Astronomical Data Analysis Software and Systems XXV
ASP Conference Series, Vol. 512
Nuria P. F. Lorente, Keith Shortridge, and Randall Wayth, eds.

Multiwavelength Variability Surveys: Reaping the Stellar Harvest

Ann Marie Cody,[1] John R. Stauffer,[2] Luisa M. Rebull,[3] Lynne A. Hillenbrand,[4] and Inseok Song[5]

[1]*NASA Ames Research Center, Moffett Field, CA, USA;* annmarie.cody@nasa.gov

[2]*Spitzer Science Center/Caltech, Pasadena, CA, USA;* stauffer@ipac.caltech.edu

[3]*Spitzer Science Center/Caltech, Pasadena, CA, USA;* rebull@ipac.caltech.edu

[4]*California Institute of Technology, Pasadena, CA, USA;* lah@astro.caltech.edu

[5]*University of Georgia, Athens, GA, USA;* song@physast.uga.edu

Abstract. Over the past decade, a number of dedicated stellar variability surveys have launched from both the ground and space. Many of these programs focus on the detection of specific events, such as exoplanet transits, microlensing amplification, or extragalactic transients. Yet the observed variability behavior encompasses a much larger range of stellar phenomena. To take full advantage of variability survey data, we must detect and classify distinct morphological features in light curves. This task has been particularly challenging for the young (1–10 million year old) stars, which are well known to vary at the 1–100% level on timescales of hours to years. Here we highlight recent progress in the identification, classification, and physical understanding of young star variability. We introduce a selection of time series of pre-main sequence stars from state-of-the-art datasets, including the Young Stellar Object Variability (YSOVAR) campaign with the *Spitzer Space Telescope*. We describe the data storage approaches and time series analysis techniques employed to extract physically meaningful information from the light curves. The lessons learned from YSOVAR and other campaigns should be broadly applicable to massive future surveys such as TESS and LSST.

1. An explosion of time series data on young stars

Young stellar objects (YSOs) in the 1–5 Myr age range are highly variable; flux changes are due to rotating starspot patterns, unsteady accretion and hot spots, as well as obscuration by circumstellar material. Variability appears at a range of amplitudes from millimagnitudes to magnitudes; likewise it is present on many timescales from hours to years. Decades of ground-based studies have revealed the broad properties of young star light curve behavior, but detailed analysis was frequently precluded by weather and scheduling constraints. Only recently have we been rewarded with an exquisite high precision (sub-1%), continuous, long time baseline (months) view of YSOs at infrared and optical wavelengths (see Figure 1). This is due to the increasing dedication of space telescopes to young star monitoring campaigns. As shown in Table 1, three major campaigns in particular have contributed greatly to our knowledge of young stars in the time domain.

The first of these campaigns- the Young Stellar Object Variability Project ("YSO-VAR"; Morales-Calderón et al. 2011, Rebull et al. 2014)- involved the *Spitzer Space Telescope*, particularly during its Warm Mission (2009 through present). Through the Guest Observer program, YSOVAR was awarded more than 500 hours of time in 2009–2011 to obtain time series photometry on 12 young star clusters in the 3.6 and 4.5 μm m bands with the IRAC camera. Each cluster was monitored for an approximately 40 day visibility window, at a cadence of twice per day. In addition, ground-based optical and near-infrared time series were obtained for a subset of the YSOs, depending on their spatial position, extinction, and observing conditions. *Spitzer* revealed that stars are not only variable in the optical, but those that host circumstellar dust disks are also highly variable in the infrared. Thus a main purpose of YSOVAR was to understand the mid-infrared time domain properties of YSOs and their disks, including variability prevalence, amplitudes, colors, and periodicities. Altogether, the program obtained ~30,000 mid-infrared light curves of objects in and around young star clusters.

With the early success of YSOVAR, an even more ambitious project was launched, focusing on the 1–3 Myr NGC 2264 cluster. The so-called Coordinated Synoptic Investigation of NGC 2264 ("CSI 2264"; Cody et al. 2014) involved not only *Spitzer* but also optical monitoring by the *CoRoT* space telescope (Baglin et al. 2006). Both facilities acquired simultaneous observations over a period of approximately one month at the end of 2011. In addition, the *Chandra* and the Microvariability and Oscillations of STars ("MOST") telescope were recruited to carry out additional monitoring, along with complementary data from the Very Large Telescope and other ground-based facilities. Like YSOVAR, the project provided high precision (0.1–1%) photometry at moderately high cadence (<2 hours). A subset of the data was obtained in very high cadence modes, every ~15 seconds for *Spitzer*, and every 32 seconds for *CoRoT*. The combined datasets were used to identify and explain several new types of variability peculiar to YSOs (e.g., Stauffer et al. 2014; Stauffer et al. 2015).

