Hunting Giant Radio Galaxies

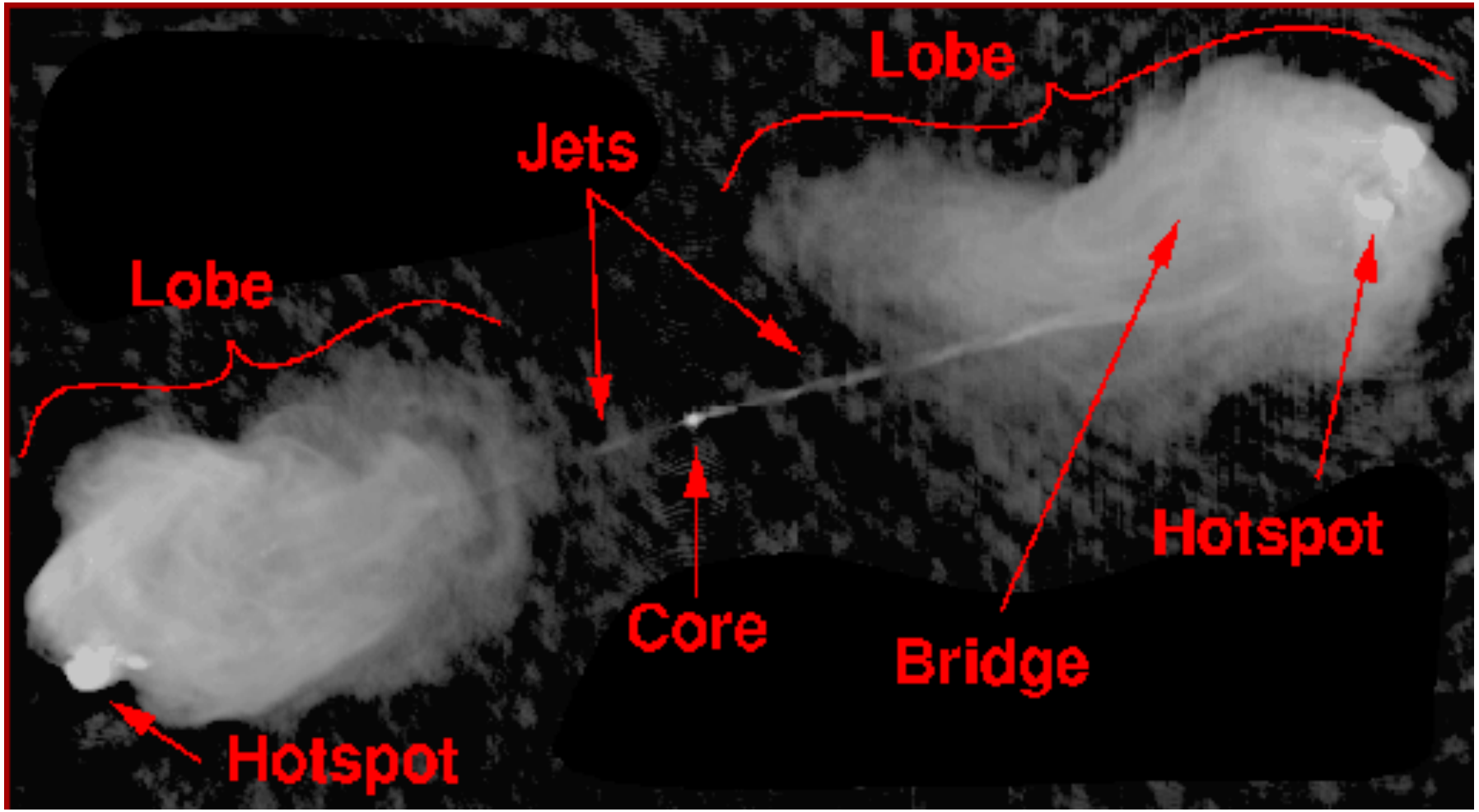
From Visual Inspection to Machine Learning

Hongming Tang

Jodrell Bank Centre for Astrophysics, Uni. of Manchester

Email: hongming.tang@manchester.ac.uk

Radio Galaxy Morphology



Credit: C. L. Carilli

Giant Radio Galaxy

Giant Radio Galaxy: If a radio galaxy has its projected linear size > 700 kpc, the object would be defined as a GRG (Dabhade+20a,b).

