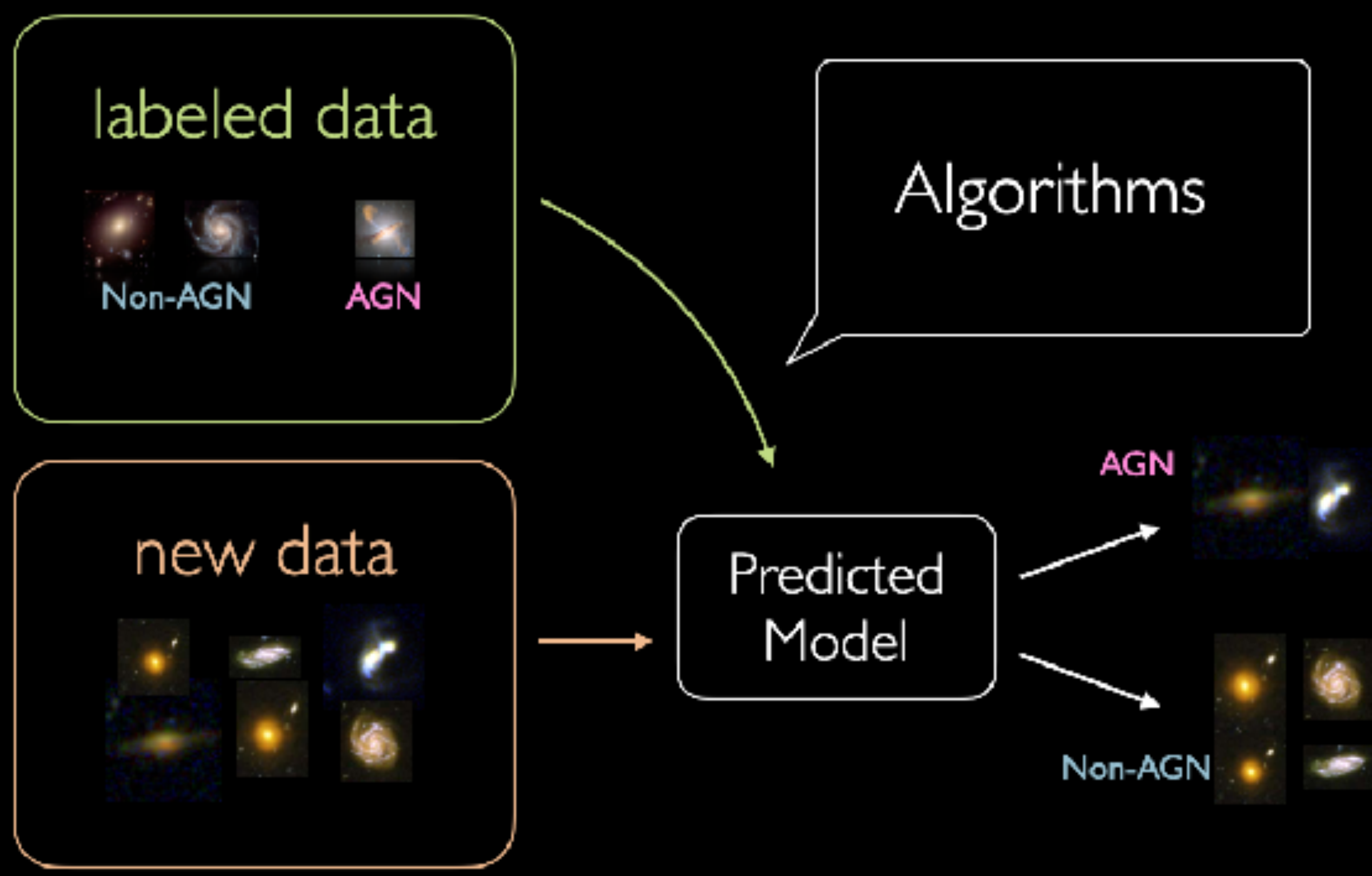
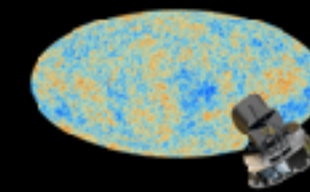
# Unveiling Galaxies and AGNs by Machine Learning



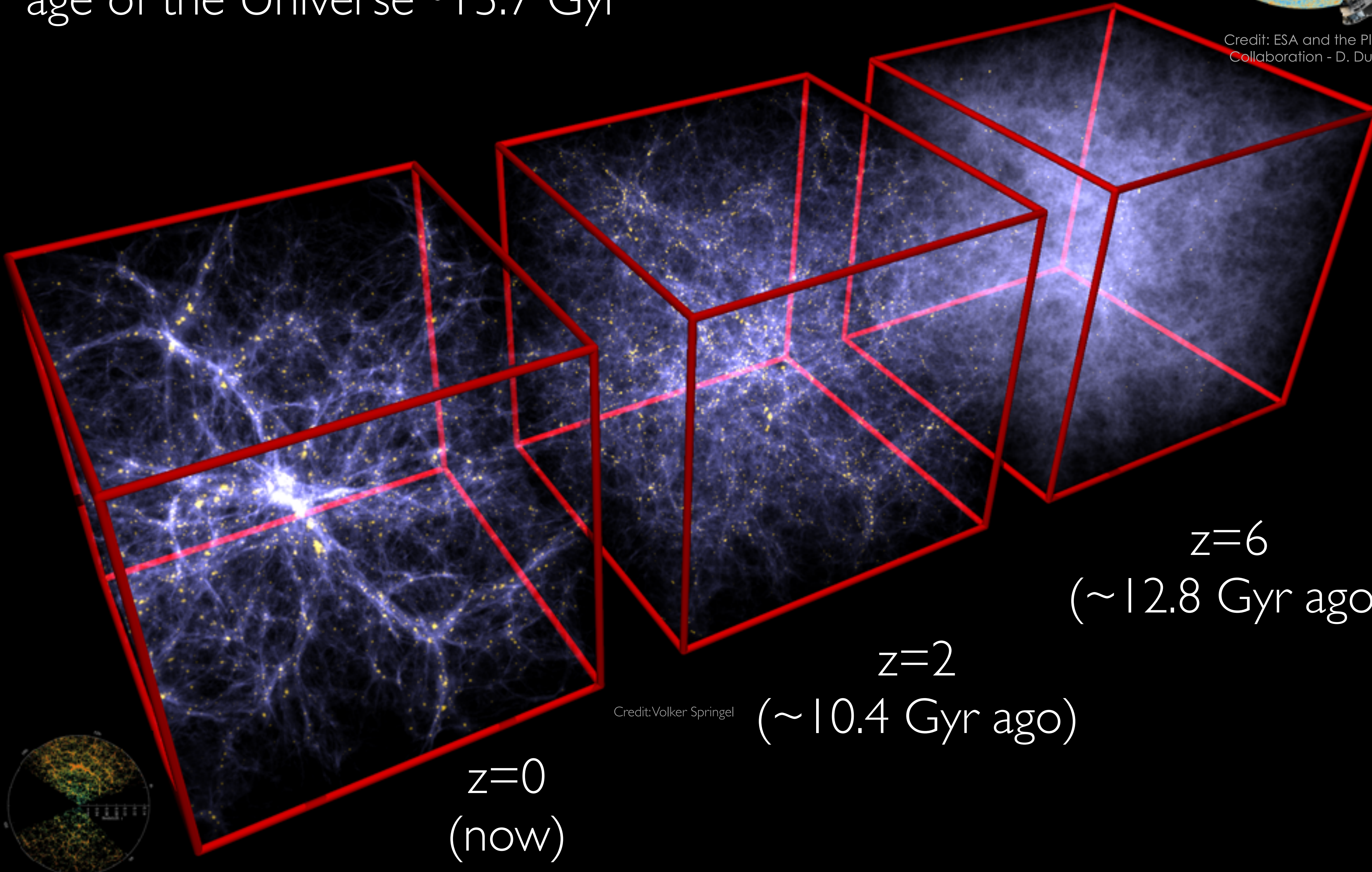

Yu-Yen Chang

IAU-IAA
11 Jul 2023

# age of the Universe~13.7 Gyr

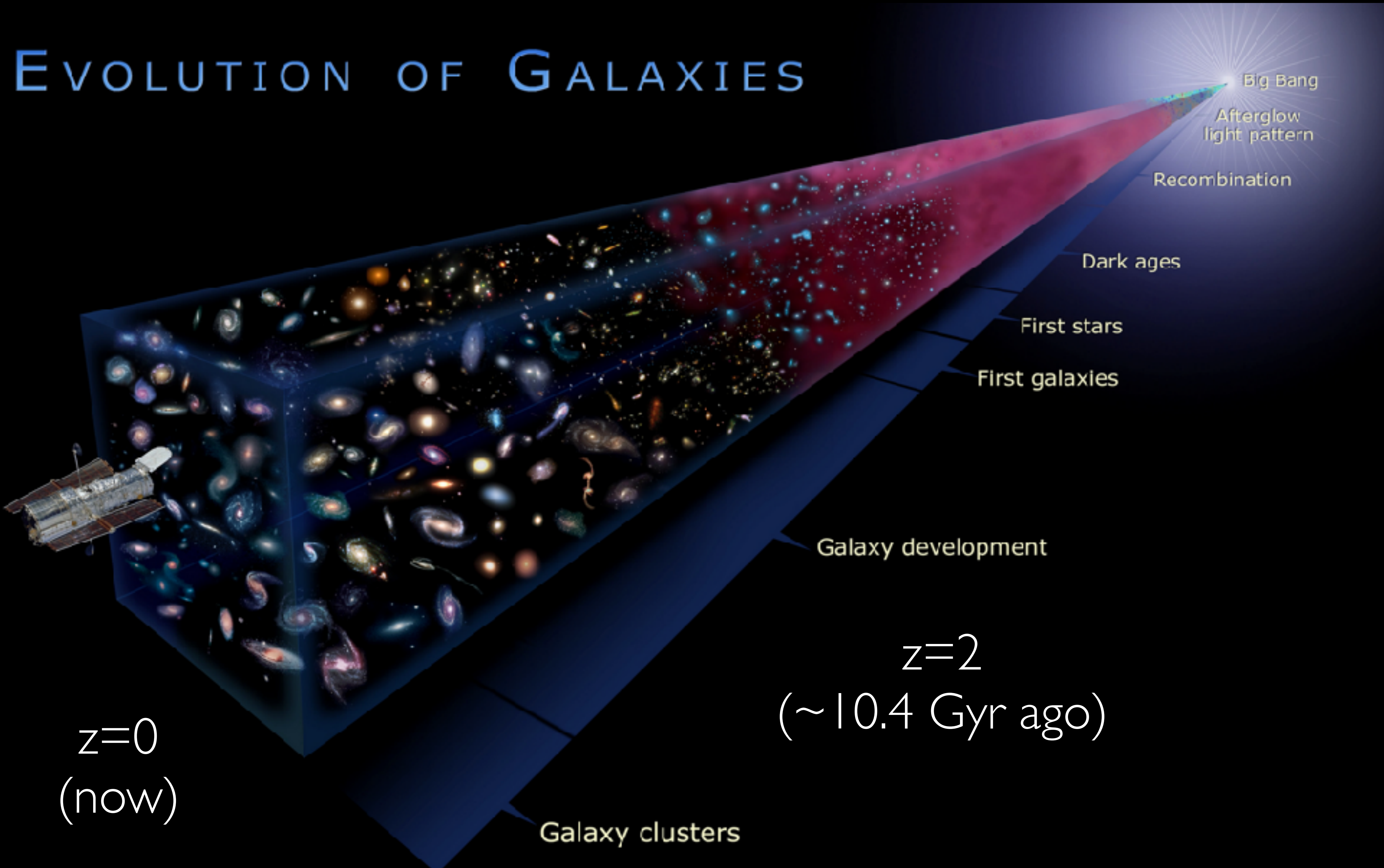


Credit: ESA and the Planck Collaboration - D. Ducros




Credit: Volker Springel

Credit: SDSS




z=0
(now)

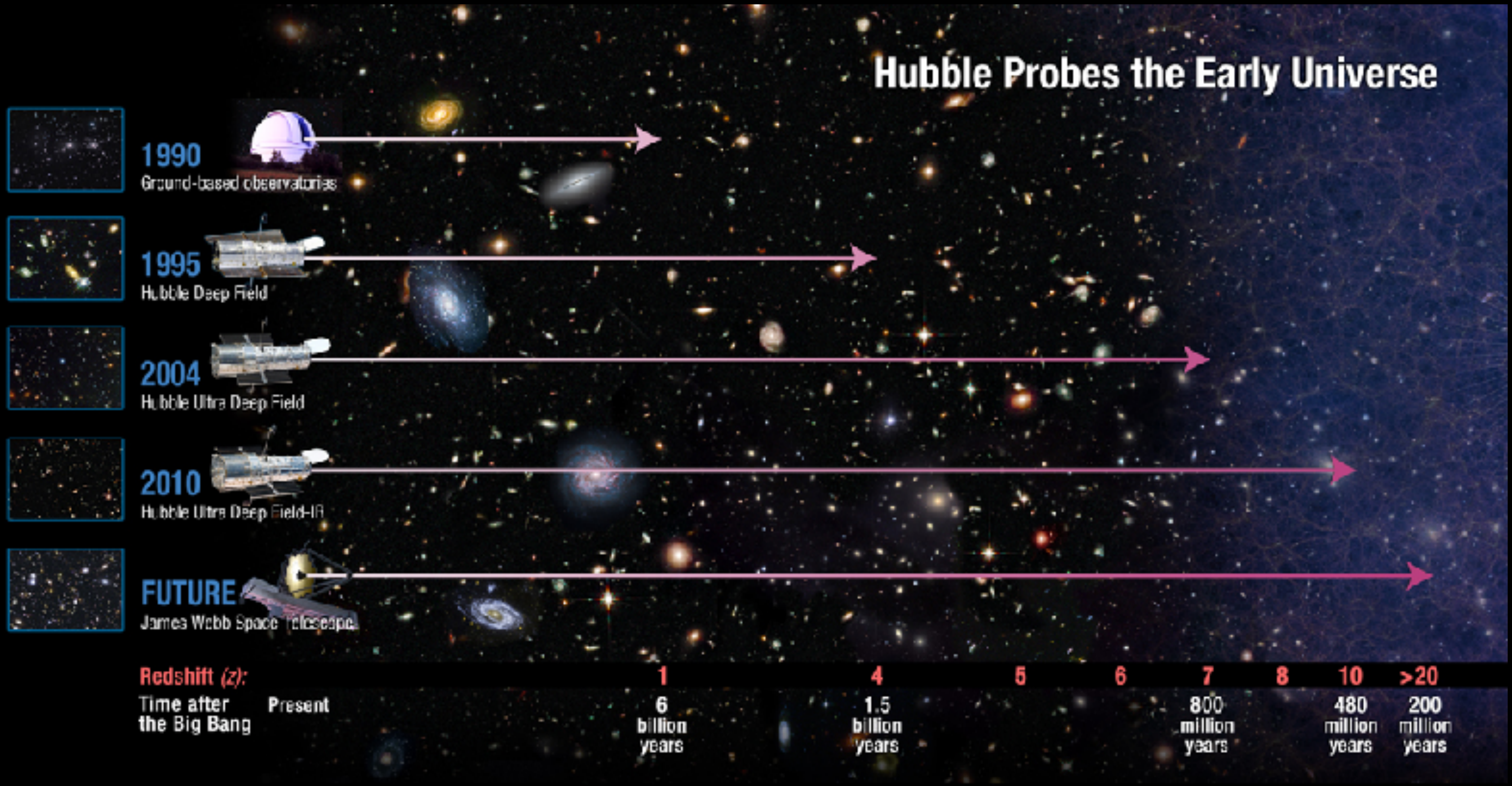
z=2
(~10.4 Gyr ago)

Credit: STScI

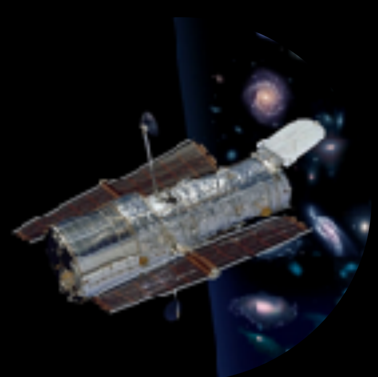
age of the Universe~13.7 Gyr

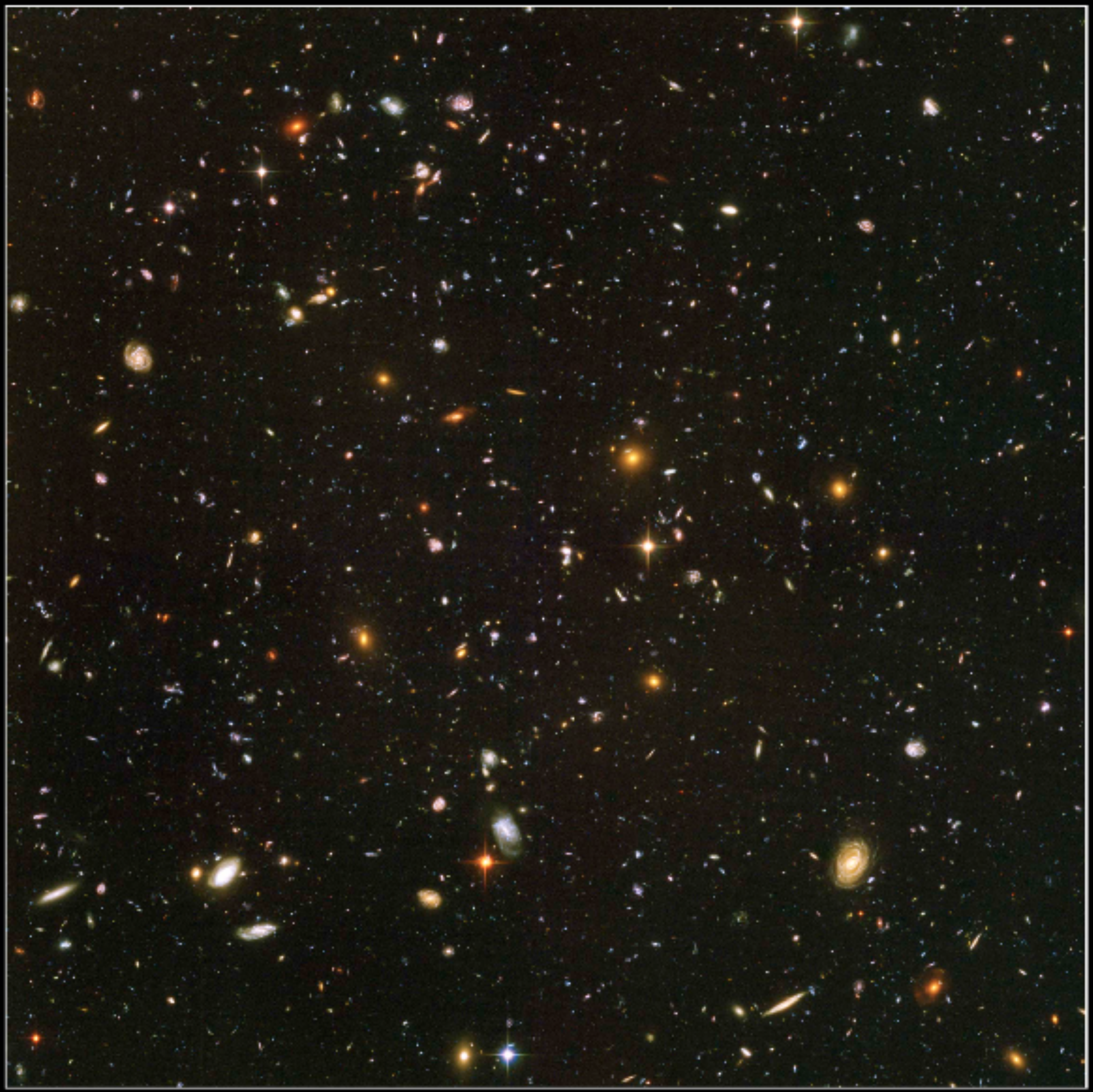
# Galaxy Evolution with Large Surveys

- Evolution of galaxies is not observable
- Evolution of population is observable



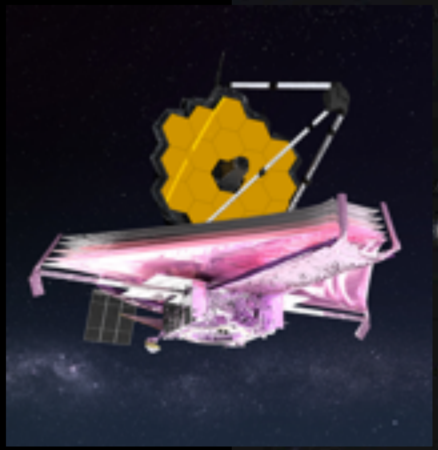

Credit: NASA





**Hubble Ultra Deep Field**
**Hubble Space Telescope** • Advanced Camera for Surveys

NASA, ESA, S. Beckwith (STScI) and the HUDF Team    STScI-PRC04-07a



Credit: NASA

# Galaxy Evolution

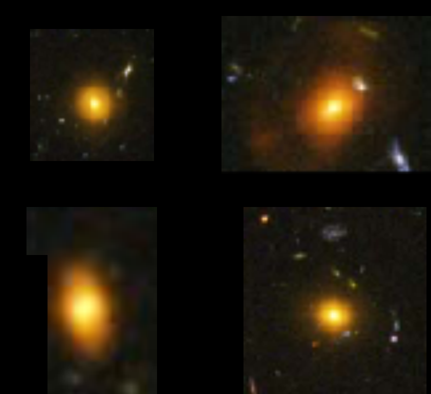
Star-forming galaxies (young.blue.structures active star formation)

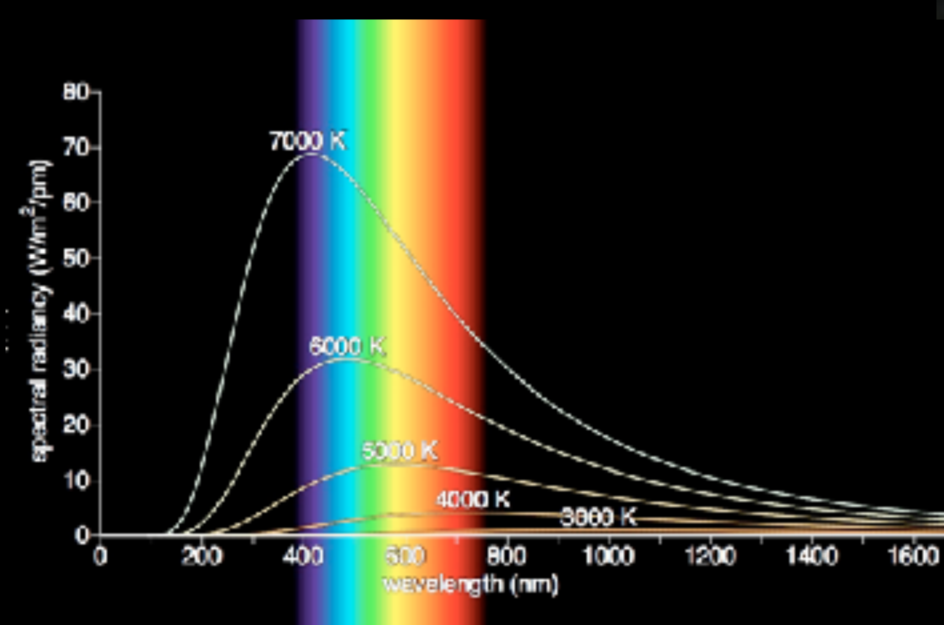
Credit: ESA & NASA

?

