

# Palomar Transient Factory Data Processing & Archive

Frank Masci (IPAC / Caltech)

iPTF Summer School, July 2016

# The Palomar Transient Factory (PTF)

---

- PTF is a robotic wide-field time-domain survey in the optical
- Carried out on 48" Schmidt telescope on Mt Palomar
- Can survey up to 3000 deg<sup>2</sup> per night to  $R \sim 20.5$  ( $5\sigma$ )
- Median seeing  $\sim 2$  arcsec.
- Cadence (repeatability): 1 min to 5 days: can discover SNe, variable stars, exoplanets, asteroids...
- PTF science operations ran from Mar 2009 – Dec 2012
- Became the *intermediate* PTF (iPTF) after this

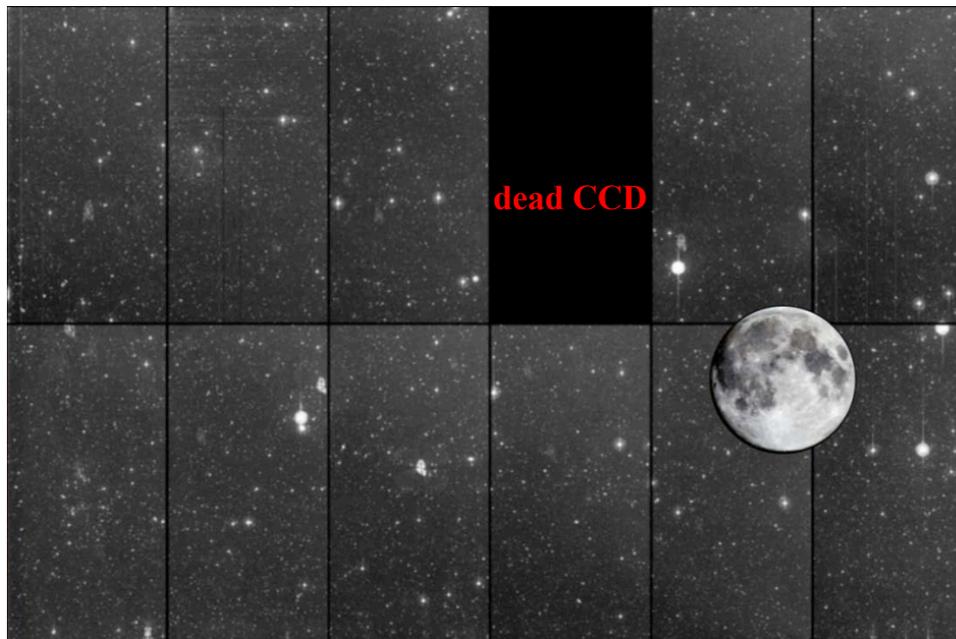


Edwin Hubble, 1949

# The PTF Survey Camera

---

- Based on the original CFHT 12K camera
- 7.26 deg<sup>2</sup> FOV, 11 working CCDs (2048 × 4096 pixels each); ~ 92 megapixels in total
- Pixel size on sky at FOV center is ~ 1 arcsec
- Typically 60 second exposures with 30 second readouts
- On average ~ 20 to 30 exposures per survey field per year
- Generates >100 TB of image data products and 10<sup>12</sup> extracted sources **per year**

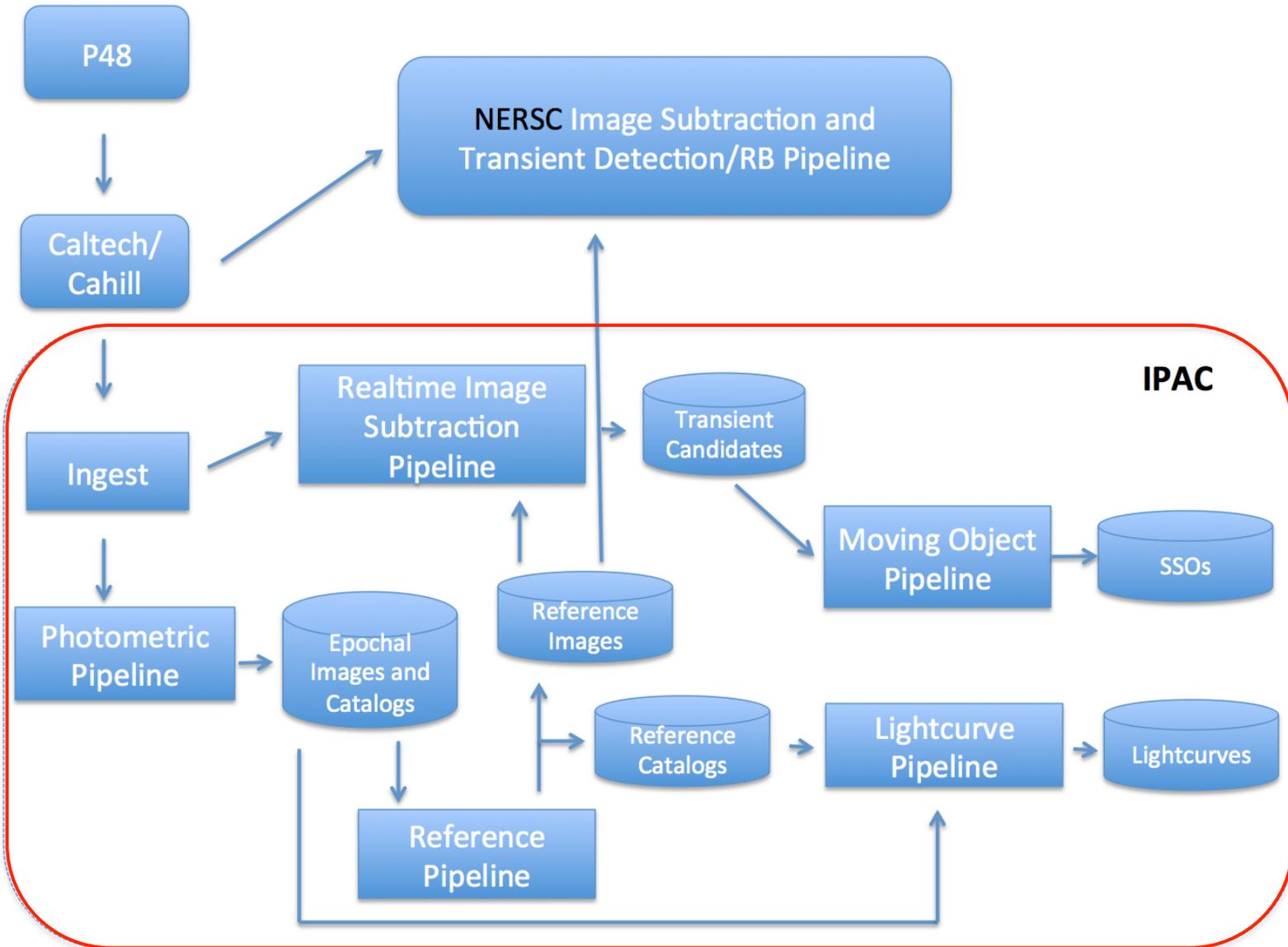


# Data processing overview

---

- Data flows through multiple pipelines, creating a variety of science products tailored for different scientific purposes. These pipelines run on different timescales.
- **Photometric (or frame processing) pipeline:** daily (end-of-night) processing to produce high quality instrumentally-calibrated images and source catalogs
- **Reference image pipeline:** combines high quality frames into deeper images (coadds) – products are used in the real-time and lightcurve pipelines below. Reference images are periodically made, depending on availability of good data for a given field/chip (more later).
- **Lightcurve (or relative-photometry) pipeline:** uses source catalogs from the photometric pipeline to create high precision photometric lightcurves. Also periodically made.
- **Real-time pipeline:** runs throughout a night to support transient-discovery via image-differencing. Outputs feed into various science marshals: extragalactic; galactic; solar system
- Interfacing with the above: an advanced data archive with exploratory tools to support long-term data curation and public distribution – storage of raw data, processed images, and source catalogs

# Data Flow



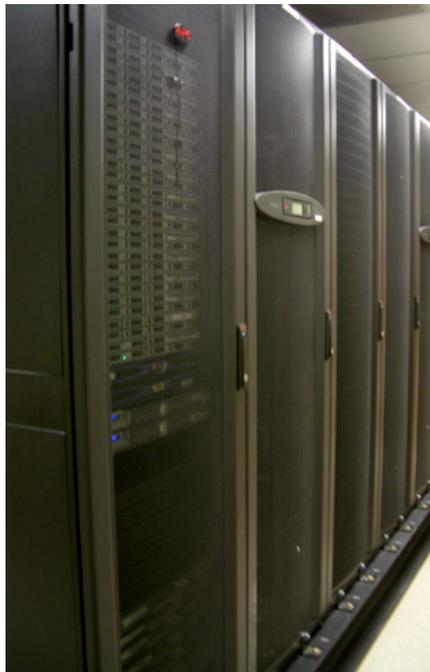
# Data Transfer

---

- Raw CCD camera-image data from individual exposures are packaged into Multi-Extension FITS formatted files at the telescope.
- First transferred via a microwave link to the San Diego Supercomputing Center; then to Cahill at Caltech.
- At Cahill the data forks to two places: **NERSC** at LBNL, and **IPAC** at Caltech:
- **NERSC**: developed first version of the realtime pipeline: image-differencing and transient-discovery with machine learned vetting; designed around SN discovery; still operating today.
- **IPAC**: all flavors of processing described on slide 4, including a reimplementaion of the realtime/transient-discovery pipeline for future use (became live ~ mid 2014);
  - also maintains the central archive for all PTF/iPTF products; served through IRSA