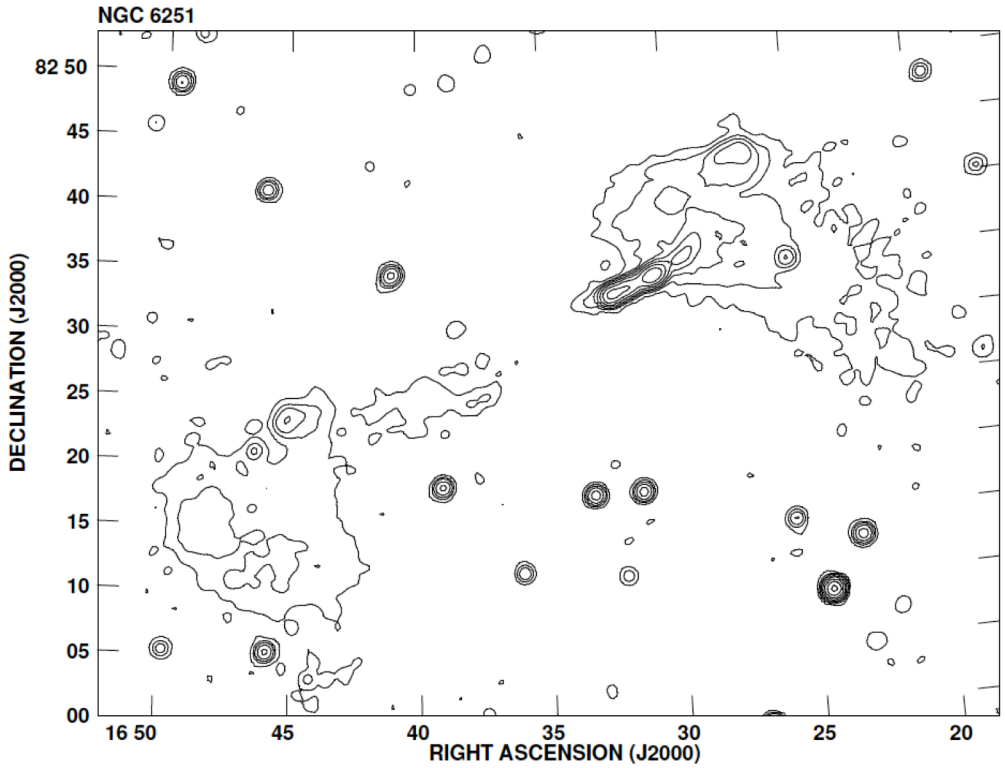
Key factors to decide source linear size:

1. Largest Angular Size (LAS): The end to end angular distance between the margins of AGN radio lobes.
2. Redshift (z): Determines the RG distance from us

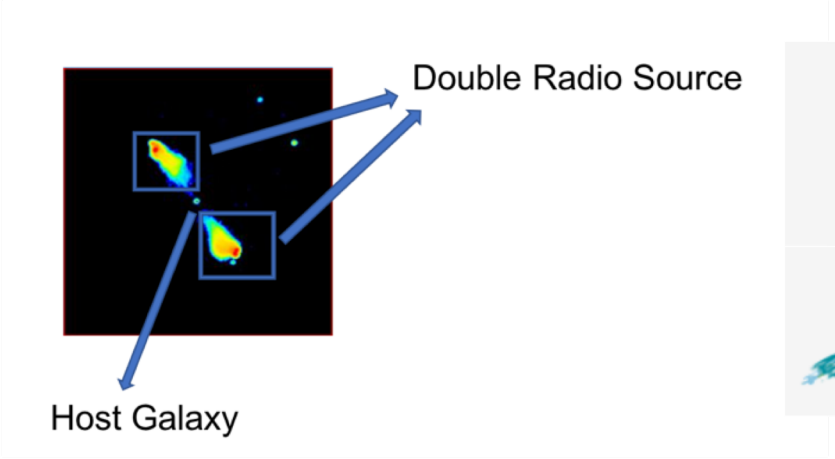
Why we hunt GRG?