While the *CoRoT* telescope has ceased to operate, many of its capabilities have been replaced by K2, the extended Kepler mission (Howell et al. 2014). Exquisite optical precision photometry is once again becoming available, for a dozen young (1–10 Myr) to intermediate age (~150 Myr) clusters on the ecliptic plane. Data from K2 began to flow in late 2014, and each campaign involves a $\sim 10^{\circ} \times 10^{\circ}$ field of view monitored for ~80 days straight. Light curves are only available for a subset of 10,000–20,000 targets chosen in advance by proposals from the community. Several campaigns include open clusters, each including several hundred to a thousand young stars with light curves. With ~3500 points per light curve and precision as good as ~100 parts per million, the K2 open cluster dataset is revealing broadband optical variability at a higher level of detail than ever seen before.

With the profusion of light curves has come new challenges in data management. While the raw data volume might not constitute "big data" status (a collection of young star light curves is typically under 1 Tb), its complexity calls for careful design of storage, viewing, and analysis tools. In the following, we detail several approaches taken to digitally organize the vast array of YSO data that has been collected from these recent campaigns.

Table 1. Recent spaced-based young star monitoring programs

Project	Primary telescope(s)	Wavelengths	# Light curves	Pts per light curve
YSOVAR	*Spitzer*	optical, near-IR, mid-IR	~30,000	100
CSI 2264	*Spitzer*,*CoRoT*	X-ray, optical, mid-IR	~25,000	100-100,000
K2 clusters	*Kepler*	optical	5,000	4,000

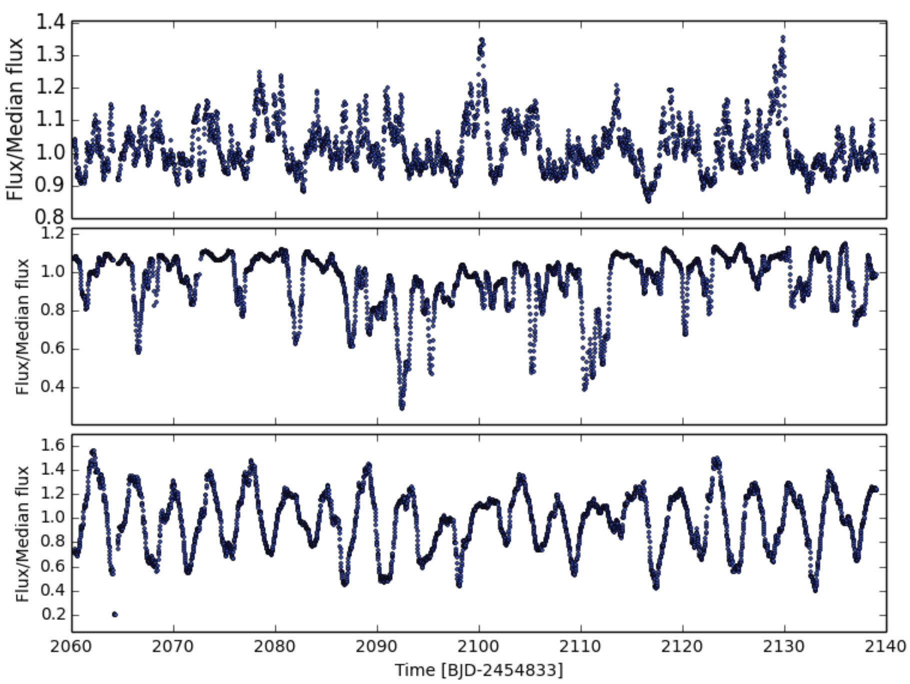


Figure 1. Light curves of young stars from the K2 Mission. These examples show various morphological features which can be classified statistically and connected with different variability mechanisms. Each has approximately 3,500 points over 80 days (30-minute cadence).