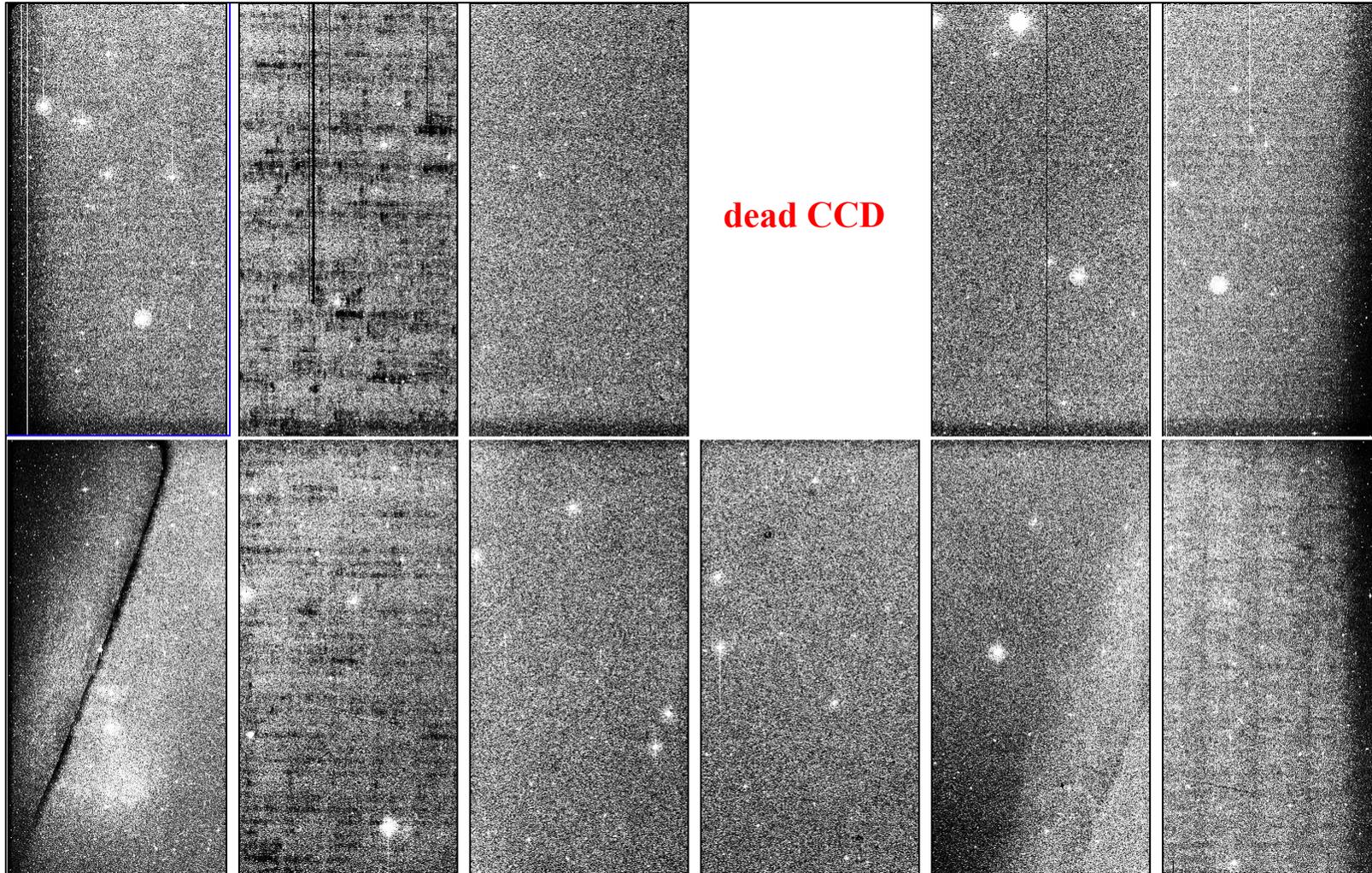
# Infrared Processing and Analysis Center

- IPAC is a Multi-mission Science Center (IRAS, ISO, Spitzer WISE, Herschel, Planck, 2MASS...)
- iPTF generates ~1TB of data every 4-5 days.
- iPTF cluster has 24 machines with 240 cores total
- Roughly 0.7 PB of spinning disk
- Databases, archive and file servers; tape backup
- Maintains archive data retrieval interfaces/services



