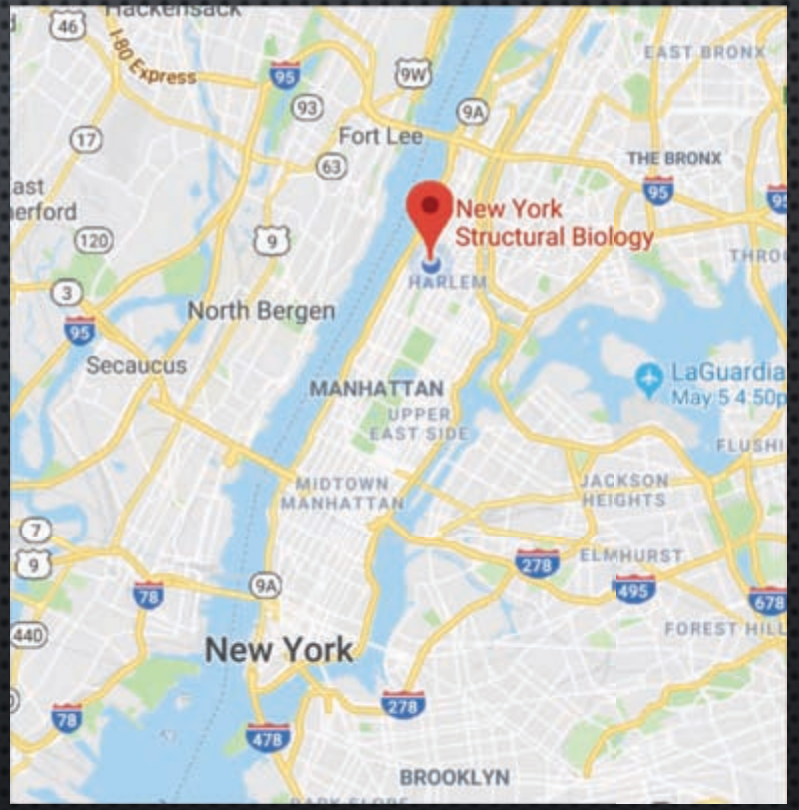
# STRATEGIES FOR PRODUCTIVE LARGE SCALE DATA COLLECTION

Laura Yen
Simons Electron Microscopy Center
New York Structural Biology Center
5/9/2019

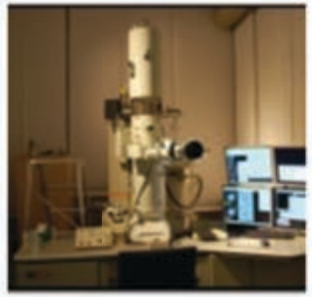
# NEW YORK STRUCTURAL BIOLOGY CENTER



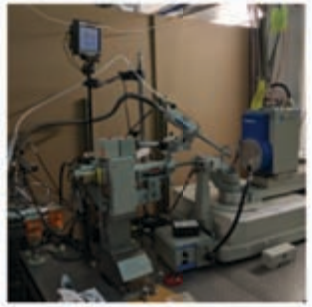
# NEW YORK STRUCTURAL BIOLOGY CENTER

NMR

X-Ray



Simons Electron Microscopy Center



Special Projects

- NYSBC is composed of 4 groups
- EM Group:
  1. NRAMM
  2. SEMC
  3. NCCAT
- 9 member institutions
- >150 users

THE ROCKEFELLER UNIVERSITY

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Memorial Sloan Kettering Cancer Center

The City University of New York

Icahn School of Medicine at Mount Sinai

NEW YORK UNIVERSITY

EINSTEIN Albert Einstein College of Medicine

Weill Cornell Medical College

Wadsworth Center NEW YORK STATE DEPARTMENT OF HEALTH

# SEMC INSTRUMENTS

Titan Krios#1
Falcon3
K2

Titan Krios#2
Falcon3
Energy Filter + K2
Cs-Corrector

Titan Krios#3
Falcon3
Energy Filter + K2

FEI Helios 650
Quorum cryostage

FEI Tecnai F20
DE20 + TVIPS CMOS

FEI Tecnai T12
TVIPS CMOS

JEOL 1230
Gatan US4000 CCD

# UPCOMING: NATIONAL CENTER FOR CRYOEM ACCESS AND TRAINING (NCCAT)

- In construction now!
- Mission:
  - Provide nationwide access to advanced cryoEM technical capabilities
  - Assist users in the development of cryoEM skills needed for independent research.
- Expansion:
  - 5000 sq ft construction
  - 3 more Titan Krios coming fall 2019!
  - Screening TEMs TBD

## BOTTOM LINE:

WE HAVE A LOT OF SCOPES AND A LARGE USER BASE WITH A VARIETY OF RESEARCH NEEDS.

WE NEED TO PROVIDE USERS WITH A DATA COLLECTION + ANALYSIS WORKFLOW THAT IS EASY TO USE, EFFICIENT, AND GIVES THEM THE METADATA THEY NEED TO COLLECT THE BEST DATA POSSIBLE!

# Overview

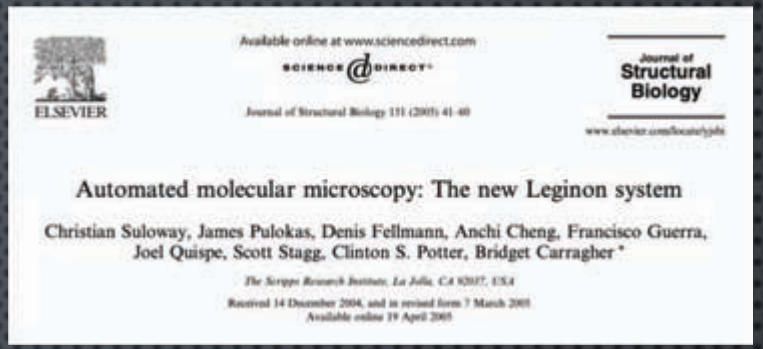
- How do we collect cryo-EM data at the SEMC?
  - Overview of Leginon + appion
- How do we optimize data collection?
  - TONS of metadata and Live feedback
  - Hardware beam tilt correction for large image shift
  - Leginon-Slack integration

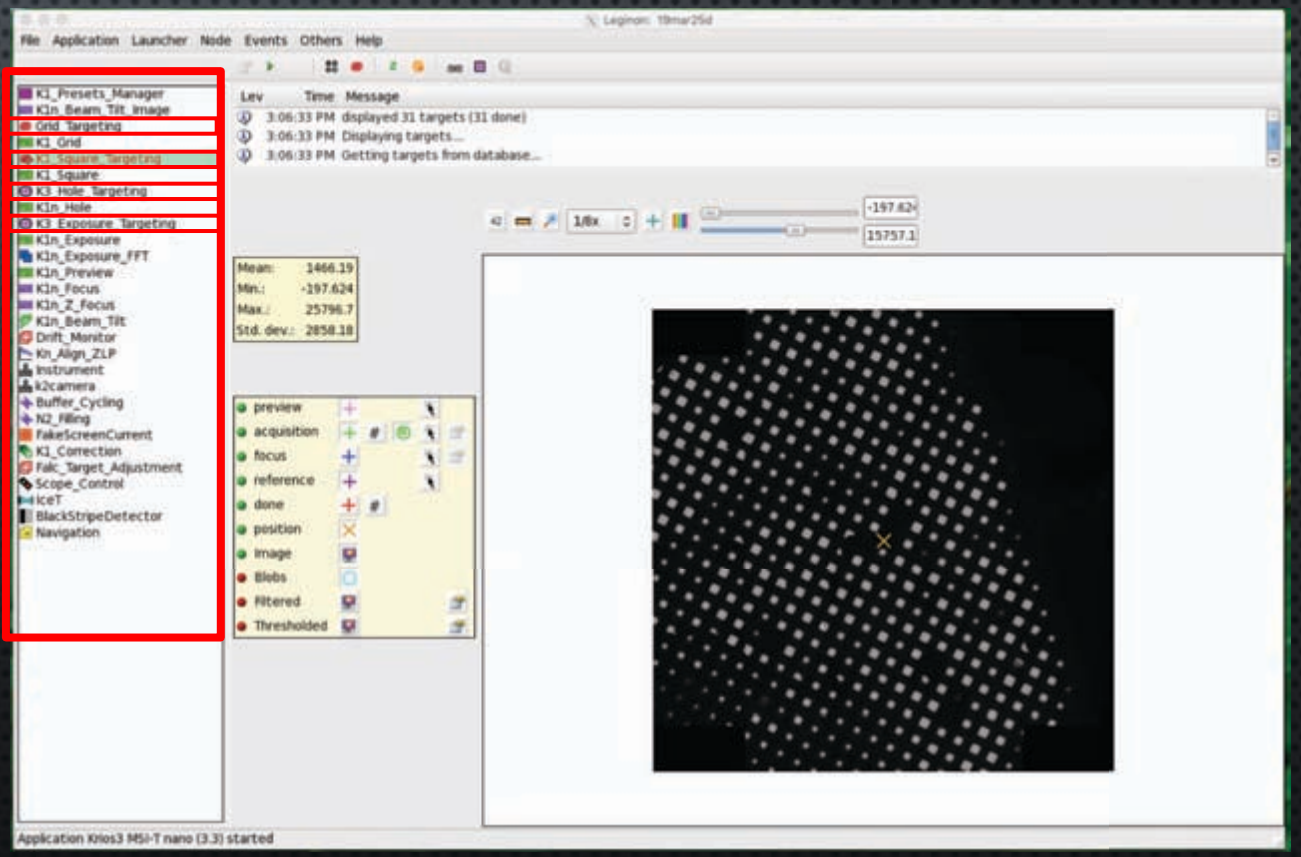
# Pipelines for data collection and on-the-fly processing

- At the NYSBC we use **Leginon** for data collection and **Appion** for on-the-fly processing
- Others pipelines available for data collection:
  - SerialEM (Mastronarde lab, University of Colorado)
  - EPU (Thermo Fisher Scientific)
- Other pipelines available for on-the-fly processing
  - Scipion (Carazo lab, Universidad Autónoma de Madrid)
  - Warp (Cramer lab, Max Planck institute)
  - Sphire (Penczek lab, UTH and Raunser lab, Max Planck institute)

# LEGINON

- LEGINON - AUTOMATED DATA COLLECTION ON TEM
- DEVELOPED IN THE EARLY 2000'S AT THE SCRIPPS RESEARCH INSTITUTE
  - BRIDGET CARRAGHER AND CLINT POTTER
- LEGINON = NO NIGEL