- Probe Warm-hot phase of IGM (WHIM; Safouris et al. +09, Peng et al. +15)
- Believed to be very old RGs (Kaiser and Alexander +97, Dabhade et al. +17)

- Allow detailed imaging, enable further investigation upon physical process within them (Willis+74)

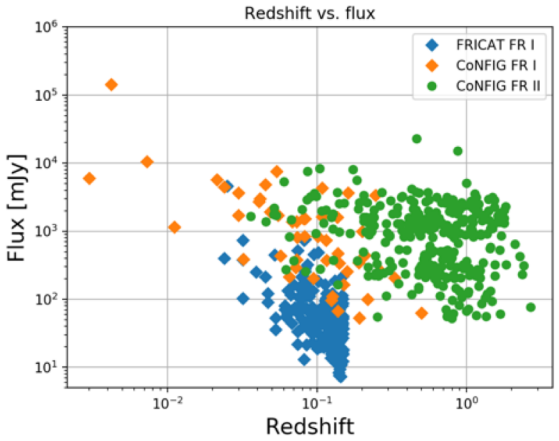


Finding GRG candidate

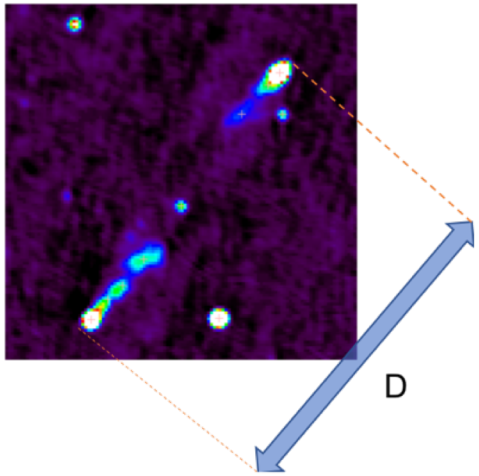


1. Source localization

2. Host Identification



3. Redshift Availability



4. Projected linear size (D) measurement

When D > 700 kpc,
We call it a GRG candidate

Notes of GRG hunting

1. Hunting GRG is usually a side project of large scale radio source/component catalogues.

2. The LAS of a candidate was usually validated by multi-frequency imaging (some have their LAS measured from survey image data).

3. Host galaxy redshifts were usually measured by taking galaxy spectroscopy. When spectroscopy is unavailable (proposed observation or archival data), photometric redshift might be used with caution.

Methods of hunting GRGs:

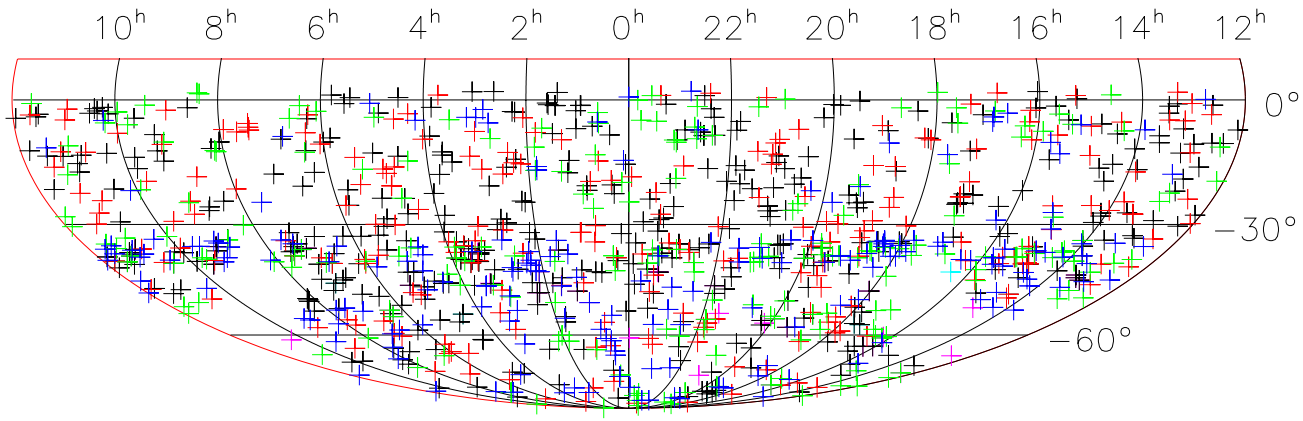
1. Expert Visual Inspection (e.g., Tang et al.+18 in prep.)