Quiescent galaxies (old. red. smooth little star formation)

Credit: NASA,ESA, STScI/AURA, J. Blakeslee






Credit: https://www.aether-force.com

# Galaxy Evolution

Star-forming galaxies (young.blue.structures active star formation)

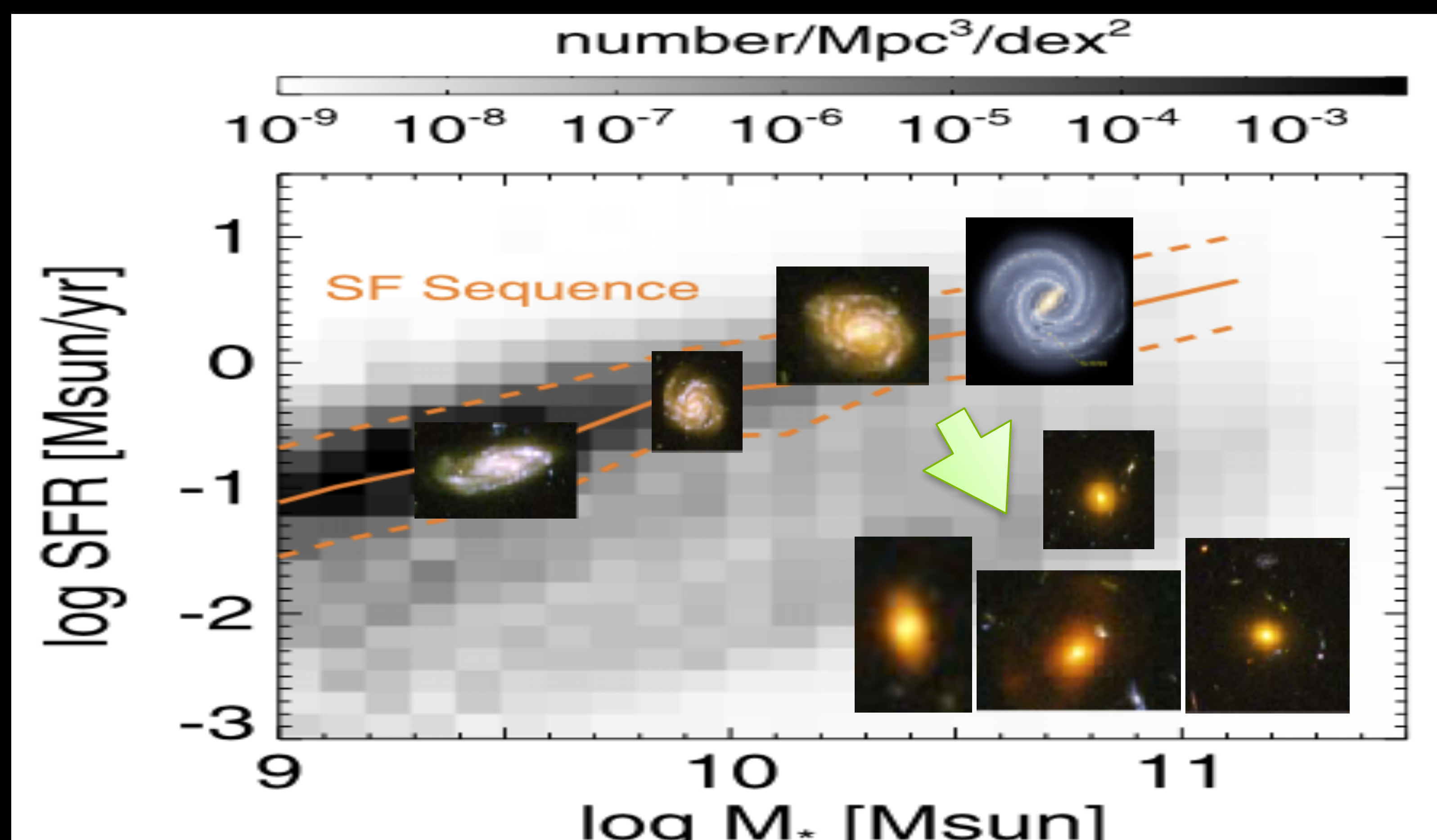
Credit: ESA & NASA

?

Quiescent galaxies (old. red. smooth little star formation)

Credit: NASA,ESA, STScI/AURA, J. Blakeslee

SFR=star formation rate




Chang+15 x858,365 galaxies Stellar Mass

# Size Evolution of Galaxies

Star-forming galaxies


Quiescent galaxies

vdWel,…, Chang+14a

also see Shen+03; Firm+07; vdW+07; van Dokkum+08; Buitrago+08; Williams+10; Newman+12

# Shape Evolution of Galaxies

## Two-population Model of Quiescent Galaxies

flatter

oblate fraction X1/3

rounder

triaxial fraction X2

z=2(~10.4 Gyr ago)

z=0(now)

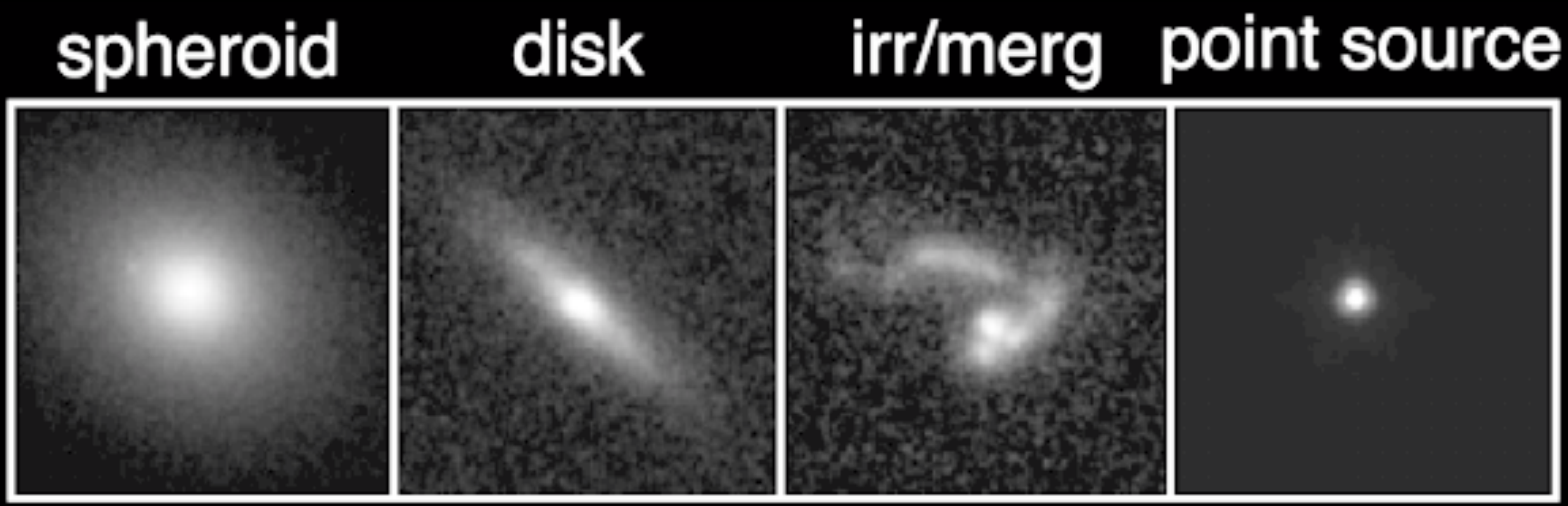
The majority of massive quiescent galaxies at z>1 are flat/disk-like.

Chang+13b

also see vdW+11
Chang+13a

# Galaxy Mergers: Visual Classification

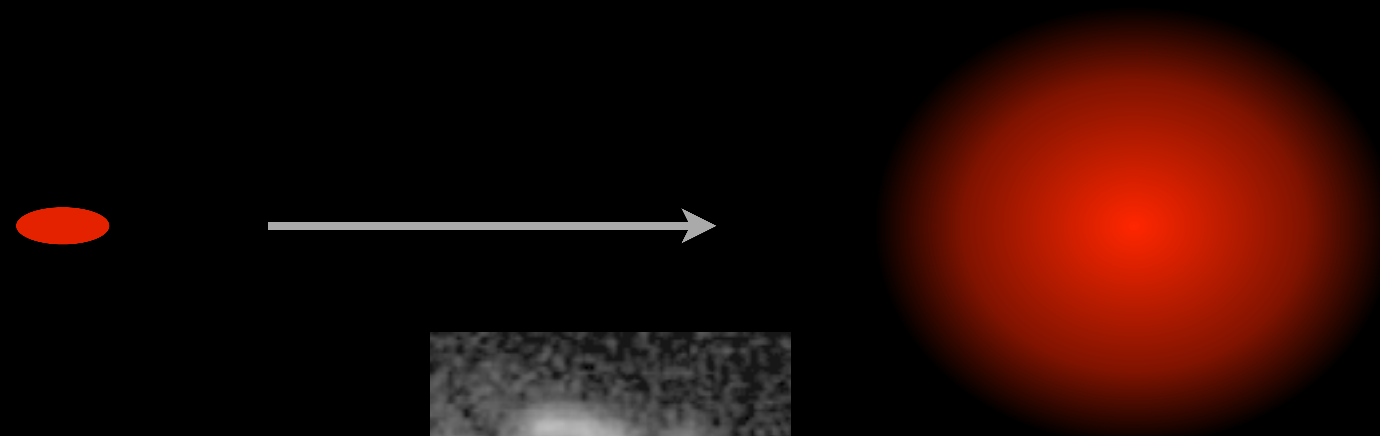

Credit: NASA, ESA, the Hubble Heritage Team (STScI/AURA)-ESA/Hubble Collaboration and A. Evans (University of Virginia, Charlottesville/NRAO/Stony Brook University), K. Noll (STScI), and J. Westphal (Caltech)




Disturbed feature fractions depend on the SFRs (IR luminosity) for AGN, dusty, star-forming galaxy sample.

Chang+17b, 18

- <u>How do Galaxies Form and Evolve over Cosmic Time?</u>
  - Quiescent galaxies are formed disk-like and become larger and rounder through galaxy mergers.



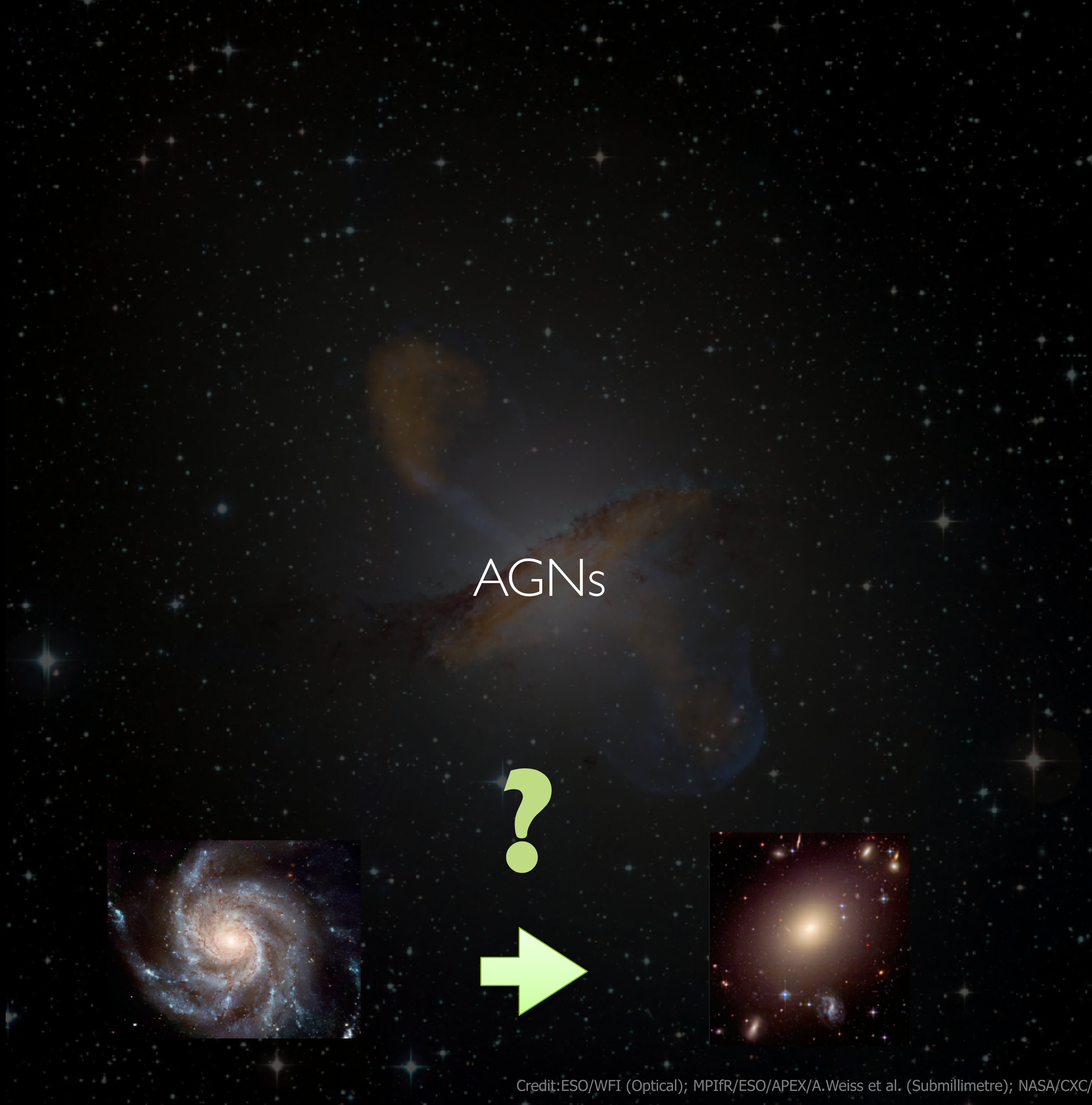
z~2
(~10.4 Gyr ago)

z~0
(~now)

Chang+13a,b

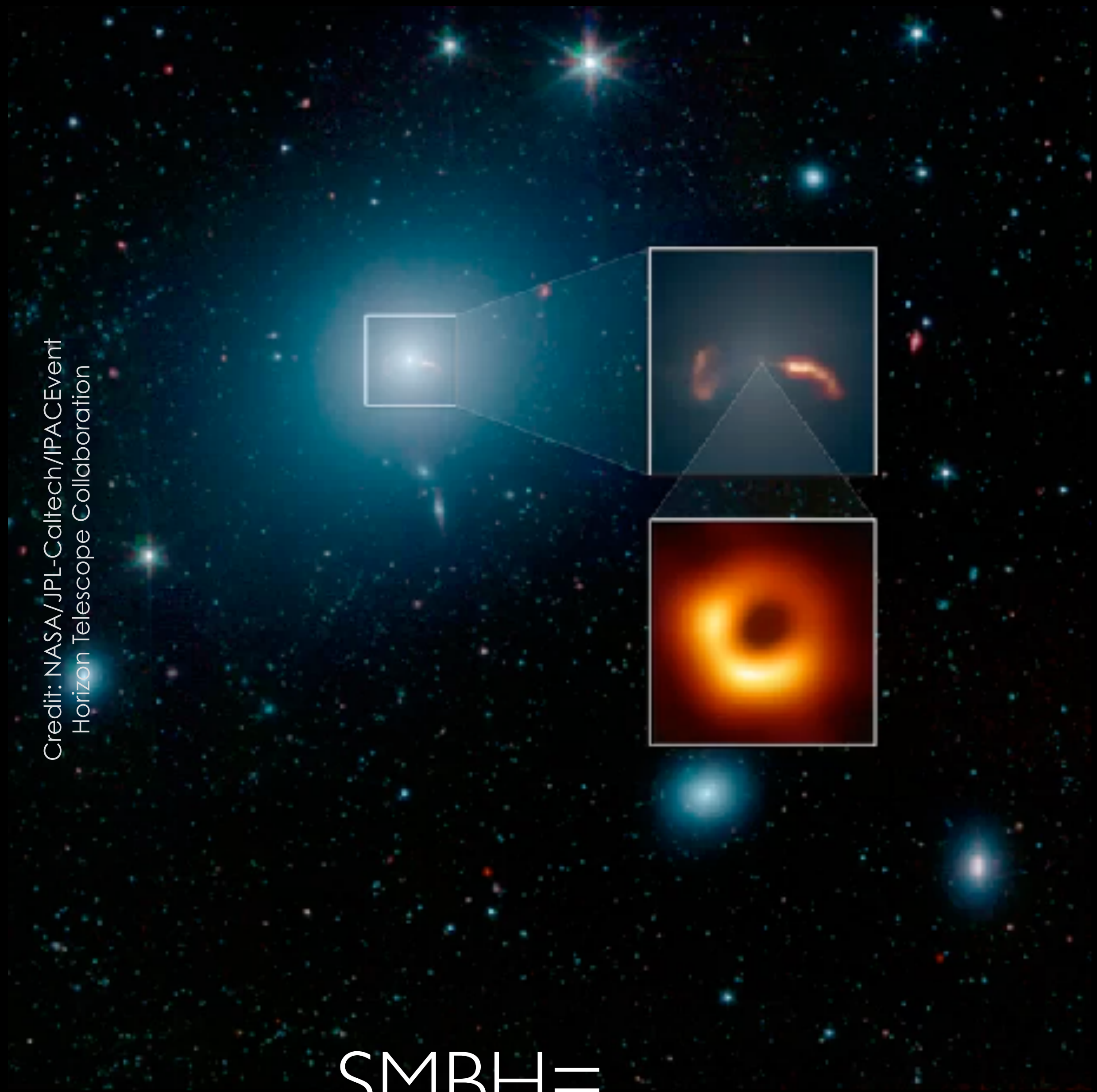
vdWel,…, Chang+14a

# AGNs

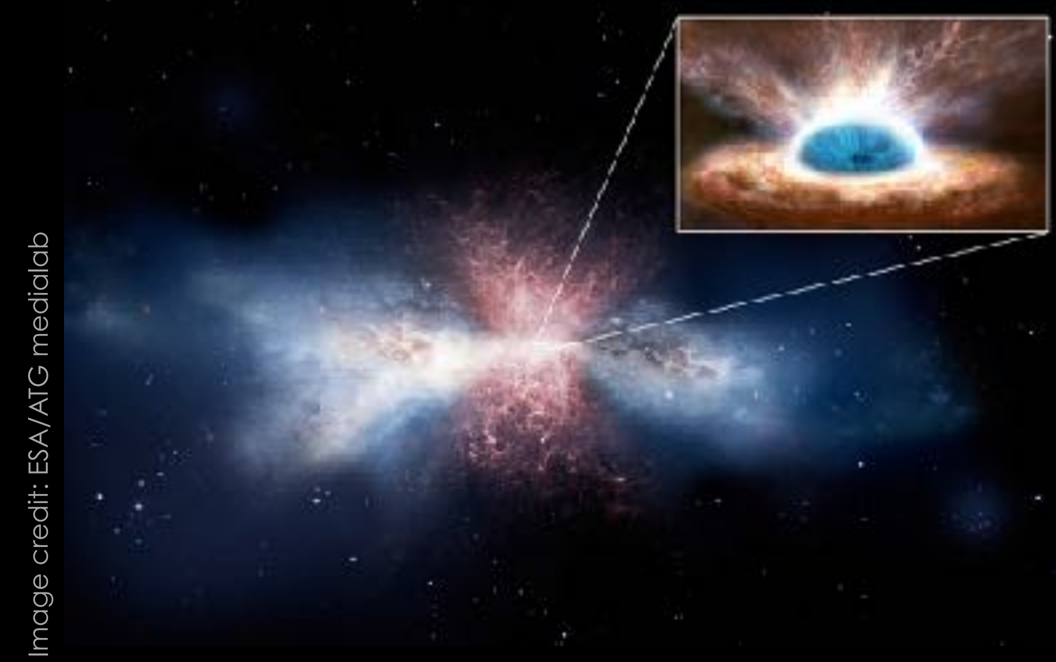


Credit:ESO/WFI (Optical); MPIfR/ESO/APEX/A.Weiss et al. (Submillimetre); NASA/CXC/CfA/R.Kraft et al. (X-ray)