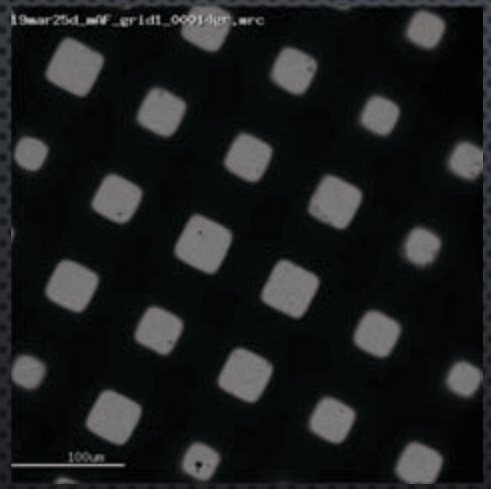



Organized by nodes

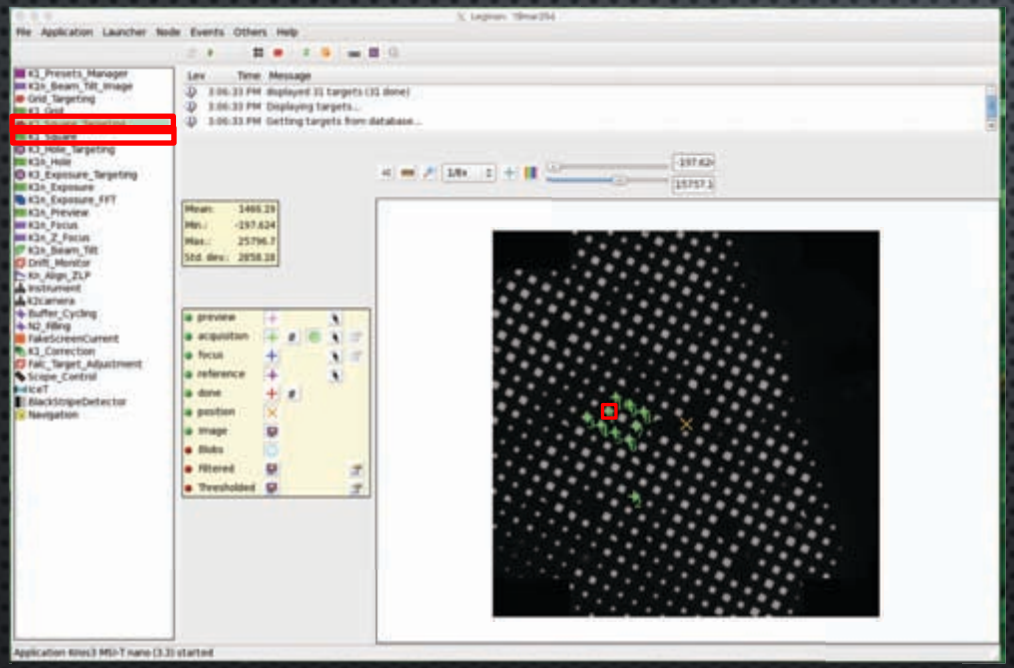
Leginon MSI application → 4 nodes/magnifications where most of the automation happens

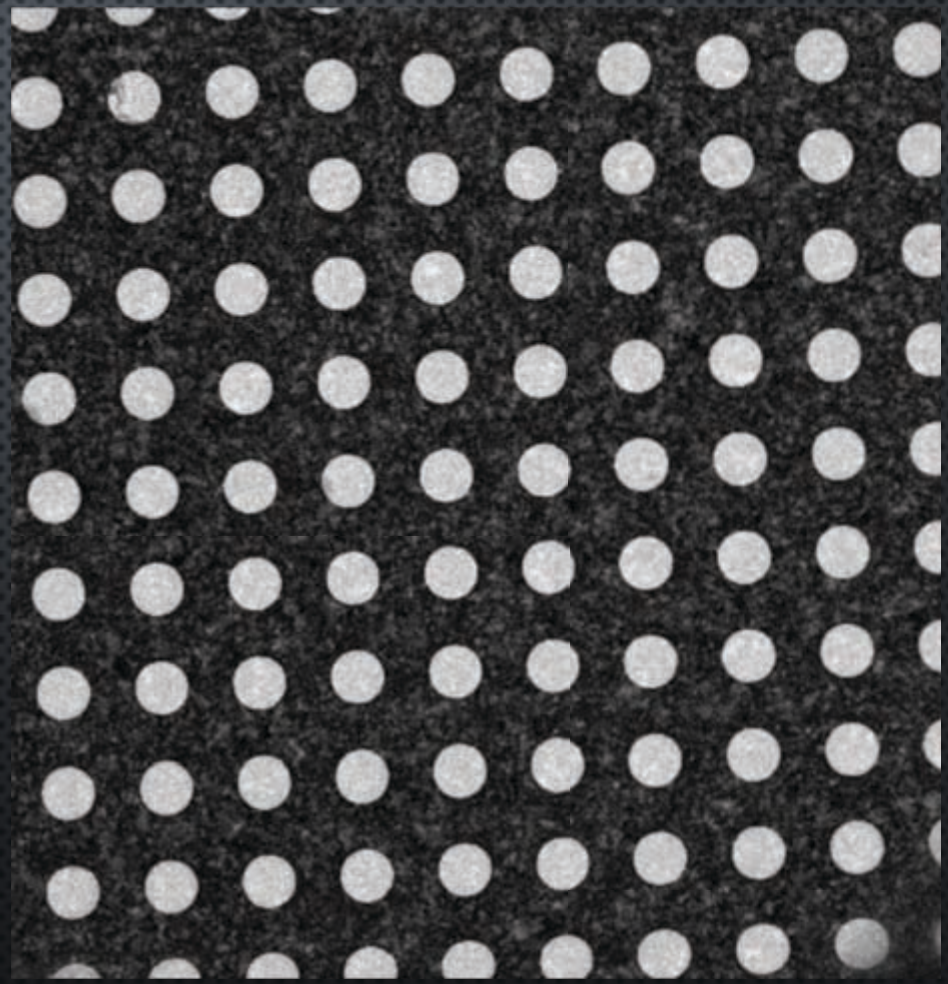
# LEGINON: THE ATLAS

X 23 →

# LEGINON: TARGET YOUR SQUARE OF INTEREST

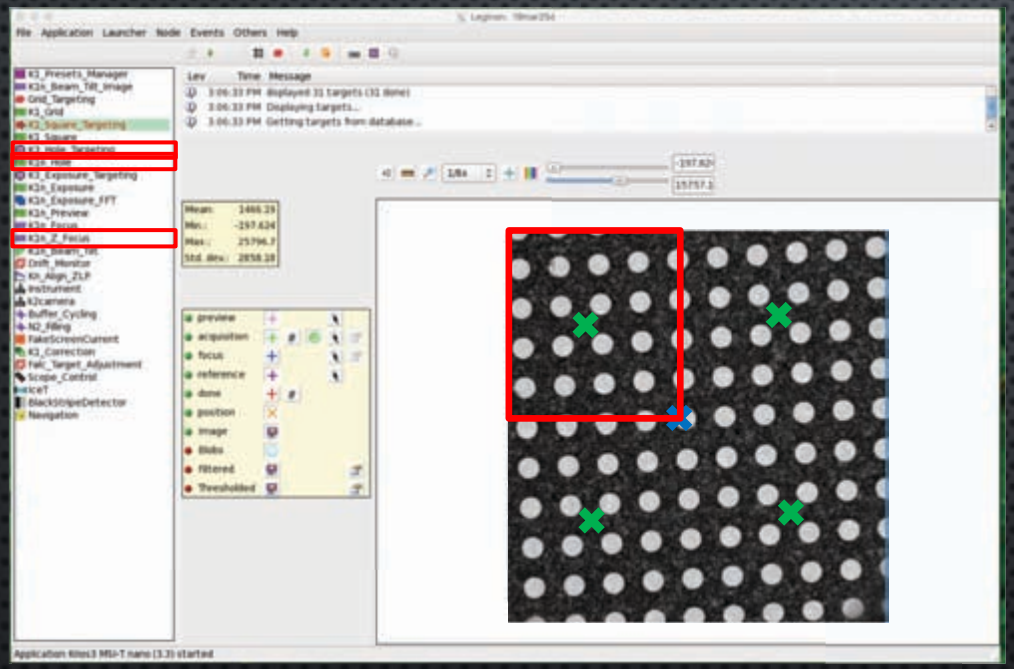


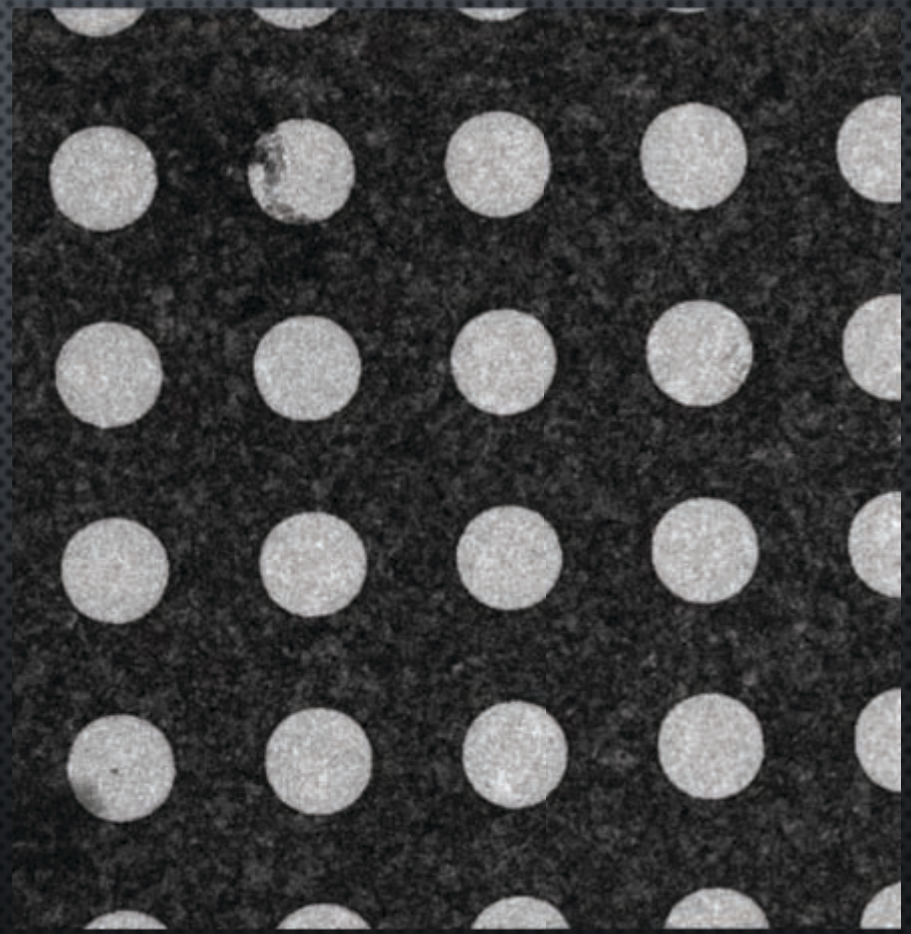


Square sequence:

1) Collect square targets 1 - 10 by stage movement

# LEGINON: CONFIRM YOUR HOLES OF INTEREST

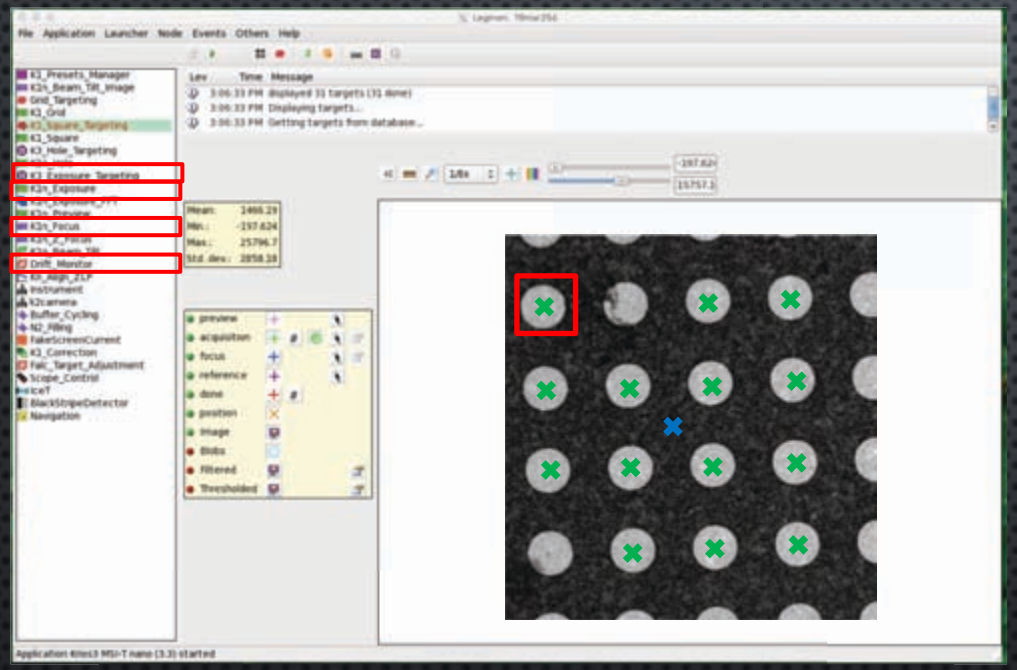


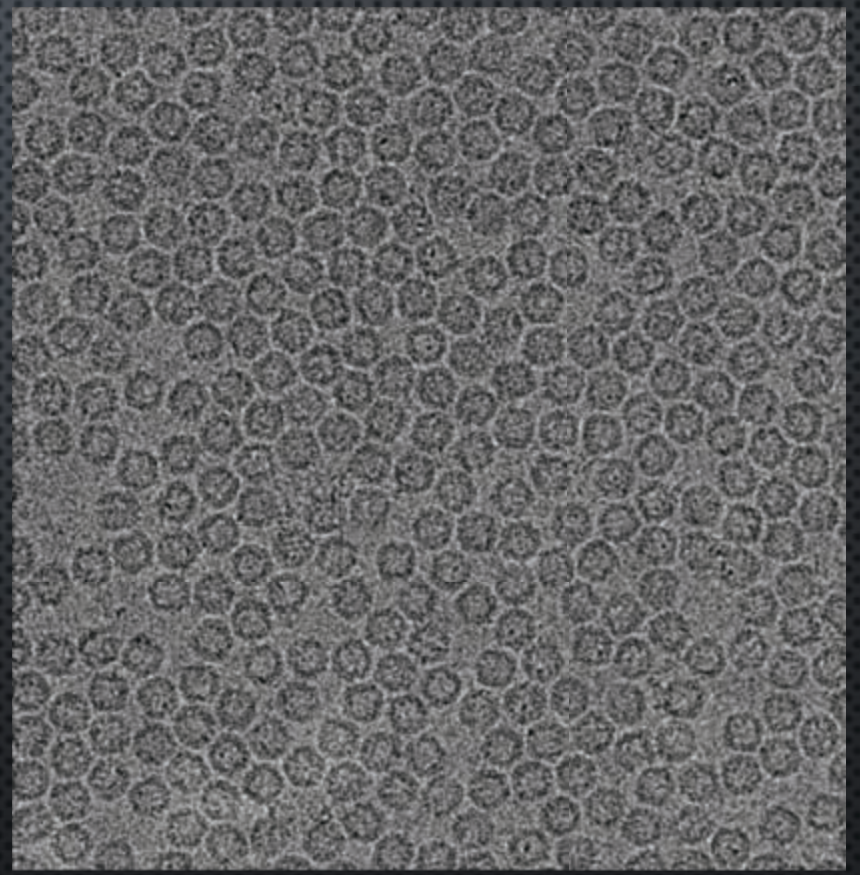


Hole sequence:

1) Eucentric height
2) Collect hole target 1 (square 1) by stage movement

# LEGINON: CONFIRM YOUR EXPOSURE TARGETS





Exposure sequence:

1) Check Drift_Monitor
2) Check Focus
3) Collect exposure targets 1 – 14 (hole 1, square 1) by beam-image shift

# LEGINON: THE WORKFLOW


Leginon has now finished collecting 14 exposures within hole 1 in square 1.

Next it will take exposures from hole 2 of square 1, and so forth.

This will continue until all 10 squares have been collected.

Automation is setup to take 10 squares, 4 holes/square, ~16 exposures/hole. This is enough for up to 640 high magnification images, which only took ~30 minutes to queue up. These are enough targets to collect for the next  6 – 8 hours.

# APPION

- Pipeline for processing EM images
- Wrapper for existing EM processing packages
- Convenient web-based viewer
  - Emgweb.nysbc.org
- Track and record pre-processing results
  - Frame alignment (motioncorr2)
  - CTF estimation (CTFFINDv4)
  - Particle picking
- Concurrent with data collection enabling quick analysis of data quality
- Used primarily for up to 2D analysis


Journal of Structural Biology

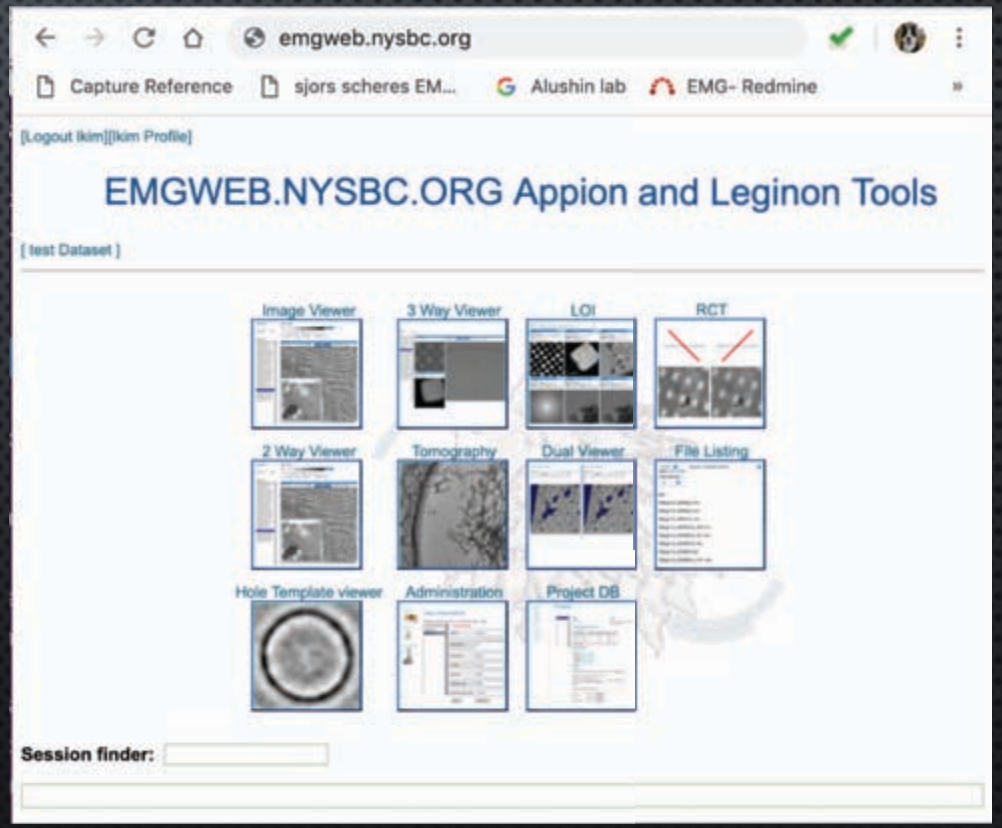
Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

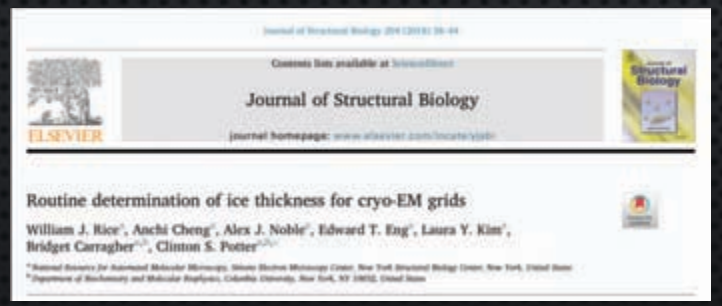
Appion: An integrated, database-driven pipeline to facilitate EM image processing