2. Citizen Scientist + Expert Visual Inspection (e.g., Banfield et al.+17; Tang et al. +20)

3. ML + Expert Visual Inspection (e.g., Proctor +16; Dabhade et al. +20b)
4. ML only ?

1. Southern Large DRAGN Survey

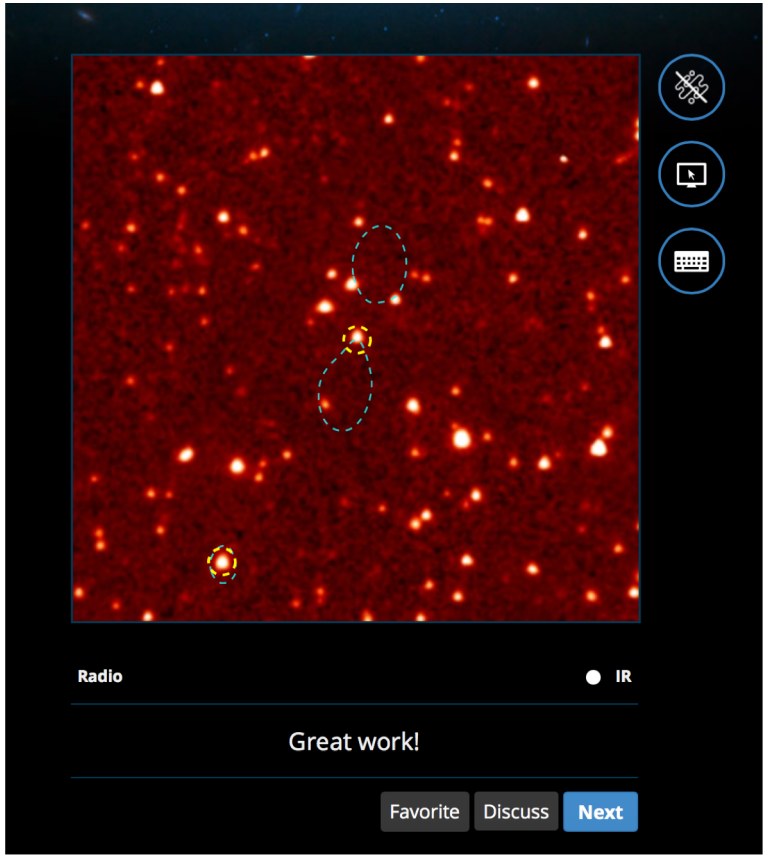
1. LAS > 4 arcmin
2. Covering over 90% of the Southern Hemisphere (|b| > $4^{\circ}, \delta < 2^{\circ}$)
3. Candidate selection via visual Inspection upon SUMSS/NVSS
4. Cross validation done upon FIRST if available
5. Host identification done with NED
6. 952 radio galaxies are identified, with over 75 newly identified GRGs.



