Automated Variability Classification and Constant Stars in the Kepler Database

J.E. Neff[1,3], M.A. Wells[1,4], S.N. Geltz[1], A. Brown[2]

[1]*Dept. of Physics & Astronomy, College of Charleston, Charleston, SC, USA 29424*

[2]*Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, CO, USA 80309*

[3]*On detail to National Science Foundation, Arlington, VA, USA 22230*

[4]*Now at Dept. of Astronomy & Astrophysics, Pennsylvania State University, University Park, PA 16801*

Abstract. We have monitored hundreds of stars at 30-minute cadence as part of our Cycles 1-4 Kepler Guest Observer programs. Our initial sample was chosen largely by their GALEX colors (i.e. "UV bright"), so we anticipated that it would contain a large number of spotted stars. In fact, the sample also contained pulsating A and F stars, pulsating red giants, and a substantial fraction of stars that showed no variability at all on any timescale. Visual inspection of the light curves alone generally is not adequate to determine the cause or even the level of variability, so we developed sophisticated tools to automatically classify stars using various windowing and power spectrum tests. We present preliminary results and discuss the detection limits for variability in the Kepler database.

1. Introduction

1.1 Kepler Light Curves

Over a 4-year period from 2009 to 2013 the Kepler spacecraft obtained total visual brightness measurements every 30 minutes for about 150,000 stars. Individual stars were observed for a minimum of 90 days, and many were observed for the entire duration of the mission. A much smaller number were observed with higher (1-minute) cadence. The observatory design and data calibration procedures were optimized for the detection of planetary transits. Stellar variations on longer timescales are difficult to recover from the data, because the overall flux calibration drifts unpredictably throughout each 90-day observing quarter, and it is radically different from one quarter to the next. Short timescale variations (like planetary transits, but also pulsations, spots, and close binary effects) can still be recovered with some kind of detrending routine. Adopting this approach completely eliminates the possibility of discovering longer timescale variations in the mean level (due to starspot evolution, activity cycles, slow pulsations, etc.).