2. Time series storage and retrieval: The YSOVAR database

One of the earliest (and largest) star formation regions to be monitored by the YSOVAR project was the Orion Nebula Cluster and surroundings ($1^\circ \times 1^\circ$ field of view). Because YSOs in Orion are not as embedded as in other regions, it was possible to obtain optical and near-infrared time series simultaneously with a number of ground-based telescopes (see Morales-Calderón et al. 2011). There was an immediate need to develop an adaptable storage system for both the Orion light curve set and future YSOVAR data holdings. In response to this need, we developed the YSOVAR database, which involves a web-based interface on top of MySQL infrastructure. The database consists of ten interlinked tables, with the bulk of data stored in the photometry table. Each observation of a photometric source is recorded as a single datapoint, with an associated set of coordinates, filter, and telescope. 119 million photometric measurements for 788,000 unique point sources in the YSOVAR project are stored in the MySQL server.

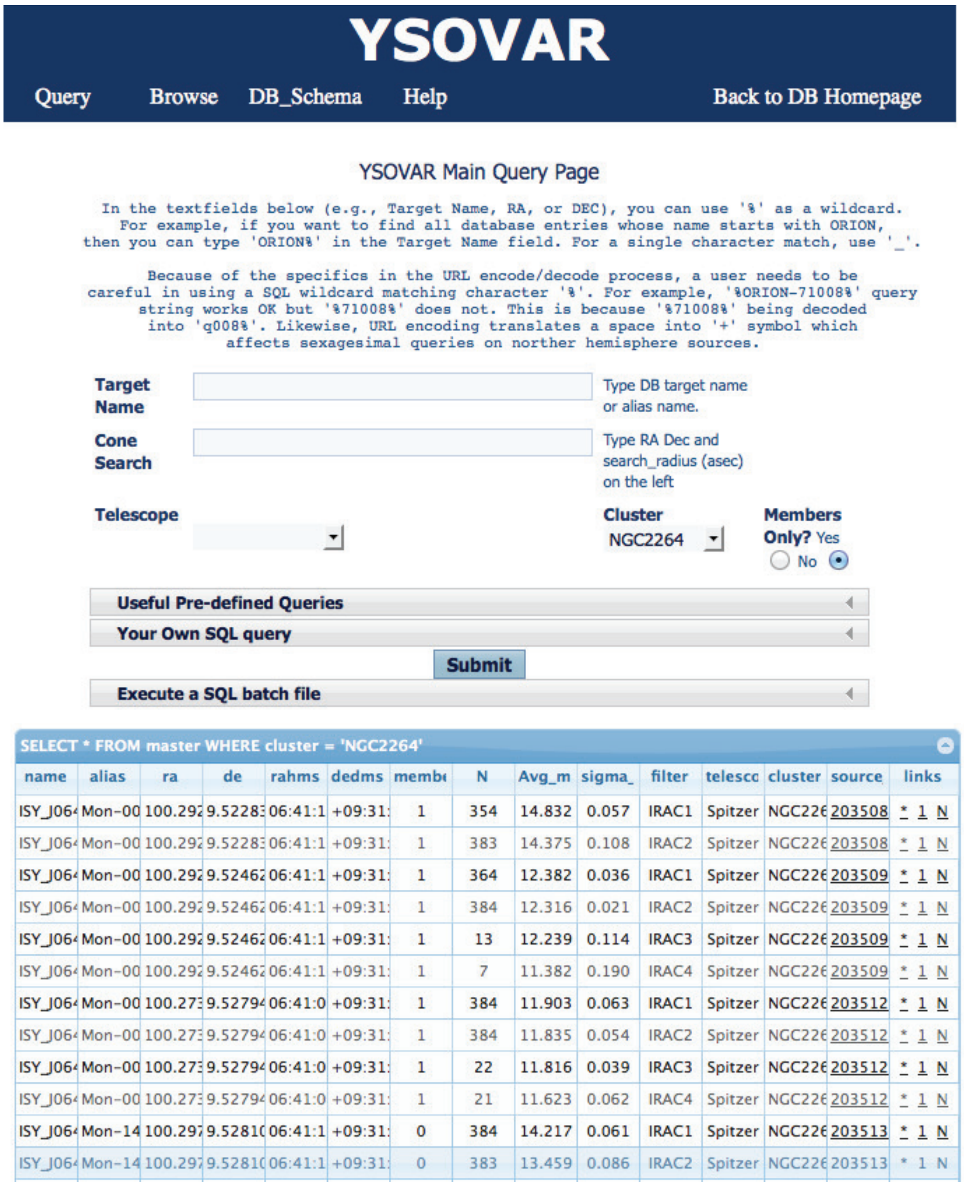
Light curves are assembled within the YSOVAR database by running a clustering algorithm to group points that are within $1''$ of each other. The center of the cluster is then taken to be the object position. Known objects are cross-matched against the list of object positions to attach names to each star. The photometry table can be queried on a per-object basis to view the light curves in any band. In addition, an interactive plotting program is linked to each source, such that one can view one or more light curves on a plot with adjustable axes. The light curve view pages are also configured to show postage stamp images of the star at all available wavelengths. These features are generated using python CGI scripts.