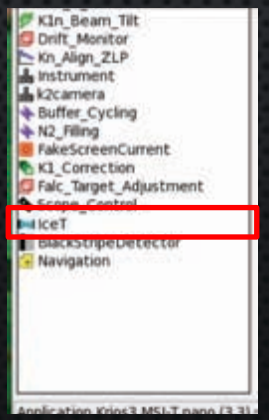
Gabriel C. Lander, Scott M. Stagg, Neil R. Voss, Anchi Cheng, Denis Fellmann, James Pulokas, Craig Yoshioka, Christopher Irving, Anke Mulder, Pick-Wei Lau, Dmitry Lyumkis, Clinton S. Potter, Bridget Carragher



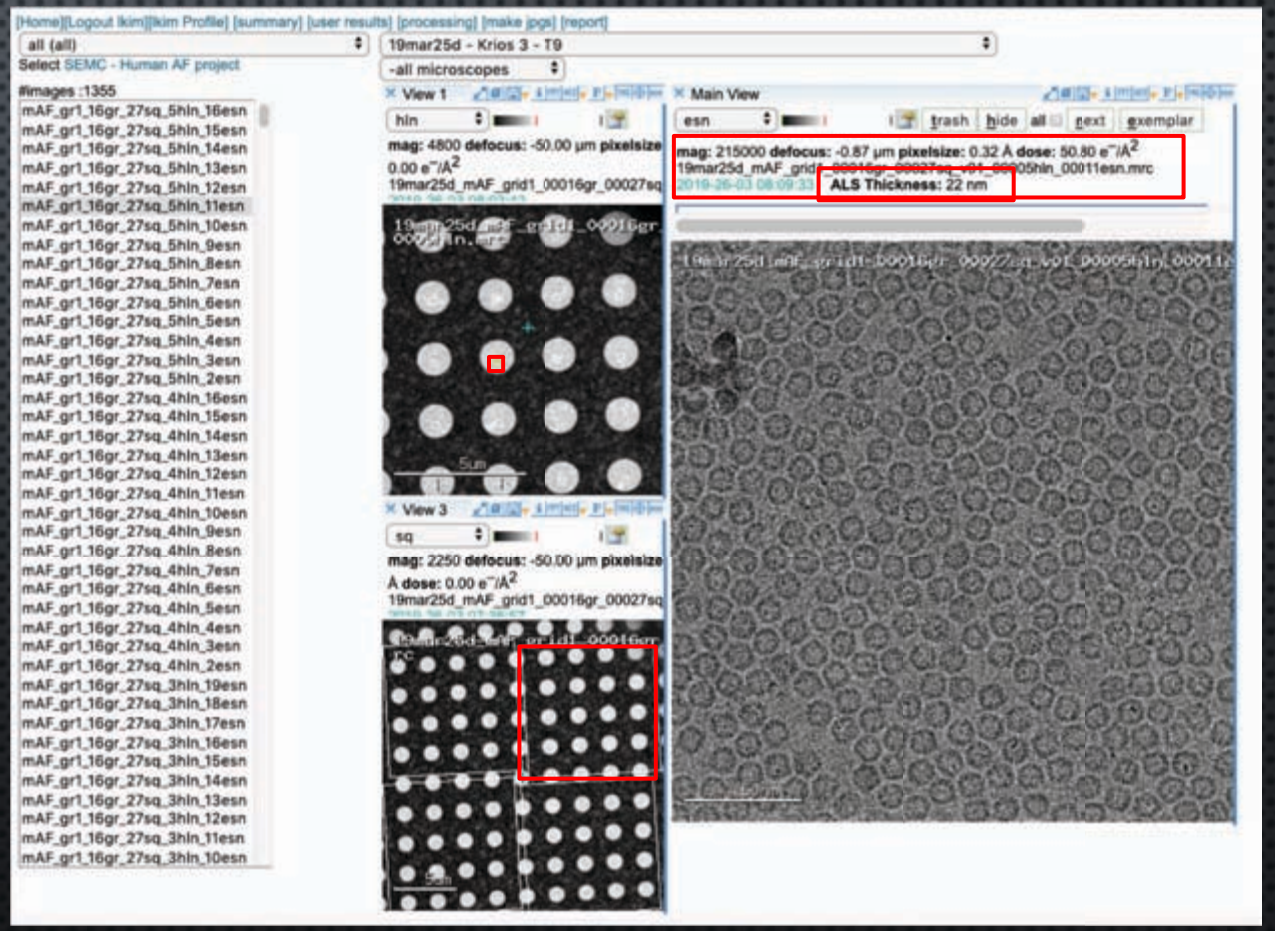
# APPION FEATURES

- Multi-scale image viewer
- Tons of metadata
- Ice thickness
  - Comparison of electron scattering intensity of sample versus reference vacuum intensity