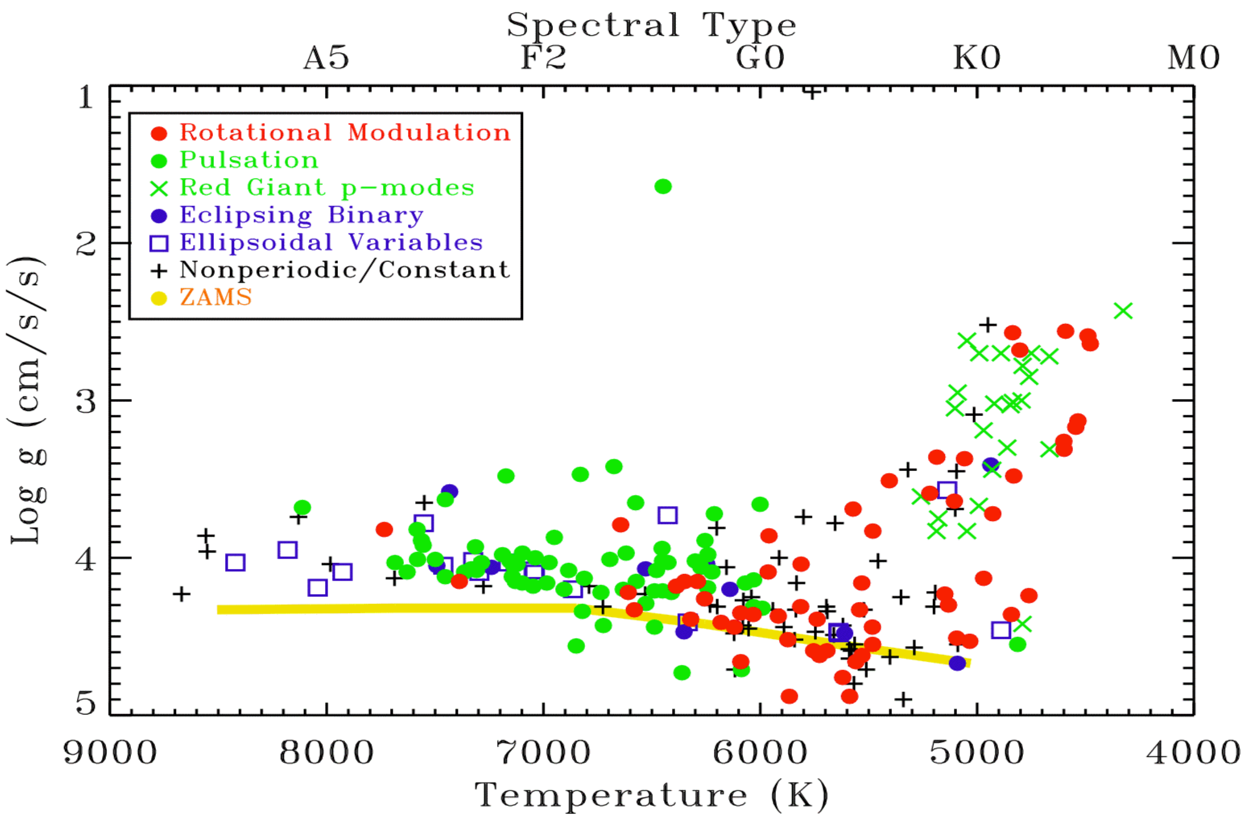

Figure .1: A pseudo HR diagram of our initial GO sample from Paper I. Identifications were made using a labor- and computationally-intensive procedure based primarily close scrutiny of the power spectra. Variability is broken into 4 color-coded classes based on the driving mechanism, but there are multiple subclasses.

Most of the *KEPLER* targets are too faint to have been observed spectroscopically before the mission, although there has been a major effort to perform spectroscopic follow-up of variable targets of many types (e.g. rotational modulation due to starspots or fixed surface patterns, pulsators, and binary effects). The *Kepler Input Catalog* (KIC) lists photometrically-determined properties for all the stars, but these are not uniformly reliable. The *KEPLER* photometric bandpass includes the entire optical spectrum; no color information is available. Given the large number of targets, limited prior knowledge of their properties, lack of reliable longterm flux calibration, how do we efficiently uncover the interesting targets? Which properties of their variability (amplitude, change in amplitude, periodicity, etc.) can

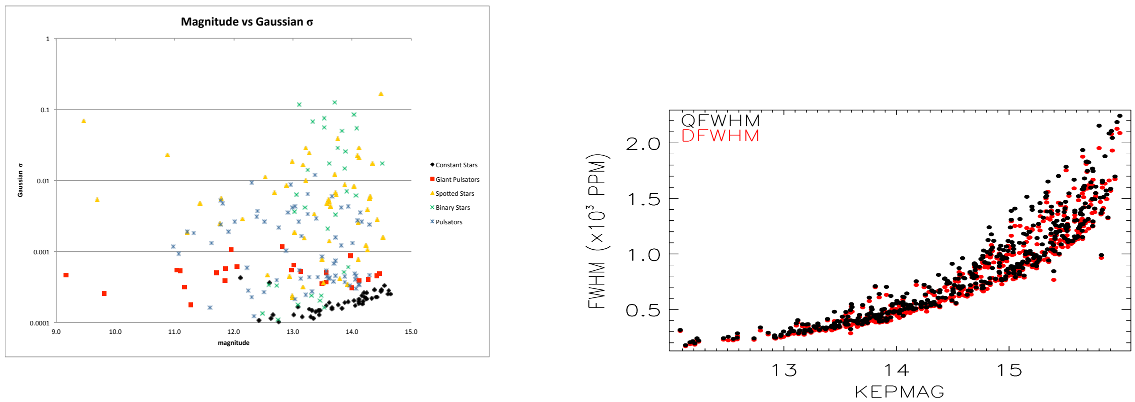

Figure .2: Our automated classification procedures rest fundamentally on a single parameter (the gaussian width of the light curve) probed over characteristic timescales. The left panel shows this measure of variability plotted versus Kepler magnitude for our entire GO sample. The right panel illustrates the effect of daily detrending (red dots) on the detectability threshold.

be parameterized automatically to discover and classify variables of these different types for more detailed follow-up studies?

1.2 Automated Classification of Light Curves

The task of automatically classifying variability based on light curves has been the focus of many efforts. Most automated schemes first obtain parameters from the light curves using a Fourier decomposition, such as the Lomb-Scargle method (Lomb 1976) or some variant. These parameters are then passed to a decision making process: random forest classifier (Richards *et al.* 2011; Dubath *et al.* 2011), multivariate Gaussian mixture classifier (Debosscher *et al.* 2007, 2009; Blomme *et al.* 2010; Debosscher *et al.* 2011), Bayesian network classifier (Eyer & Blake 2005; Sarro *et al.* 2006, 2009; Blomme *et al.* 2011; Debosscher *et al.* 2011). Eclipsing binaries, due to their ability to yield useful stellar parameters such as radius and mass, have seen attention from Prša *et al.* (2011), Slawson *et al.* (2011), Matijevič *et al.* (2012).