Virtually any subset of light curve data can be extracted from the YSOVAR database via SQL query. This is convenient, for example, for extracting data on objects at particular locations or with particular observational parameters. Nevertheless, our experience is that many database users do not have the time or desire to learn basic SQL. Therefore, we have also built in a web-based query form, from which one may search for objects by position or name, download all data from a single cluster, or isolate data from particular filters or telescopes (see Figure 2). Query results are displayed on the webpage through the jqGrid plugin to the JQuery package which allows on-the-fly sorting, filtering, and page navigations. For example, one might want to select photometric datapoints from a particular time window or with a maximum uncertainty value. There is then an option to download all selected data as a tarfile.

The YSOVAR database currently contains photometry for all young clusters observed during the YSOVAR project, as well as Spitzer light curves for the mid-infrared portion of the CSI 2264 project. Overall, we have found that having an established database ensures that all team members have access to same level of light curve processing. Furthermore, the database can be frozen at any time to provide a stable release to refer to within the literature.

3. Literature mining efforts

To make physical sense of the flood of time series data on young stars, it is imperative to assimilate a wide variety of literature data, from spectral energy distributions to stellar spectroscopic properties. Parameters such as effective temperature or spectral type, Hα equivalent width, and multiplicity can be collected in a separate table of the YSOVAR database, and then linked up to the photometric data by object name. We took this approach for the YSOVAR/Orion dataset. However, it soon became clear that a more visual format was needed to assess the available background information on stars of interest. We therefore generated a set of quasi-static webpages, each of which was dedicated to an individual star in the YSOVAR or CSI 2264 sample.



name	alias	ra	de	rahms	dedms	membe	N	Avg_m	sigma_	filter	telesco	cluster	source	links
ISY_J064	Mon-00	100.292	9.52283	06:41:1	+09:31:	1	354	14.832	0.057	IRAC1	Spitzer	NGC226	203508	* 1 N
ISY_J064	Mon-00	100.292	9.52283	06:41:1	+09:31:	1	383	14.375	0.108	IRAC2	Spitzer	NGC226	203508	* 1 N
ISY_J064	Mon-00	100.292	9.52462	06:41:1	+09:31:	1	364	12.382	0.036	IRAC1	Spitzer	NGC226	203509	* 1 N
ISY_J064	Mon-00	100.292	9.52462	06:41:1	+09:31:	1	384	12.316	0.021	IRAC2	Spitzer	NGC226	203509	* 1 N
ISY_J064	Mon-00	100.292	9.52462	06:41:1	+09:31:	1	13	12.239	0.114	IRAC3	Spitzer	NGC226	203509	* 1 N
ISY_J064	Mon-00	100.292	9.52462	06:41:1	+09:31:	1	7	11.382	0.190	IRAC4	Spitzer	NGC226	203509	* 1 N
ISY_J064	Mon-00	100.273	9.52794	06:41:0	+09:31:	1	384	11.903	0.063	IRAC1	Spitzer	NGC226	203512	* 1 N
ISY_J064	Mon-00	100.273	9.52794	06:41:0	+09:31:	1	384	11.835	0.054	IRAC2	Spitzer	NGC226	203512	* 1 N
ISY_J064	Mon-00	100.273	9.52794	06:41:0	+09:31:	1	22	11.816	0.039	IRAC3	Spitzer	NGC226	203512	* 1 N
ISY_J064	Mon-00	100.273	9.52794	06:41:0	+09:31:	1	21	11.623	0.062	IRAC4	Spitzer	NGC226	203512	* 1 N
ISY_J064	Mon-14	100.297	9.52810	06:41:1	+09:31:	0	384	14.217	0.061	IRAC1	Spitzer	NGC226	203513	* 1 N
ISY_J064	Mon-14	100.297	9.52810	06:41:1	+09:31:	0	383	13.459	0.086	IRAC2	Spitzer	NGC226	203513	* 1 N

Figure 2. The YSOVAR database houses optical and infrared time series data from the *Spitzer* Space Telescope. It includes light curves for some 30,000 distinct sources, all of which can be queried by target name, position, cluster, or more complex constraints via SQL. The database is useful for providing a stable repository of light curves, as well as selecting out subsets of data based on observational set-up.