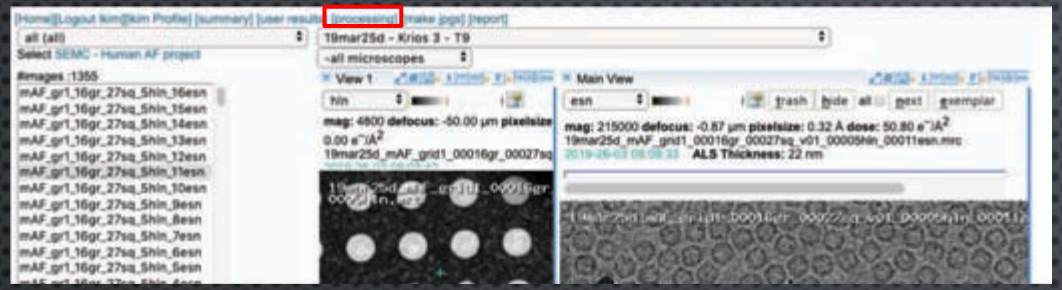


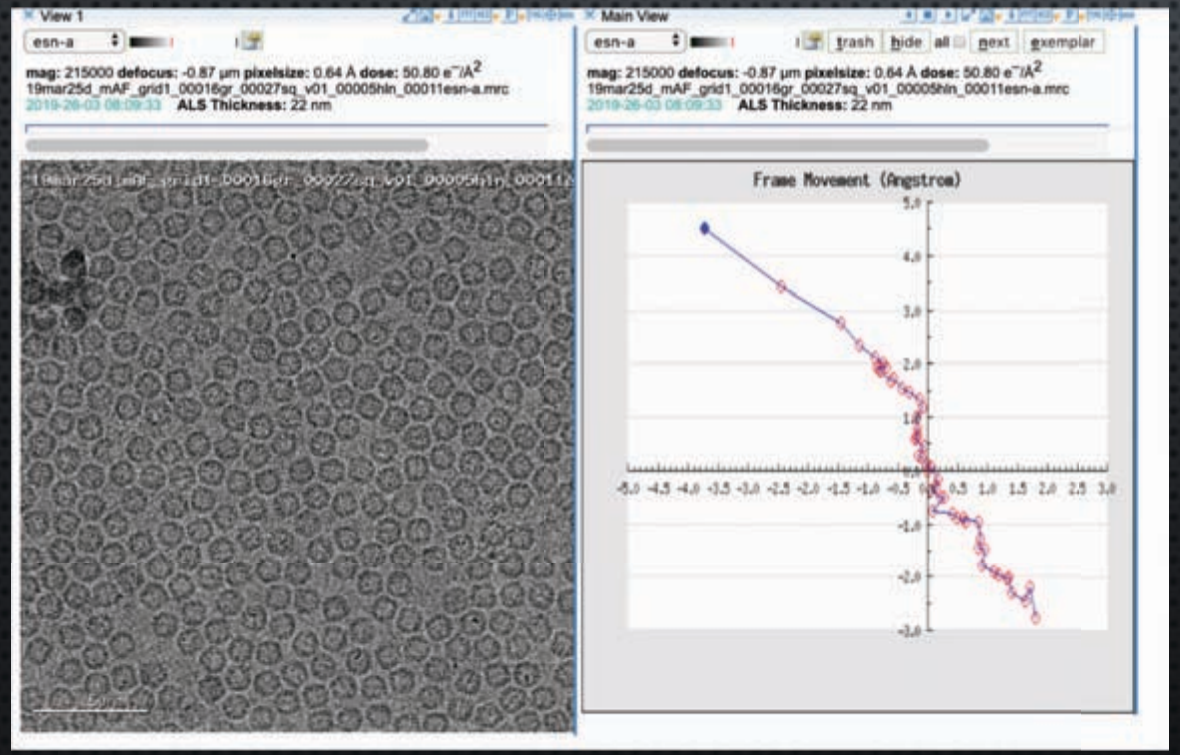




3 Way Viewer

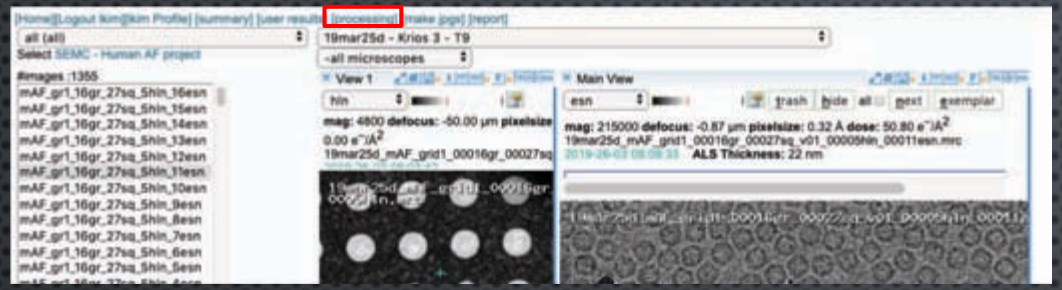


# APPION: FRAME ALIGNMENT

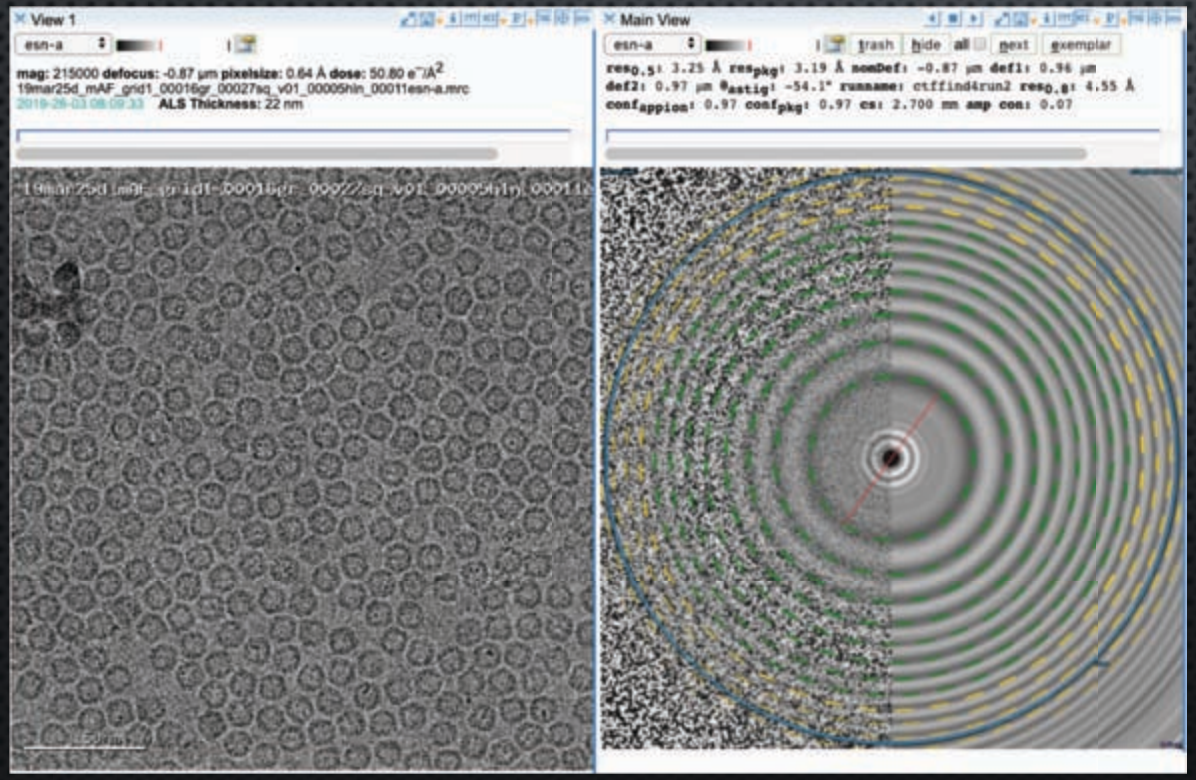


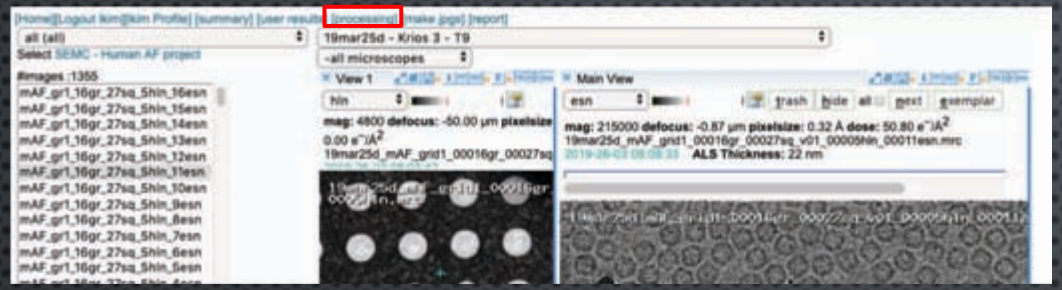

# APPION: CTF ESTIMATION

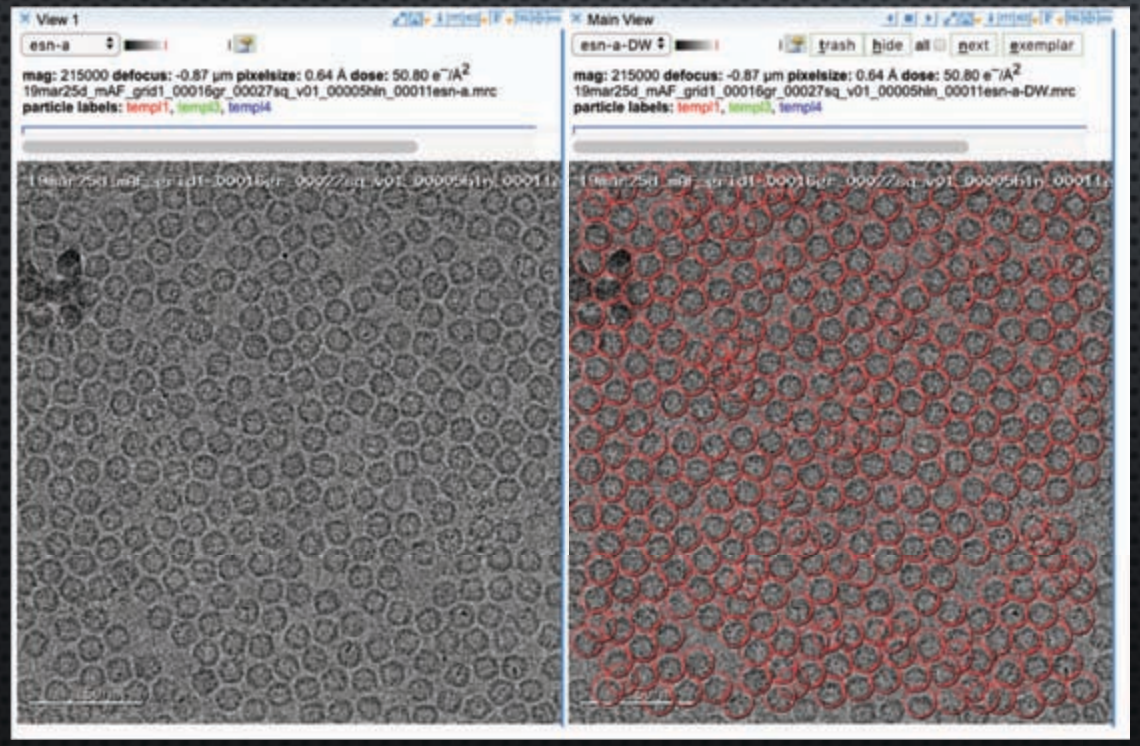






# APPION: PARTICLE PICKING

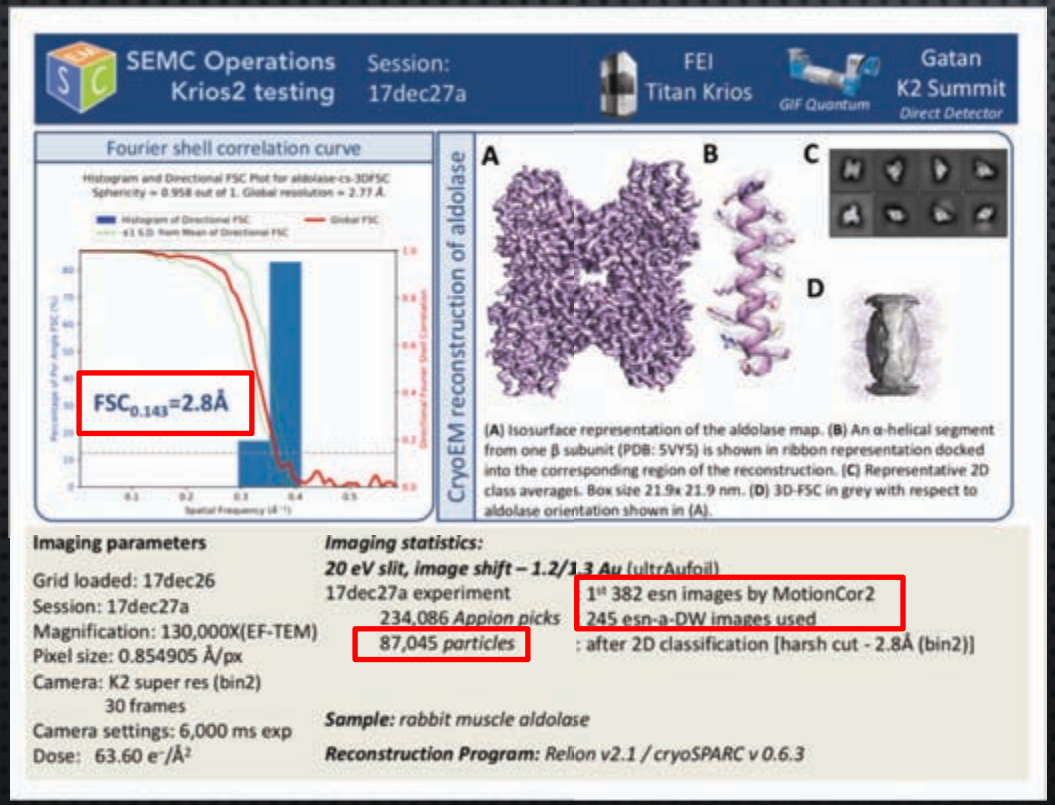


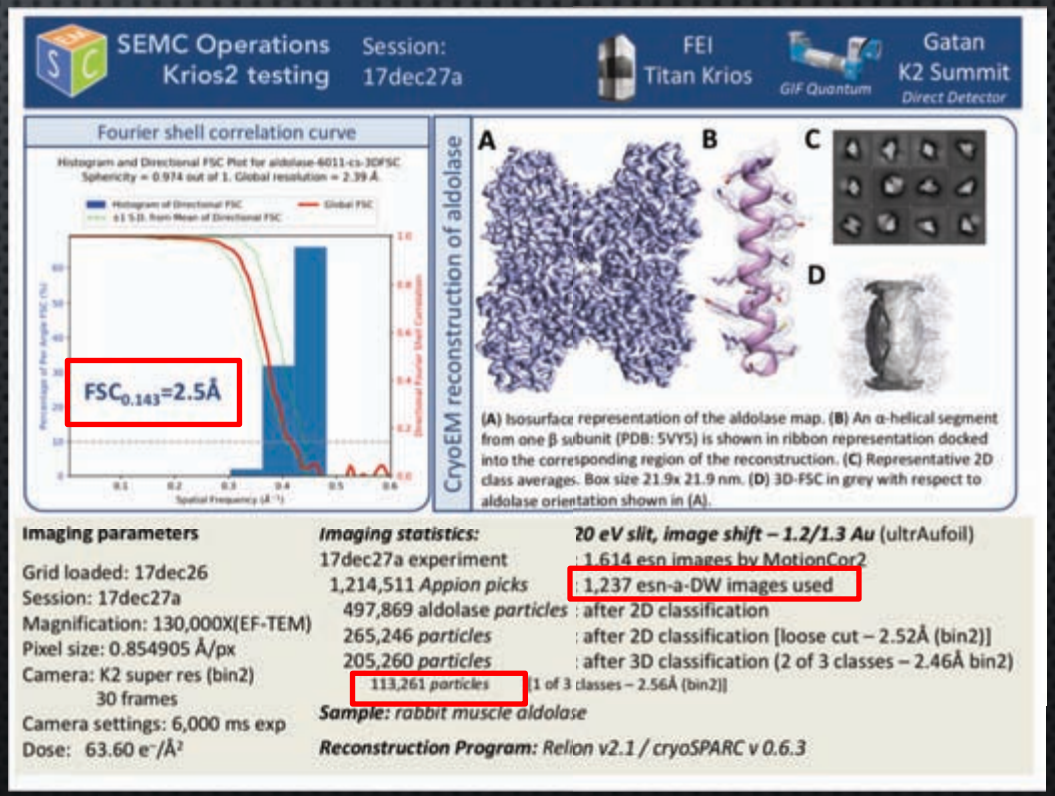





# POST PROCESSING ON-THE-FLY

- After a few hours of data collection we can start 3D analysis
- After ~4 hours of data collection, we can get a <3 Å result within 12 hours
  - Using cryosparc and relion3
- After the full data collection we can do a final pass at 3D analysis

