

Variable object classification for the third Gaia Data Release (DR3)

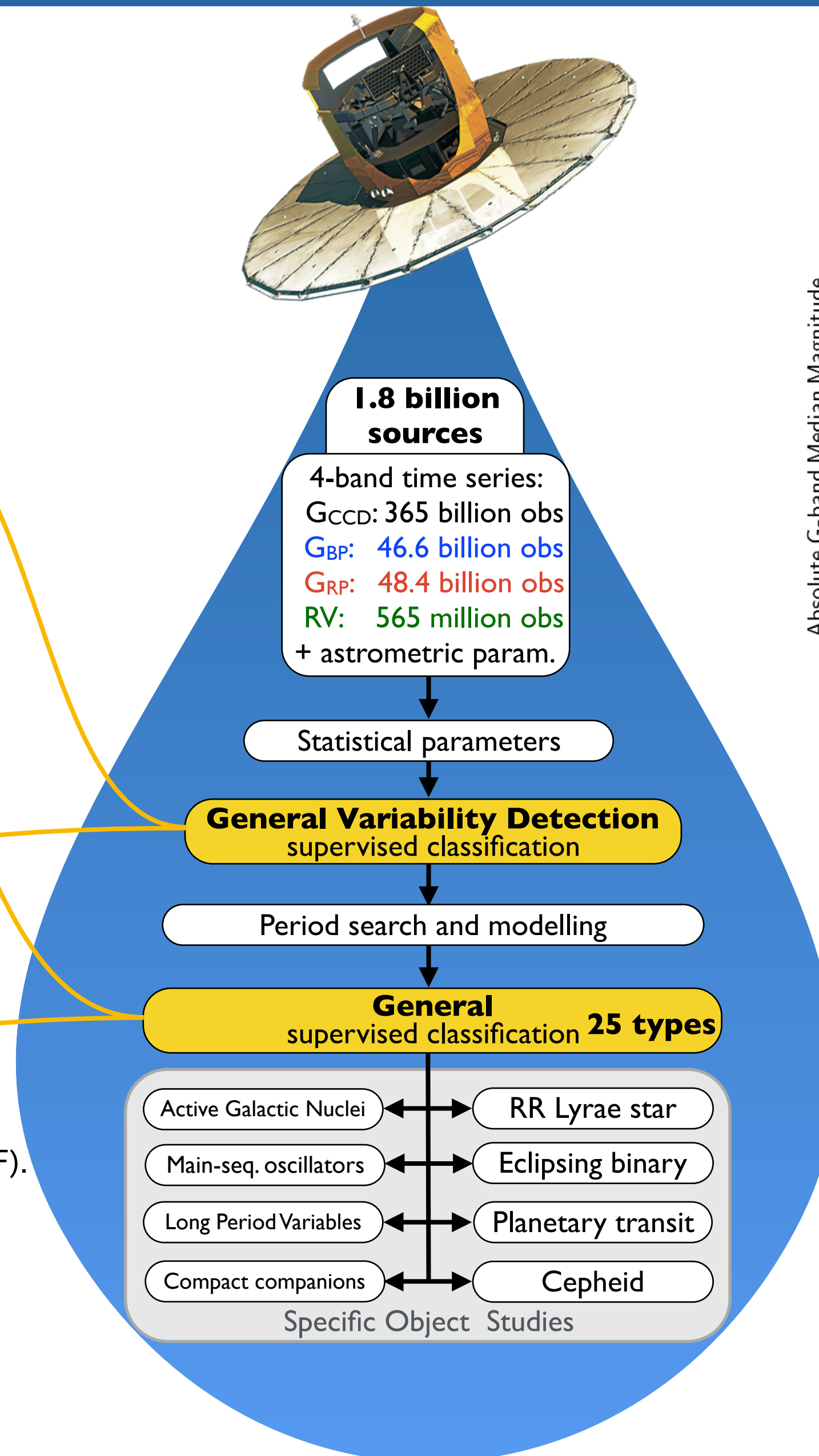
B. Holl, L. Rimoldini, M. Audard, P. Gavras, K. Nienartowicz, L. Eyer, N. Mowlavi, K. Nienartowicz and Gaia CU7/Geneva Data Processing Center
University of Geneva, Switzerland - RHEA for European Space Agency (ESA), Madrid, Spain - Sednai sàrl, Geneva, Switzerland