# SMBH-AGN-Galaxy coevolution

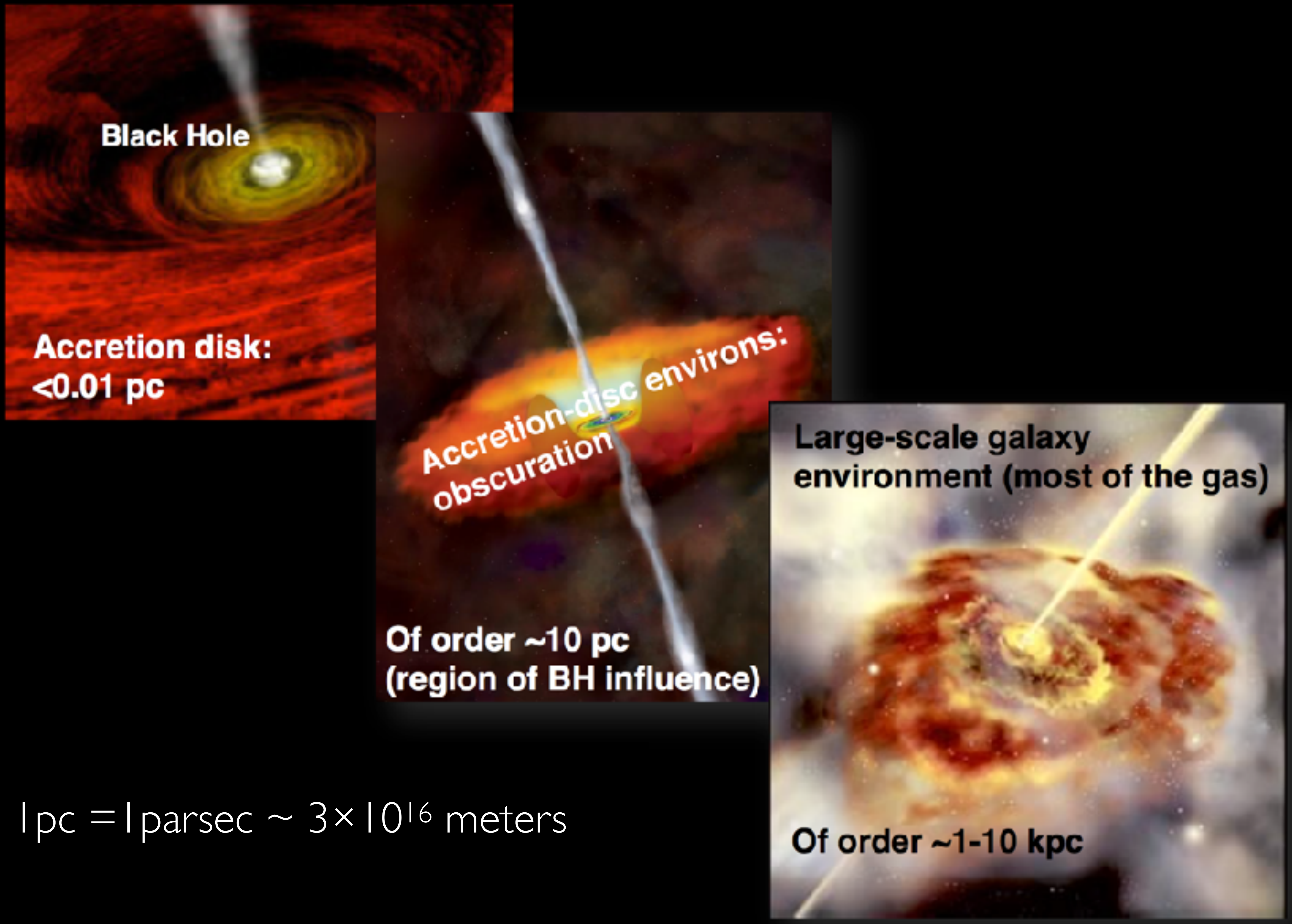



SMBH=
Super Massive Black Hole




AGN=
Active Galactic Nucleus

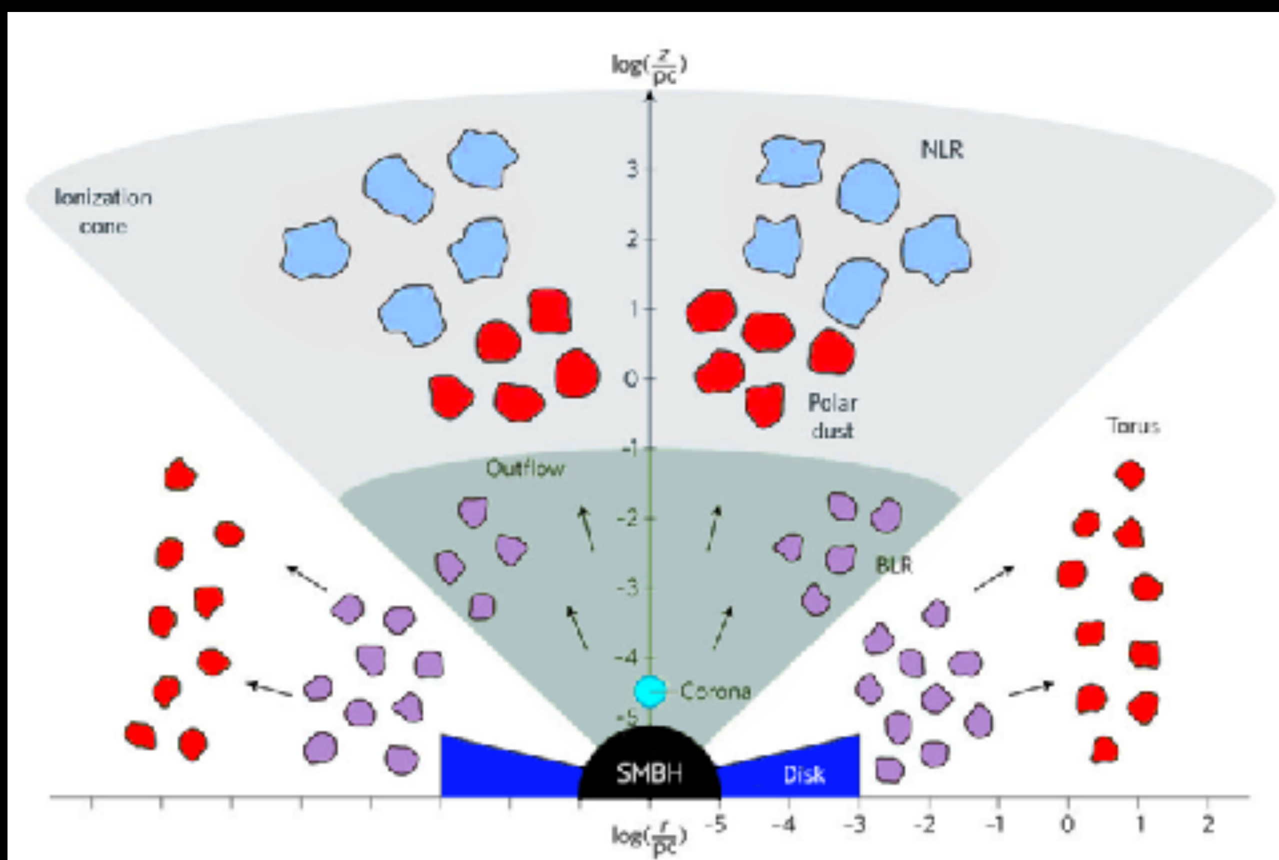
# Black Hole-AGN-Galaxy connections



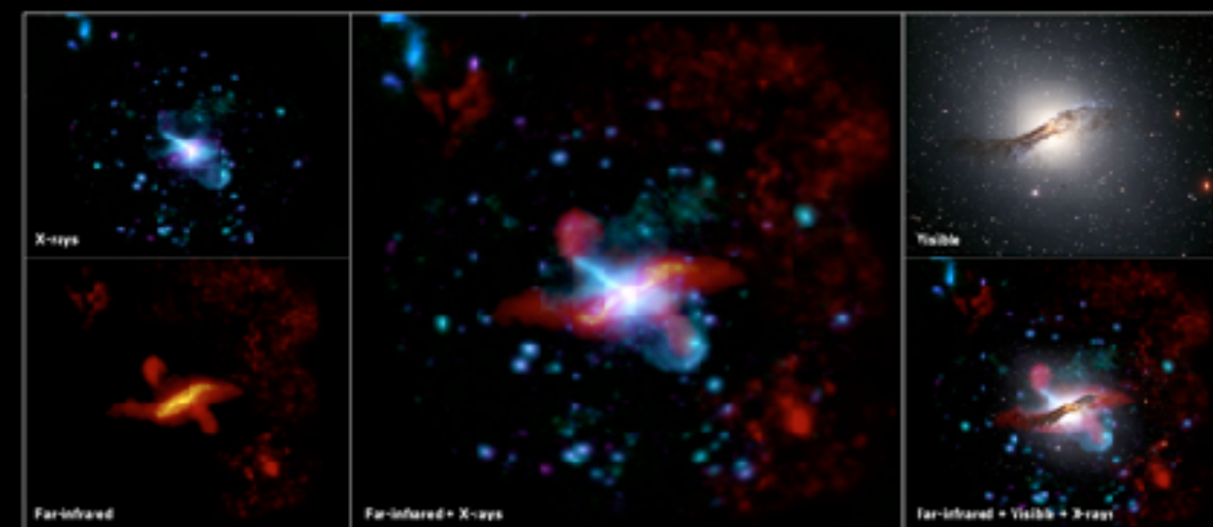

1pc = 1parsec ~ $3 \times 10^{16}$ meters

Credit: David Alexander

# AGN=Active Galactic Nucleus

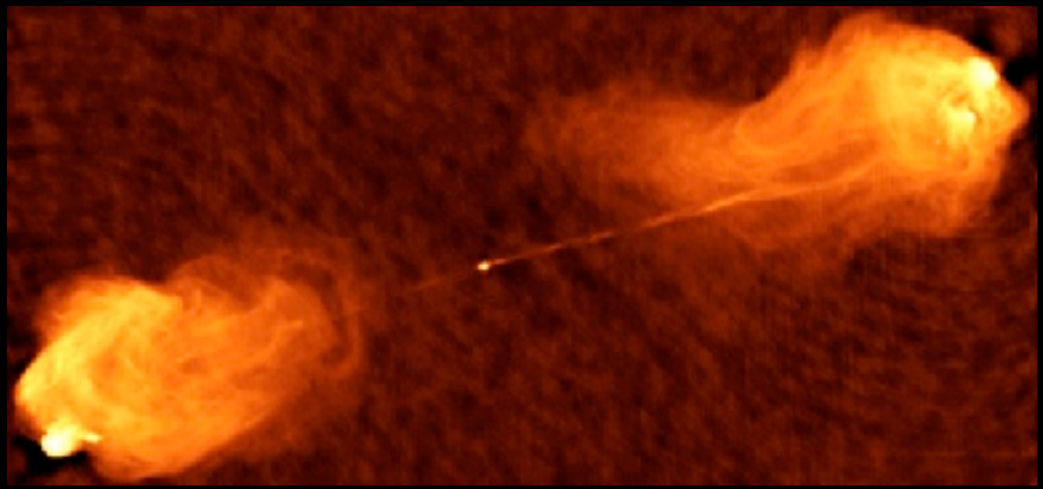



Credit: Hickox+2018




Credit: ESA/NASA, the AVO project and Paolo Padovani

# AGNs by multiwavelength study

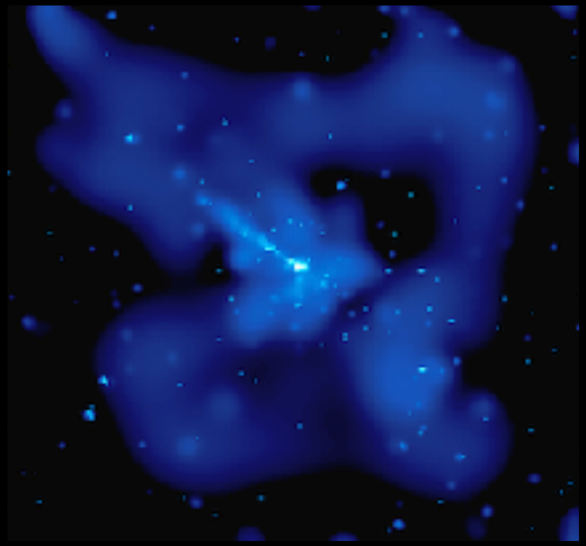


Credit: John Conway and Philip Blanco (VLA)

radio: Cygnus A

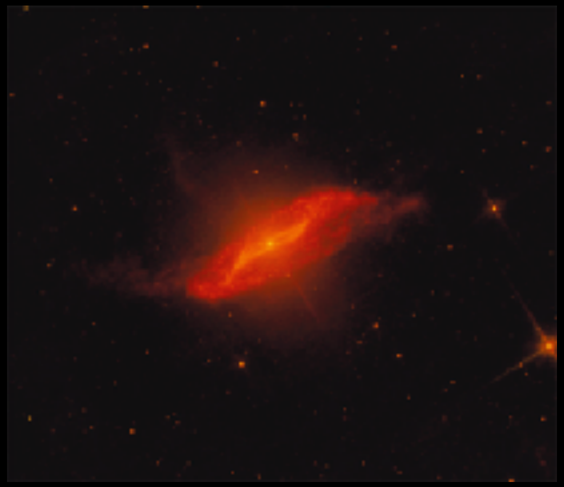
Credit: NASA and Hubble Heritage Team (StScI/AURA)

optical: M87



Credit: NASA/CXC/M.Karovska+

X-ray: Centaurus A

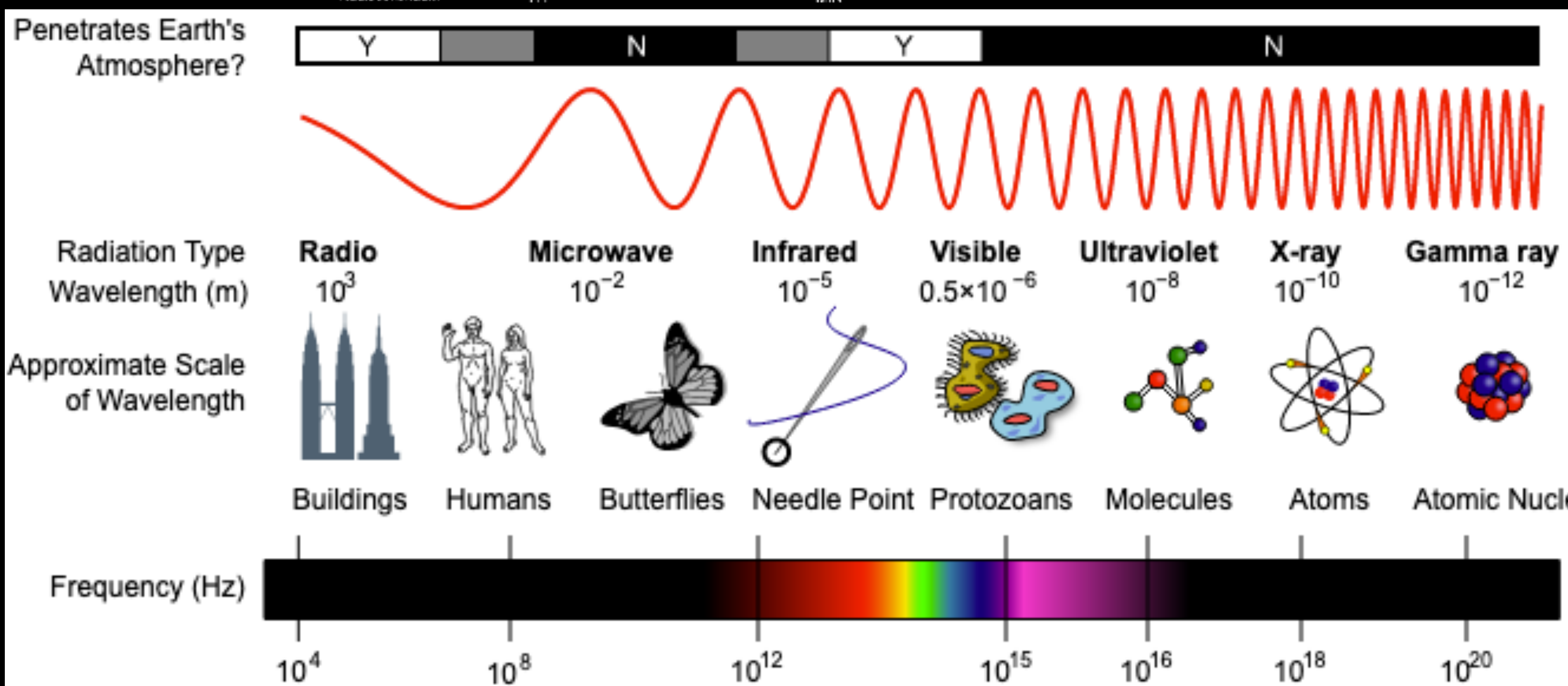


Credit: NASA/JPL-Caltech/J. Keene

IR: Centarus A

# AGNs by multiwavelength study


Credit: A.R.I.-S.



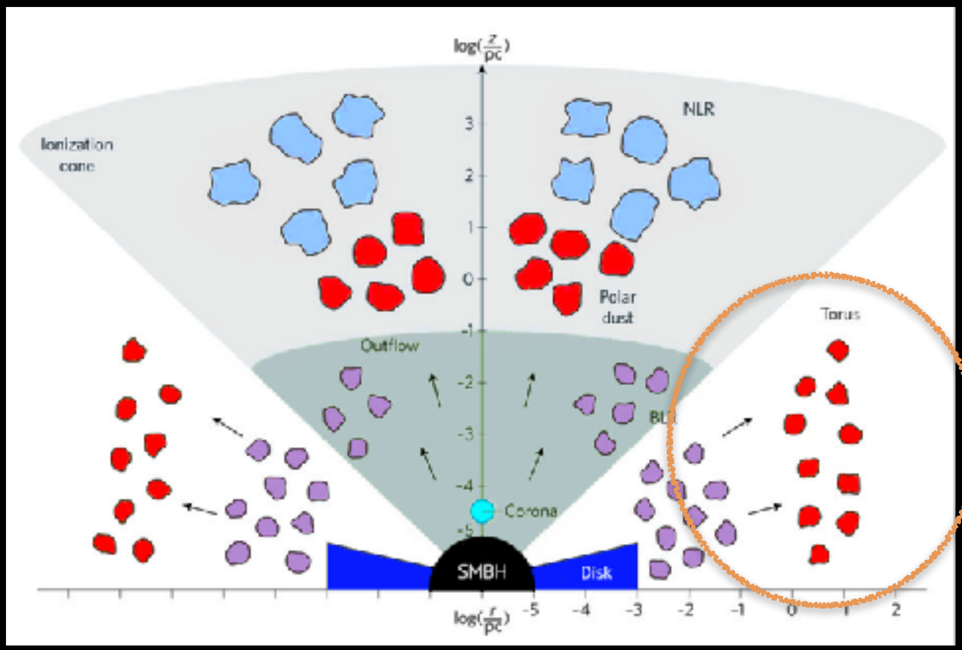

Credit: NASA

# Infrared Selected AGNs: SEDs

(=Spectral Energy Distribution)

MAGPHYS+AGN

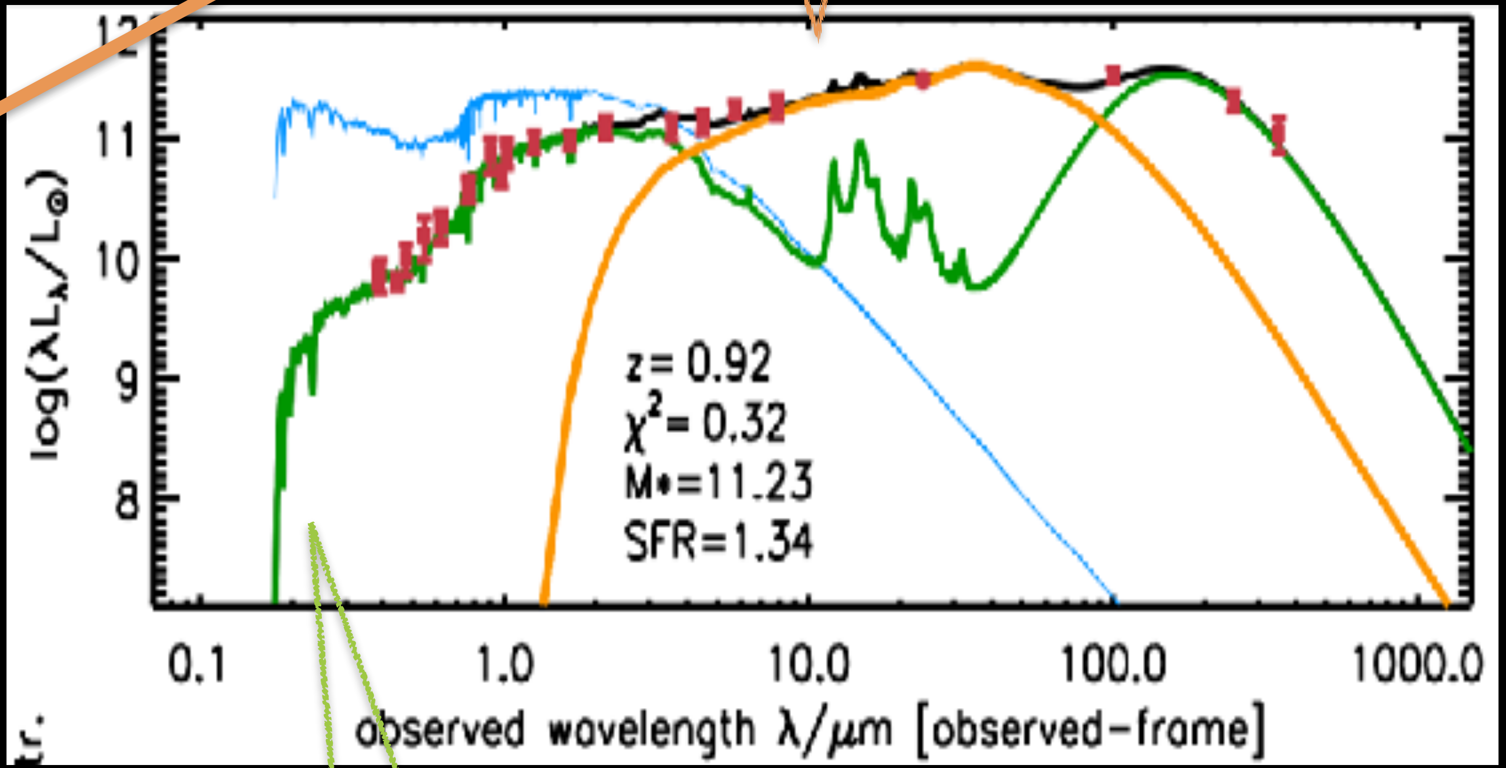
da Cunha+08,15,in prep.
Juneau+, in prep.
Chang+17b




Credit: Hickox+2018

torus emission

luminosity

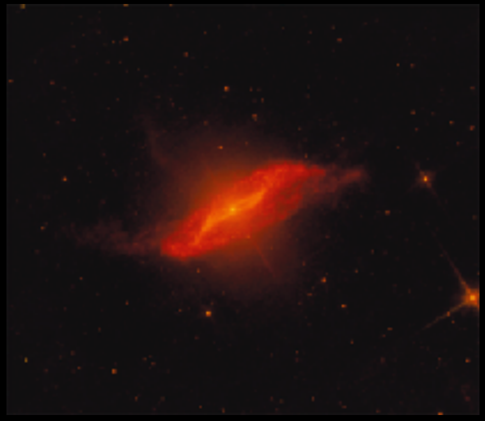



Chang+17a

star-forming galaxies

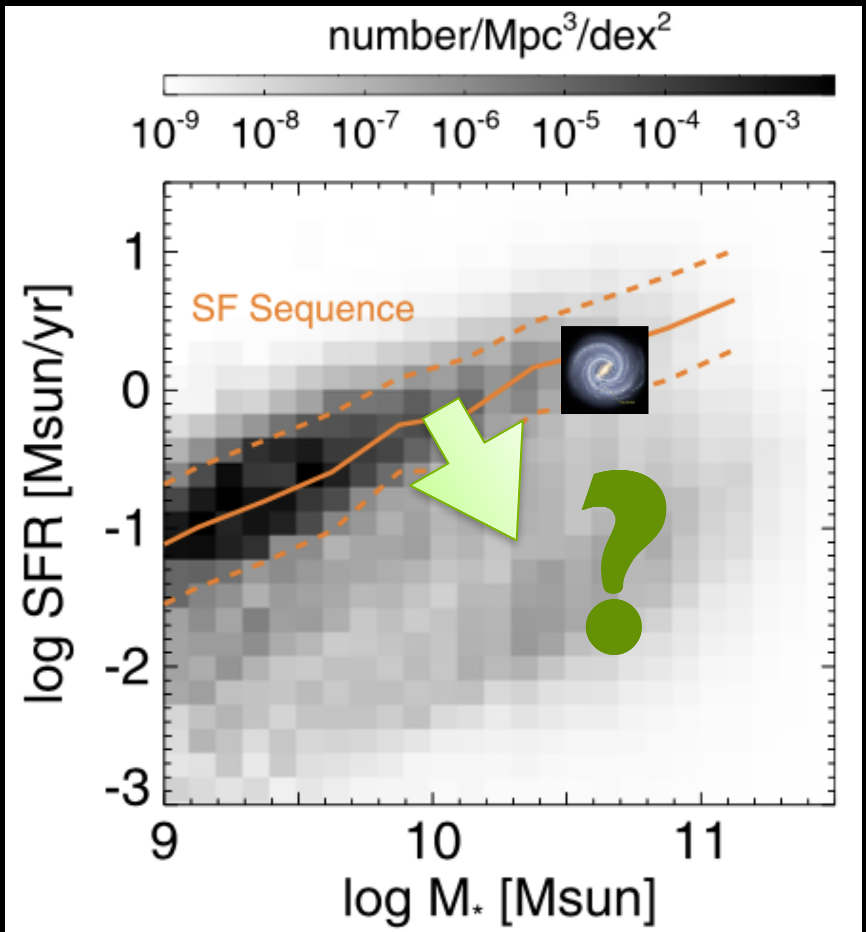
→z~2.5 (11 Gyrs ago)