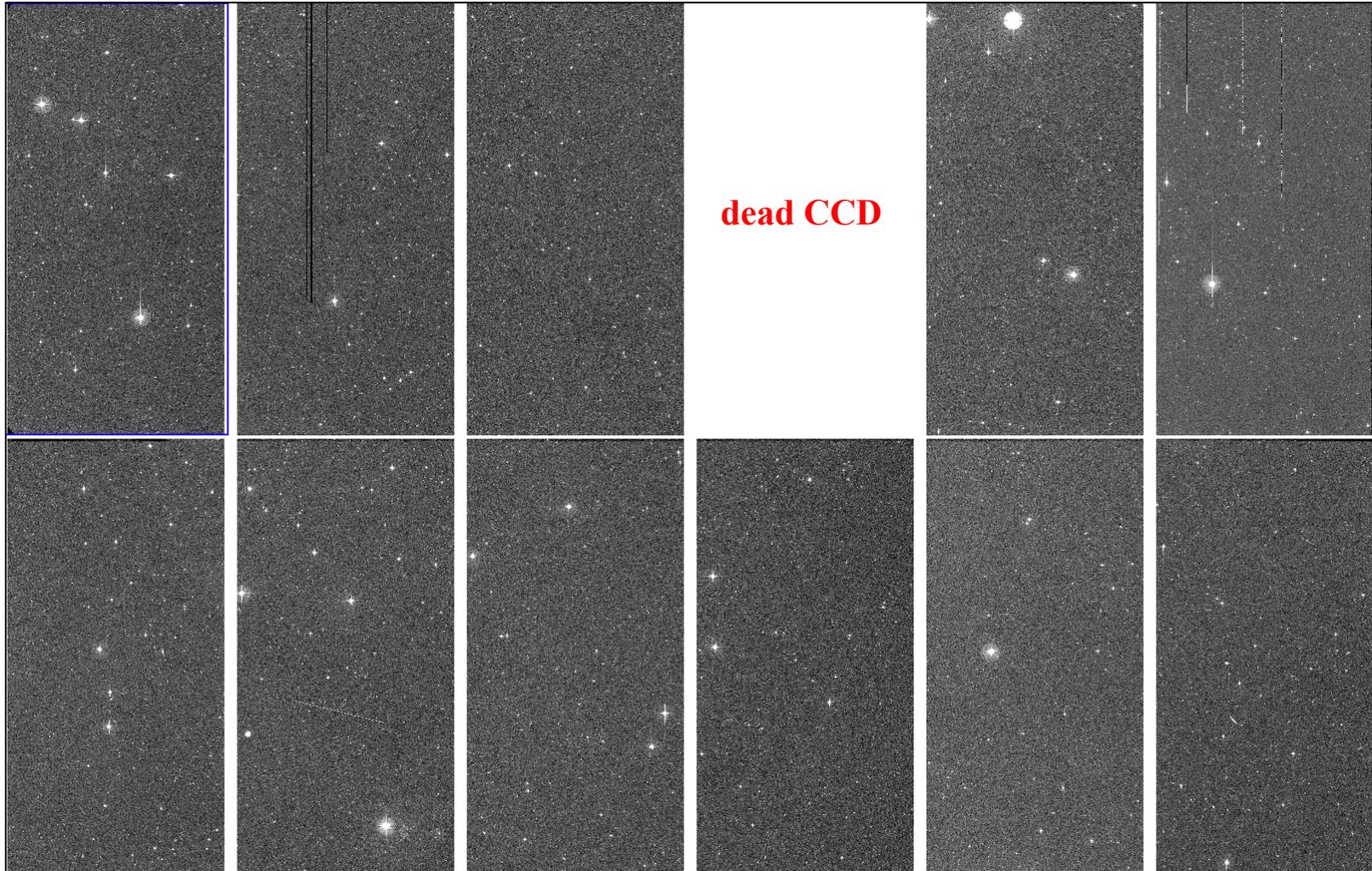
# Raw CCD images from one exposure

---



## Processed CCD images from same exposure

---

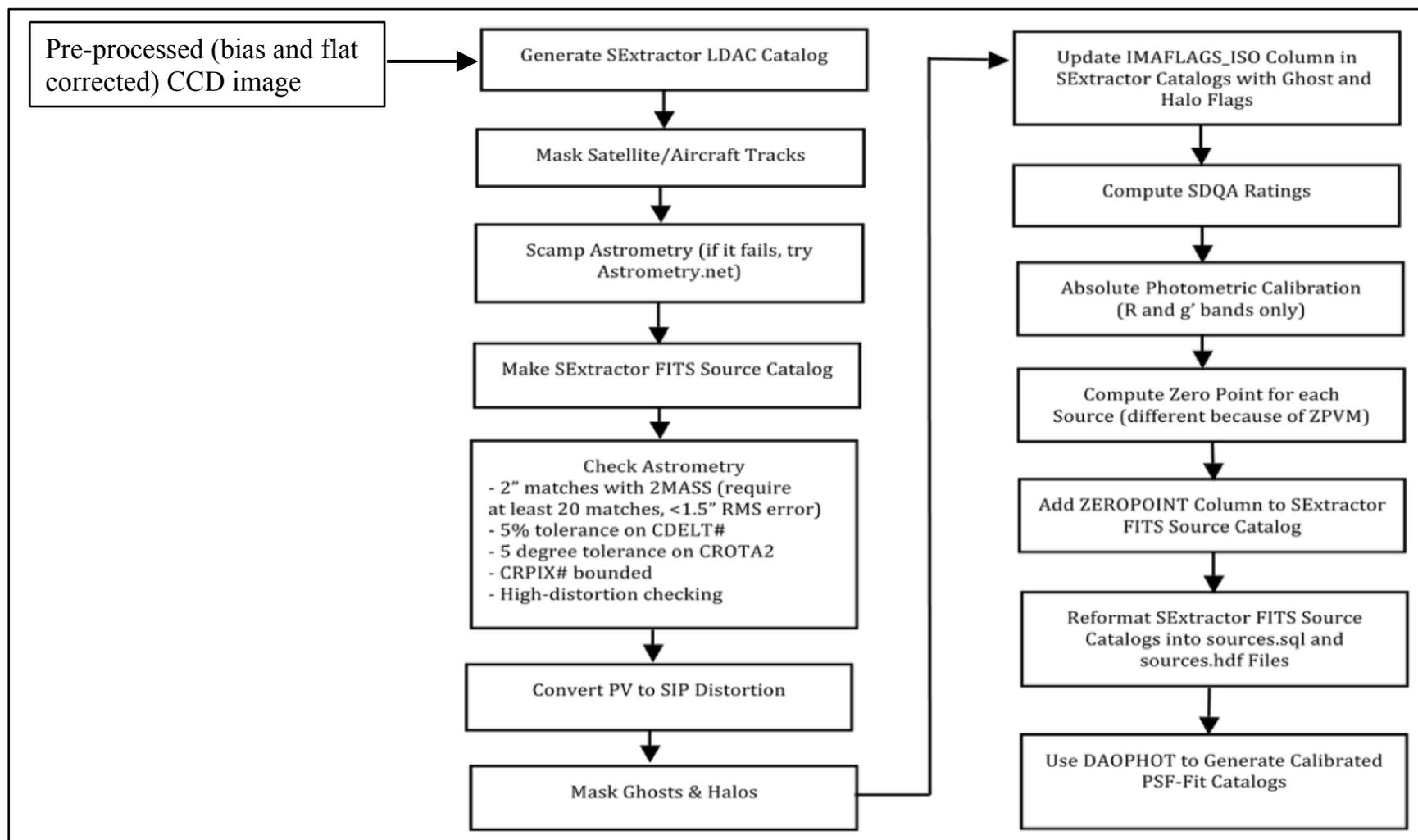


# Photometric (frame-processing) pipeline

---

- Triggered at the end of the night, after all the data has been received.
- Instrumental calibrations are derived from an entire night's worth of data. More specifically, flat-field maps are derived from the on-sky data.
- Photometric calibration is from a nightly model-fit using the SDSS overlap region: fits for color and airmass terms, spatial and time-dependent throughput variations. Accuracy is ~ few percent.
- Astrometric (and distortion) calibration is done at the individual CCD-image level against a combined SDSS and UCAC4 catalog. Typically good to 0.15 arcsec in unconfused regions.
- Outputs are calibrated single-CCD FITS images with bit-masks and accompanying source catalogs in FITS binary table format – both aperture and psf-fit photometry is provided.
  - These products are archived at IPAC and available 1 – 3 days after observation.
  - Subsets are made publically available periodically (more later).

## Photometric (frame-processing) pipeline steps

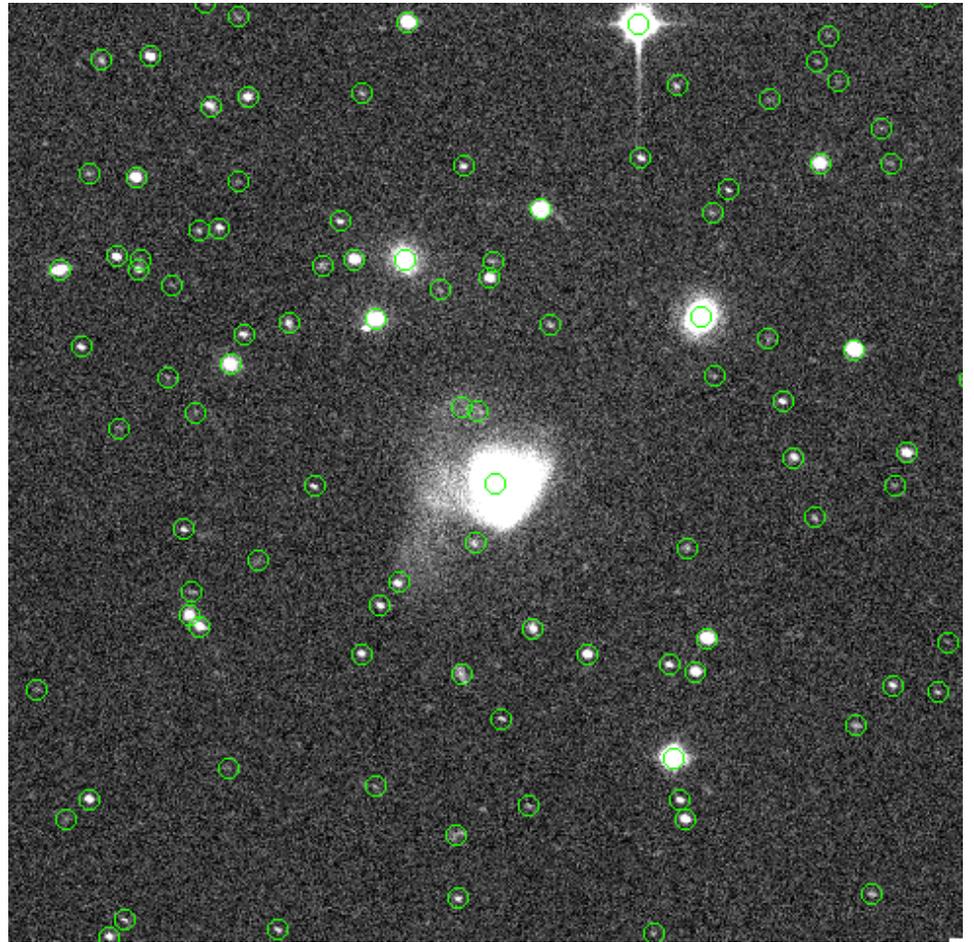


*from Laher et al., 2014, PASP, 126, 674*

# Photometric (frame-processing) pipeline output

---

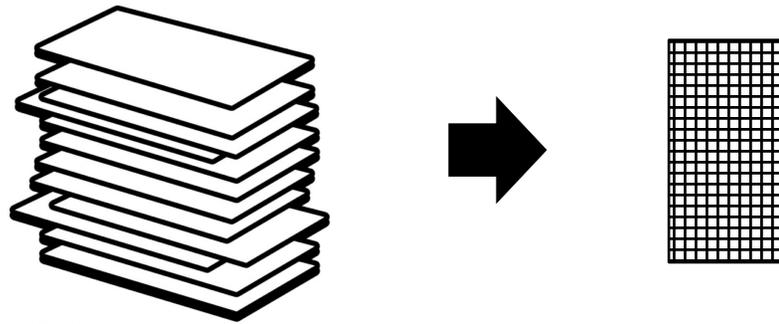
- Zoom-in on *R*-band image of Arp 220 (galaxy merger); 8 arcminutes across.
- Extractions from aperture photometry catalog are overlaid.
- Single-epoch (CCD-) based products archived:
  - processed CCD image with metadata
  - bit-mask image identifying bad pixels
  - calibrated aperture photometry catalog
  - PSF-fit photometry catalog
  - raw image data also available
  - all products above are in FITS format
  - nightly processing log



## Reference image (co-addition) pipeline

---

- When enough individual CCD exposures accumulate, the “reference image” pipeline is triggered
- This pipeline coadds (combines) the “best” image data for a given CCD, field, and filter: i.e., with best seeing, photometric conditions, and astrometry
- Images are first reprojected, then combined using pixel-stack averaging with outlier trimming.

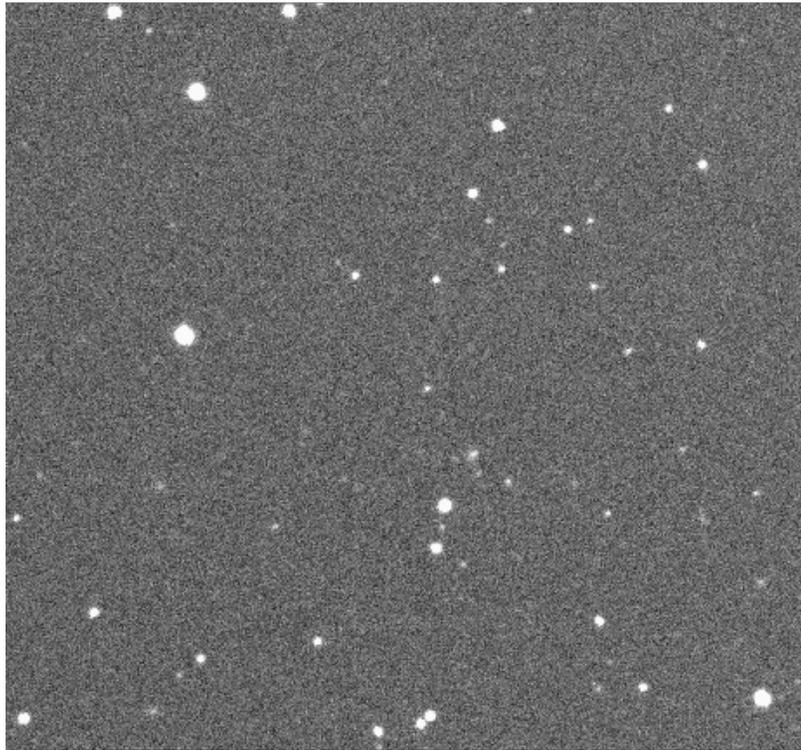


- The coadds are images of the “static” sky as represented by the state of the input CCDs used
  - deeper than the individual exposures: currently, stacks are ~ 5 to 50 images deep.
- Source catalogs are also generated from these images: both PSF-fitting and aperture (SExtractor)
- Reference image products: images, coverage maps, and catalogs are publically available.
- Products support the real-time (image subtraction) and light-curve pipelines