Preparation

- **Crossmatching**
 - ▶ >150 catalogs
 - ▶ ~8 M sources
 - ▶ ~100 (sub) types
 - ▶ based on positional and magnitude difference
- **Training set for General Variability Detection**
 - ▶ 2 types: 60 K 'variables' and 63 K 'others'
- **Training set for General classification**
 - ▶ 40 type groups: 60 K variables

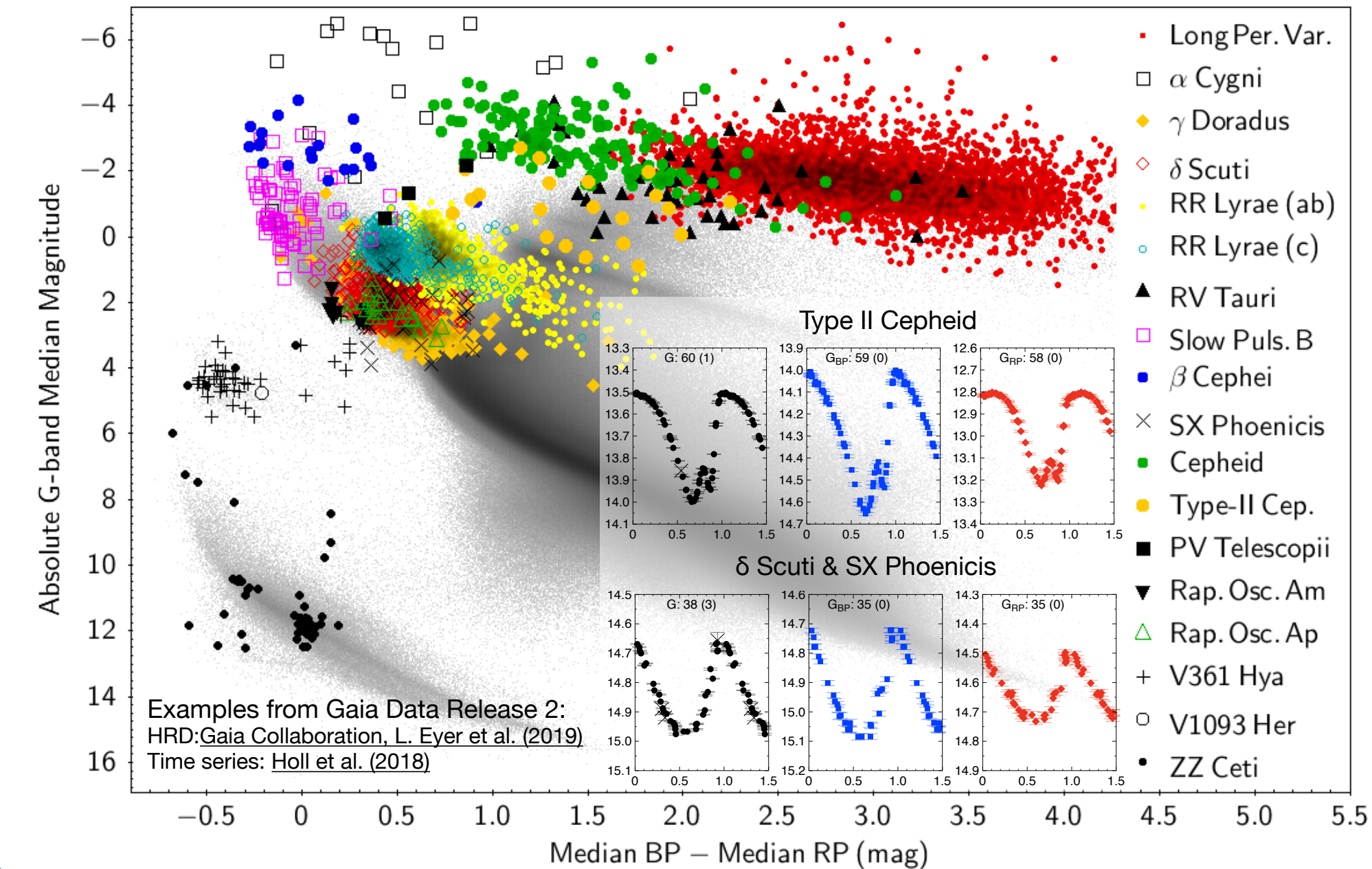
Execution

- **Classifier framework H2O** (www.h2o.ai)
- **General Variability Detection**
 - ▶ XGBoost: binary classifier
 - ▶ Completeness: 99%
 - ▶ Contamination in variables: 0.4%
- **General classification**