(Universe~13.7 Gyrs)

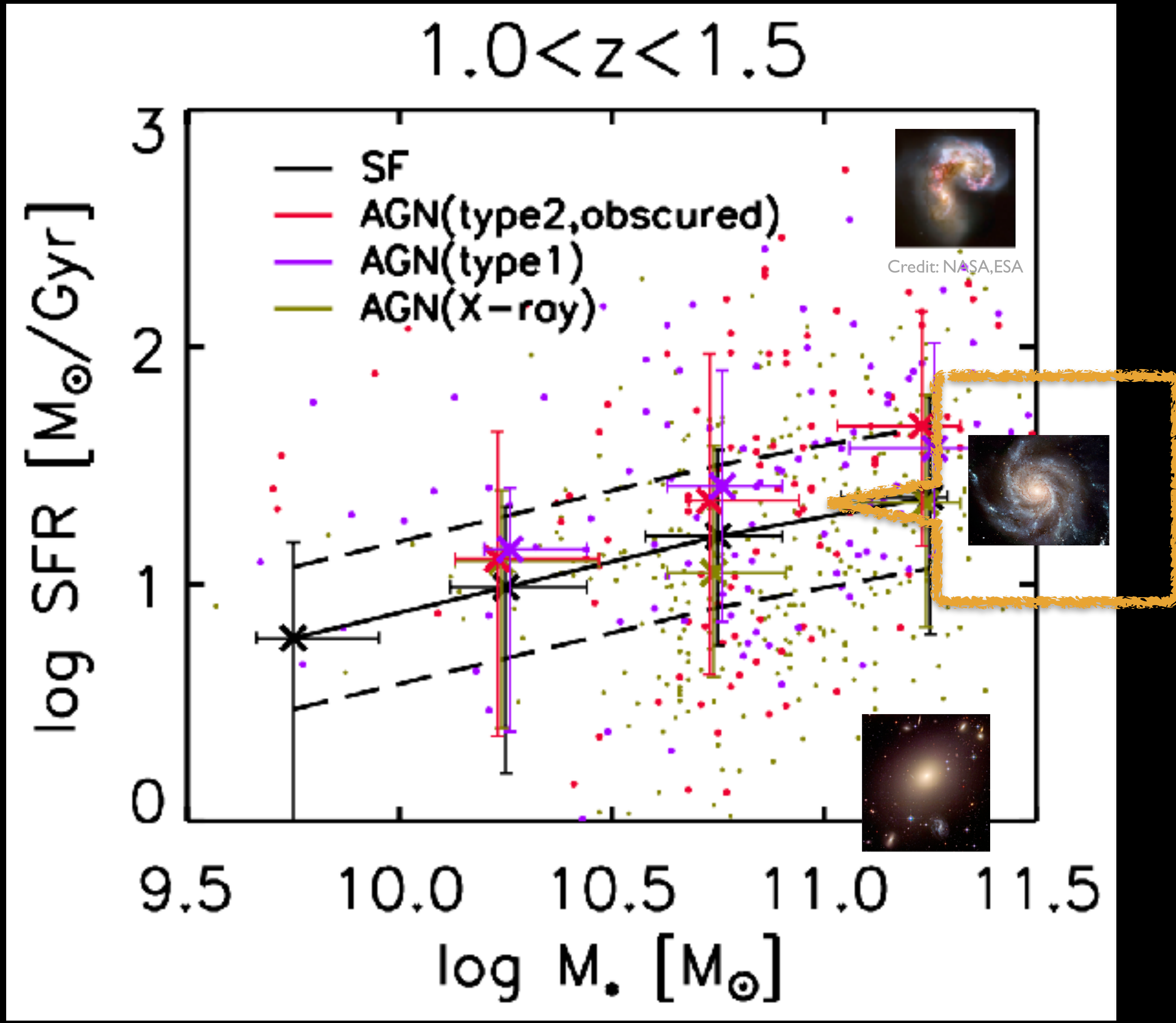


# Star-forming Relation (Main Sequence)

Obscured IR-AGN hosts lie in the SF sequence.




Chang+15




Chang+17b

z=0(now)
z=1(~7.8 Gyr ago)

SFR=star formation rate

# AGN sizes by stacked HST/F814W images

without AGNs

with AGNs

z=1(~7.8 Gyr ago)

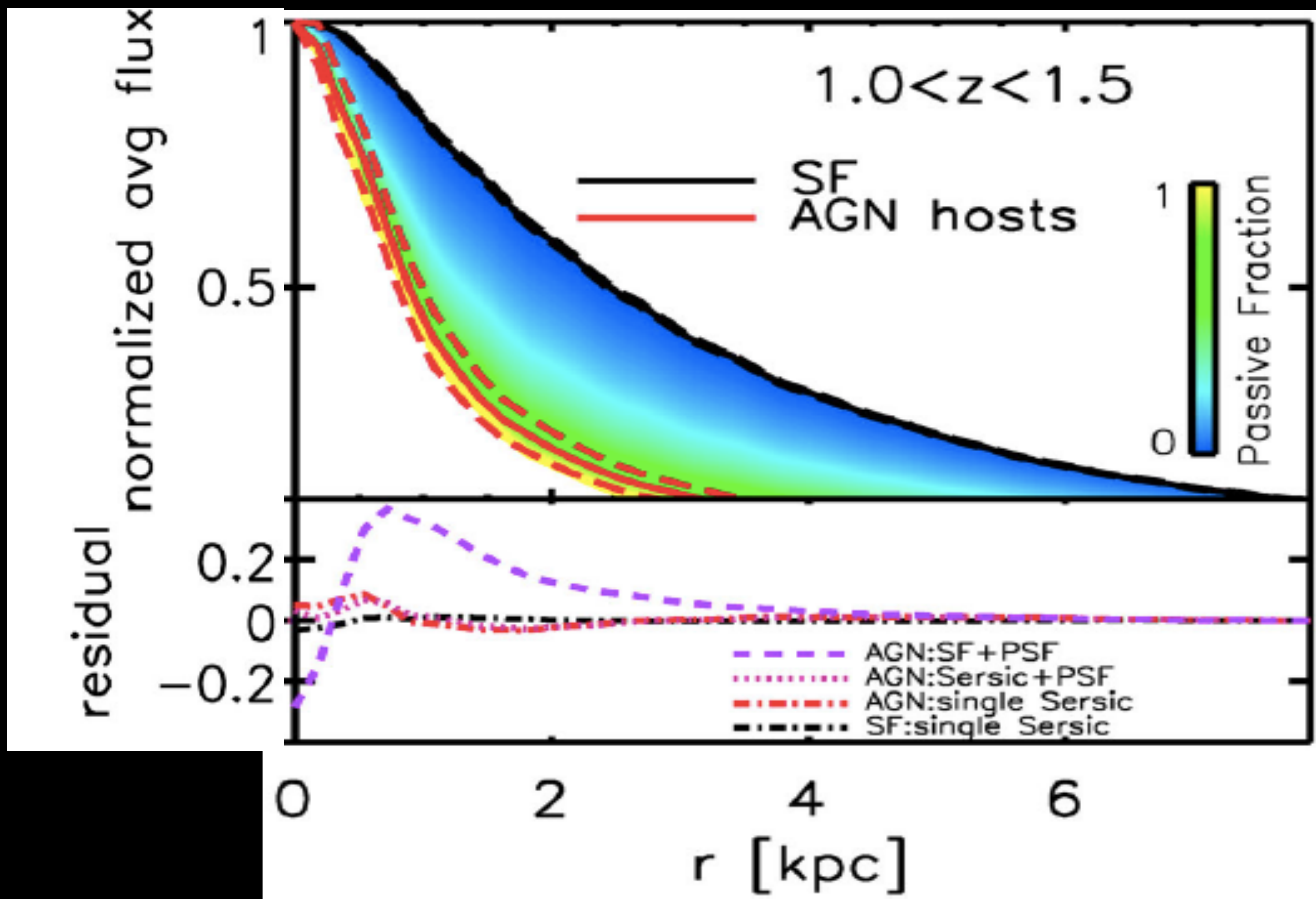
Obscured IR AGN hosts are rounder and smaller

Chang+17a

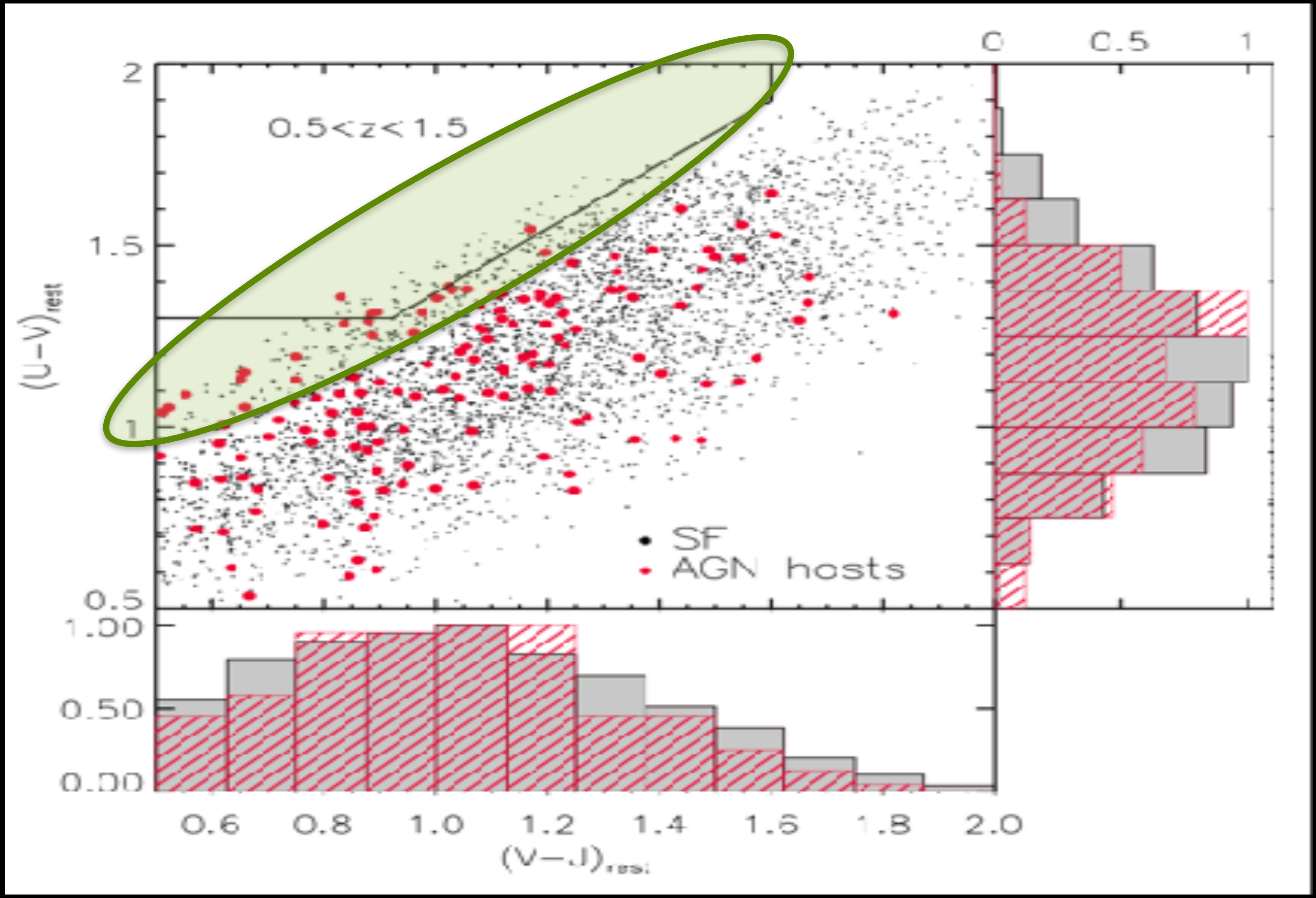
# Average Light Profiles

Obscured IR-AGNs
MIR (SF) galaxies

—— images
- - - uncertainty



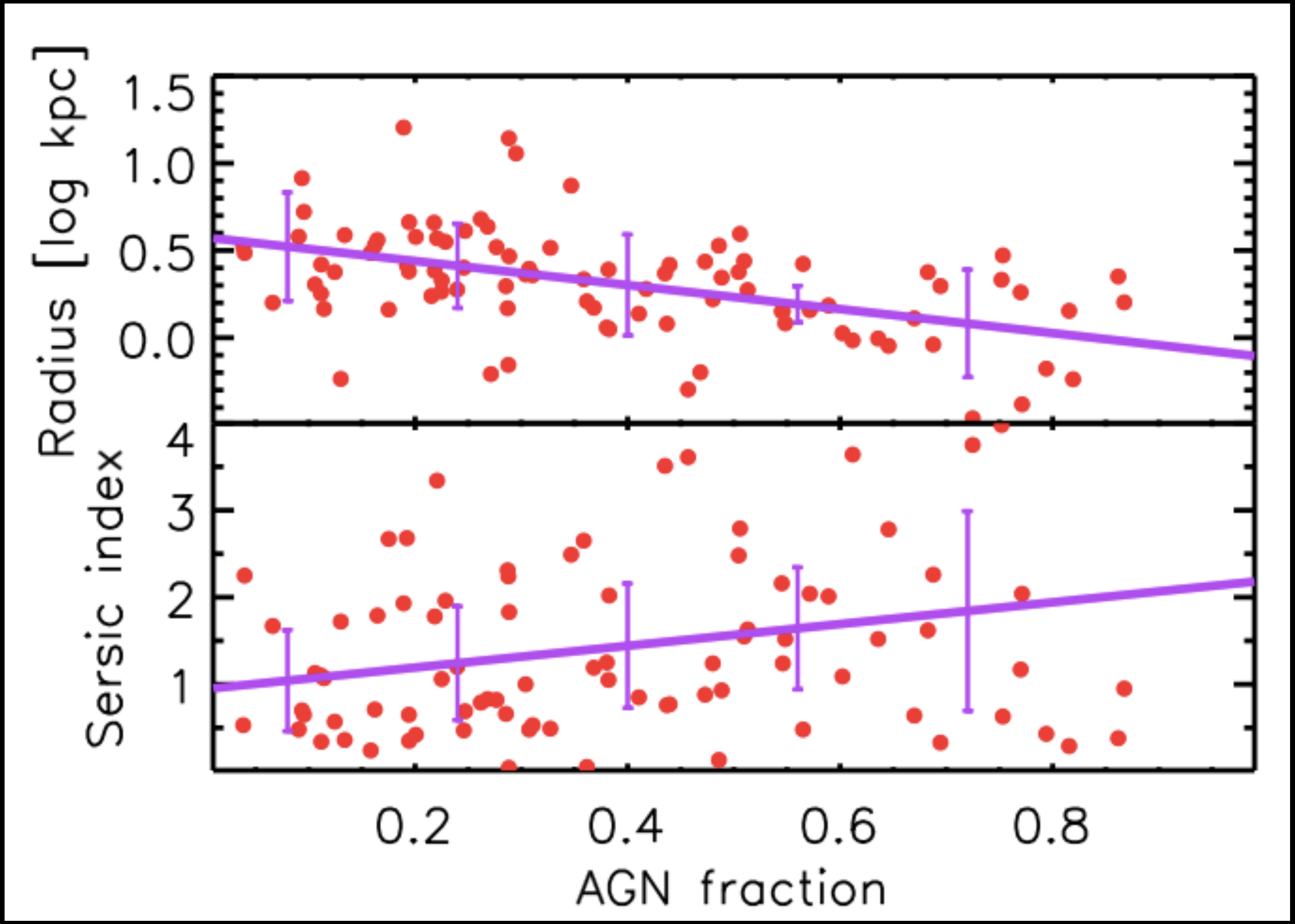
# Decomposition UVJ plot: quiescent or star-forming?




Chang+17a

- Not dominant by passive population
- Compact star-forming component

# Size vs. AGN Fraction




20-50% AGN →25-50% size smaller

Chang+17b

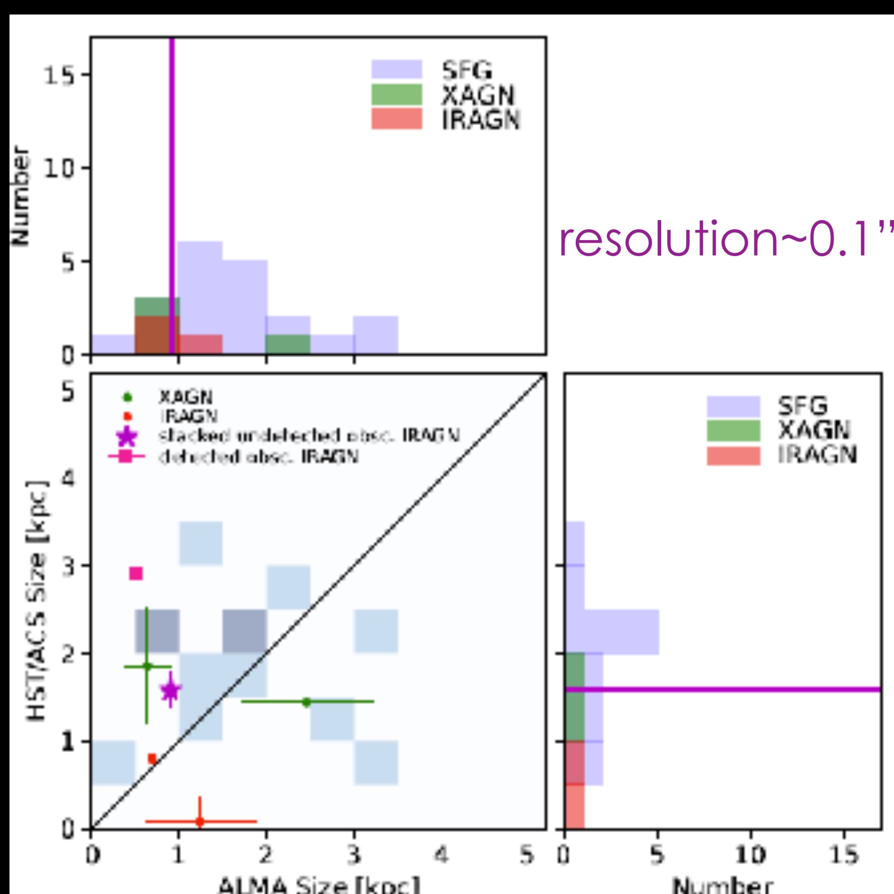
# ALMA Observation

PI: Chang

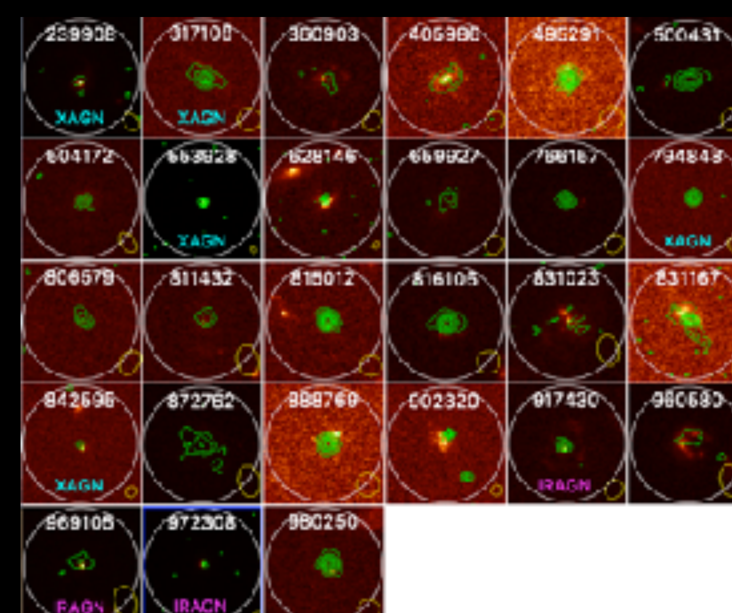
Credit: ESO/C.Malin

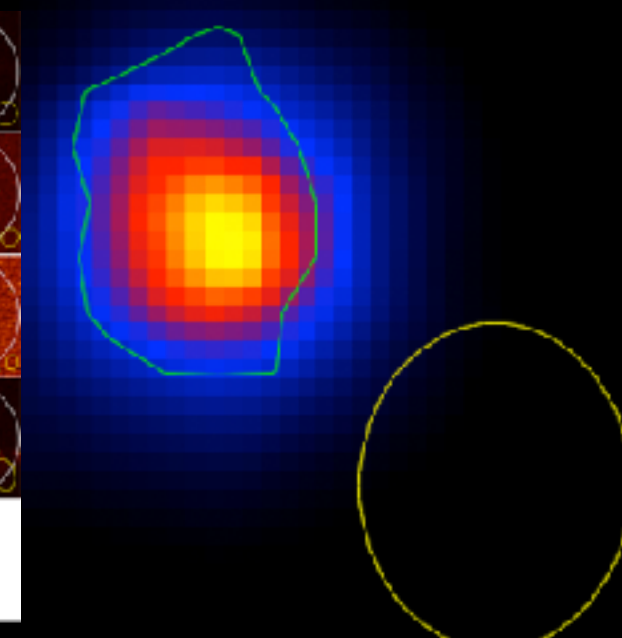
The sizes measured from both ALMA/1mm (dusty) and HST/F814W (stellar) images show that obscured AGN host galaxies are more compact than non-AGN SFGs.