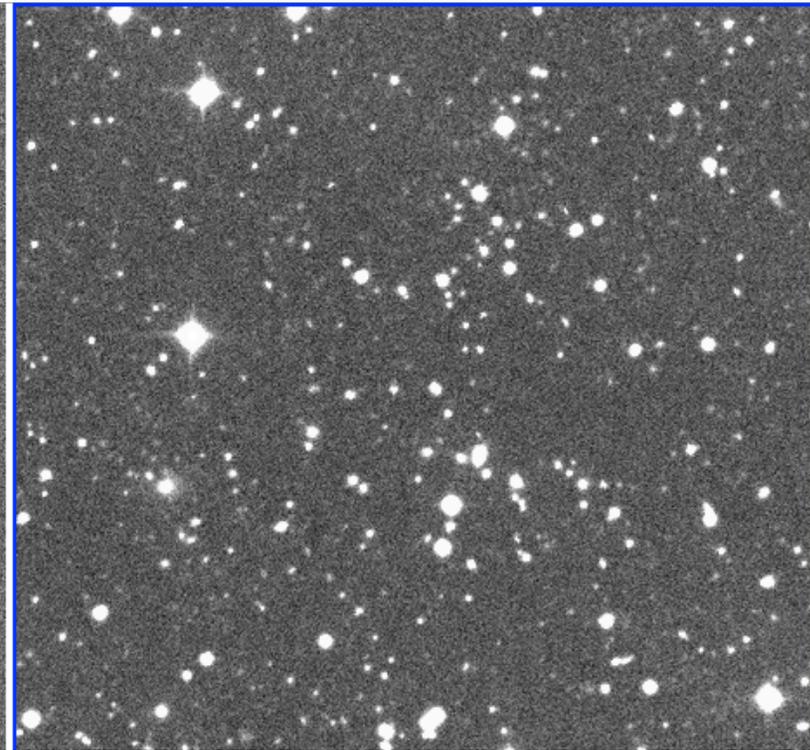
# Reference image example

---

Single image, 60 sec exposure in R filter



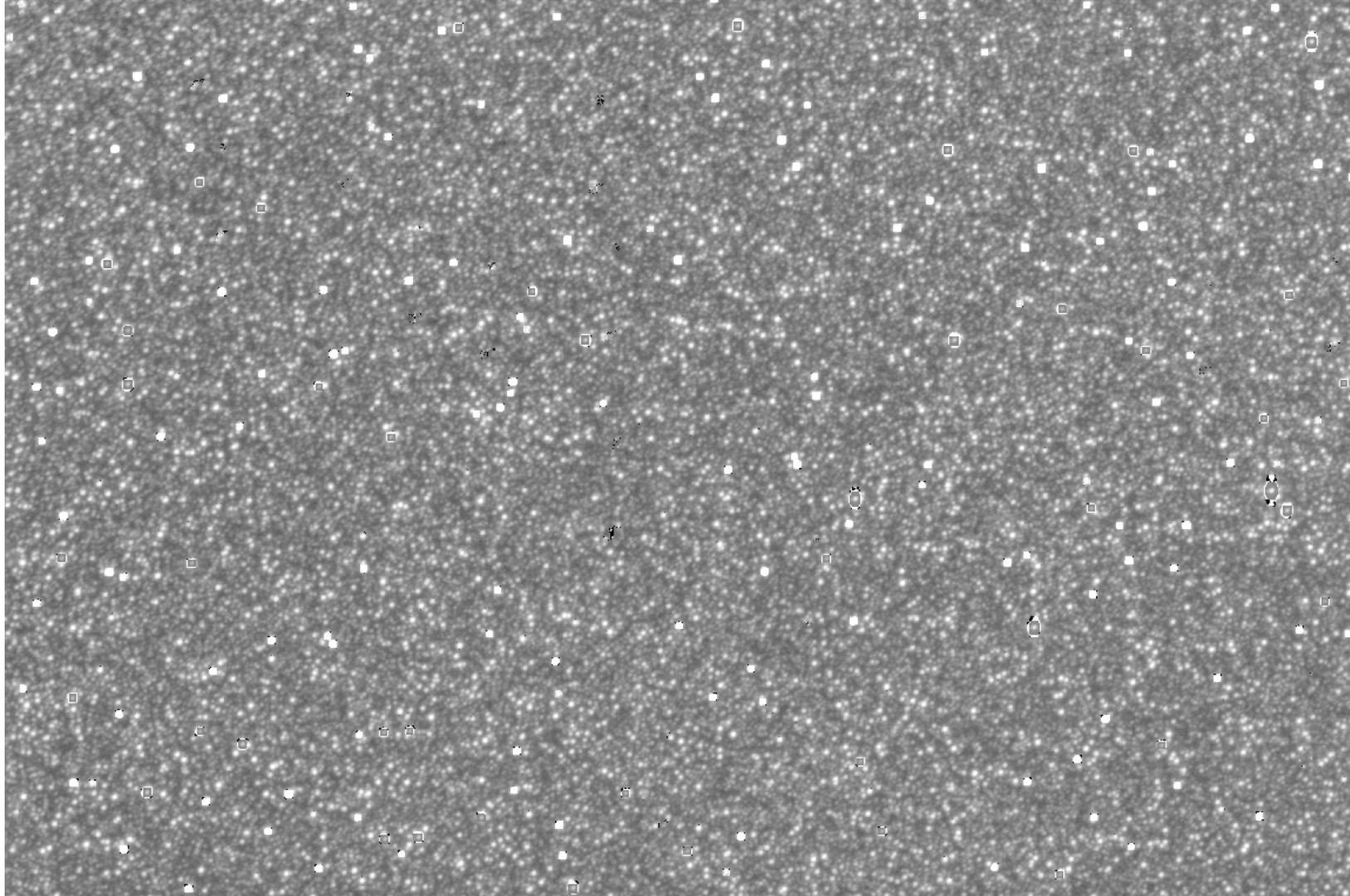
Stack of 34 images (field 5257, CCD 7)



# Reference image example (near galactic center)

---

Stack of 10 images (field 1549, CCD 5, R filter)



# Deeper image co-adds (courtesy: Jason Surace)

---



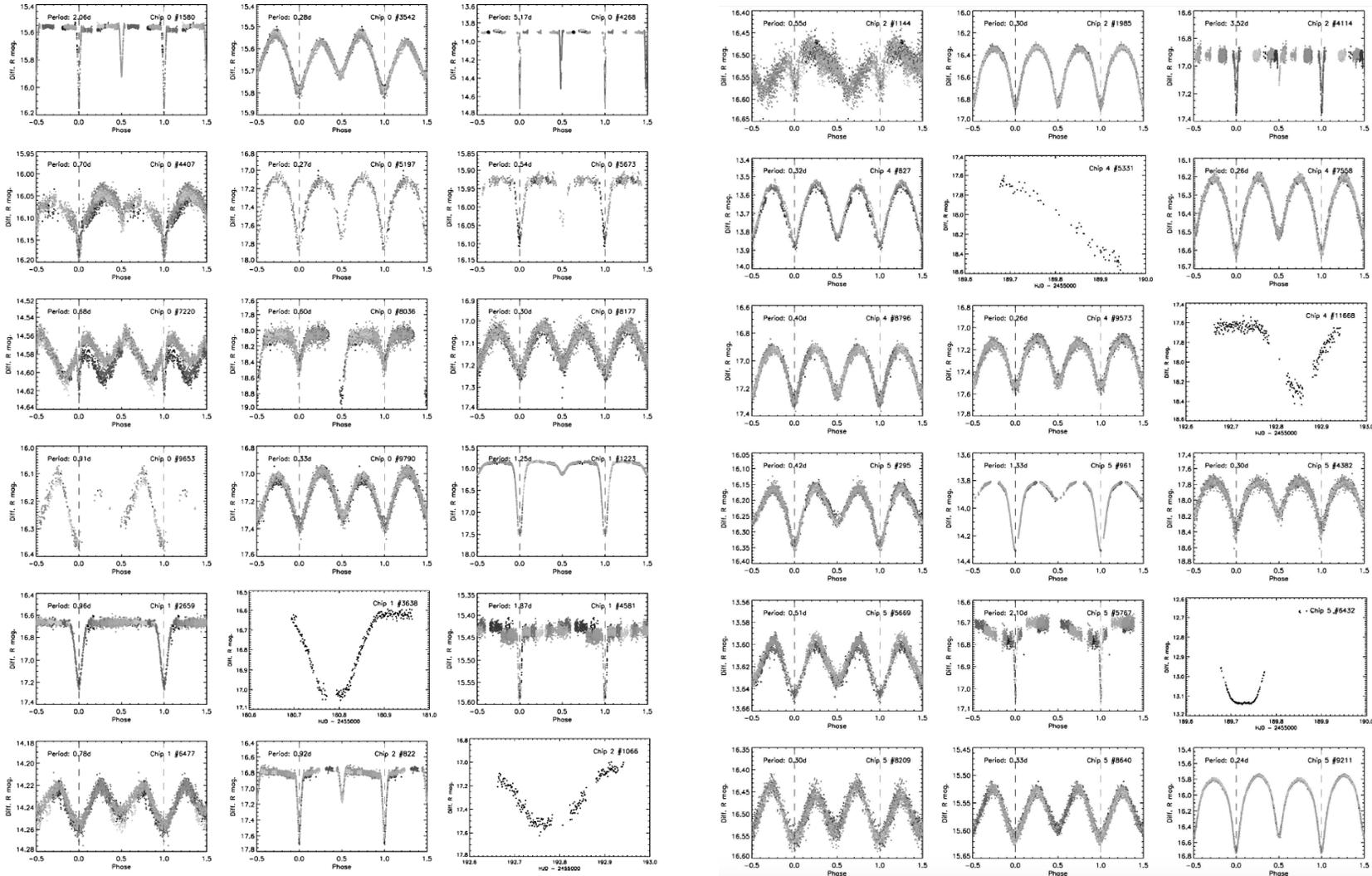
## Lightcurve (relative photometry) pipeline

---

- At the end of each night, all detected sources from the photometric pipeline are matched against the reference-image source catalog for a given field, CCD, and filter
- The “cleanest” least variable sources are used as anchors for the relative photometric calibration
- Individual image gain-correction factors are computed using a least-squares fitting method
- Application of these refined gain-correction factors improves the overall relative calibration to a few millimag for bright sources
- This pipeline is triggered on timescales of typically 1 to 2 weeks
- A lightcurve database for a subset of fields and epochs will be publically available in Dec 2016

# Example PTF lightcurves from the Orion project

Binary star lightcurves; from Van Eyken et al. (2011)



