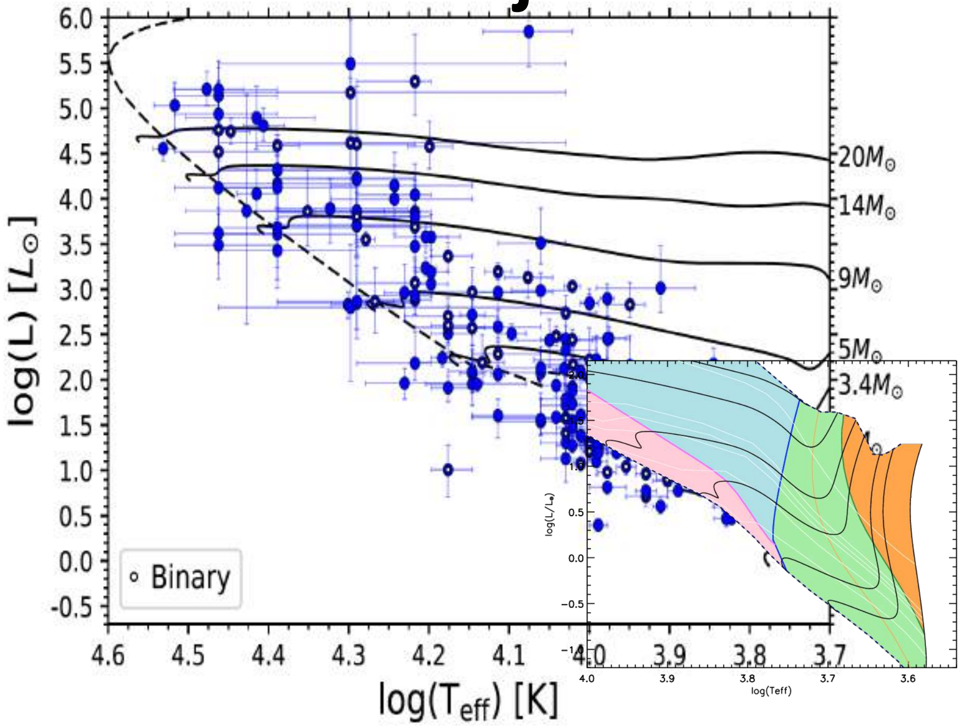
NTT

+

+

Populate HR diagram

260 objects



+ ~3000 objects

Results

- We retrieve **8452 new PMS candidates. 3131 (682) potential Herbig Ae/Be stars.**
- We retrieve **636 new Classical Be stars candidates.**
- We retrieve **1266** candidates of belonging to either one of the two categories.

Completeness
$\mathbf{78.8 \pm 1.4\%}$

Completeness
$\mathbf{85.5 \pm 1.2\%}$