Credit: ALMA

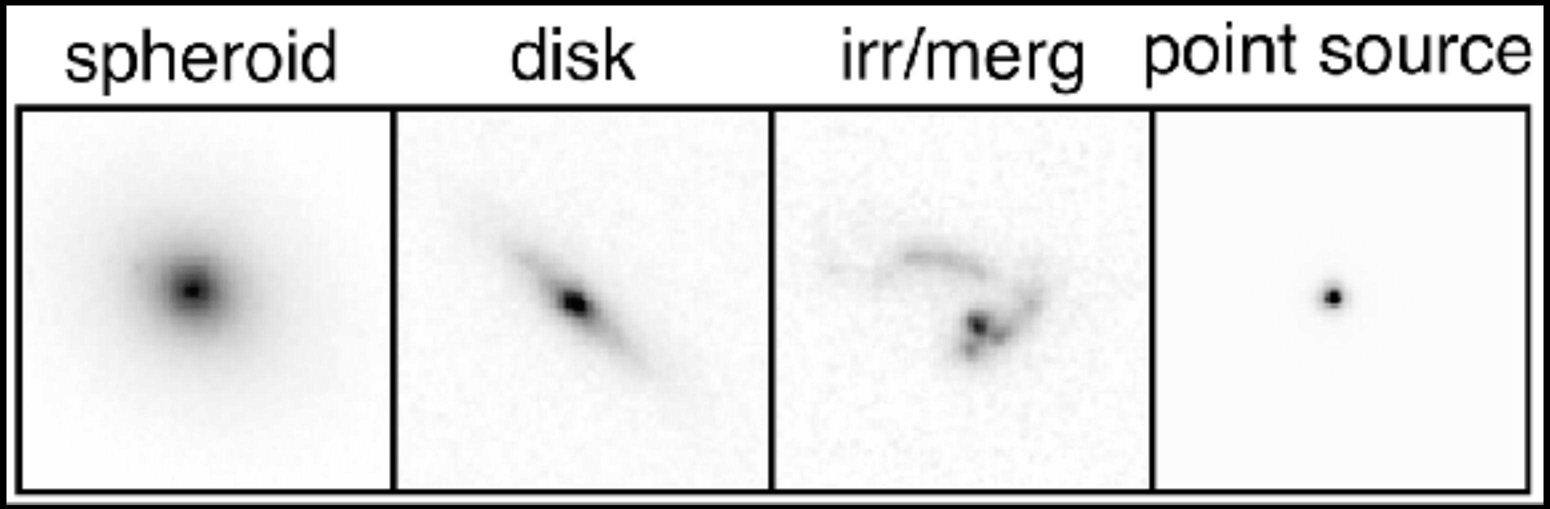



Chang+20



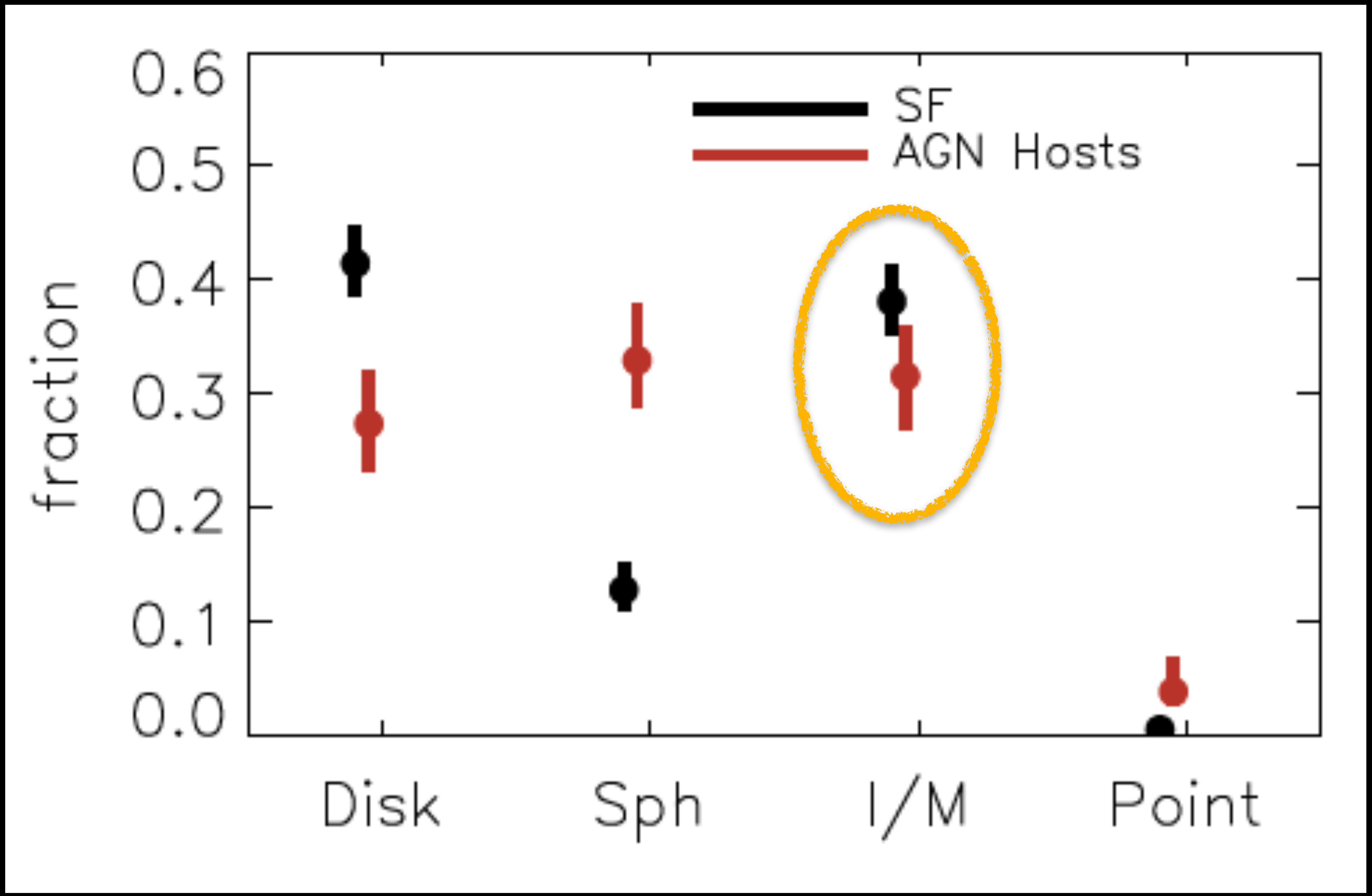


# Visual Classification (AGN)

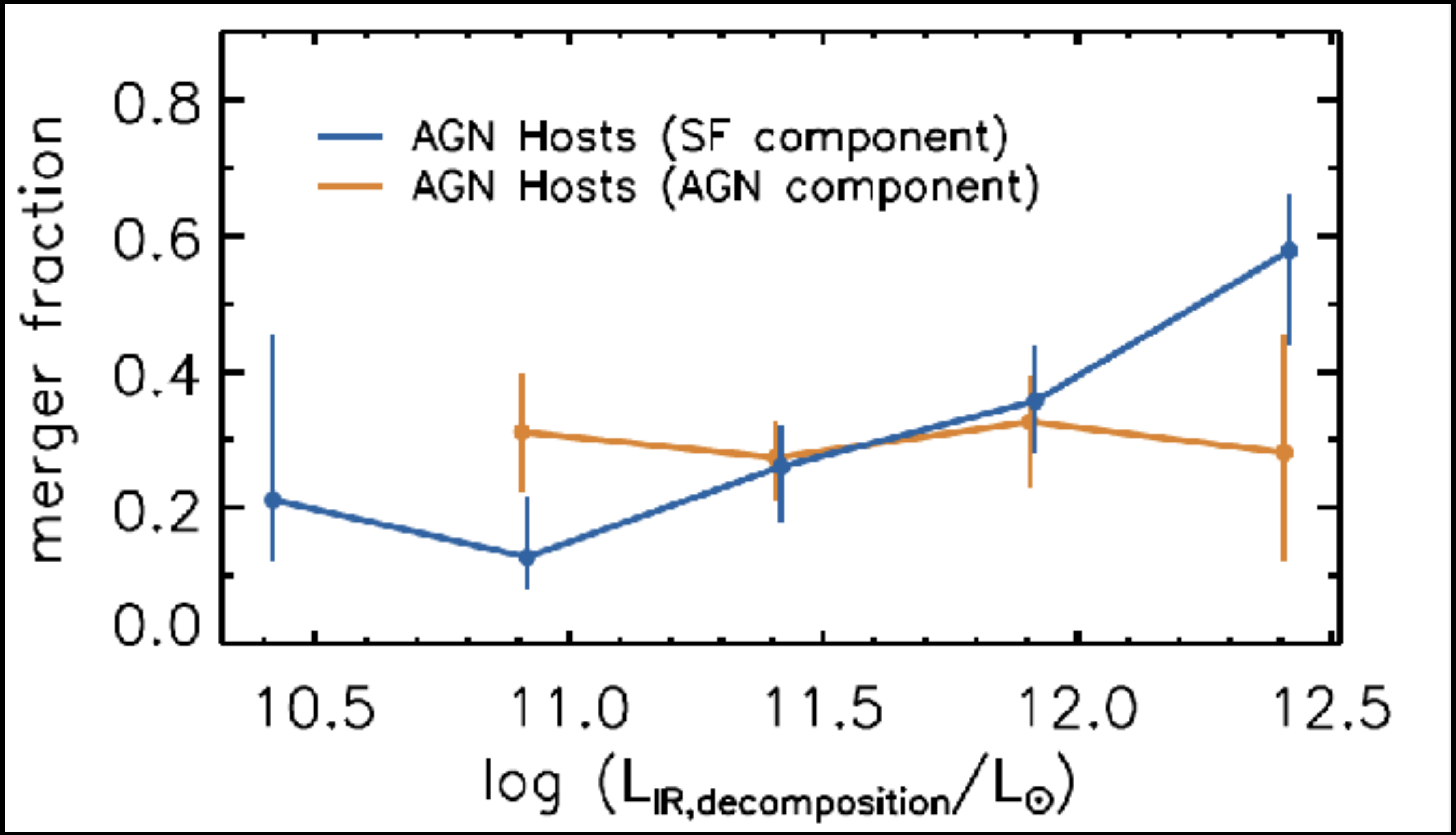



$\log M_{star} > 10.5$
$0.5 < z < 1.5$

- major merger is not the main driver of obscured IR-AGN activities.



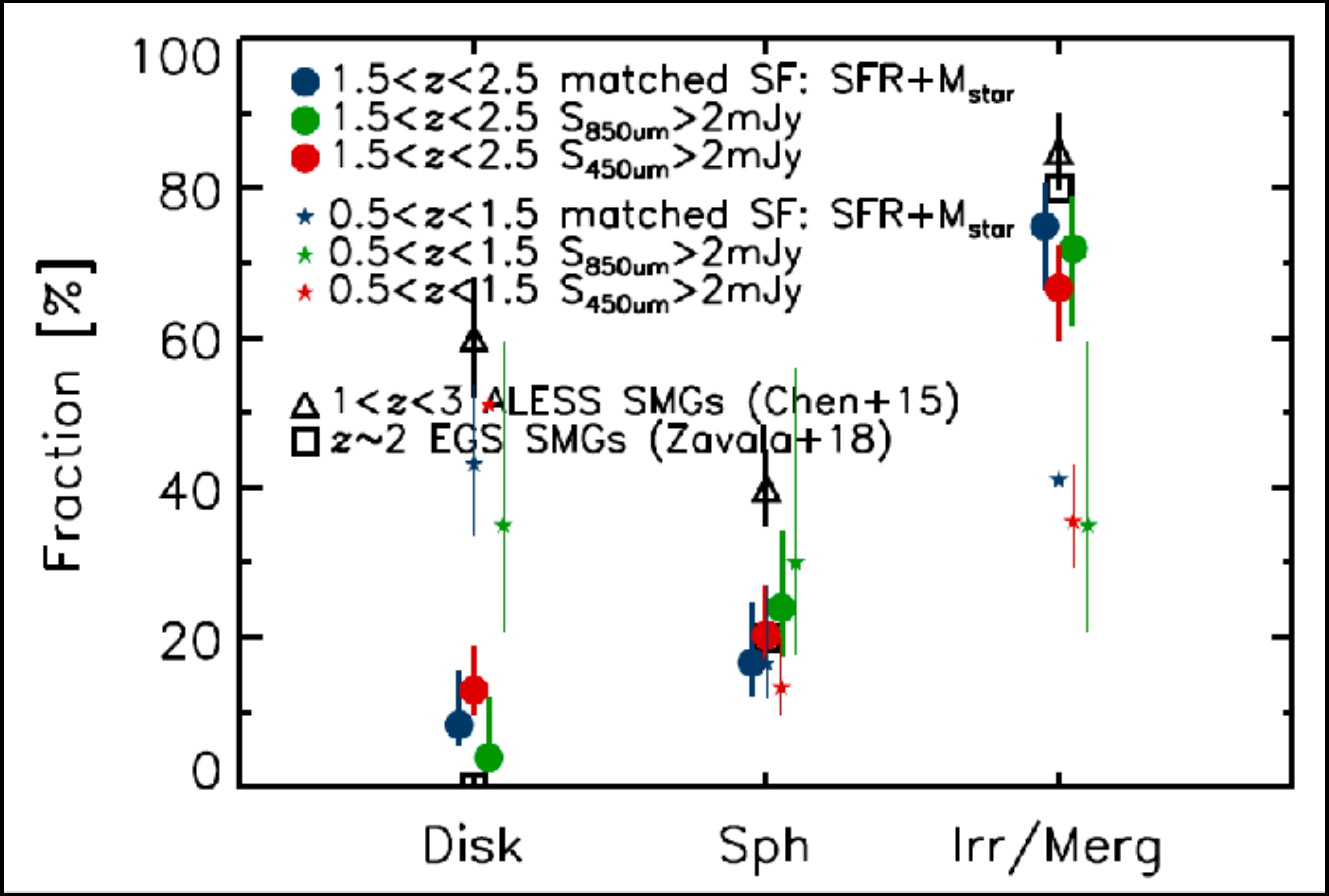
# Merger Fraction (AGN)




- Merger fractions@z~1
  - depend on total or decomposed SF IR luminosity (SFRs).
  - rather than decomposed AGN IR luminosity.

Chang+17b

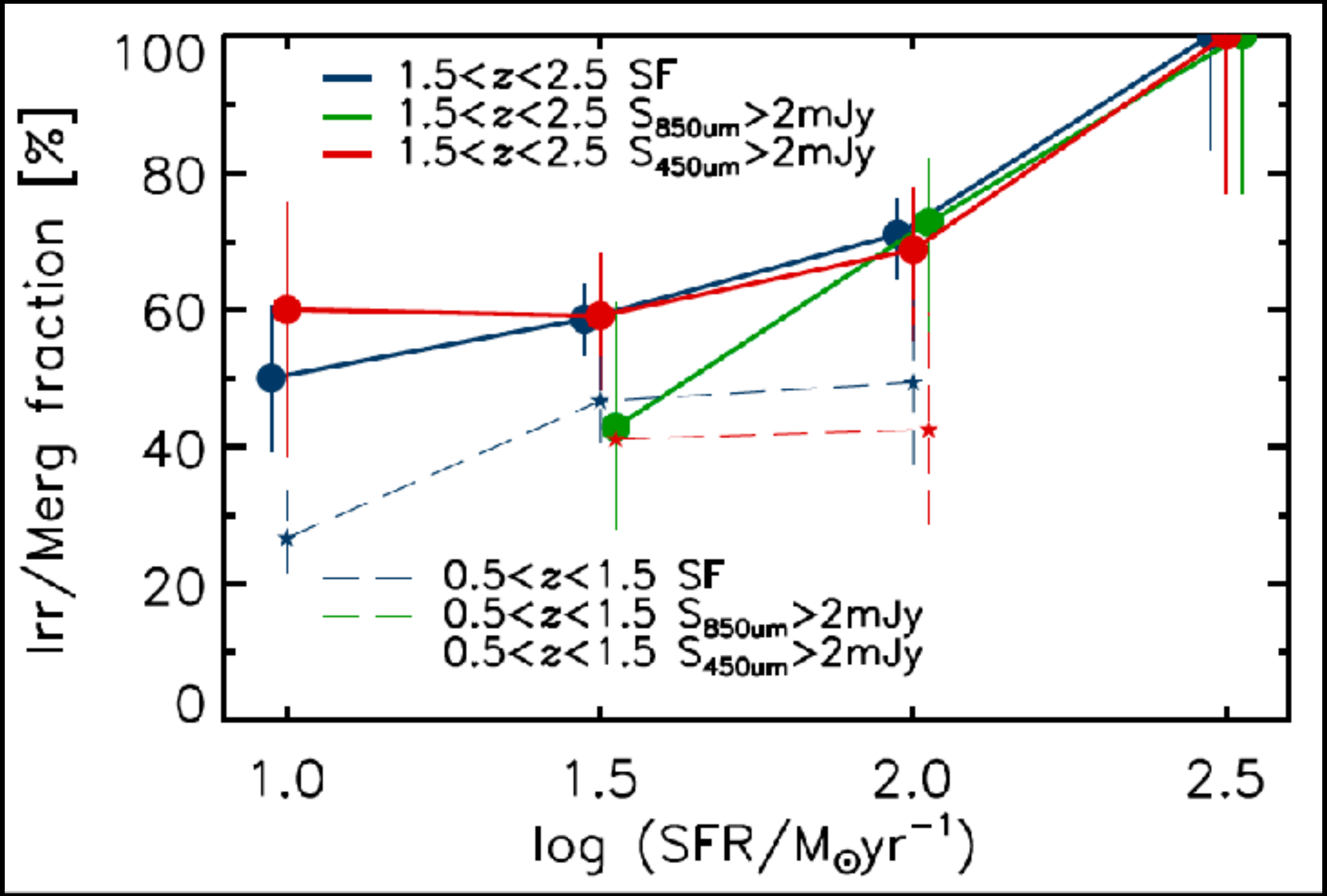
# Visual Classification (SMGs)




Most of the STUDIES SMGs (70%) show disturbed features.

Chang+18

# Irr/Merg fraction vs SFR (SMGs)




Disturbed feature fractions depend on the SFRs for both dusty and star-forming galaxy sample.

Chang+18

# How do Black Holes Affect Galaxy Evolution?

- IR selected AGNs @ z~1 are compact star-forming AGN hosts.

AGNs!

- On the Star-forming sequence — Chang+17b
  - Similar to normal galaxies
- Structures of AGN host galaxies — Chang+17ab, 20
  - Compact star-forming AGN hosts
- Merger fractions — Chang+17b, 18
  - Depend on SF IR luminosity (SFRs)

# Galaxies/AGNs by Machine Learning

- Increase sample size
  - AGN identification
  - Other rare populations
- Investigate galaxy structures
  - Merger stage classification
  - Star-forming, quiescent galaxies, and others
- Derive physical parameters
  - Similar to SED fitting
  - Unsupervised learning

# Identify AGNs by Machine Learning Method

labeled data

Non-AGN AGN

COSMOS,Chang+2017b
(>20-bands, 2 deg$^2$)

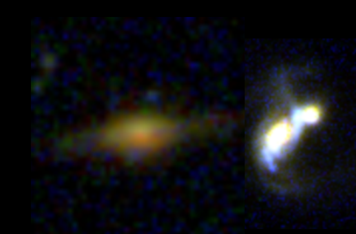
Python:
XGBoost
Keras,
scikit-learn

new data

HSC+WISE
(5~9-bands, 160 deg$^2$)