1.3 Manual Classification of Our Guest Observer Sample

Brown et al. (2014; hereinafter Paper I) report on the classification of a sample of 239 stars observed in a *KEPLER* multi-year Guest Observer (GO) program. This sample was chosen by *GALEX* colors and x-ray emission, so it was expected that it would contain a large fraction of magnetically-active, spotted stars. We classified the stars in that sample using primarily the power spectrum distribution with some aid from visual inspection of the light curves. The sample included pulsating A and F stars, pulsating G and K giants, and close binaries, as well as spotted stars. We discounted using the parameters in the KIC. This classification process was very time-intensive, and these procedures became impractical as our sample continued to expand as more data became public and as more ground-based follow-up observations were obtained. Calculating the power spectra, in particular, became a limiting factor.

1.4 Semi-Automated Classification of Kepler Light Curves

We therefore used the knowledge gained from our original sample to develop semi-automated and, eventually, fully-automated procedures to classify *KEPLER* light curves without the need for visual inspection, calculation and analysis of power spectrum, period finding, or excessive operator input. The resulting pipeline, testing it with a larger sample with supporting ground-based data, and its application to a randomly-selected sample of 1000 Kepler light curves is fully described in Wells *et al.* (2014). Many of the tools and modules are of general utility. For example, a single command operates on an input directory of any number of fits files and automatically classifies light curves into about 10 physically-distinct mechanisms, including non-variable stars (which comprise a surprisingly large fraction). Plots can be created automatically for each step or generated later with modular tools that follow standardized calling sequences. The tools developed for file manipulation and plotting can be applied to any kind of two-parameter data tables in fits or ascii format; nothing is specific to *KEPLER* data.

Our methodology is relatively simple, extremely fast, and fairly robust. Rather than perform frequency analysis of signals, we analyze the properties of the variability within each light curve. We also employ detrending techniques over a range of timescales. Several relationships were drawn using the well-studied sample in Paper I as a training set. These relationships were then applied to a test sample to validate the technique before being applied to a randomly selected sample of *KEPLER* targets. We discuss these results and several implications for *KEPLER* calibration and data analysis.

2. Technique

As an initial step, we classified our 239 GO targets (the "GO sample") by variability type, shown in Fig. 1 as a pseudo "HR Diagram" (Paper I). The ultraviolet brightness of rotational modulators and close binaries is presumably due to enhanced magnetic activity. Our sample also included a large number of pulsators. Stars with no detectable variation occur throughout the HR diagram. Our labor-intensive classification procedure involved computing and visually analyzing power spectra for all the stars. This became computationally prohibitive the Kepler database continued to grow. In addition, our ground-based followup programs were extended to find additional rotational modulators (the "MMT sample"). It therefore became necessary to develop an automated system to classify the stellar variability. We used the GO sample as a "training set" and the MMT sample as a trial set. When the procedures were robust, we used them to automatically classify a randomly-selected sample of Kepler targets (the "Random Sample").

Our automated classification procedure (Wells et al., in preparation) is constructed as a modular "pipeline". Starting with a directory full of Kepler data files, it automatically generates a large number of optional ancillary plots and tables and provides a single list of Kepler ID v. variability type. It is written so generally that it will work on any kind (not just Kepler) of Fits tables, allowing the user to plot or extract any of the data fields into IDL vectors and ascii tables. It was designed to replicate the results of our intensive analysis of the GO sample, so it does not classify all possible types of variability (though these can be