Content on each page includes past photometric measurements (both average magnitudes and any periodic timescales detected), basic stellar properties, “postage stamp" images of the object from public image servers, and plots of light curves and other data gathered as part of the monitoring campaign. These individual star pages have proved to be a crucial and frequently used reference, as all data for each object are assembled in a single place, in a much more visual manner than provided by the YSOVAR database interface. Ideally, the pages would be directly connected to the database, so as to dynamically update when new data is added.

The process of assembling the individual star pages for YSOVAR and CSI 2264 targets required significant manual mining of the young star literature to gather published data on each object. This tedious process has motivated in part the development of a comprehensive database of YSO properties in a much larger set of young clusters than encompassed by the YSOVAR project. The database, developed by Hillenbrand & Baliber (2015), has the ambitious goal of housing *all* measurements made on YSOs within 1 kpc. While not directly connected with variability, its content will prove highly valuable in correlating observed light curve properties with stellar and disk parameters. Top level data tables include photometric and spectroscopic data, positions and motions, membership information, multiplicity, as well as derived properties (e.g., extinction, luminosity, stellar mass). The database is expected to eliminate repeated work by young star researchers around the world.

4. Statistical approaches

After collecting ancillary data from the literature, the next step is interpretation of variability behavior in the context of (circum-)stellar properties. What space-based programs with the *Spitzer*, *CoRoT*, and *Kepler* telescopes have shown is that light curve morphologies are incredibly diverse (e.g., Figure 1). They appear to populate different categories of behavior as seen by eye, but ideally these classes should be selectable using robust statistics. Thus statistical metrics are an additional set of quantities that we have assembled for YSOs and incorporated into the YSOVAR database.

A challenge in using statistical light curve properties to separate different variability classes is that standard metrics such as medians, peak-to-peak amplitudes, standard deviations, and higher order moments do not appear to correlate uniquely with light curve behavior. To statistically select the distinct YSO variability types, we explored a parameter space of 50–100 statistical metrics, measured for each photometric band. Ultimately, none of these adequately separated the variability detected by eye. We therefore generated several entirely new metrics, as shown in Figure 3 and explained in Cody et al. (2014). The flux asymmetry metric, also known as “M", is a measure of the tendency of a light curve to preferentially take excursions above the median (“bursters") or below the median (“dippers"). The stochasticity metric, also known as “Q", is a measure of how periodic a light curve is. It is based on the residuals left over when a phased light curve is subtracted from the raw version. Light curves that are highly periodic have Q values near 0, while those that are completely stochastic have Q values near 1.

The great variety of light curve behavior among YSOs is somewhat unique for a single type of astronomical object. The computation of a large set of statistics, including some creatively motivated new ones such as M and Q will be an important process to carry out for many different types of astronomical variables that are harvested from future time domain surveys.