Predicted
Model

AGN



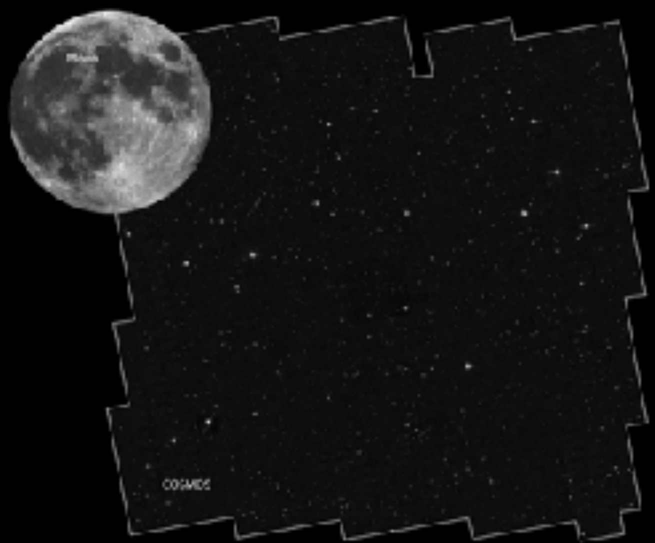
Non-AGN

Chang+ 21

Credit: HSC/SSP

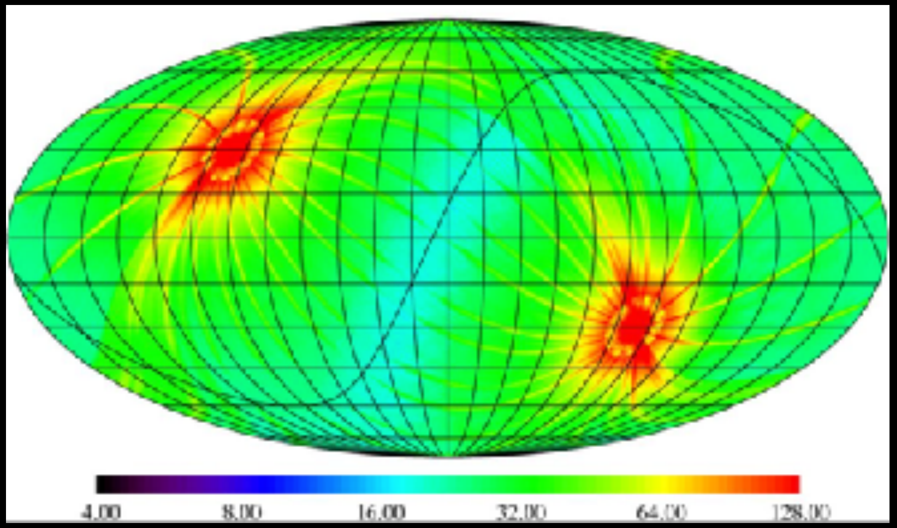
# Survey Fields

COSMOS

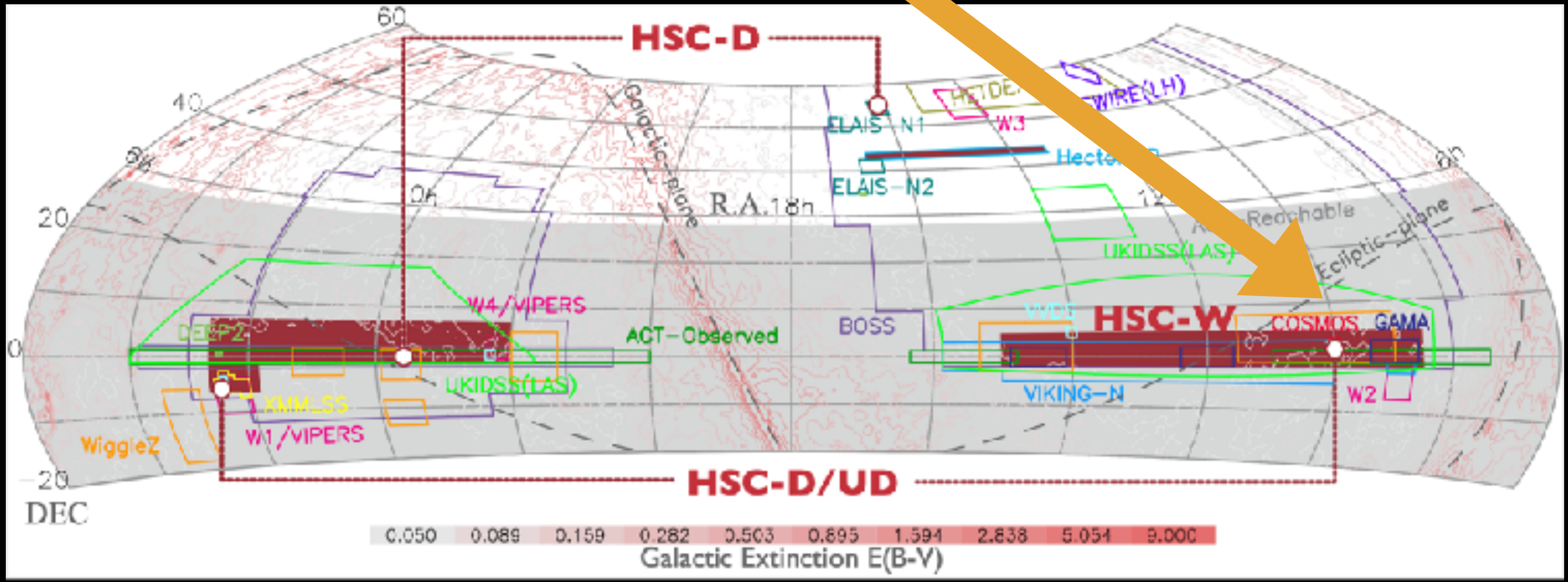


Credit: NASA, ESA, and Z. Levay

ALLWISE




HSC



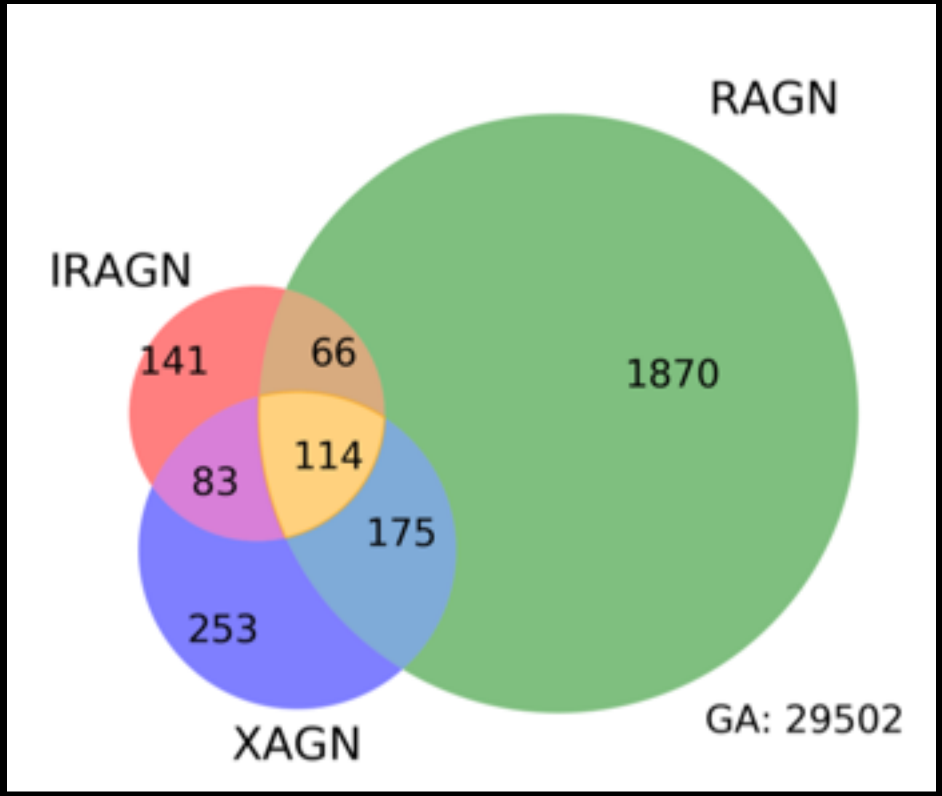
# Data

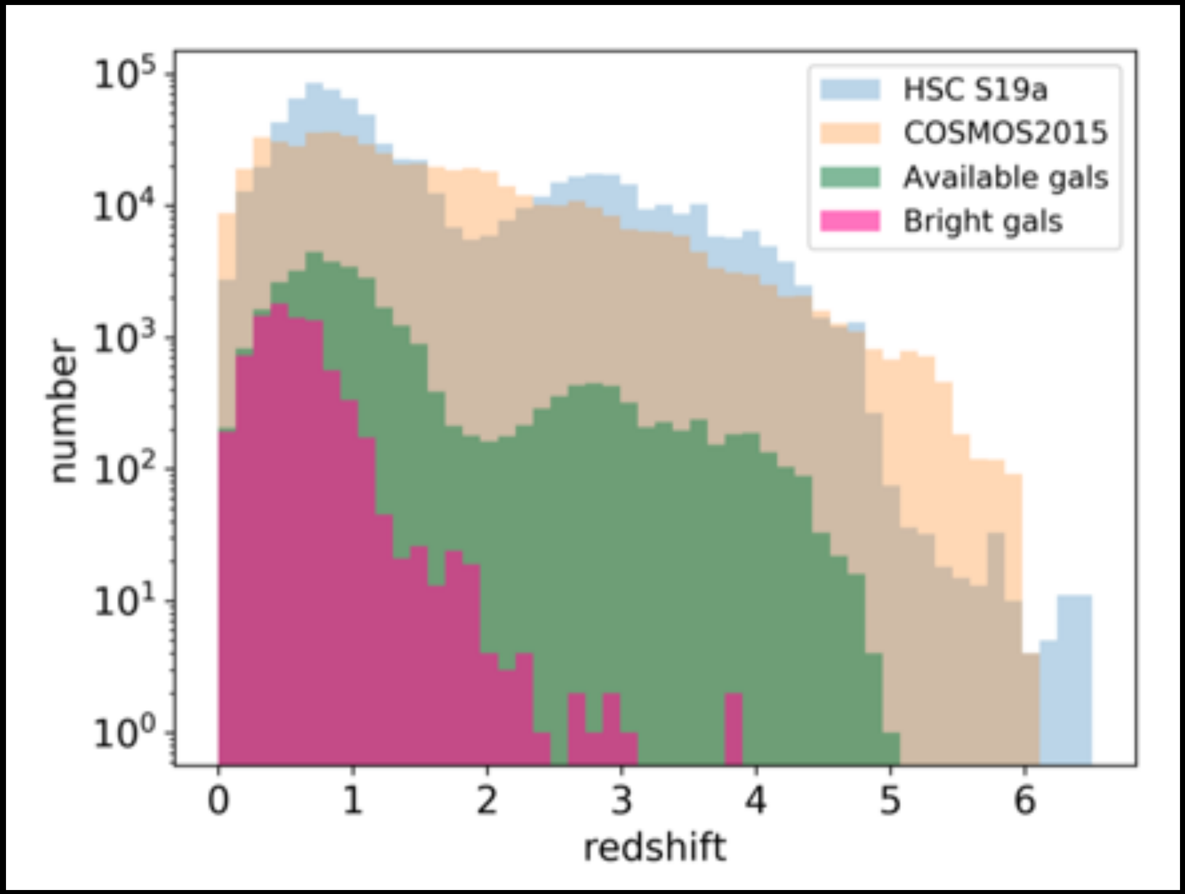
- HSC-SSP S19a (grizy+photo-z)
- WISE (band-1+band-2)

- COSMOS field
  - 625 XAGNs (X-ray)
  - 404 IRAGNs (infrared)
  - 2225 RAGNs (radio)
  - 29502 GAs (non-AGNs)

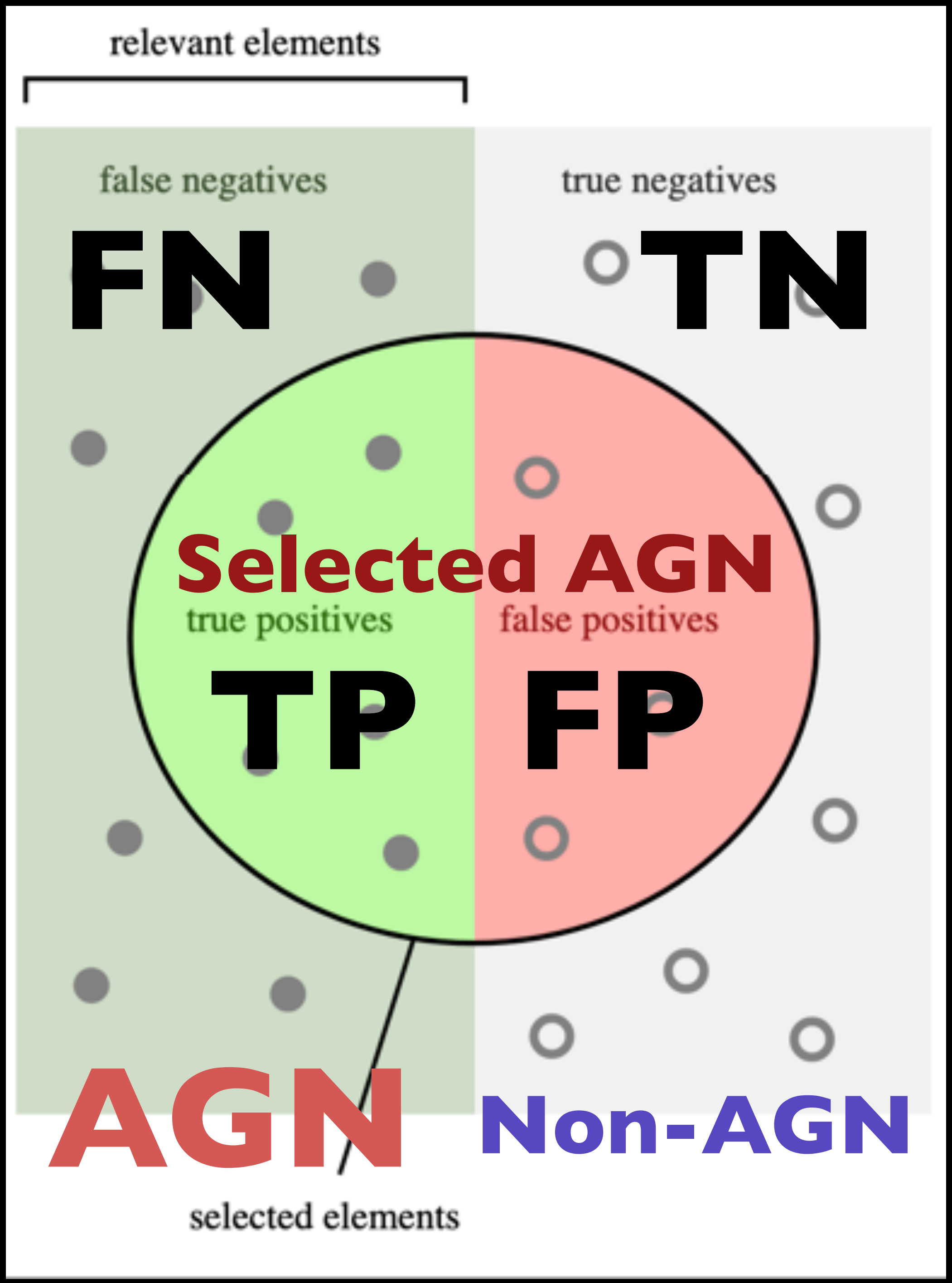
Civano+2016; Marchesi+2016; Chang+2017a,b;
Smolcic+2017; Delvecchio+2017

- 2/3 training +1/3 test sample
- In-balance data (over-sampling)



# Performance (AGN)

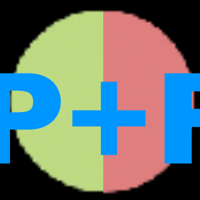



Credit: Walber

- TP: an AGN source which is classified as an AGN — Ideal case: large
- TN: a non-AGN source which is classified as a non-AGN — large
- FP: a non-AGN source which is classified as an AGN — small
- FN: an AGN source which is classified as a non-AGN — small

# Performance: Precision, Recall, F1 Score

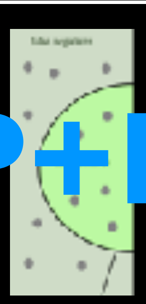
(Purity)

$$\text{Precision} = \frac{TP}{TP+FP}$$

High

(Completness)

$$\text{Recall} = \frac{TP}{TP+FN}$$

High

$$F1 = 2 \times \frac{P \times R}{P + R}$$

Ideal case: F1=1

# Performance: F1 Score

$$\boxed{F1 = 2 \times \frac{P \times R}{P + R}}$$

Ideal case: F1=1

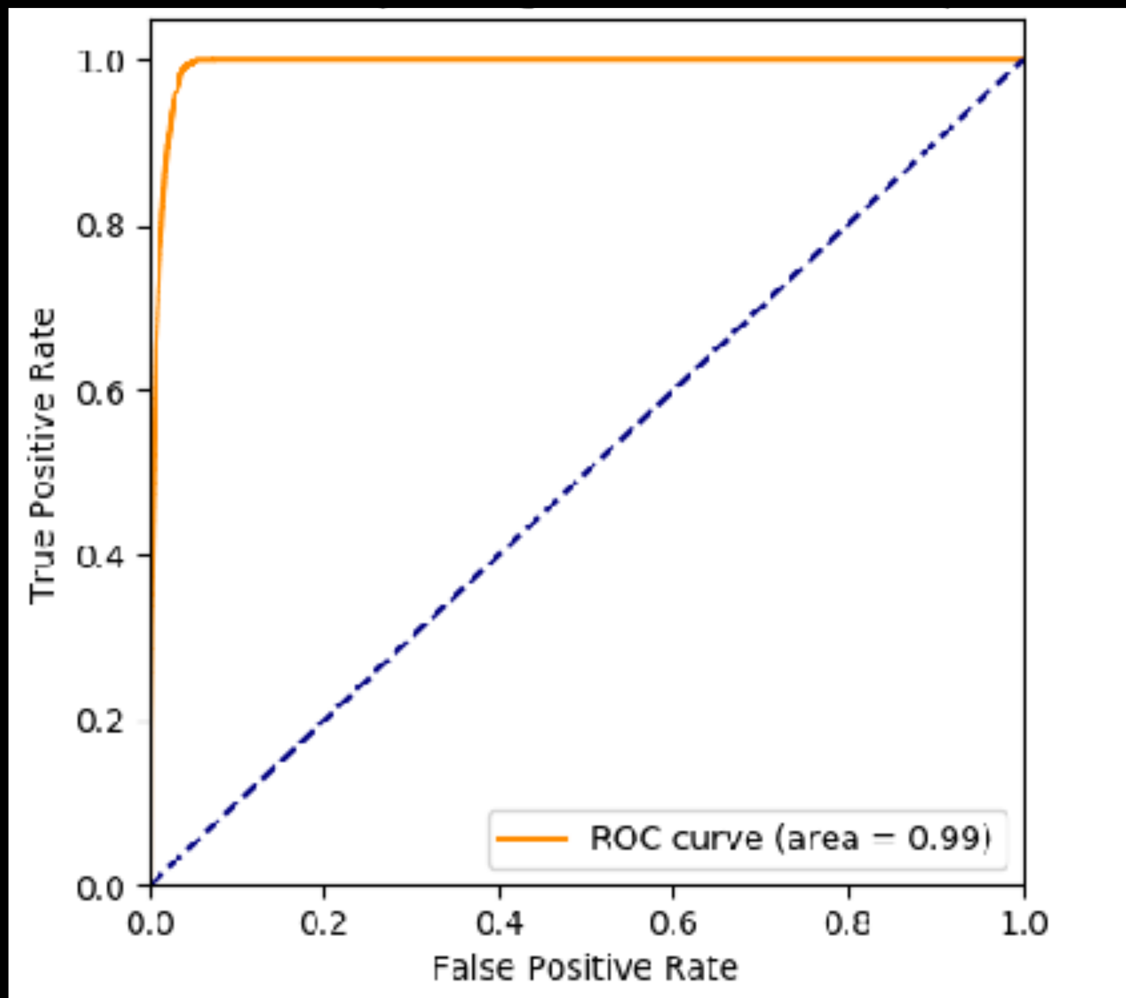
- Logistic Regression>0.45
- Random Forest>0.65
- Sequential Model in Keras>0.60
- XGBoost>0.65

Chang+ 21

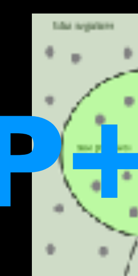
# Performance: ROC Curve

ROC: Receiver Operating Characteristics

TPR



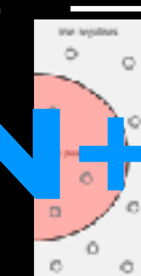

FPR

True Positive Rate= $\frac{TP}{TP+FN}$

TPR

High

False Positive Rate= $\frac{FP}{TN+FP}$

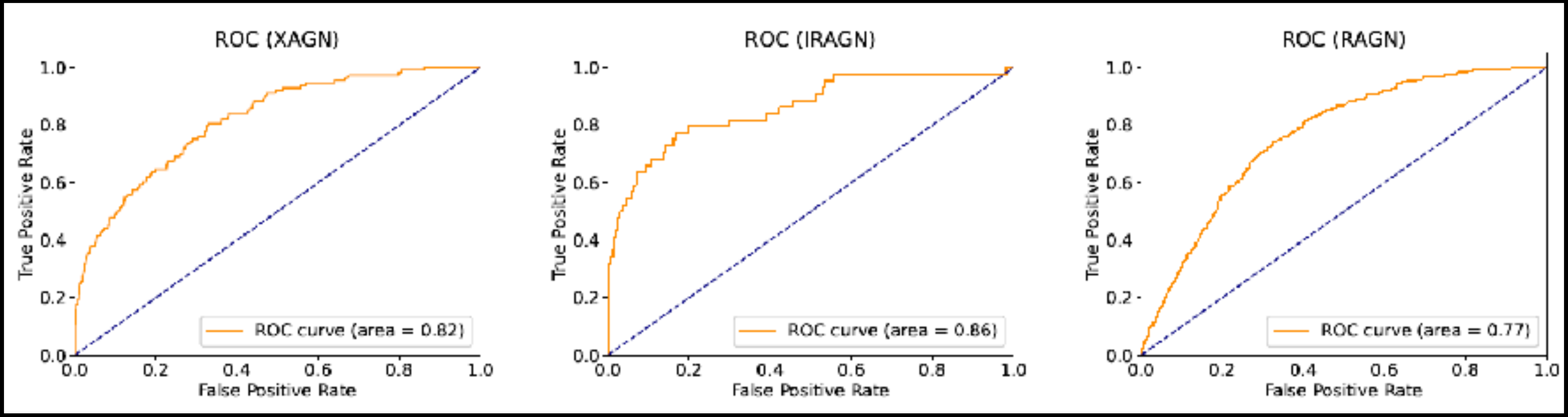
FPR

Low

AUROC: Area Under the Curve of ROC

Ideal case: AUROC=1

# Performance: AUROC

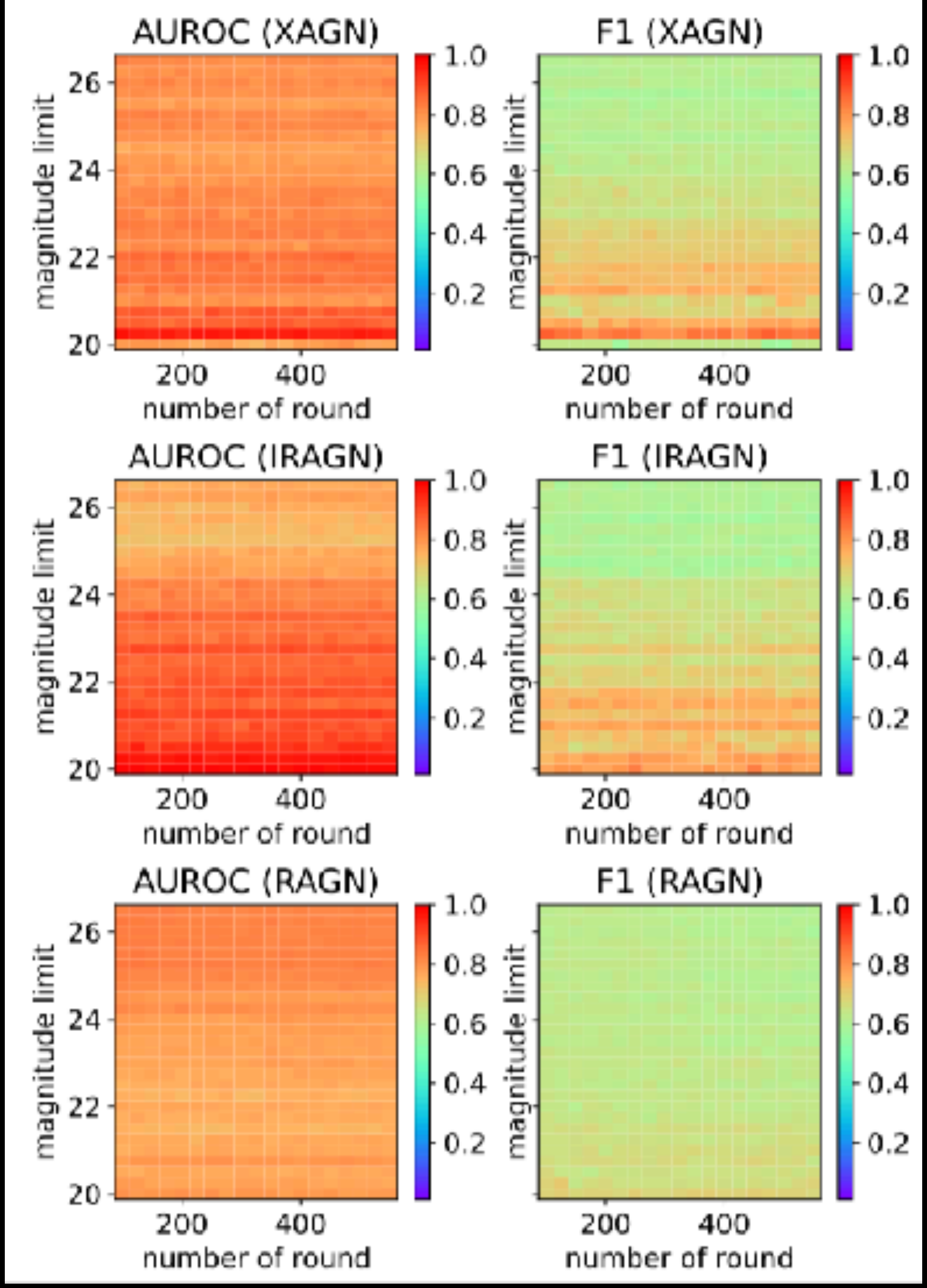



Chang+ 21

Ideal case: AUROC=1

- Logistic Regression>0.65
- Random Forest>0.55
- Sequential Model in Keras>0.70
- XGBoost>0.75

# Grid-Searching (XGB)




- The performance is high for bright XAGN and IRAGN hosts.
  - F1: >0.70
  - AUROC: >0.80

# Type-1 vs. Type-2 AGNs (XGB)

- XAGNs
  - broad-line (type-1)
  - non-braord-line (type-2)
- IRAGNs:
  - unobscured (type-1)
  - obscured (type-2)

- F1: >0.70 (type-1) vs. >0.50 (type-2)
- AUROC: >0.95 (type-1) vs. >0.70 (type-2)

# Take-home Message

- HSC (optical) with WISE (near-infrared) data perform well to identify AGN hosts.
  - F1 >0.65 and AUROC > 0.75.