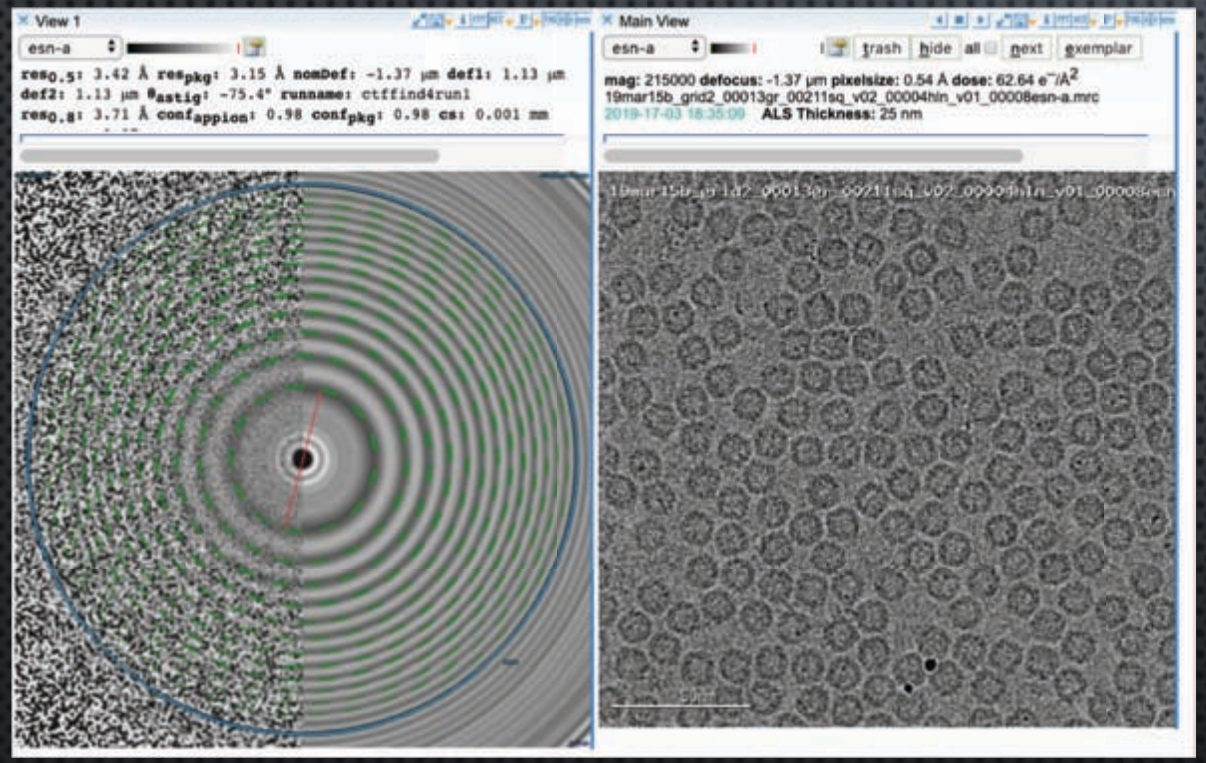


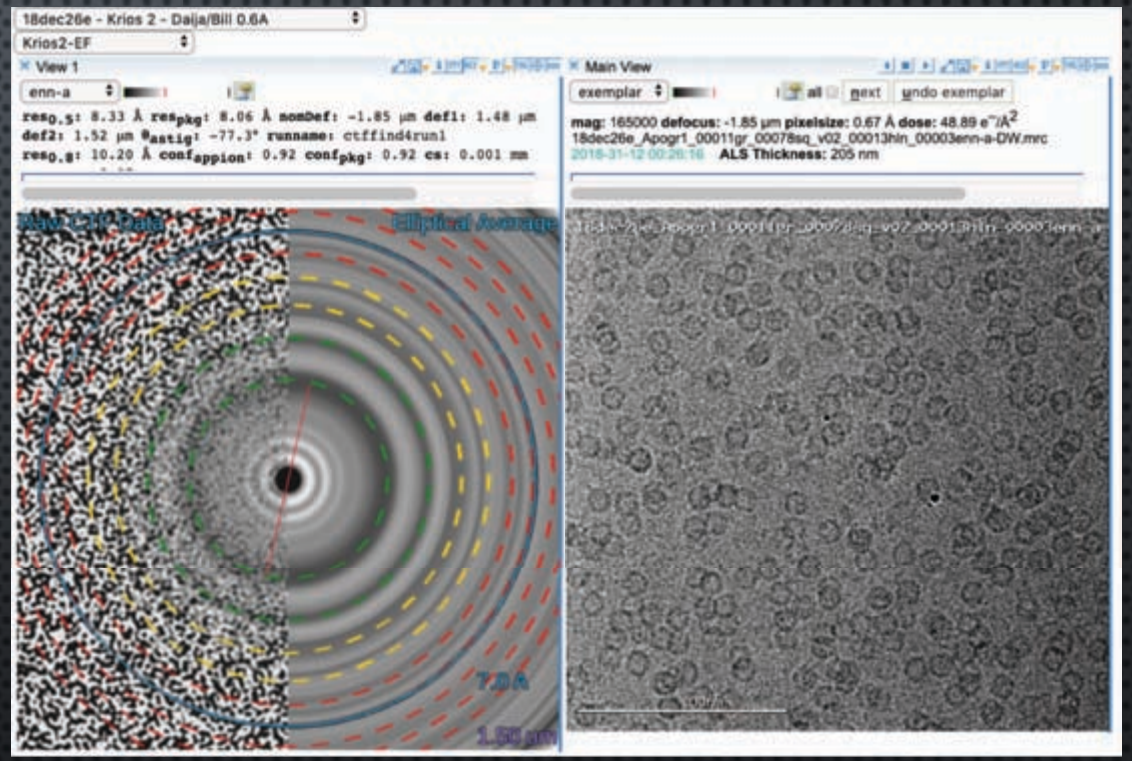



# HOW DO WE OPTIMIZE DATA COLLECTION?

- EMGWEB- Leginon + Appion are integrated to provide real-time monitoring of data quality in a webviewer format
- Hardware beam tilt correction for large image shift
- Leginon-slack integration