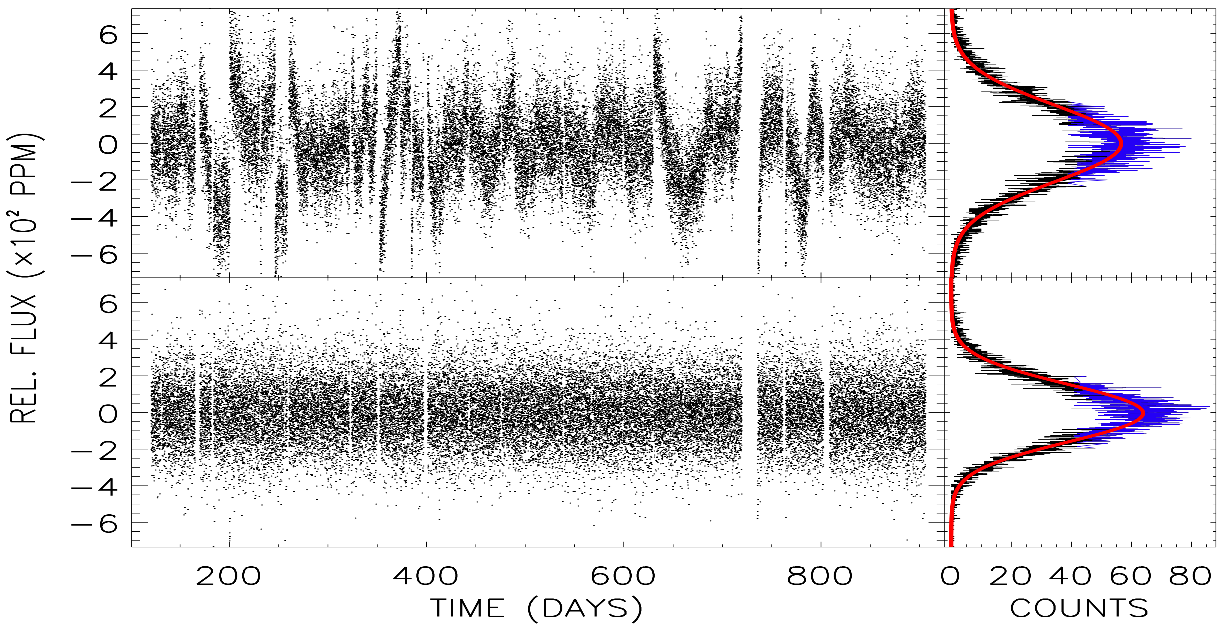

Figure .3: We use as a basic variability parameter the width of a histogram of the Kepler flux distributed about a mean value over some characteristic timescale. The top panel shows a raw Kepler light curve normalized to each quarterly mean (which is then set to 0). None of this variability is real, as illustrated in the lower panel, which is normalized to the daily mean. The right-hand panels show the histogram with a gaussian fit superimposed. The blue portion highlights the part of the histogram above a half maximum value that is determined without the additional step of fitting a gaussian. Widths determined either way are consistent so long as the histogram is symmetrical (not always the case; binary stars, for example, show a dual peaked histogram).

added as modules). We then applied it to the larger MMT sample and verified (using visual inspection of power spectra and light curves, as well as outside information like the Kepler input catalog) that it yielded correct classifications. We then selected a random sample of Kepler data and fed it through the pipeline. The basic principles are illustrated in Figs. 2 through 4.

3. Classification Results for Three Samples

Approximately 73% of the targets in our ultraviolet/x-ray selected GO sample are variable, compared to only 40% of the stars in our randomly-selected sample (see Table 1). The subsequent selection from our MMT sample was much more successful at identifying spotted star candidates and eliminating pulsators. Fig. 5 shows the result of our classification of a random sample of Kepler stars. Delta Scuti pulsators and pulsating K giants are unambigously identified, while Gamma Dor pulsators and Gamma Dor/Delta Scuti hybrids are somewhat intermingled with low-level rotational modulation of F-type stars. Three types of