## Real-time pipeline overview

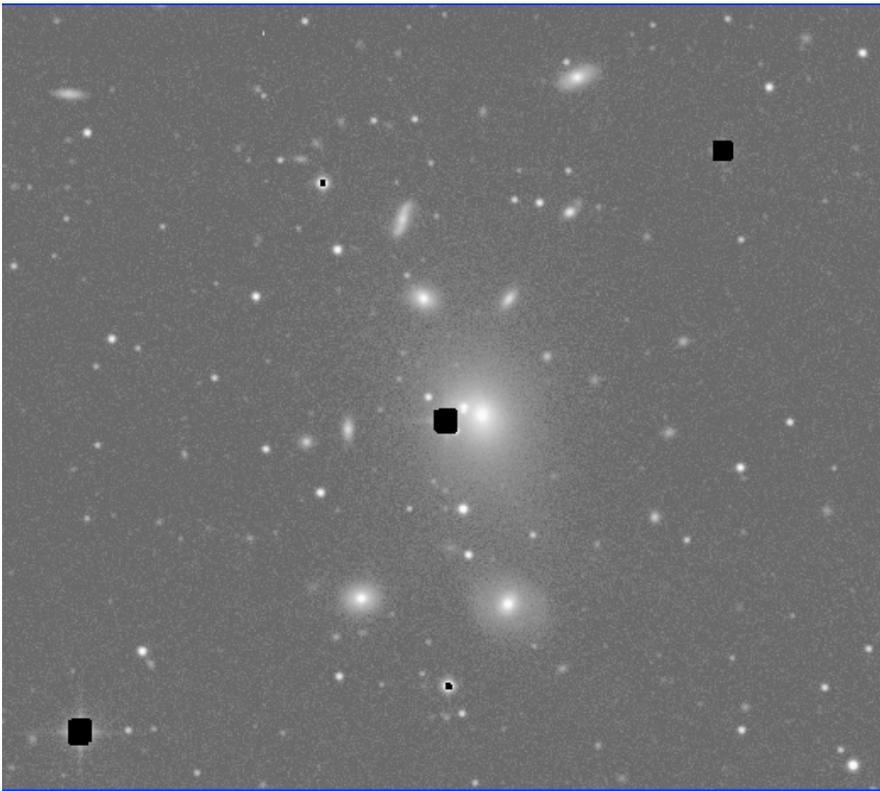
---

- Uses image-differencing against the reference-image library to extract transient candidates.
- Candidates are then automatically “scored” using machine learning.
- Data is processed in near real-time as it’s received; turnaround is 10-25 minutes from telescope to vetted transient candidates
- Outputs are used for same-night follow-up of “interesting” candidates
  - pushed to an external gateway for pickup by the science marshals: galactic, extragalactic, solar-system, and generic ToO alerts
- Difference images and transient-source catalogs are astrometrically and photometrically calibrated
  - both aperture and PSF-fit photometry is performed
- Transient sources from this pipeline also feed a “streak detection” module to find fast-moving objects and a moving-object pipeline to construct moving-object tracklets
- Products from this pipeline are not publically available at this time.

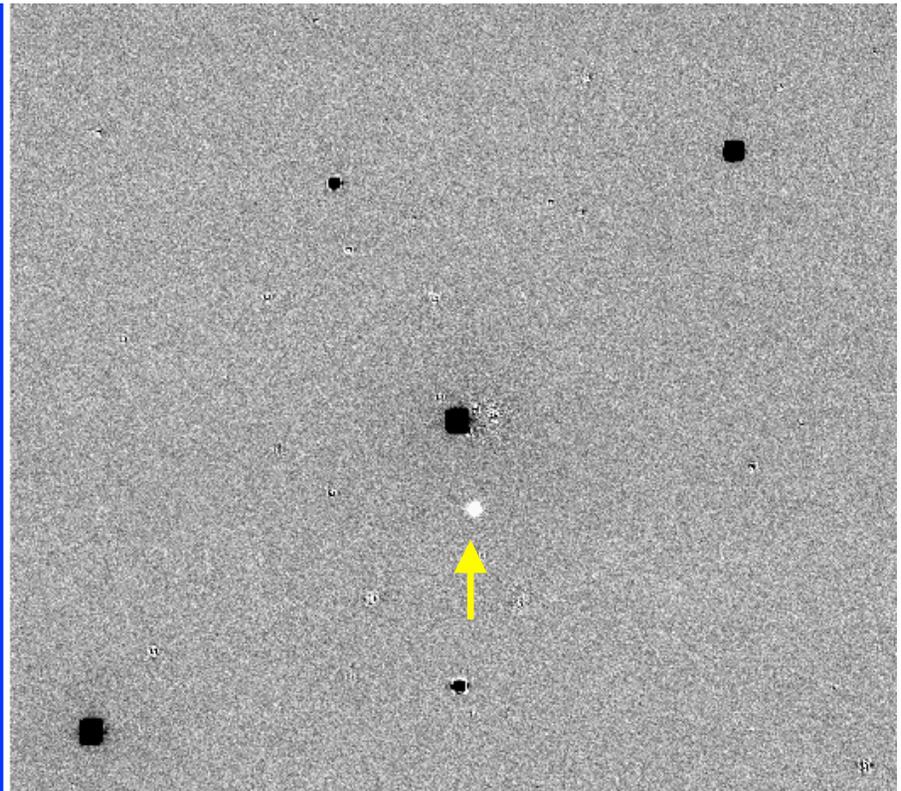
## Example difference image

---

*science image exposure (zoom)*



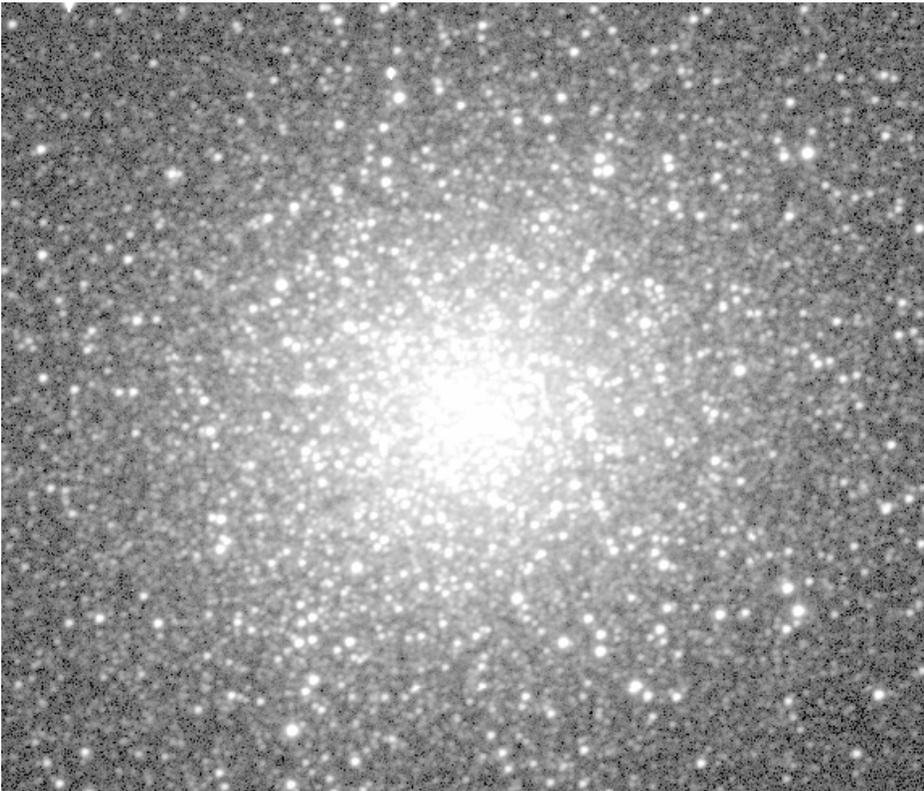
*“exposure minus reference”*



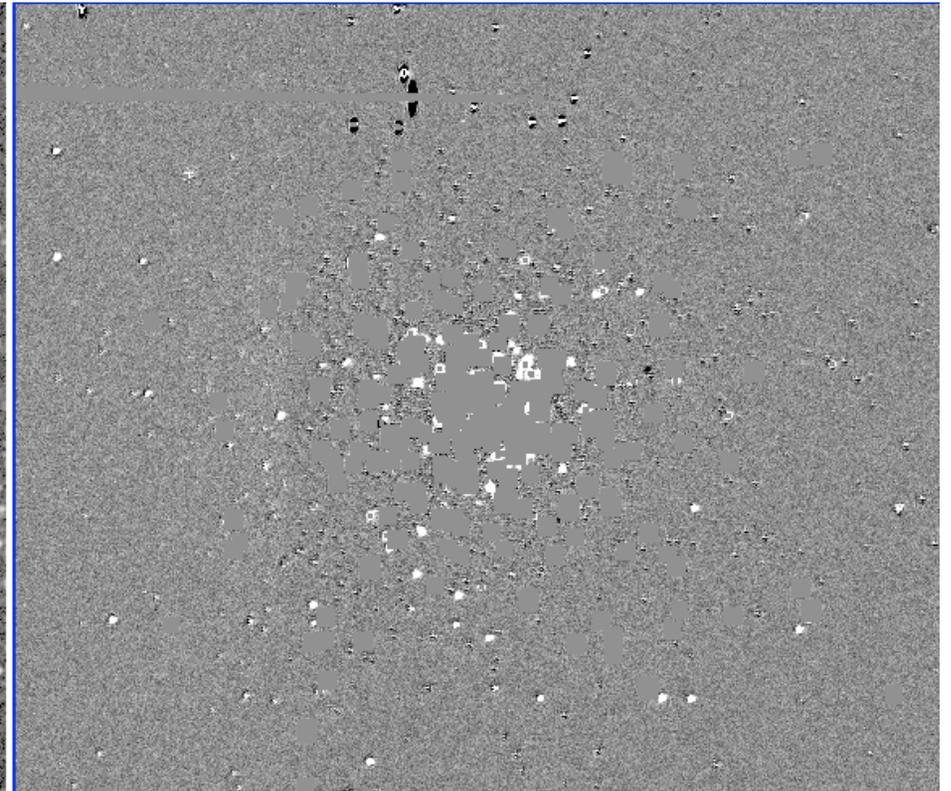
# Example difference image: zoom on M13 globular cluster

---

*science image exposure (~ 9' x 9' zoom)*



*“exposure minus reference”*



Lots of RR-Lyrae variables!