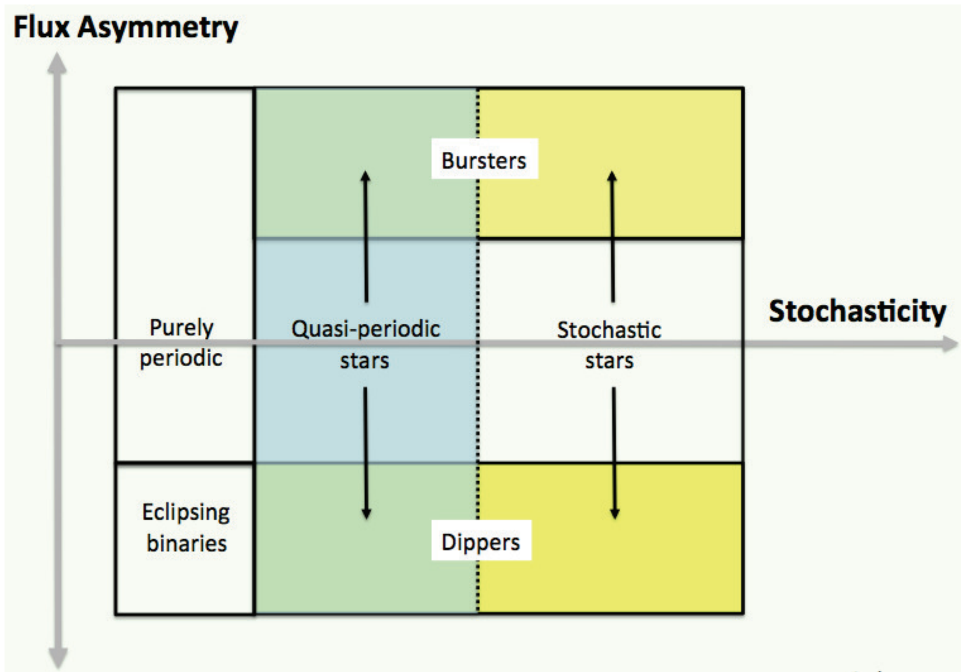


Figure 3. New statistical metrics have helped to classify young star light curves into distinct morphological categories. Here the flux asymmetry is a measure of the tendency of extreme points to lie above ("bursters") or below ("dippers") the median light curve value. The stochasticity is a measure of how periodic (or not) a light curve is. Typically many different statistics must be computed on light curves to find the few that best distinguish different types of behavior.

5. Archival data products

A number of variability analyses have now been carried out on both the YSOVAR project data and the CSI 2264 data (e.g., Rebull et al. 2015, Poppenhaeger et al. 2015, Stauffer et al. 2015, Sousa et al. 2016). To promote reproducibility as well as new science opportunities, we have begun a series of public data releases. These were originally intended to occur through the YSOVAR database, but the lack of visual display (other than interactive light curves for individual sources) decreased its appeal. We therefore opted instead to release the CSI 2264 Spitzer and *CoRoT* light curves through the Infrared Science Archive (IRSA). IRSA provides data archiving for a number of missions including Spitzer, and a stable long-term service.

The CSI 2264 data release[1] included separate master light curve tables for each of the *CoRoT* and *Spitzer* series photometry sets. A third table presents the gathered

[1] http://irsa.ipac.caltech.edu/data/SPITZER/CSI2264

multiwavelength single-epoch photometry that can be used to assemble spectral energy distributions. Each catalog can be queried via a basic search interface. Visual examination of large numbers of light curves is also possible, as we have delivered jpeg time series plots as part of the data release.

A very similar release[2] has also been carried out for a subset of the YSOVAR data that are published as of late 2015. Of note, the associated tables for this set include a number of the statistical quantities computed for YSOVAR light curves.

6. Lessons learned

The flood of time series data from YSOVAR and CSI 2264 has opened up a new realm in the study of YSOs in the time domain. As the K2 mission gets underway, we are now poised to take advantage of the lessons learned from previous campaigns. These include the following:

- Databases with multiple linked tables can be helpful for storage of multifaceted time series and corresponding observations and ancillary data. However, multiple display formats –including those that are more visually illustrative– are needed. Many scientists remain uncomfortable with SQL and prefer web interfaces with information in easily digestible format.
- Many researchers are mining the same data from the literature; combining efforts with comprehensive public databases (e.g., Hillenbrand & Baliber 2015) can help alleviate this problem.
- Uniform standards for literature tables are direly needed to eliminate manual extraction and examination of data from publications.
- Creative statistical metrics are very useful for extracting all of the variability properties of young stars.
- The best platform for dissemination of final, published time series data remains to be determined. However, it is helpful to include visual views of large numbers of light curves in additional to the usual table formats.

Acknowledgments. This work is based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

References

Baglin, A., et al. 2006, in Proc. The CoRoT Mission Pre-Launch Status- Stellar Seismology and Planet Finding (ESA Special Publication), edited by M. Fridlund, et al. (Noordwijk: ESA), vol. 1306, 33

Cody, A. M., et al. 2014, AJ, 147, 82

Hillenbrand, L. A., & Baliber, N. 2015, IAU General Assembly, 22, 2253943

---

[2] http://irsa.ipac.caltech.edu/data/SPITZER/YSOVAR

Howell, S. B., et al. 2014, PASP, 126, 398
Morales-Calderón, M., et al. 2011, ApJ, 733, 50
Poppenhaeger, K., et al. 2015, AJ, 150, 118
Rebull, L. M., et al. 2014, AJ, 148, 92
— 2015, AJ, 150, 175
Sousa, A. P., et al. 2016, A&A, 586, A47
Stauffer, J., et al. 2014, AJ, 147, 83
— 2015, AJ, 149, 130