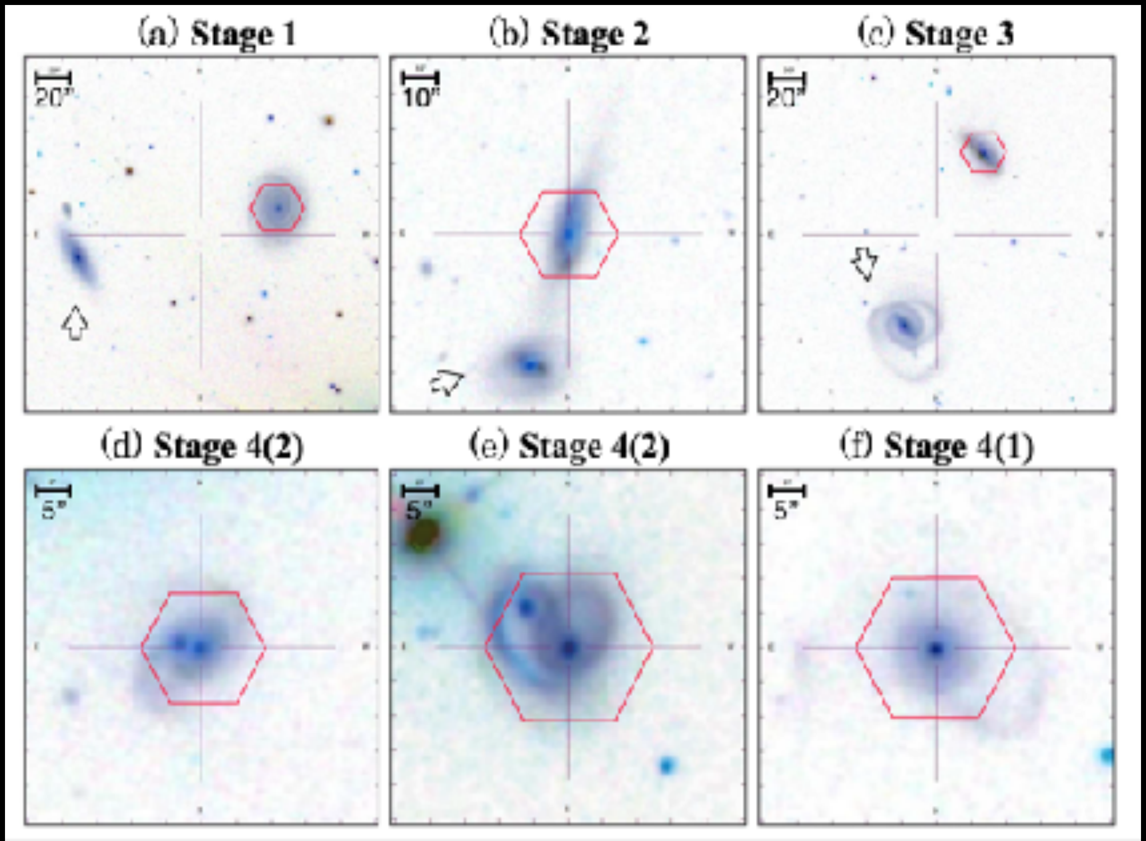
Chang+ 21

# Galaxies/AGNs by Machine Learning

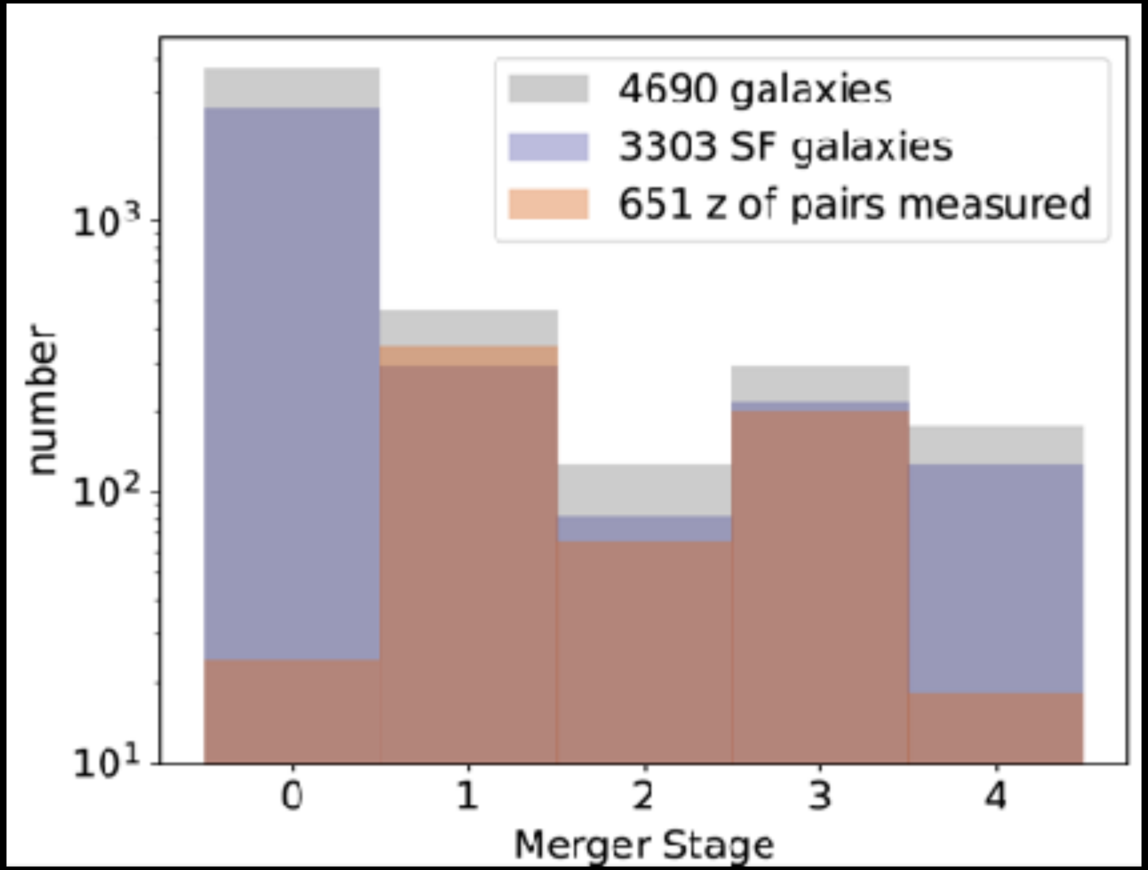
- Increase sample size
    - AGN identification
    - Other rare populations
- Investigate galaxy structures
    - Merger stage classification
    - Star-forming, quiescent galaxies, and others
- Derive physical parameters
    - Similar to SED fitting
    - Unsupervised learning

# Galaxy Merger Stages

- MaNGA (Mapping Nearby Galaxies at APO)
- Sample Size: N=4690 (3630/468/125/293/174)



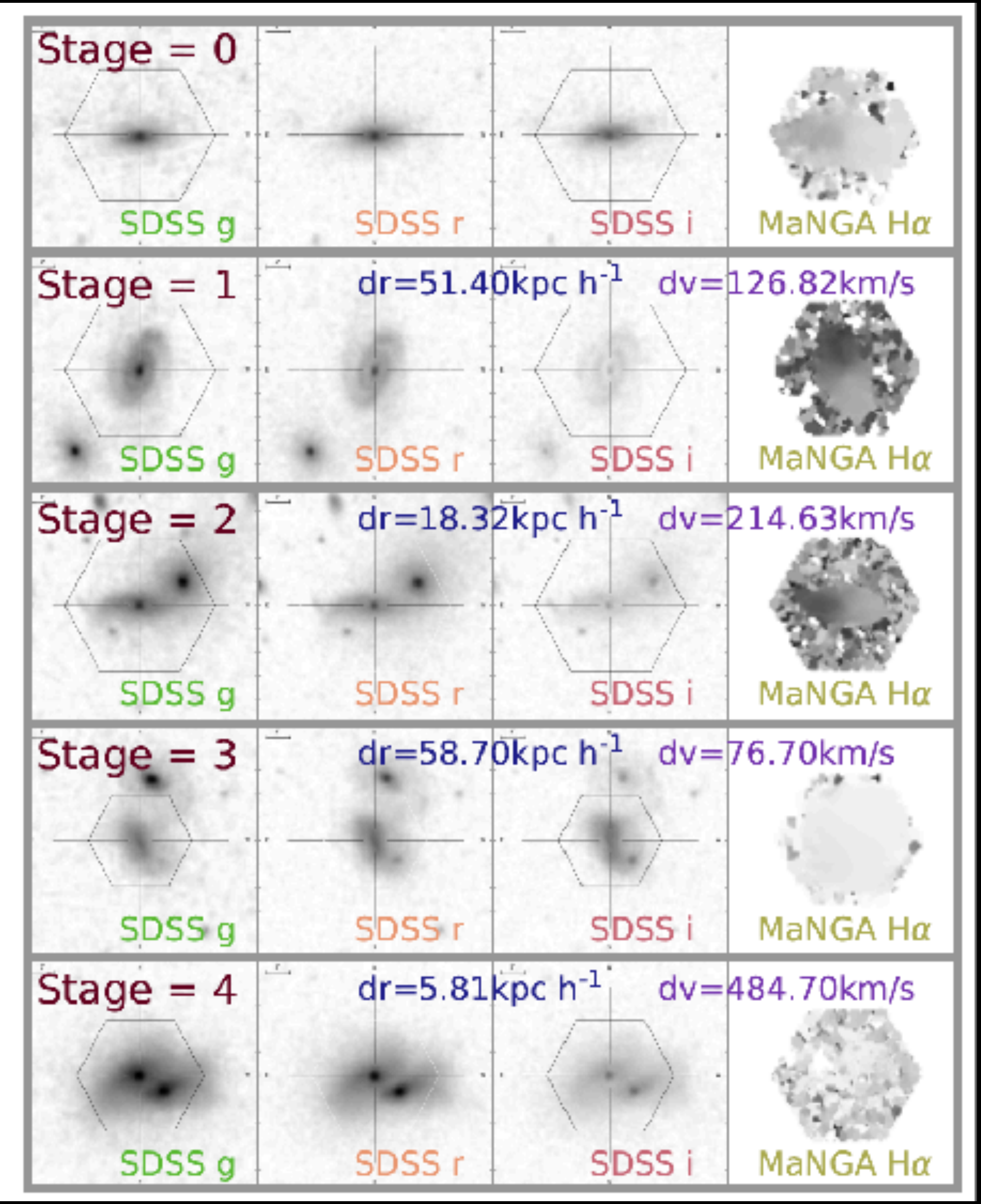

Pan+19



# Input Data

Chang+ 22

- catalog
  - dr [PROJSEP]: projected separation
  - dv [DELTAV]: line-of-sight velocity
- SDSS gri image
- MaNGA
  Ha velocity map

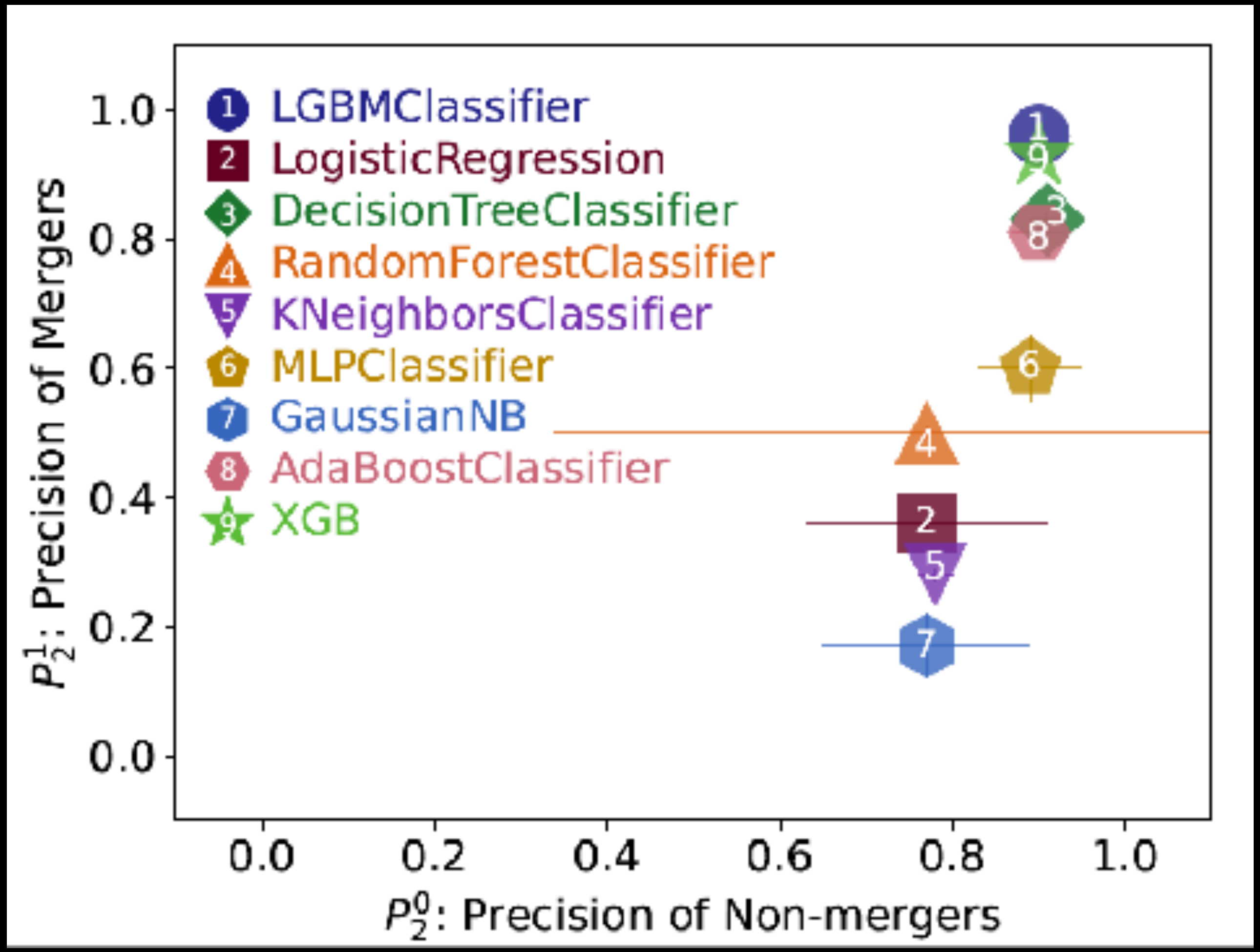


# Classifications and Classifiers

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \text{(purity)}$$

Ideal case: high

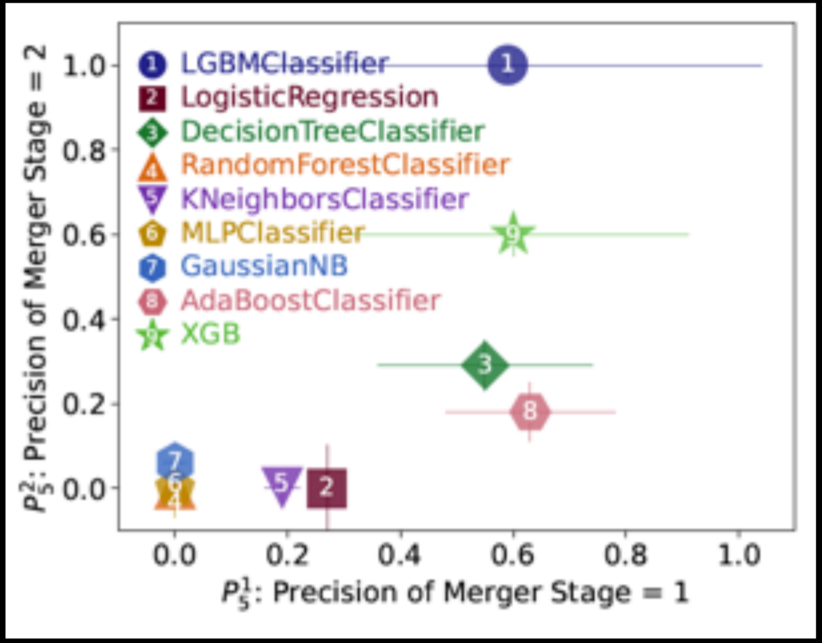
- 2-phase classification
  - Non-merger/Merger
  - Precision > 0.9
- 3-phase classification
  - (0)/(1+3)/(2+4)
  - Precision > 0.8
- 5-phase classification
  - 0/1/2/3/4
  - Precision > 0.6

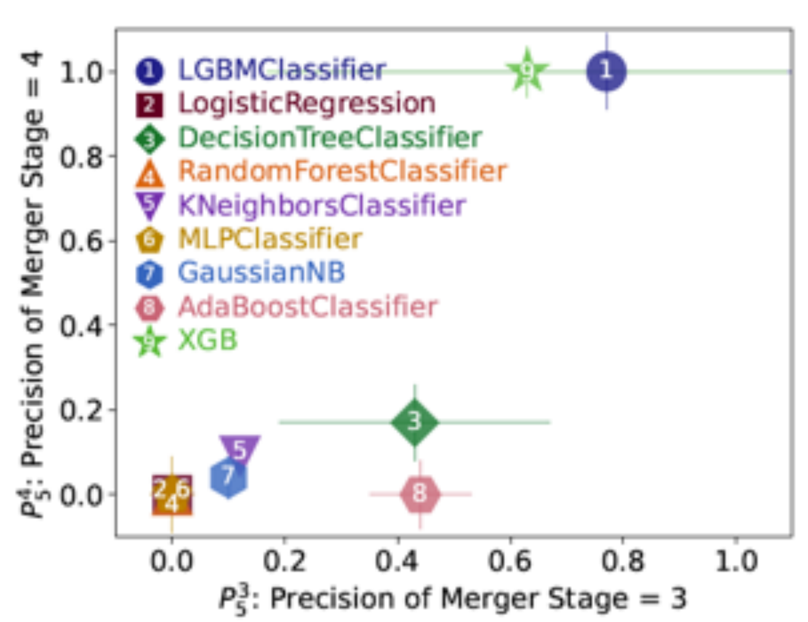



Chang+ 22

# Increase the Precision

- Sample Size: N=4690 (3630/468/125/293/174)
- Subsample:
  - SF subsample (N=3303)
  - Precision +5~10%
- Data process:
  - original+rotate X2 (N=9380)
  - 5-phase classification: precision>0.85