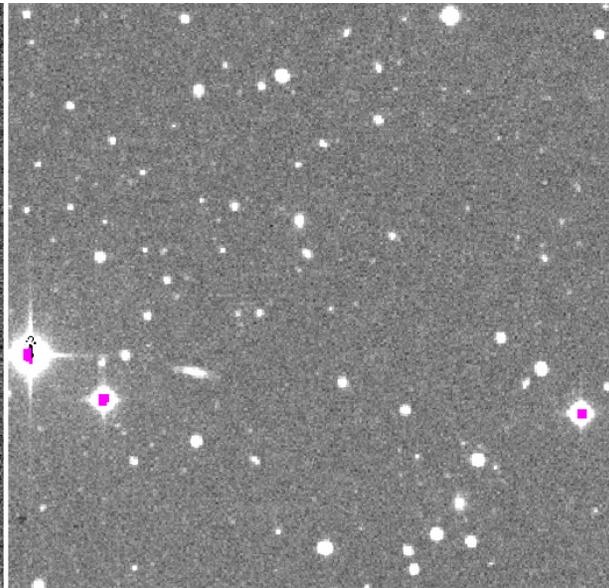
# Example difference image with a streaking asteroid

---

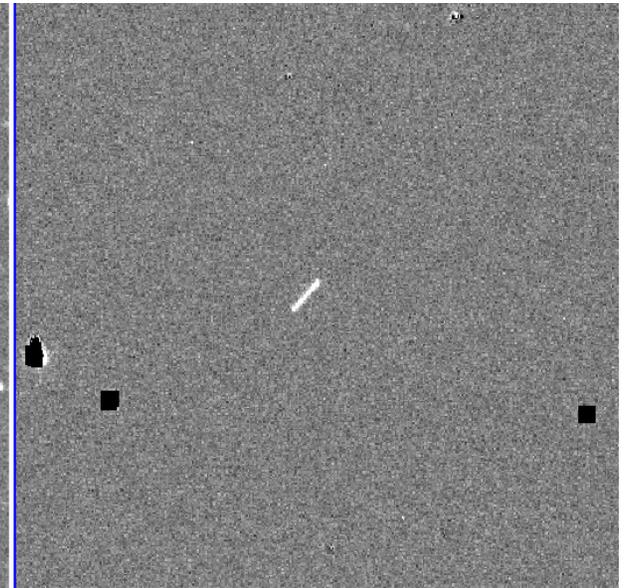
*science image exposure*



*reference image*



*exposure minus reference*



asteroid 2009 HK73

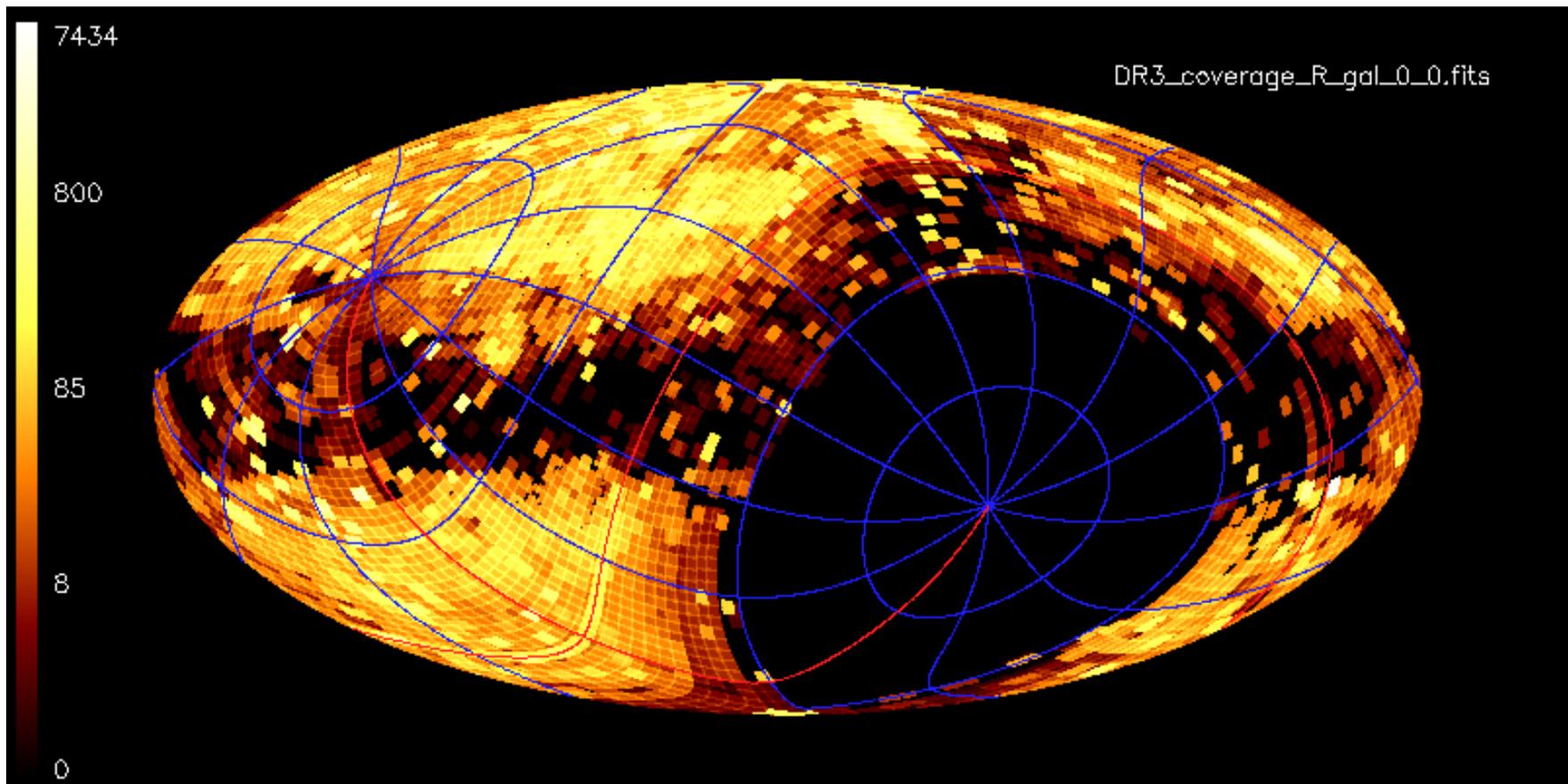
# What products are (will be) available

---

- Science-quality products available now to the public (**Data Release 2**):
  - Epochal images in *R* and *g* bands with QA metadata acquired **Mar 1, 2009 – Dec 31, 2012**
  - Accompanying source-catalog table files for these epochal images
  - Reference images (co-adds) for fields with enough epochs observed during the same period
  - Accompanying source-catalog table files for these reference images
- In early September, 2016, **Data Release 3** will add to the above:
  - ~ 650,000 more epochal images and catalogs from data acquired **Jan 1, 2013 – Jan 28, 2015**
  - ~ 9,600 more reference images (co-adds) with accompanying source catalogs.
- In early December 2016:
  - A lightcurve database, accessible via a user-interface; ~ 600 million LCs (demo later)
  - A source database based on epochal-image extractions, also accessible via a user-interface
  - A source database based on reference-image (co-add) extractions
  - These databases will contain a subset of the epochal data from **Mar 1, 2009 – Jan 28, 2015**

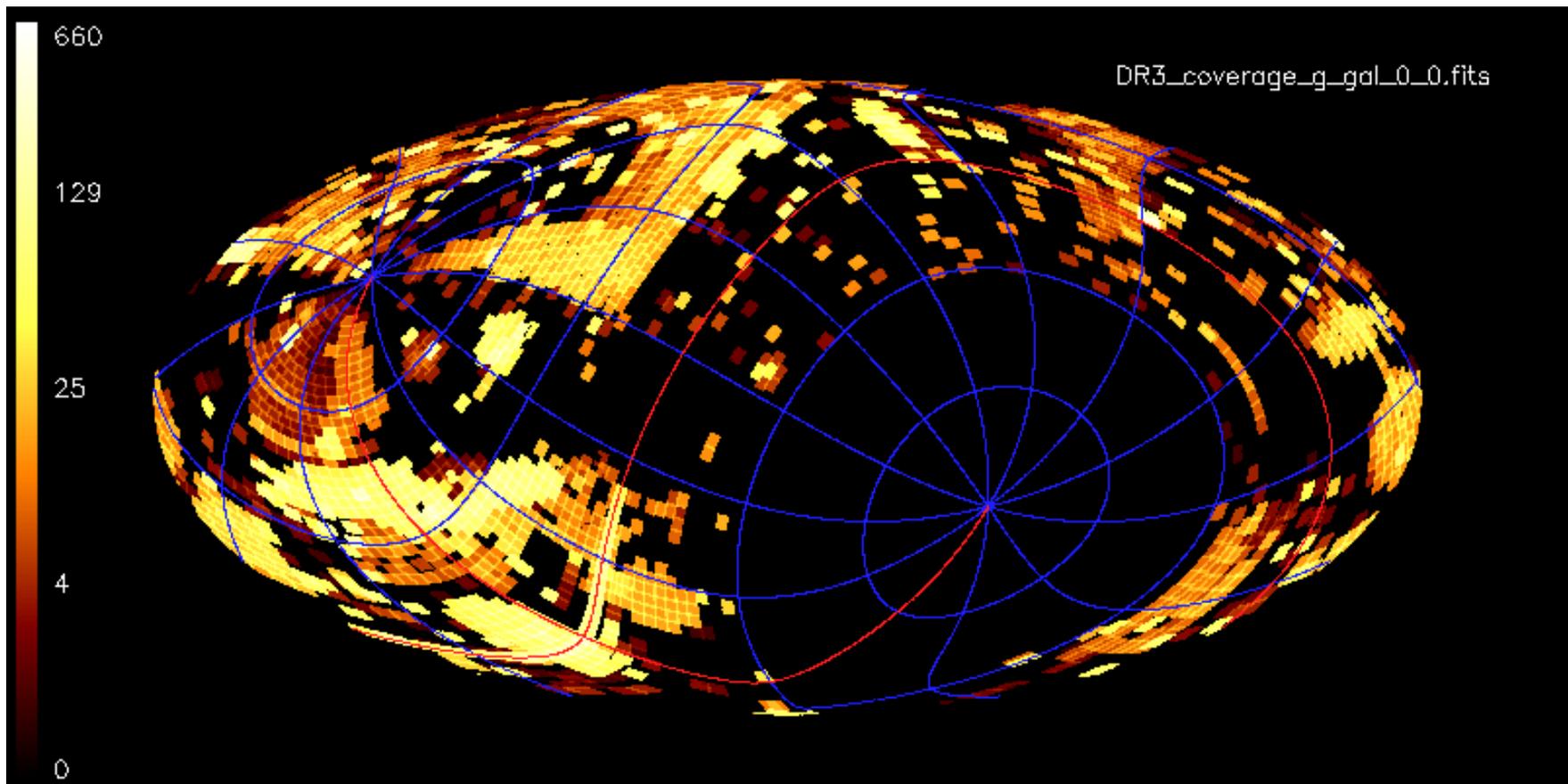
# Sky depth-of-coverage for epochal image products (R-band)

All epochs from the archive that will be publically available following the Fall 2016 release:  
~ **3.47 million R-band images**. Projection is galactic, centered at  $l, b = 0, 0$ .