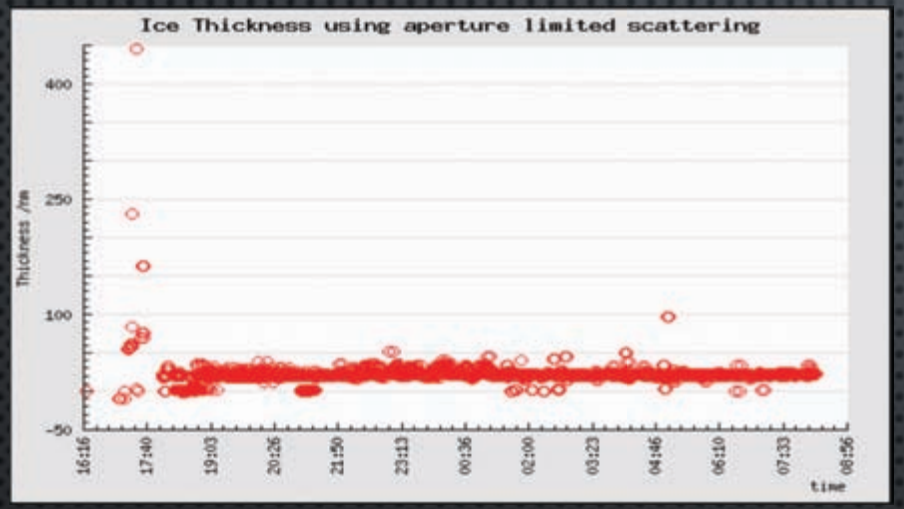
# USING THE METADATA: "GOOD" VS "BAD" DATASET OF APOFERRITIN

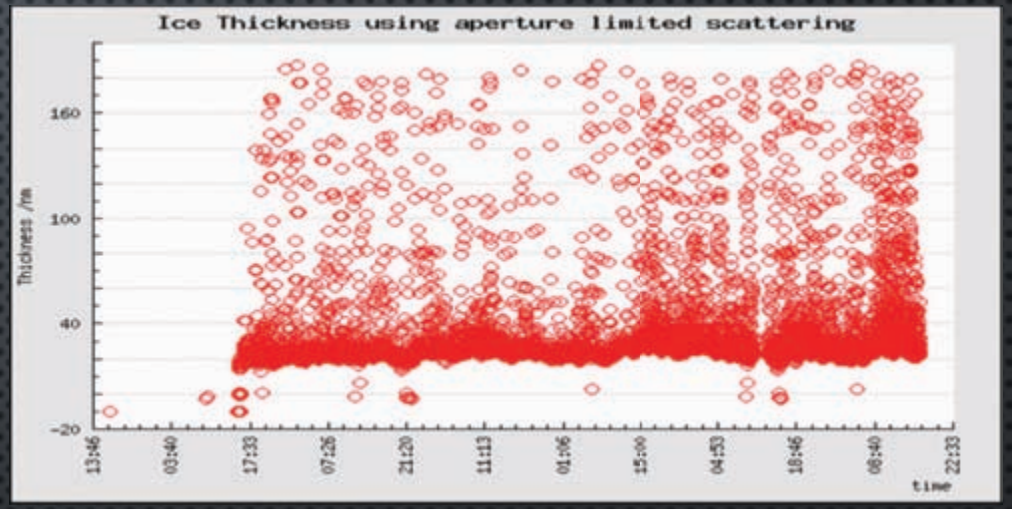


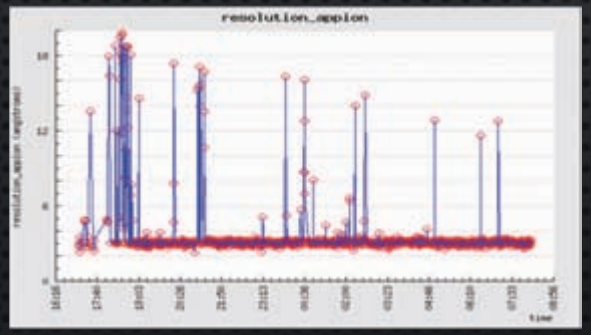



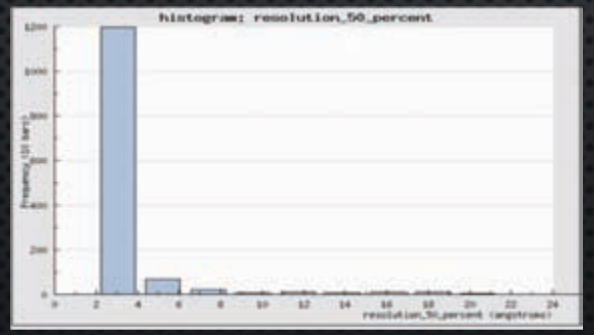

- Better particle distribution & concentration
- Thinner ice (56 nm vs 205 nm)
- Better resolution measurement from CTF estimation (3.4 Å vs 8.3 Å)

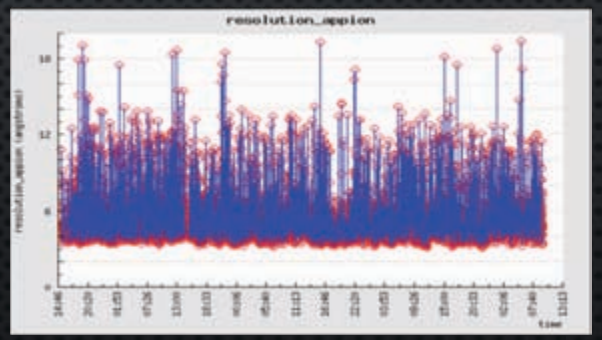
# USING THE METADATA: "GOOD" VS "BAD" DATASET OF APOFERRITIN

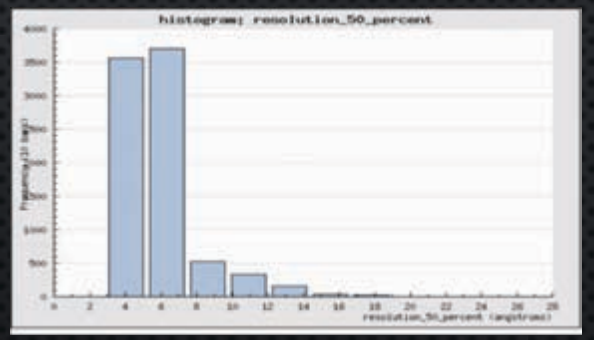


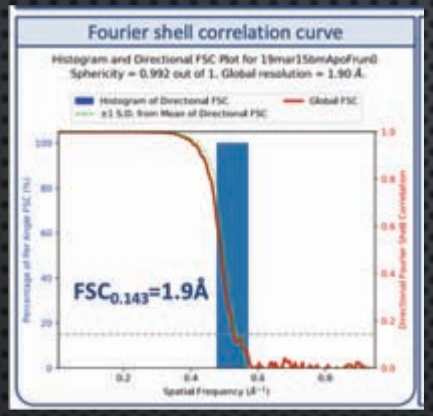



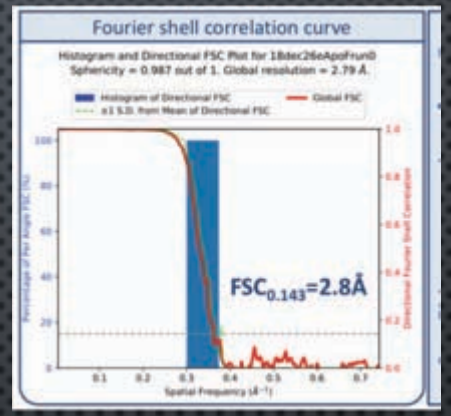

- Good dataset had narrow distribution of ice thickness in the ~30 nm range
- Good dataset had better resolution estimates

# USING THE METADATA: "GOOD" VS "BAD" DATASET OF APOFERRITIN



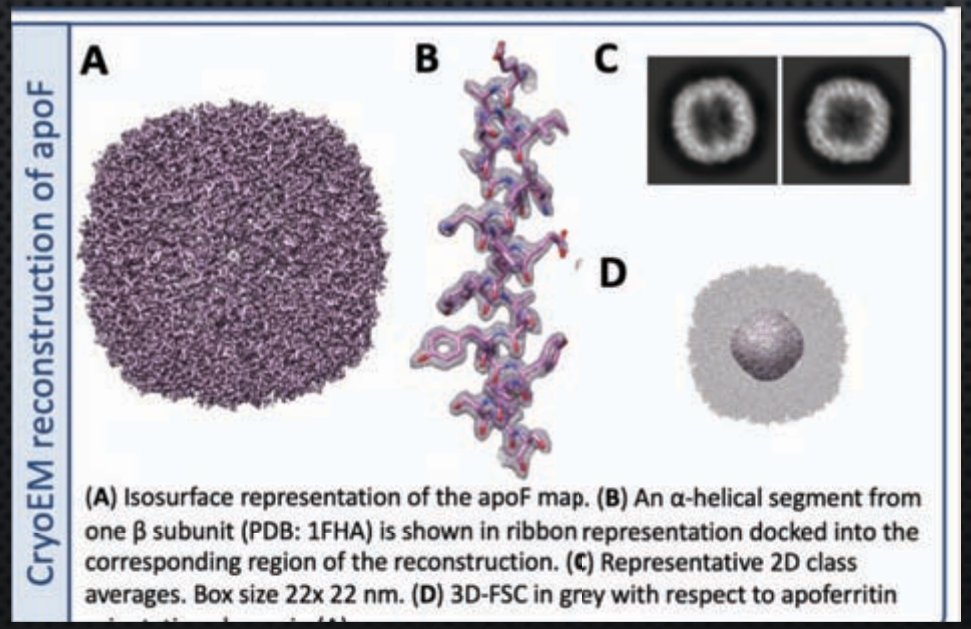

***Imaging statistics: image shift, 20 eV slit***

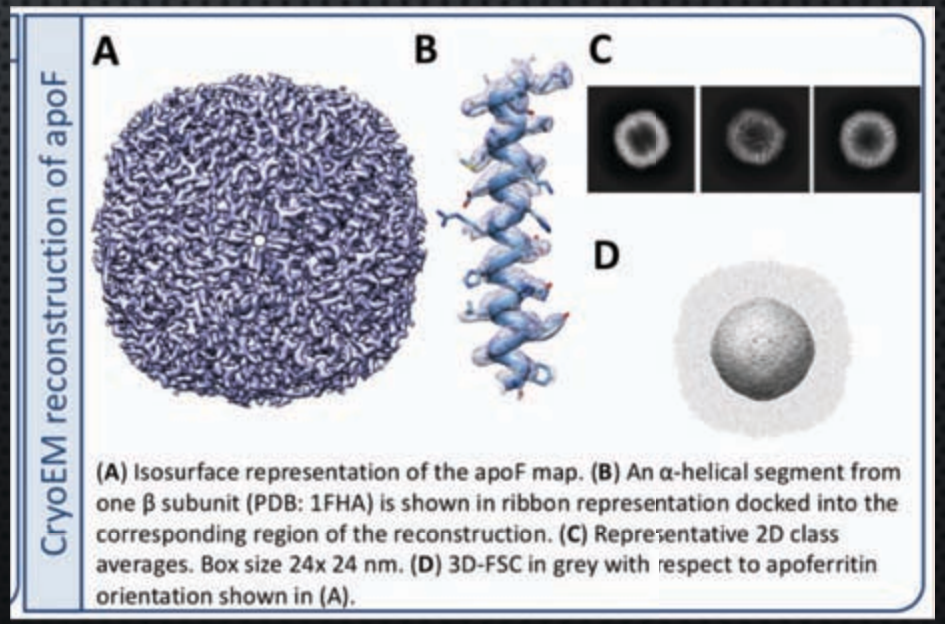
19mar15b experiment : 4828 *enn-a-DW* images by MotionCor2
549,346 particles : Appion template picking esn-a-DW images
316,980 particles : after 2D classification (2.1Å)
313,744 particles : after 3D classification – 1 of 3 classes (1.9Å)




***Imaging statistics: image shift, 20 eV slit***

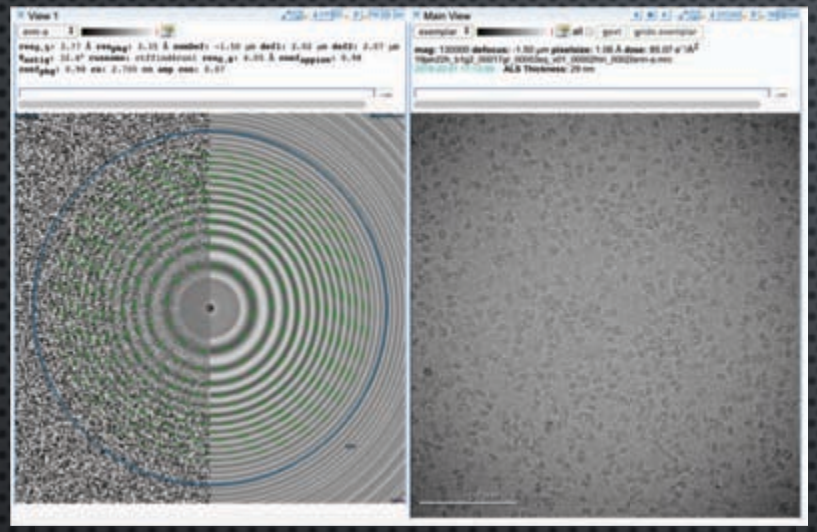
18dec26e experiment : 3276 *enn-a-DW* images by MotionCor2
269,463 particles : Appion picking from 1947 enn-a-DW images
213,216 particles : after 2D classification (2.8Å)
163,149 particles : after 3D classification (2.7Å)