Tang et al. in prep +18

2. Radio Galaxy Zoo

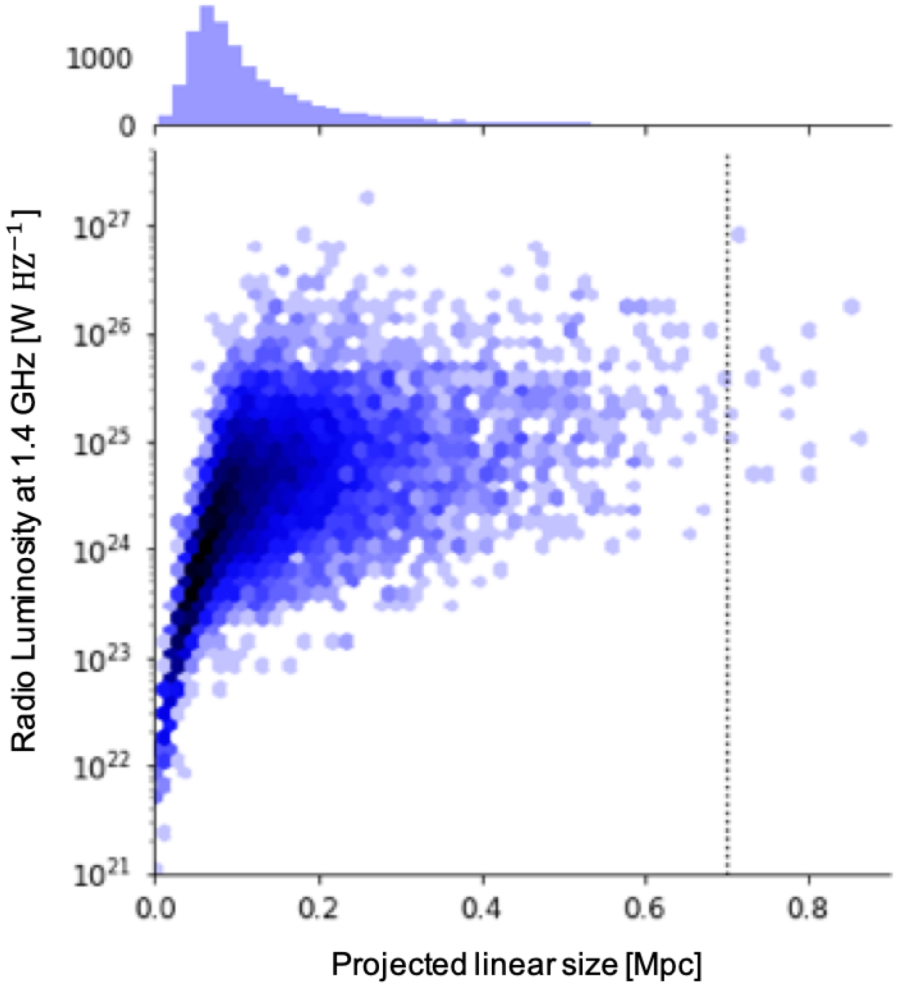
Classification



Forum

Banfield et al.+16

RGZ Data release 1



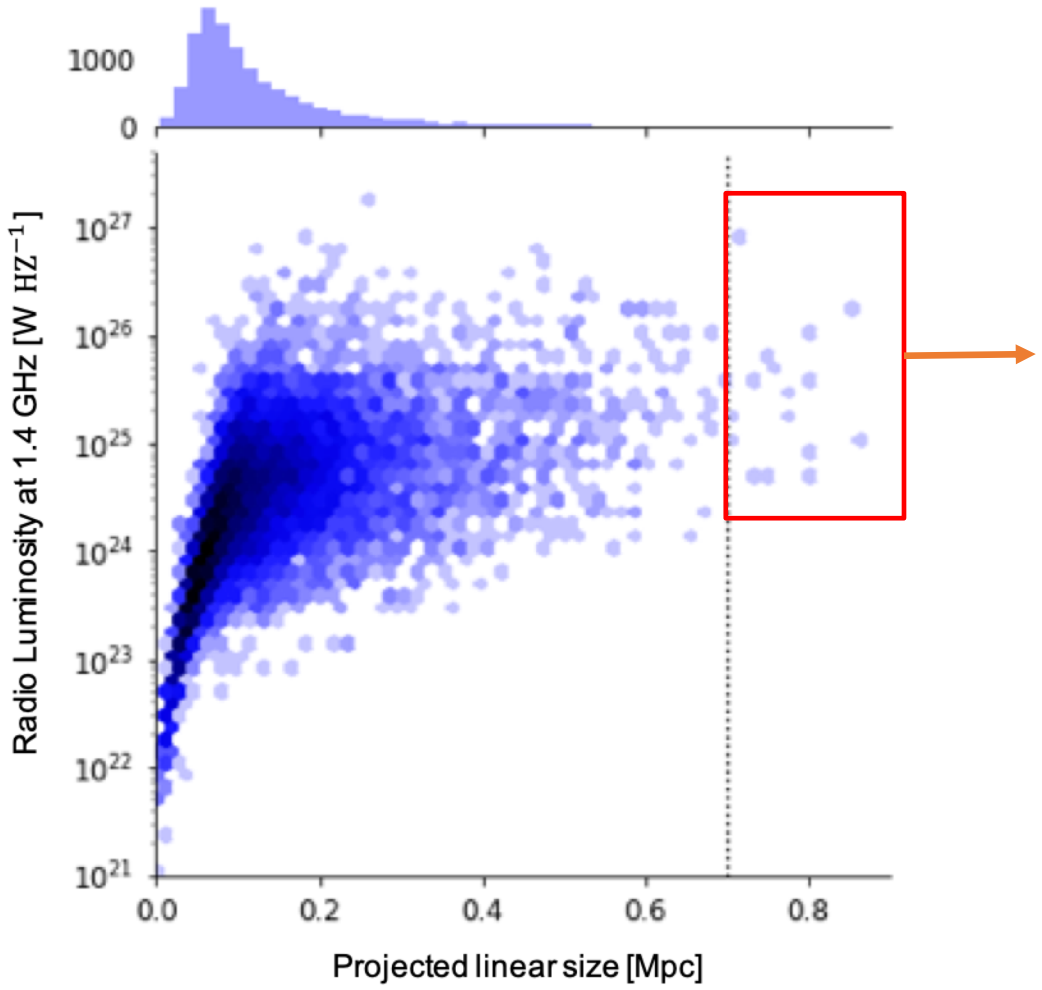
First 2.75 years data since RGZ launched:

- 11,214 volunteers (latest: 11,964)
- 1,692,415 classifications (latest: 2.2 million)
- Each source is classified by 20 users
- If only one component in the image, then only require 5 user classifications.
- Consensus level > 0.65

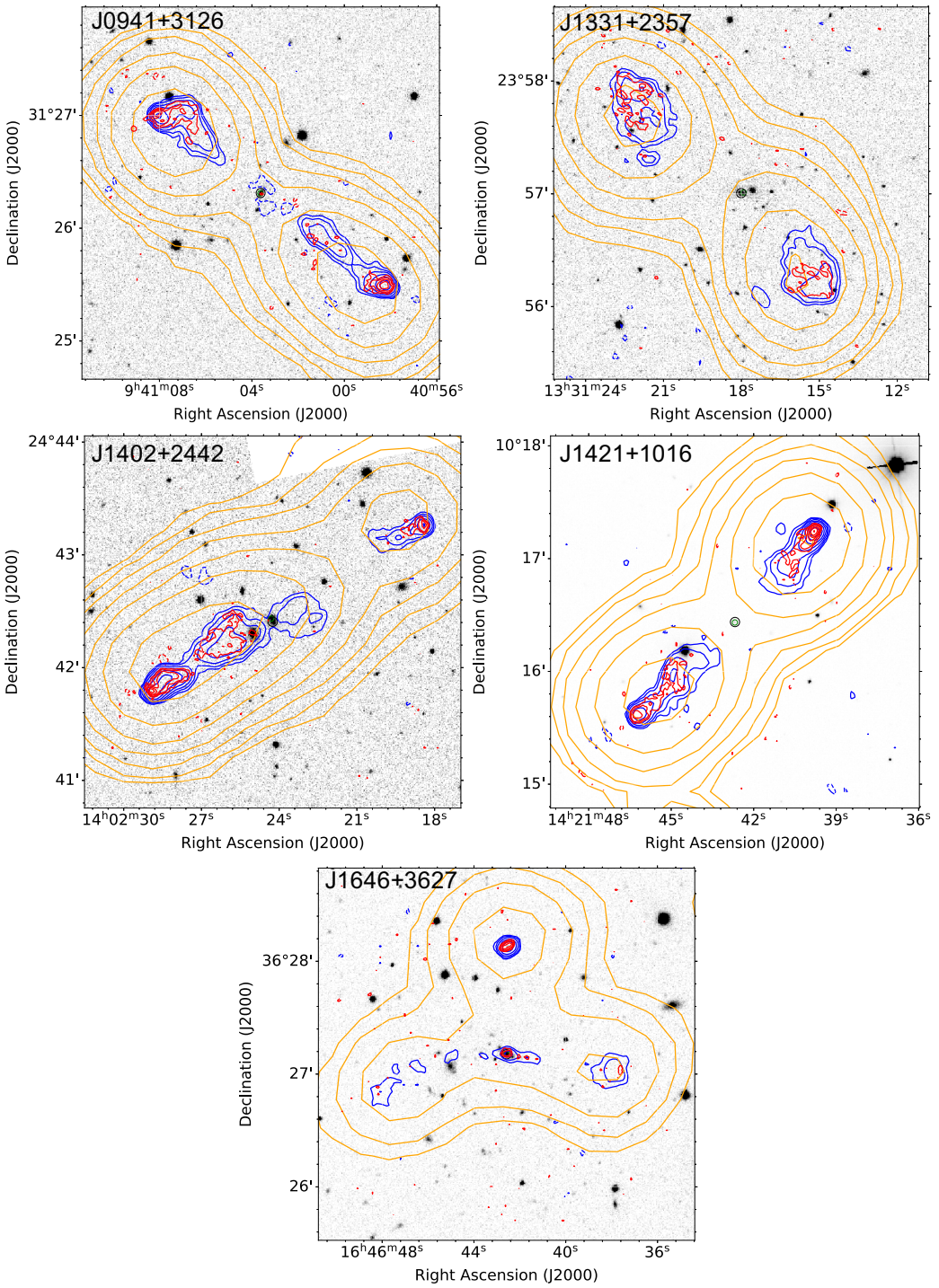
71,981 sources

Wong et al. in prep.