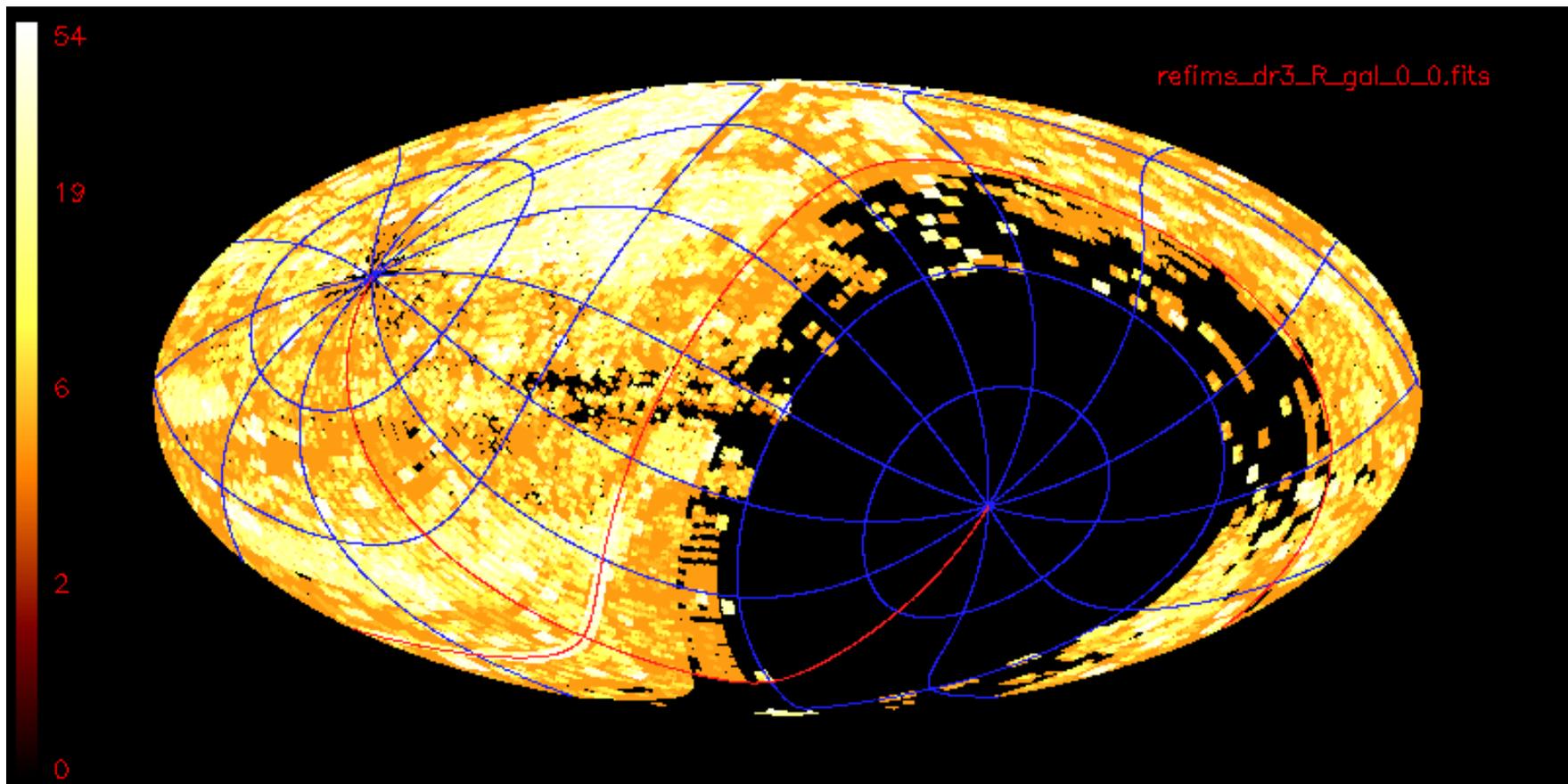
# Sky depth-of-coverage for epochal image products (g-band)

All epochs from the archive that will be publically available following the Fall 2016 release:  
~ **0.72 million g-band images**. Projection is galactic, centered at  $l, b = 0, 0$ .



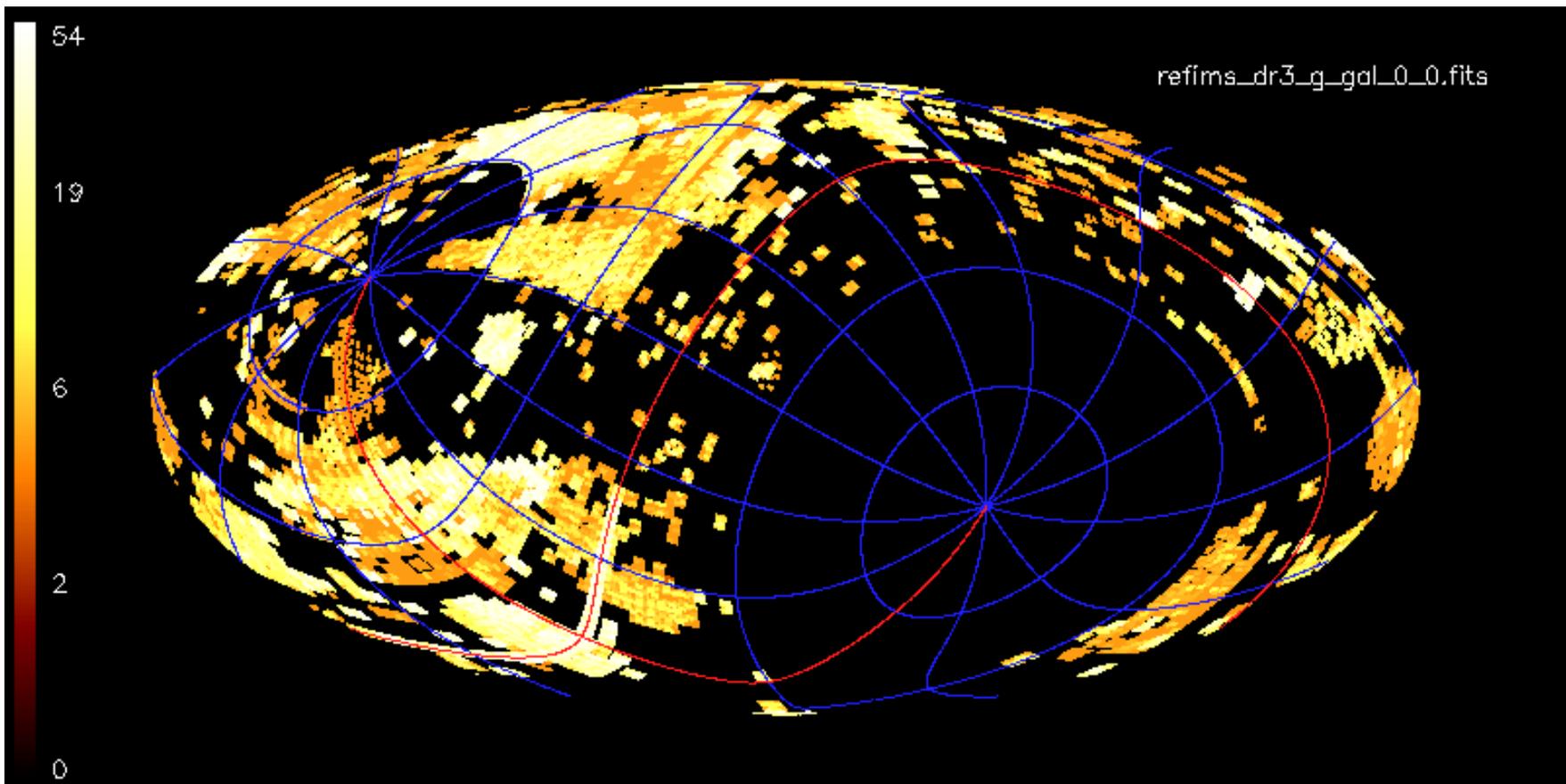
# Sky depth-of-coverage for co-add products (R-band)

All co-add images from the archive that will be publically available following the Fall 2016 release:  
**41,200 R-band co-add images.** Projection is galactic, centered at  $l, b = 0, 0$ .



# Sky depth-of-coverage for co-add products (g-band)

All co-add images from the archive that will be publically available following the Fall 2016 release:  
**15,240 g-band co-add images.** Projection is galactic, centered at  $l, b = 0, 0$ .



# Where to access PTF data products

<http://irsa.ipac.caltech.edu/frontpage/>

The screenshot displays the NASA/IPAC Infrared Science Archive (IRSA) website. At the top, the IRSA logo is on the left, and the text "NASA/IPAC INFRARED SCIENCE ARCHIVE" is on the right. Below the logo, navigation links for "IRSA", "DATA SETS", "SEARCH", "TOOLS", and "HELP" are visible. The main content area includes a search section with a "Search for Source" heading, a text input field for "Name or Coordinates", a "Search" button, a "Radius" input set to "10" with a unit dropdown set to "arcsec", and a "Search Catalog" dropdown set to "WISE" with another "Search" button. To the right, a "Featured Image: M33 with PTF" section shows a colorful image of the M33 galaxy with red, green, and blue spots. Below the image, text reads: "This Palomar Transient Factory image shows M33 in H-alpha (red), R (green), and g (blue). Image credit: A. Waszczak". Below the featured image are links for "Past News" and "Featured Images". At the bottom, a grid of icons represents various instruments and tools: "Catalogs", "Images", "Finder Chart", "VO/API", "Spitzer", "WISE", "Herschel", "Planck", "2MASS", "IRAS", "COSMOS", "AKARI", "BLAST", "BOLOCAM", "IRTS", "ISO", "MSX", "PTF", and "SWAS". The "PTF" icon, which depicts a hard hat, is circled in red.