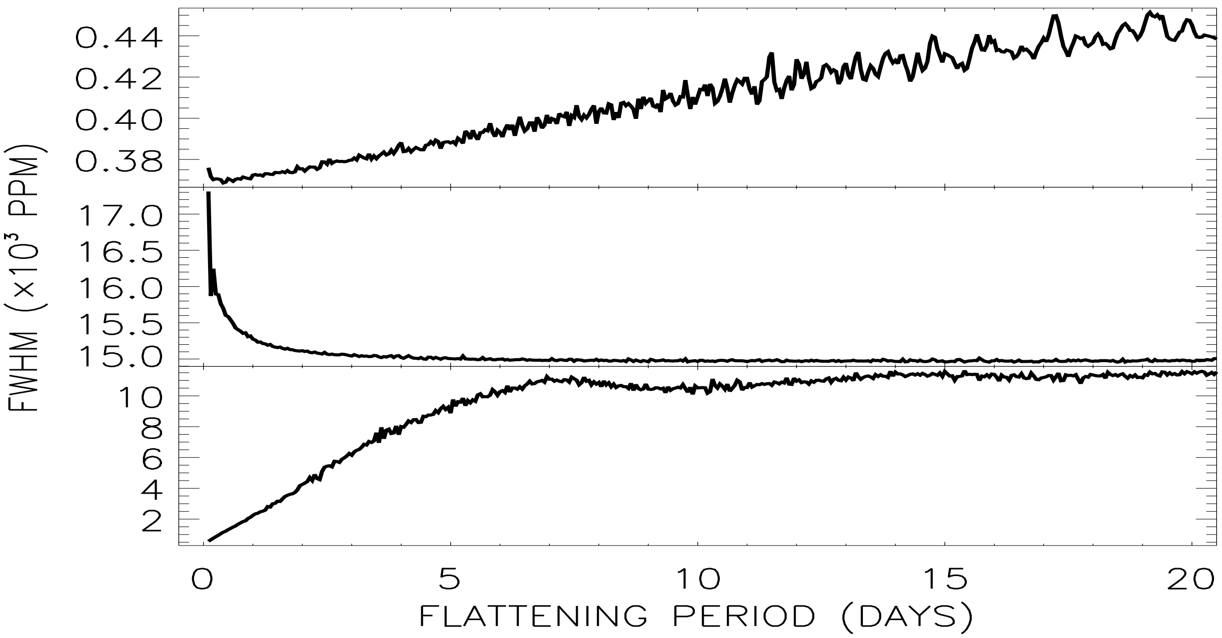

Figure .4: We calculated the variability parameter over a range of timescales for each lightcurve. This figure illustrates the power of this technique to discriminate constant stars (top) from Delta Scutis (middle) and spotted stars (bottom). These curves are linear outside the the intrinsic variability timescales of the target. In the absence of instrumental variation, we would expect constant stars to show the same width regardless of the detrending timescale. We found random variation in the Kepler light curves on timescales longer than a day.

Table .1: Results from Automated Classification Procedure

Classification	GO Sample (239 stars)	MMT Sample (451 Stars)	Random Sample (950 stars)
Rotational Modulation	24%	31%	18%
Pulsation	41%	6%	19%
Binarity	8%	0.7%	1.5%
Not Variable	27%	63%	62%

rotational modulation are clearly separated: active late-type stars, fixed-pattern Ap stars, and pre-mainsequence stars. Some binary systems include pulsating or spotted stars.

4. Constant Stars in the Kepler Database

We were initially surprised at the high number of "constant" stars in our GO sample. If they are ultraviolet/x-ray bright, they should be active and should show rotational modulation. One possibility is that they could be bright in the ultraviolet because of low metallicity. Alternatively, some could have entered a Maunder minimum phase since the GALEX and x-ray data were obtained. As the top panel of Fig. 3 illustrates, you can not tell if a star is constant by looking at its light curve. We have begun a longterm program for undergraduate researchers at CofC to monitor these constant stars from quarter to quarter, looking for evidence that stars might be entering or leaving Maunder minimum phases. The Kepler mission design and calibration pipeline is designed to search for short-timescale transits. The overall flux level drifts quasi-randomly (though with some patterns that repeat every 4 quarters). Constant stars were initially identified by their complete lack of periodic variation seen in the power spectra. We then identified a "threshold" level of detectability (Fig. 2) and characterized the nature of the drifting mean flux level (Fig. 4). In the end, we found that detrending the data using daily mean levels removes the artificial variability induced by the instrument or calibration routines and produces a narrow, symmetric distribution of flux measurements about the mean. Unfortunately, doing so removes all information about longer term variability. In general, the quarter-to-quarter standard deviation of the width of the flux distribution is less than 10%. In all cases where it was substantially higher, visual inspection of the light curves revealed a clear instrumental signature in some quarters that was not present in others. In order to see solar-like variability (low level over 10s of days) or cycle variability, greater attention needs to be focused on removing the randomly drifting instrumental signature. The procedures and results are described in more detail in Wells *et al.* (2014).

Acknowledgements. This work contains results obtained using the NASA Kepler satellite and from the Apache Point Observatory and the MMT (using NOAO community access time). Funding was provided by NASA Kepler grants NNX10AC51G, NNX11AC79G, and NNX12AC85G to the University of Colorado, by NSF grant AST1109695 to the College of Charleston, and by the South Carolina Space Grant Consortium. JEN received additional support through NSFs Independent Research and Development program available to Program Officers.