GRG found from DR1



Tang et al. +20



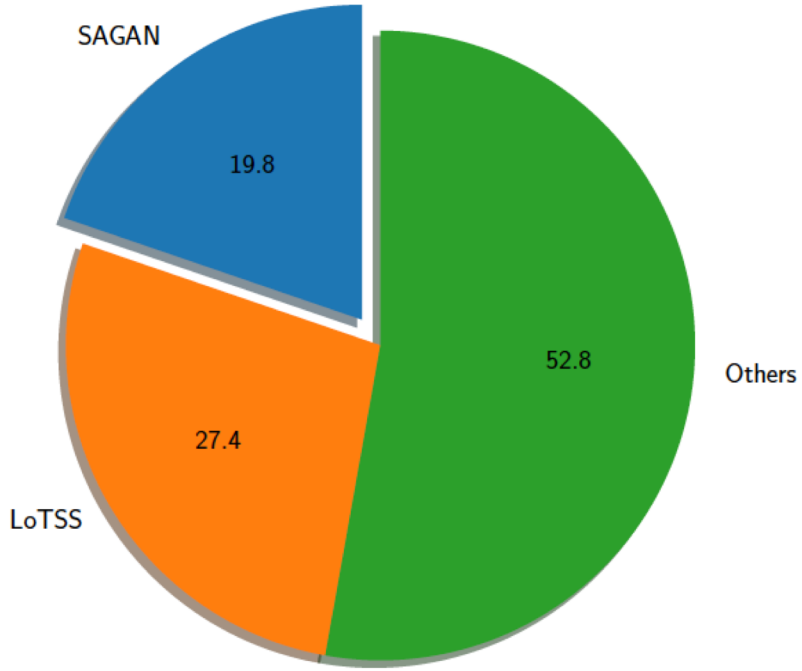
3.1 Proctor 2016: A decision-tree approach

- GRS training sample: 48 GRS with only 2 radio components from Lara et al. 2001
- Non-GRS training sample: 51,147 source pairs extracted from NVSS catalogue north of 60°
- Each source pairs had a max separation of 20’ and a minimum component area of 1.87 square arcmin at 1.4 mJy.
- Features: component semi-major axis, semi-minor axis, peak flux
- Result: The author visual inspected the samples with high probability of being GRS, and provide a candidate list of 1,616 GRS (LAS > 4’)

3.2 SAGAN I: a Proctor+2016 follow up

- Confirm candidates by looking at their NVSS/FIRST/TGSS/VLASS survey images if available
- Host galaxy identification was done using the data of SDSS/Pan-STARRS, WISE (By looking at the overlaid optical/IR and radio maps)

- Result: Among the 1,616 samples, 165 of them are known GRGs, and 151 of them are identified as New GRGs



Dabhade et al. +20b; 820 GRGs in total

Lessons learn from Proctor+16/Dabhade+20b

- 1. Training samples need to have their radio lobes / host galaxy confirmed.
- 2. The object class label (GRG, NOM) should be identified with confidence.
- 3. RG host galaxy redshift should be provided.

Q: If we are training another GRG/non-GRG classifier, what should else be considered ?

4. Tang et al. +20 & in prep.

- 1. Sample image size should be identical (e.g., 3 x 3 arcmin)
- 2. Sample LAS should be measured on the same radio survey (keep data consistency)
- 3. Image based classifier might be necessary, as RG have complex morphology, not limited to objects with double radio components
- 4. The classifier perhaps would need to import both image input (estimate LAS) and parameter input (redshift).

Summary

- 1. Traditional way of finding GRG has high consensus level, but required deep imaging and spectroscopy observation

- 2. Citizen Science could help finding objects of rare species, while these projects themselves cannot tackle the barrier of Big Data. The ML dataset built from these projects might partly solve the issues.

- 3. The foundation of such ML datasets should both consider the nature of given scientific problems, and the inner rules of ML.