# PTF Image Service at IRSA

- CCD-based image and catalog files can be searched by position or object name falling therein via a GUI
- Known Solar-System objects also recognized
- Can also search and retrieve data via a command-line application program interface

The screenshot displays the IRSA PTF Image Service interface. At the top, there is a navigation bar with 'IRSA', 'DATA SETS', 'SEARCH', 'TOOLS', and 'HELP'. A search bar contains 'Eq-J2000: Image Pixel:'. Below this is a toolbar with various icons for navigation and image manipulation. The main content area shows search results for 'M83; Type=CENTER; Region=0.0083 deg; Image Size=0.2778 deg; Product Level=l1'. A table lists 20 rows of data with columns for expid, obsdate, crval1, crval2, filter, ccdid, ptffield, seeing, and a11. The first row is highlighted in green. To the right, an 'Image Preview' window shows a grayscale image of a galaxy (M83) with a blue circle around the center. The interface also includes a 'Background Monitor' window and a 'View Options' menu.

expid	obsdate	crval1	crval2	filter	ccdid	ptffield	seeing	a11
19123	2009-05-25 05:59:15.975000	204.5342815	-30.2159701	G	6	100006	3.36	2
19435	2009-05-27 04:31:09.425000	204.6302496	-29.9765794	R	6	100006	2.25	2
19499	2009-05-27 06:06:42.925000	204.6302877	-29.9764442	R	6	100006	2.44	2
51514	2009-12-30 13:29:56.358000	204.4431954	-30.6655180	R	6	100006	4.44	2
51651	2009-12-31 13:35:51.759000	204.4432088	-30.6654161	R	6	100006	3.90	2
51843	2010-01-01 13:20:49.858000	204.4434360	-30.6653985	R	6	100006	2.78	2
52350	2010-01-02 13:35:30.259000	204.4432805	-30.6654429	R	6	100006	3.11	2
52756	2010-01-03 13:14:24.108000	204.4432014	-30.6654362	R	6	100006	2.90	2
53520	2010-01-04 13:09:43.809000	204.3623019	-30.6769362	R	6	100006	null	2
53929	2010-01-05 13:05:51.809000	204.4432884	-30.6654451	R	6	100006	3.40	2
54294	2010-01-06 13:21:41.105000	204.4432339	-30.6654538	R	6	100006	3.78	2
55670	2010-01-11 12:43:17.405000	204.4433324	-30.6653873	R	6	100006	2.89	2
55748	2010-01-11 13:48:52.705000	204.4433366	-30.6653770	R	6	100006	3.42	2
56298	2010-01-15 13:55:58.154000	204.4432781	-30.6654010	R	6	100006	4.44	2
56723	2010-01-17 13:31:45.155000	204.4433075	-30.6656140	R	6	100006	2.90	2
57098	2010-01-25 12:03:55.055000	204.4433013	-30.6654146	R	6	100006	2.59	2

## More information...

- More details on PTF image processing and data archiving at IPAC:  
Laher et al., 2014, PASP, 126, 674  
<http://iopscience.iop.org/article/10.1086/677351/pdf>
- Documentation on how to access and use the PTF public data products:  
[http://www.ptf.caltech.edu/page/data\\_access](http://www.ptf.caltech.edu/page/data_access)

INTERMEDIATE PALOMAR TRANSIENT FACTORY

HOME NEWS IMAGES VIDEOS ABOUT DATA ACCESS PUBLICATIONS MEETINGS

PEOPLE

Data Access  
First Data Release  
Second Data Release  
**Documentation**  
Data Processing  
Astrometric Calibration  
Photometric Calibration  
Data Quality Flags  
File Names and Types  
File Headers  
Papers  
Getting Survey Data  
Asteroids  
Public Light Curves and Spectra  
Code

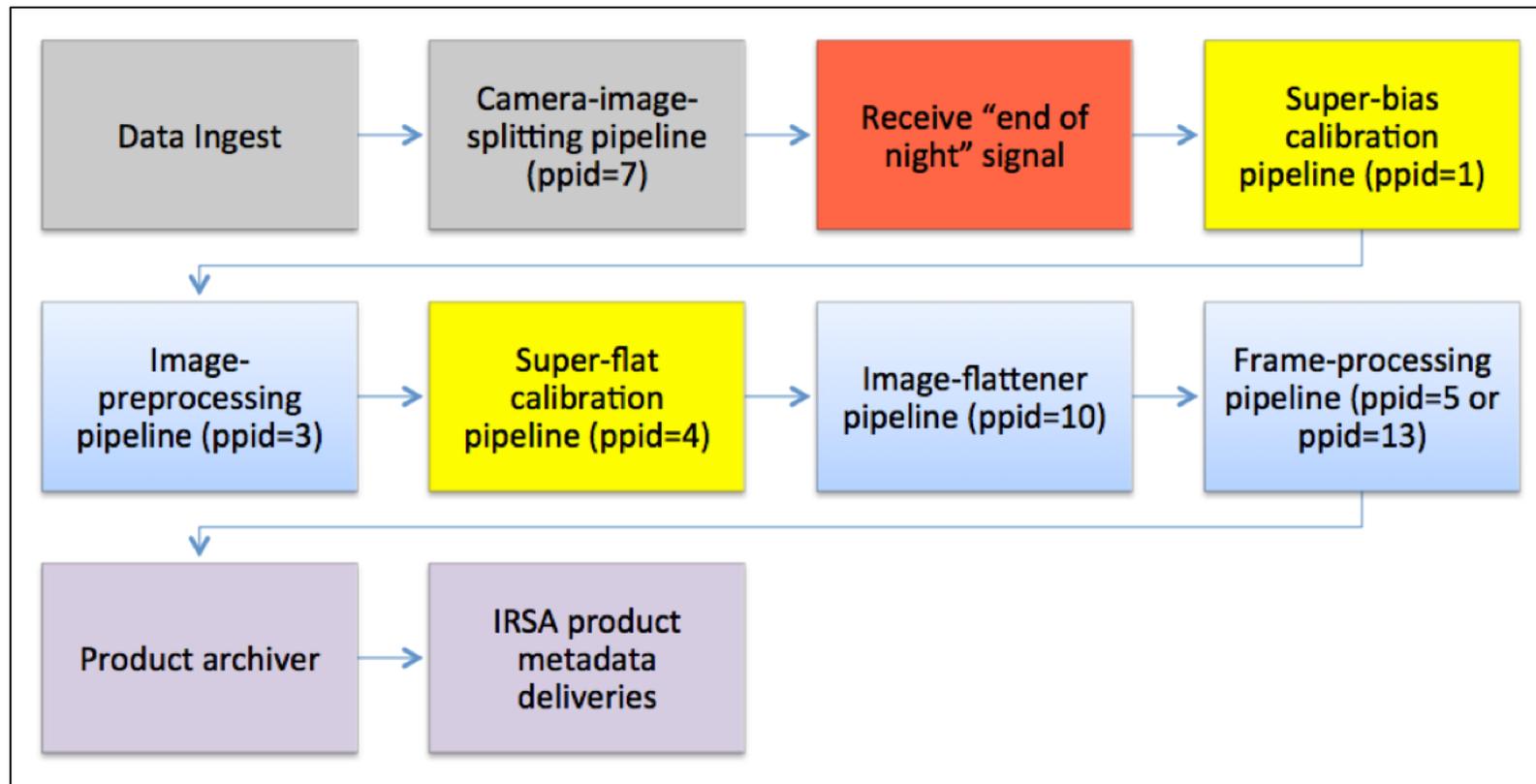
The PTF and iPTF data are reduced and archived by the [Infrared Processing and Analysis Center](#). The publicly accessible survey data products include astrometrically and photometrically calibrated single frame exposures for selected regions of the sky, as well as source catalog files for those same regions. The details of the first data release, documentation of the products, and access methods, are detailed at left.

Science subgroups within the collaboration have also made accessible certain additional data, which are also linked at left.

PTF and iPTF collaboration members have additional data access, and can contact [Jason Surace](#) for further details.

Back up slides

# PTF daily processing steps



# Photometric calibration of PTF using SDSS

- Described in Ofek et al., 2012, PASP, 124, 62
- Uses frames that overlap with SDSS footprint to fit a global linear model for nightly data
- Enables calibration of all CCD images observed during a night
- Absolute precision (with respect to SDSS) is  $\sim 0.02 - 0.04$  mag.
- Primary outputs: a global ZP value per image and a spatially-binned ZP residuals map (ZPVM)
- **These ZP estimates are only applicable to *mag\_auto* instrumental magnitudes**
- $R^{inst}$  and  $g^{inst}$  below are *mag\_auto* (SExtractor) instrumental magnitudes

details). For observations taken using the  $R$ -band<sup>18</sup> filter, we fit the following model:

$$\begin{aligned} r_{SDSS} - R_{PTF}^{inst} = & ZP_R + \alpha_{c,R}(r_{SDSS} - i_{SDSS}) + \alpha_{a,R}AM \\ & + \alpha_{ac,R}AM(r_{SDSS} - i_{SDSS}) + \alpha_{t,R}(t - t_m) \\ & + \alpha_{t2,R}(t - t_m)^2 - 2.5 \log_{10}(\delta t), \end{aligned} \quad (1)$$

while for  $g$ -band observations we fit

$$\begin{aligned} g_{SDSS} - g_{PTF}^{inst} = & ZP_g + \alpha_{c,g}(g_{SDSS} - r_{SDSS}) + \alpha_{a,g}AM \\ & + \alpha_{ac,g}AM(g_{SDSS} - r_{SDSS}) + \alpha_{t,g}(t - t_m) \\ & + \alpha_{t2,g}(t - t_m)^2 - 2.5 \log_{10}(\delta t). \end{aligned} \quad (2)$$

# Photometric performance