 - ▶ About 100 one-vs-rest classifiers *for each variable type* using XGBoost (XGB) and Distributed RandomForest (DRF).
 - ▶ Several multi-class classifiers using XGB and DRF.
 - ▶ Several DRF *meta-classifiers* combining the one-vs-rest or multi-class classifier results.
- **Main challenge: very unbalances variability types**

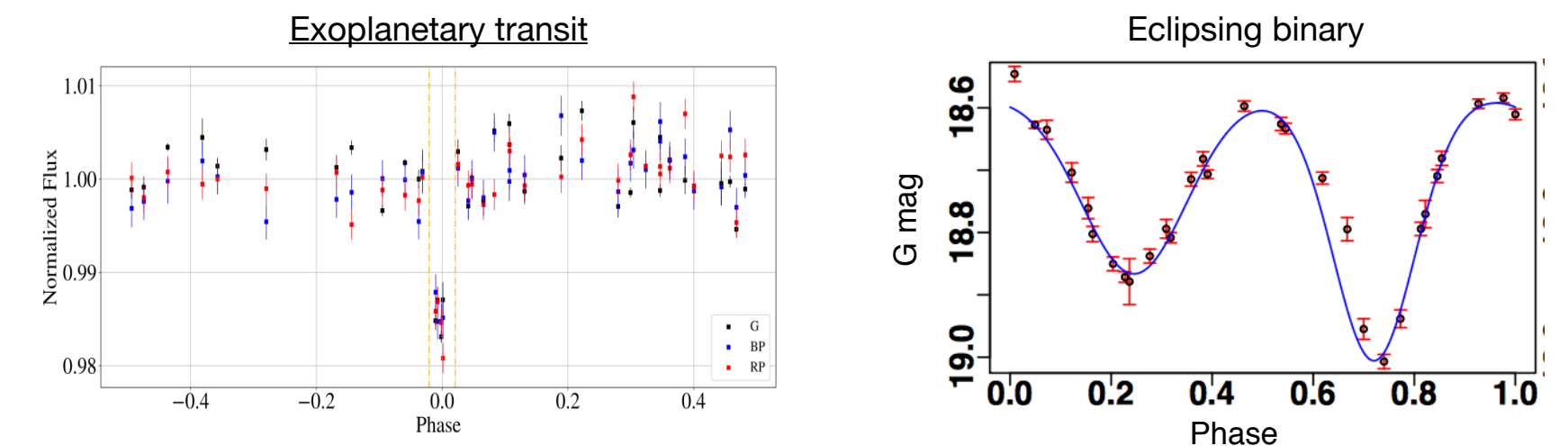


Verification & Validation

Systematic semi-automated procedure for each class (incl. literature comparison)



Example light curves from Gaia images of the week:



Results

- To be published in Gaia DR3 (mid 2022)
- About 11 million variables
- 24 variability types + galaxies