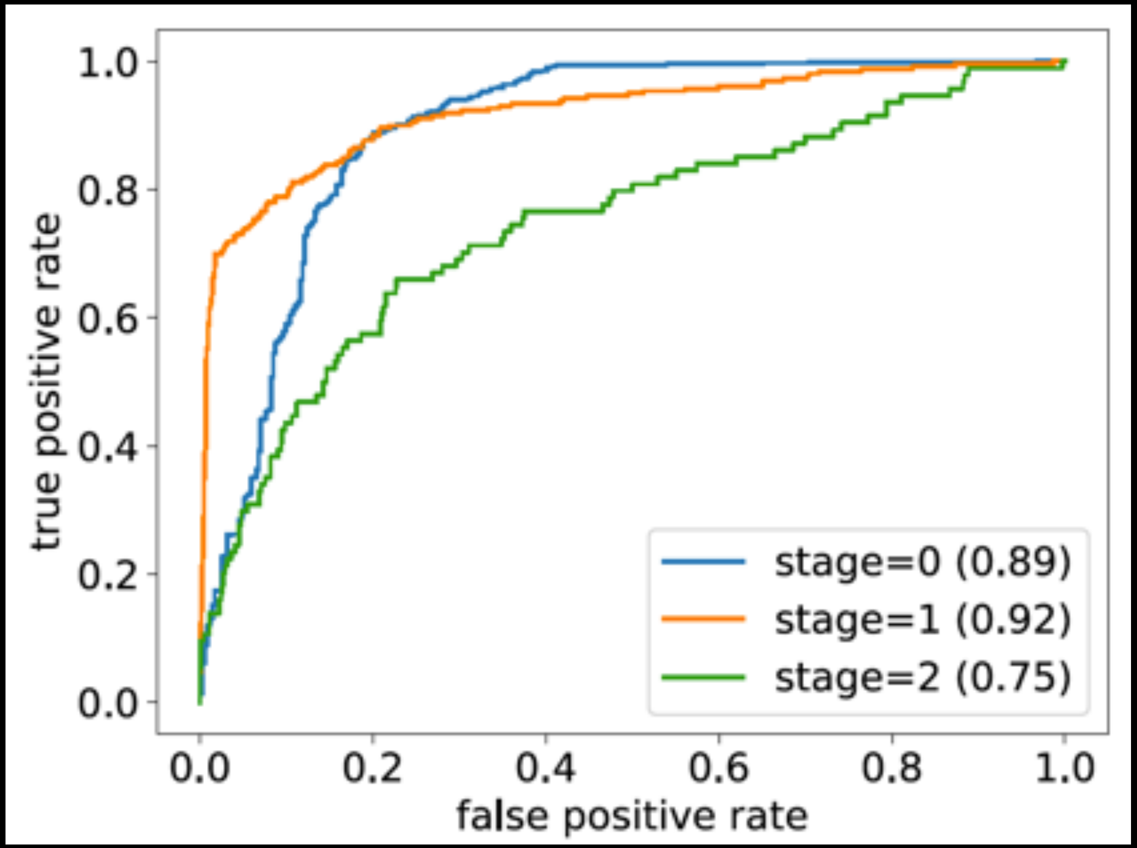


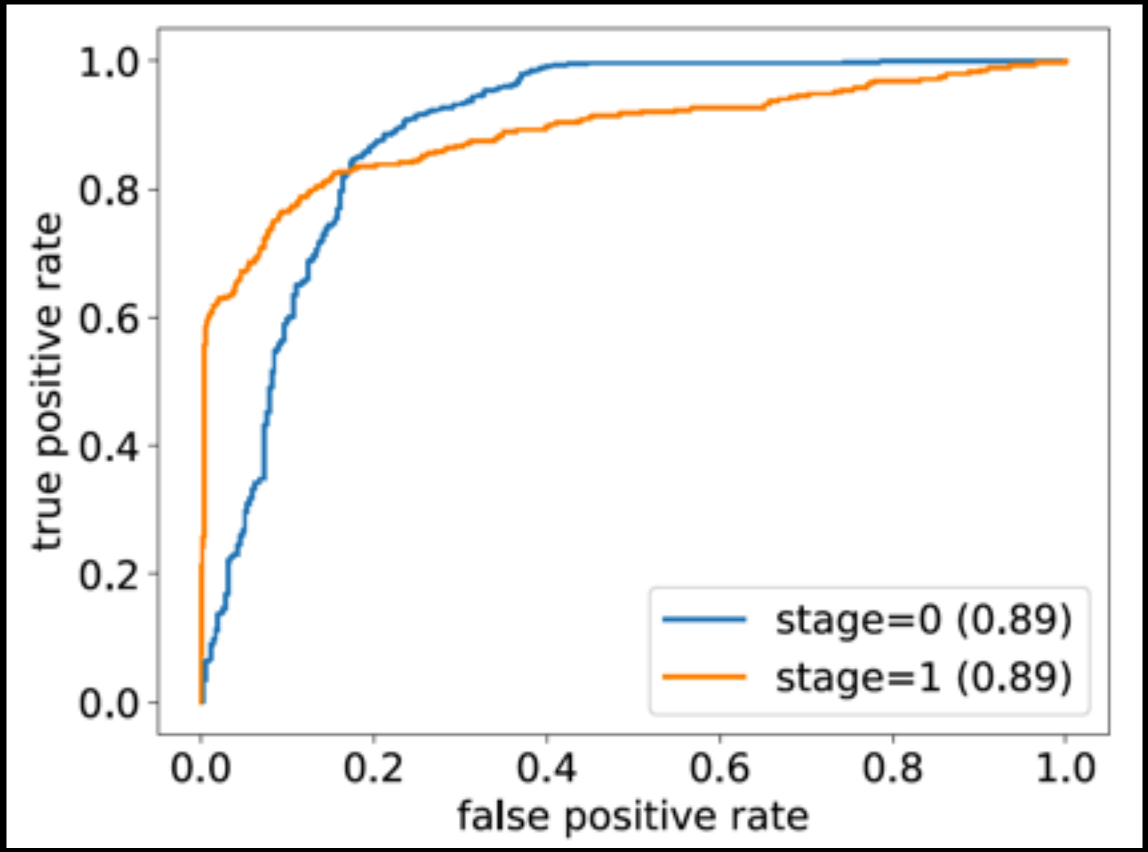

Chang+ 22

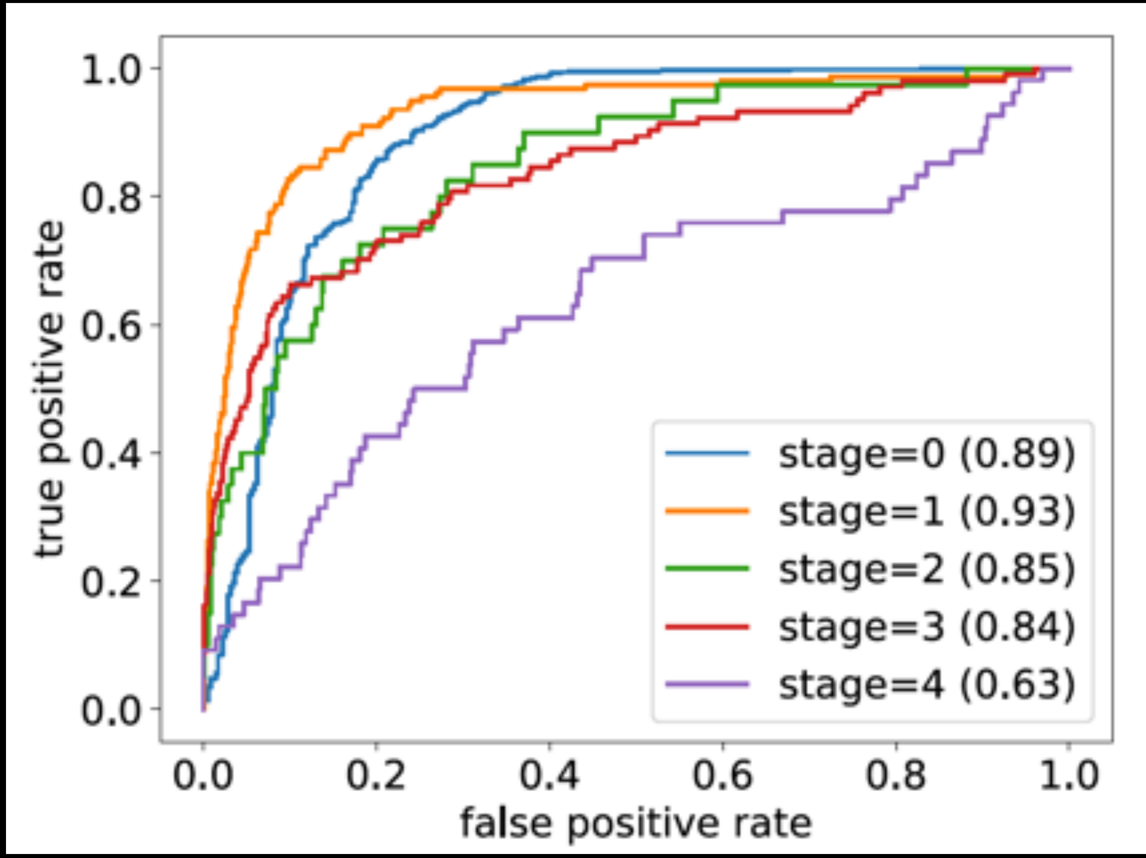
# AUROC scores

3-phase




2-phase

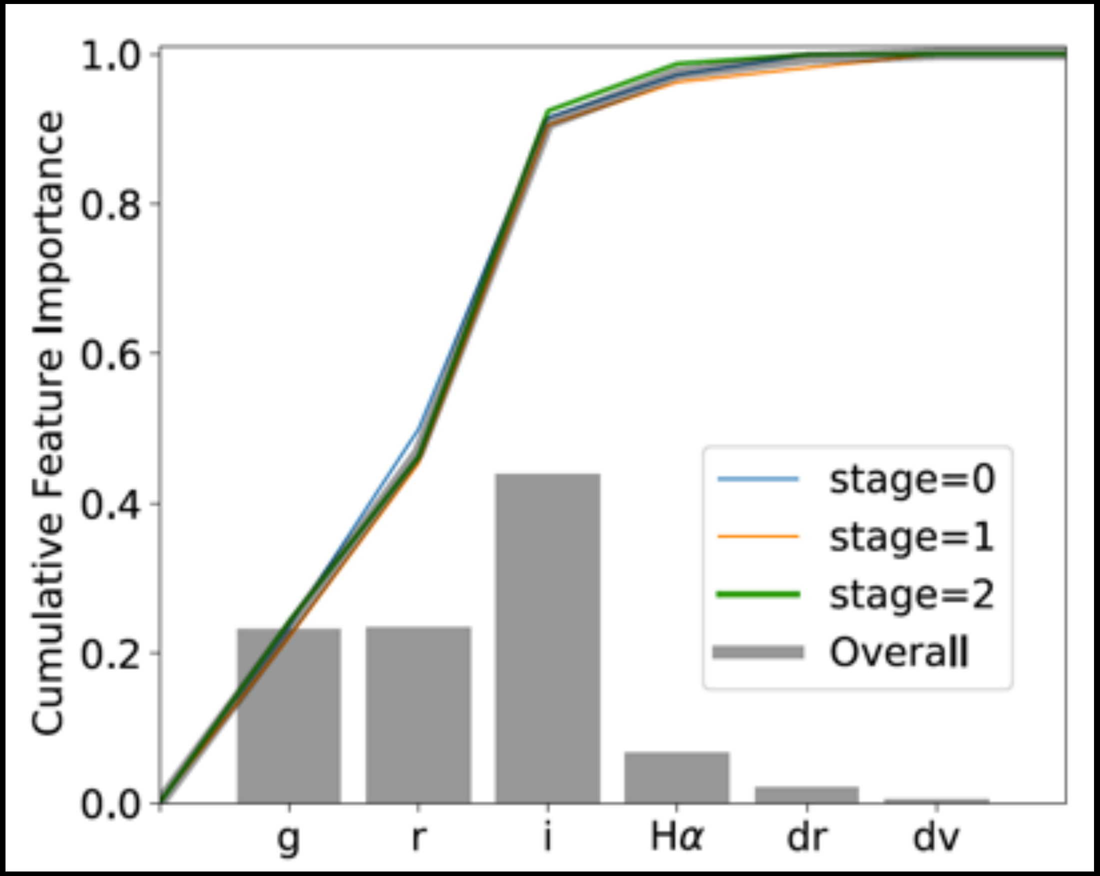


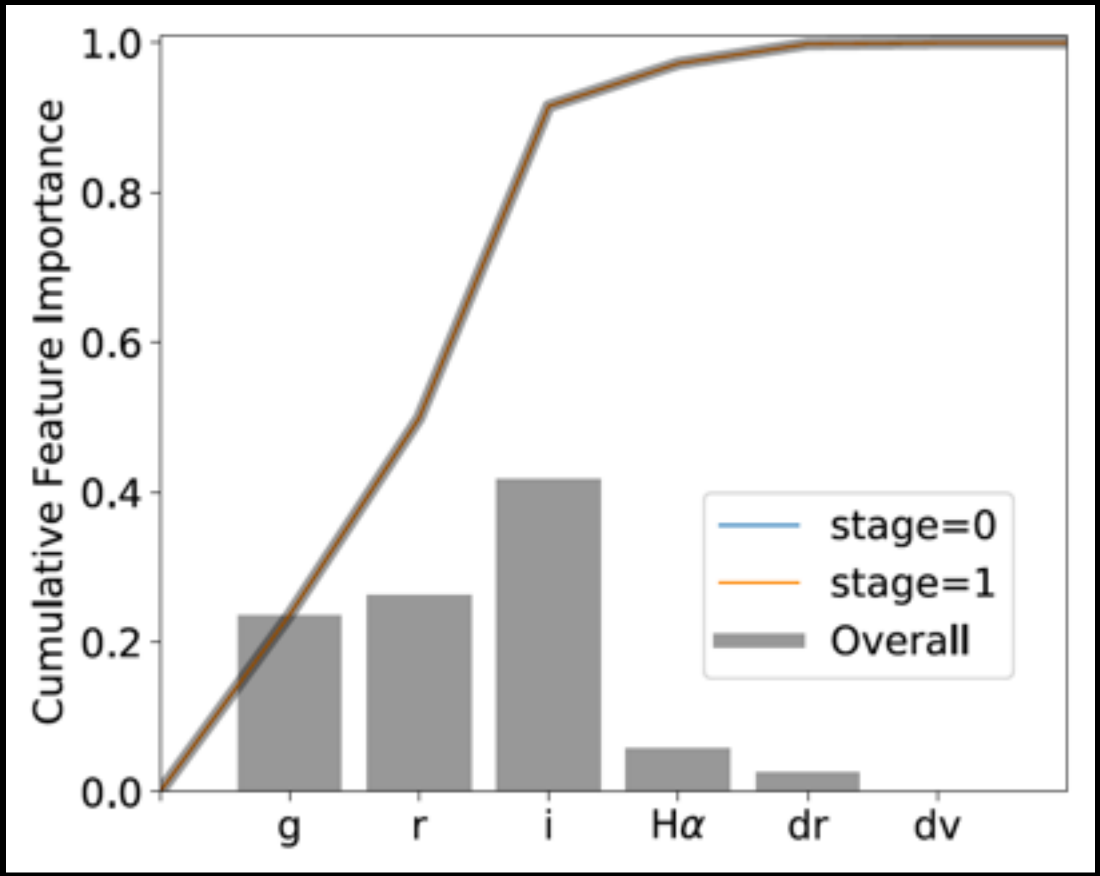

5-phase

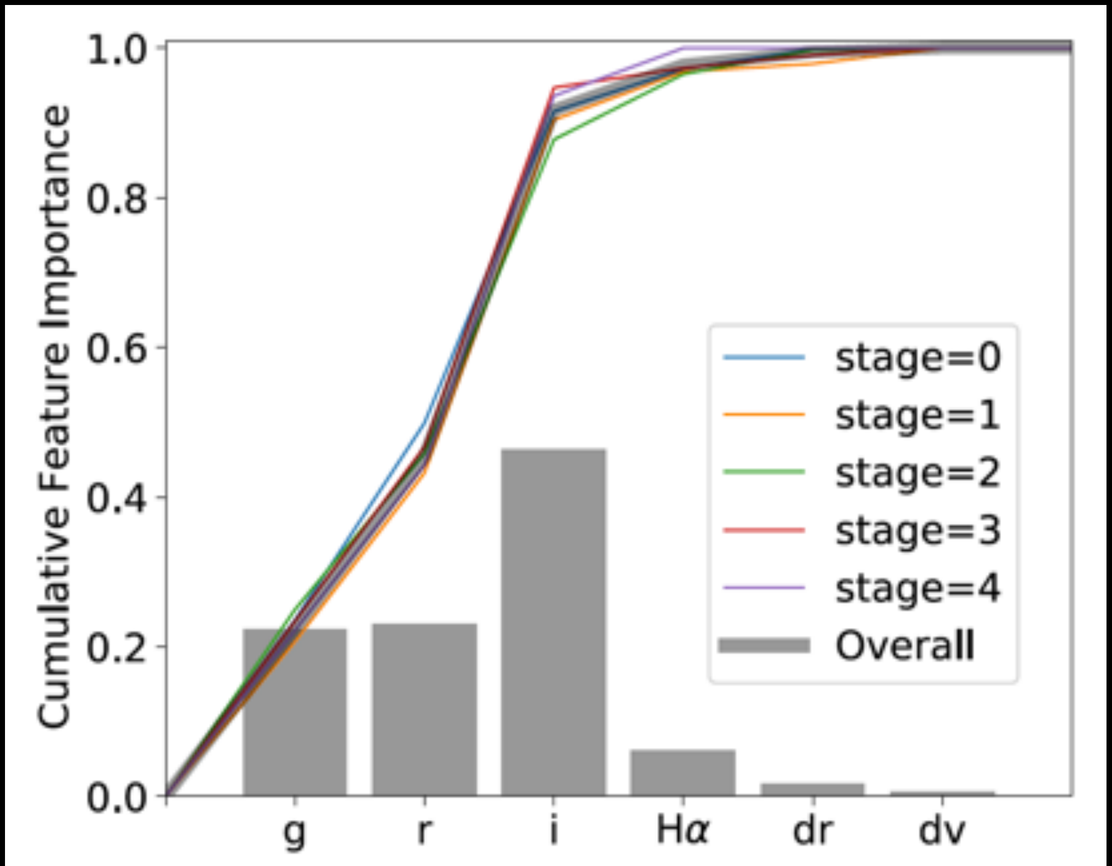
# Feature Importance

3-phase




2-phase

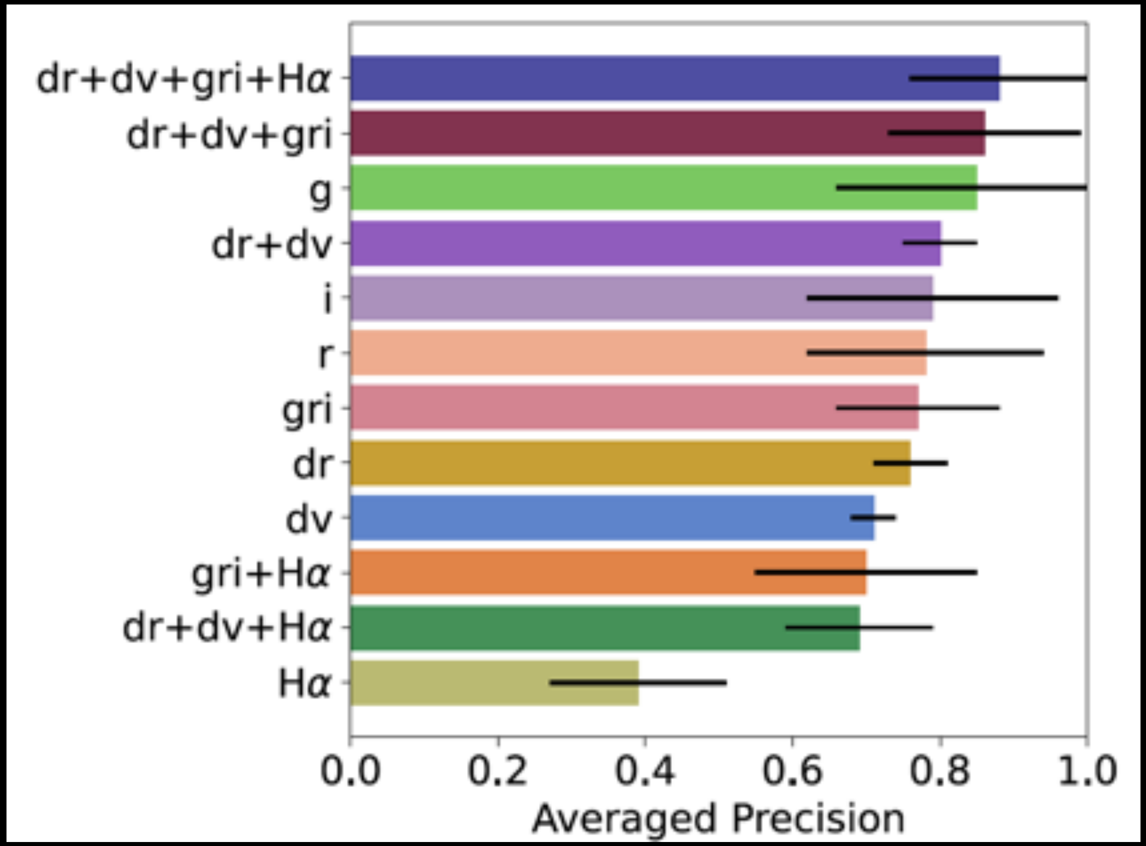


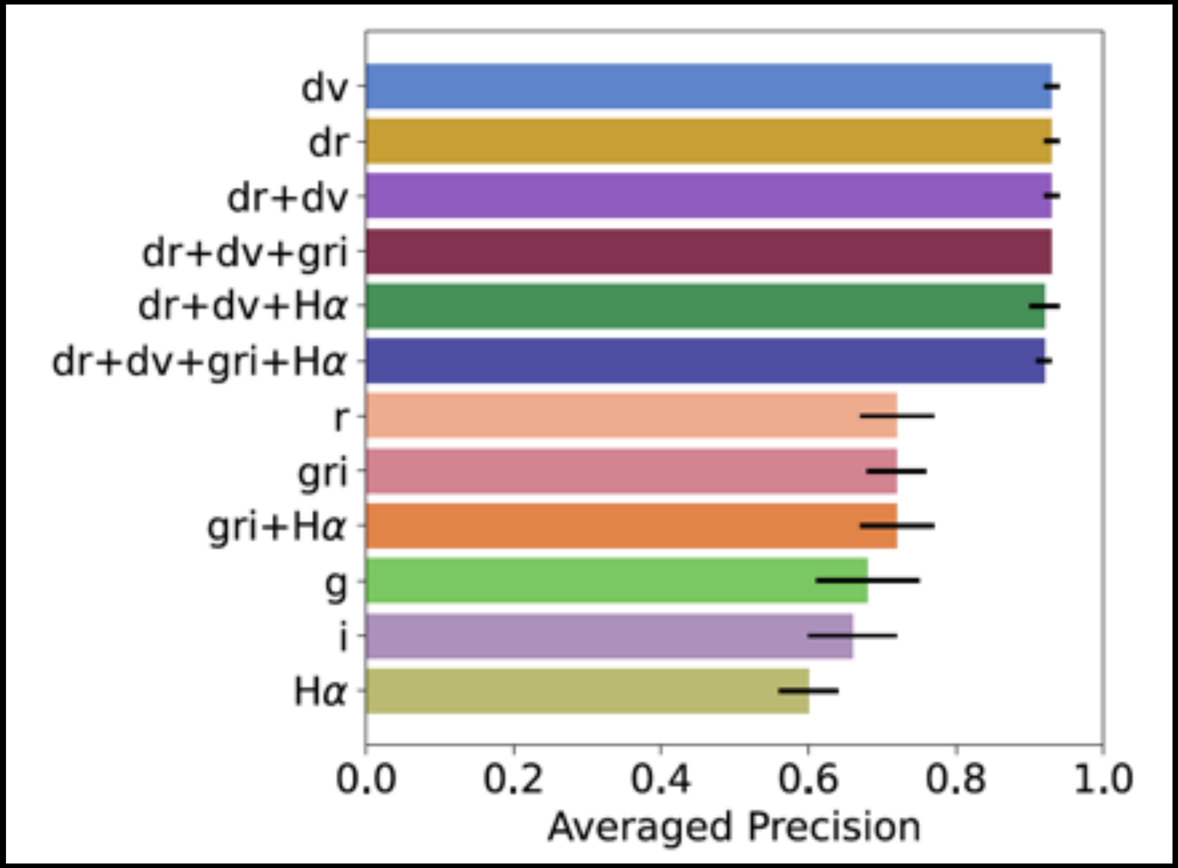

5-phase

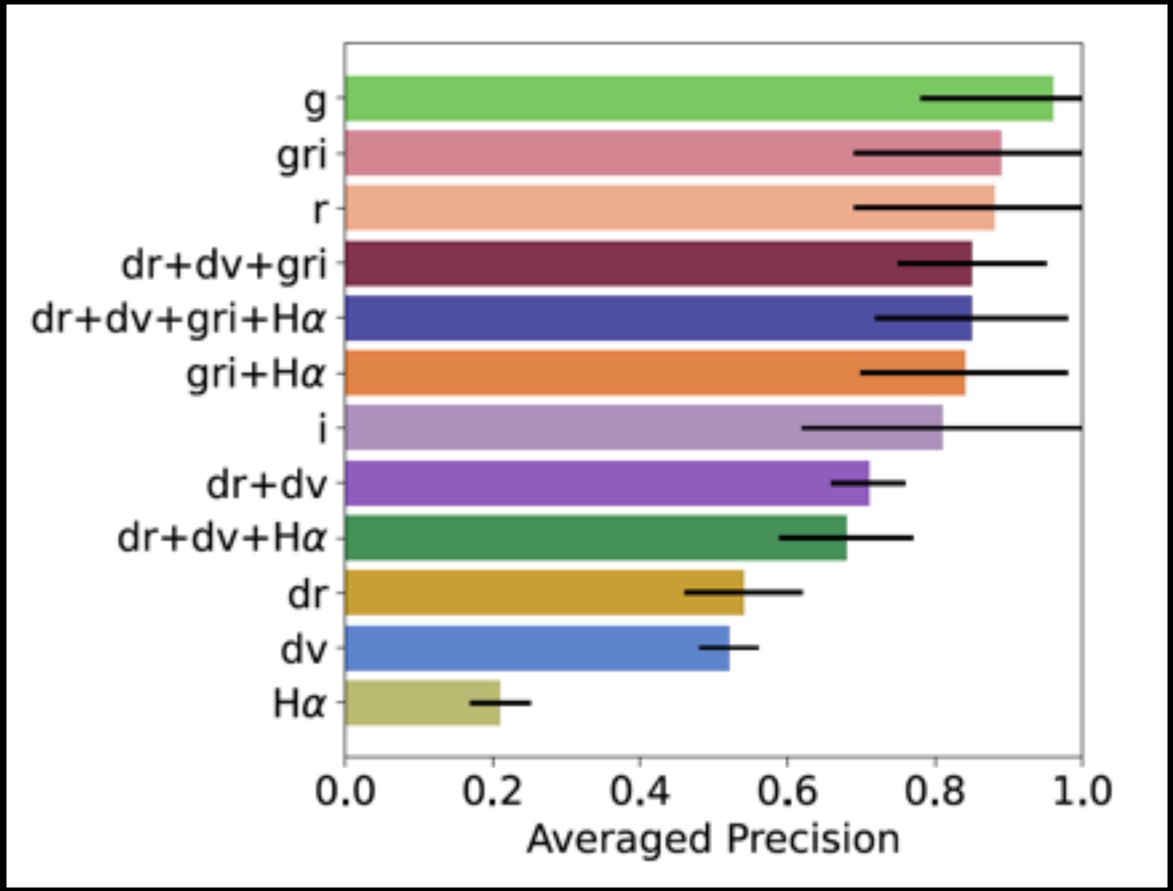
# Feature Combinations

3-phase




2-phase




5-phase

# Take-home Message

- HSC (optical) with WISE (near-infrared) data perform well to identify AGN hosts. Chang+ 21
  - F1 >0.65 and AUROC > 0.75.

- Non-merger/Merger classification: P>0.90. Chang+ 22
- Sample size is increased by rotation. Then the 5-phase classification can be good (P>0.85).

- ML results can apply to future all-sky surveys.

Thank you!

# Galaxies/AGNs by Machine Learning

- Increase sample size
  - AGN identification
  - Other rare populations
- Investigate galaxy structures
  - Merger stage classification
  - Star-forming, quiescent galaxies, and others
- Derive physical parameters
  - Similar to SED fitting
  - Unsupervised learning