References

Blomme, J. *et al.*2010, ApJ, 713, L204

Blomme, J. *et al.*2011, MNRAS, 418, 96

Brown, A., Neff, J.E., Wells, M.A., Ayres, T.R., 2014, AJ, submitted

Debosscher, J. *et al.*2007, A&A, 475, 1159

Debosscher, J. *et al.*2009, A&A, 506, 519

Debosscher, J. *et al.*2011, A&A, 529, A89

Dubath, P., Rimoldini, L., Süveges, M., *et al.*2007, MNRAS, 414, 2602

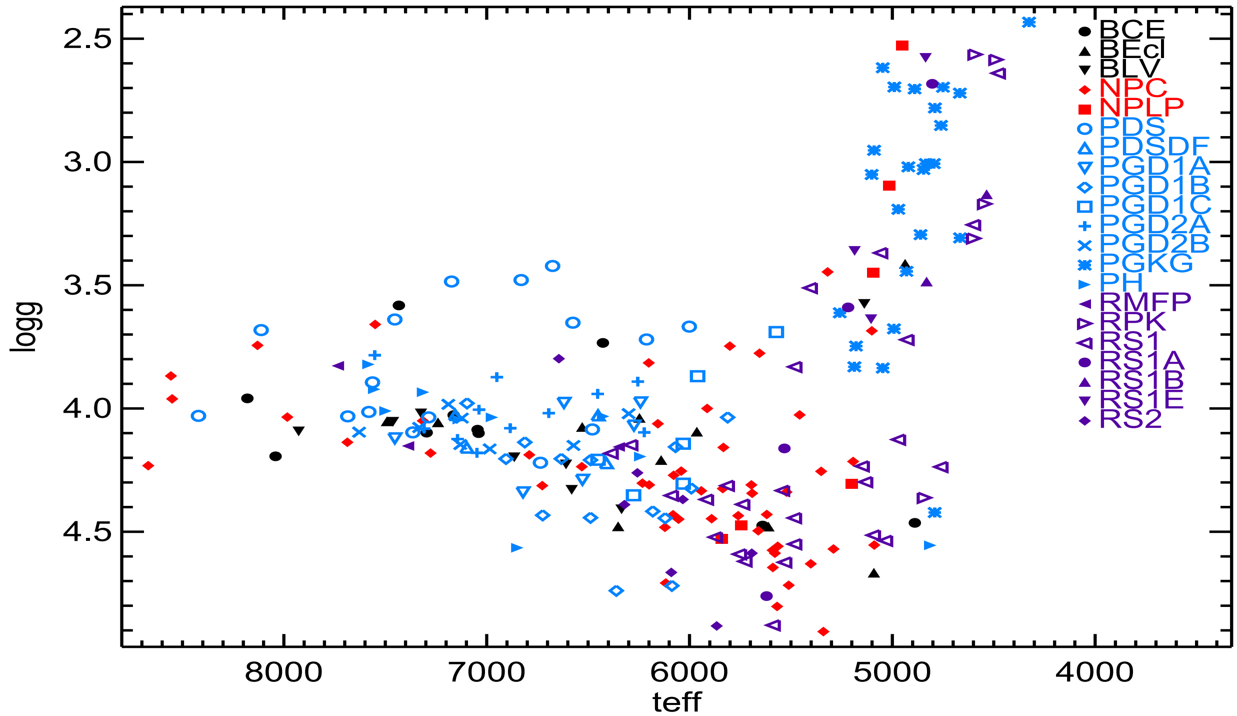

Figure .5: A pseudo HR diagram of our Random Sample of 950 Kepler targets. Identifications were made using our automated classification routine. Variability is broken into 4 color-coded classes based on the driving mechanism, but multiple subclasses are identified by the classification pipeline.

Eyer, L., & Blake, C. 2005, MNRAS, 358, 30

Lomb, N. R. 1976, Ap&SS, 39, 447

Matijevič, G. *et al.*2012, AJ, 143, 123

Prsă, A. *et al.*2011, AJ, 141, 83

Richards, J. W., Starr, D. L., Butler, N. R., *et al.*2011, AJ, 733, 10

Sarro, L. M., Sánchez-Fernández, C., & Giménez, Á. 2006, A&A, 446, 395

Sarro, L. M., Debosscher, J., López, M., & Aerts, C. 2009, A&A, 494, 739

Slawson, R. W. *et al.*2011, AJ, 142, 160

Wells, M. A., Neff, J. E., Brown, A. et al. 2014, in preparation.