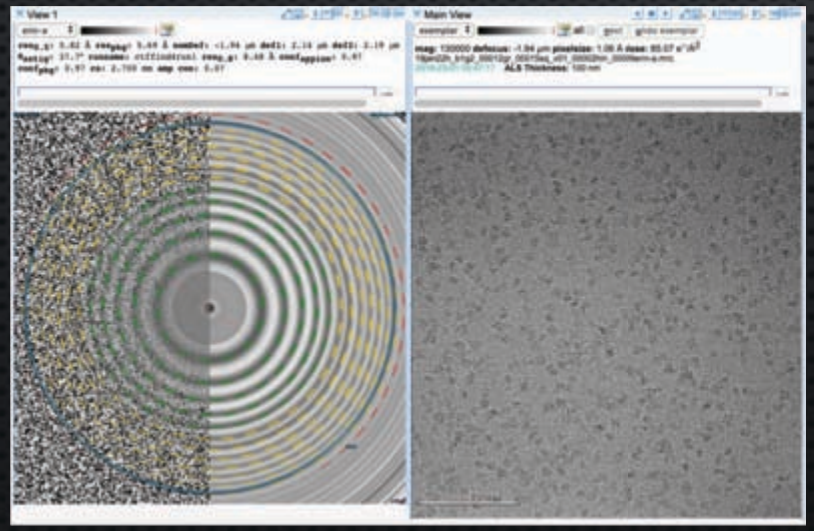


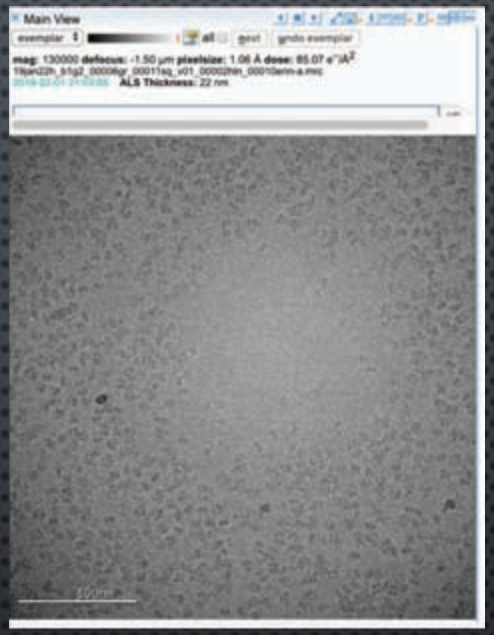



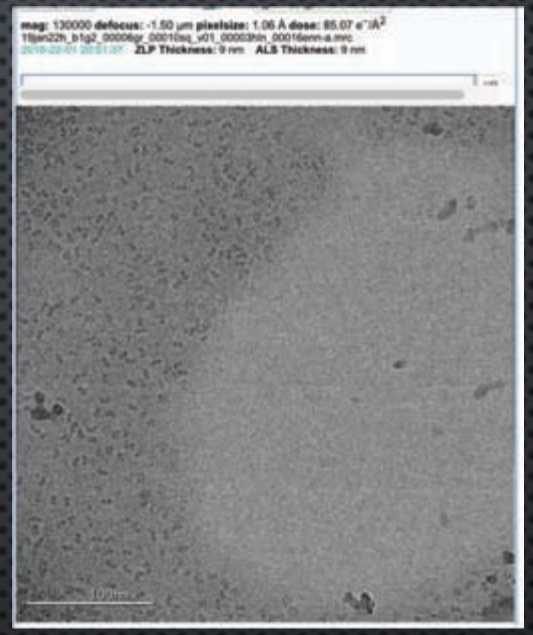
# BUT WHAT ABOUT NON-IDEAL TEST SPECIMENS?

# USING THE METADATA: CHALLENGING SAMPLES



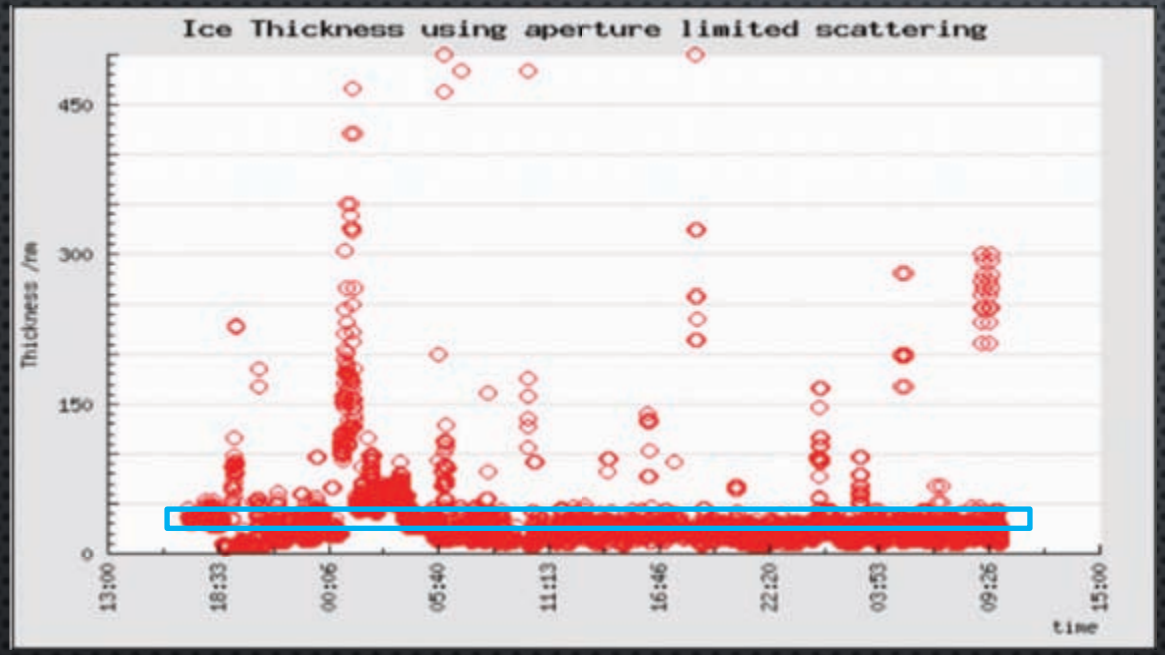






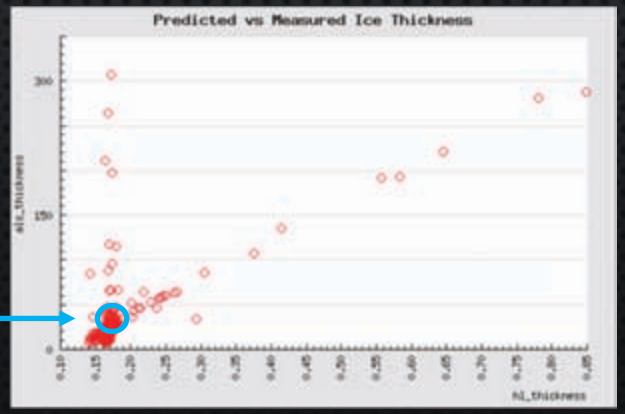
- Membrane protein that is unstable
  - Falling apart at the air-water interface?
- At ~30 nm ice thickness, good particle density, morphology, and good resolution
- At ~100 nm ice thickness, particle density and morphology still good, but resolution decreases
- At < 25 nm ice thickness, particle falls apart

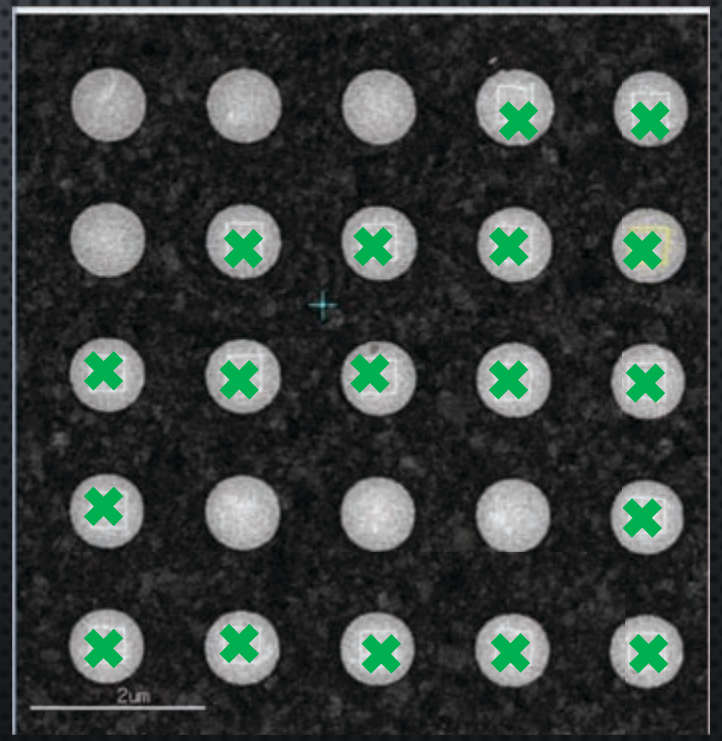
# USING THE METADATA: CHALLENGING SAMPLES




Use ice thickness measurements in combination with hole thickness measurements from Leginon to control the exposures that you want o take!

The sweet spot =
25 – 40 nm ice



# INCREASING THROUGHPUT

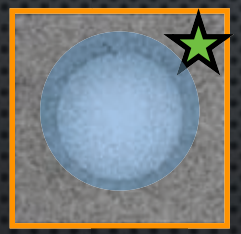
# INCREASING THROUGHPUT: MORE BEAM-IMAGE SHIFT AND FEWER STAGE MOVEMENTS

- It is known that beam-image shift up to 2 um is acceptable and will produce a high quality map
  - Beam-image shift up to 10 um is also possible but requires software correction (beam tilt refinement in relion)


- Beam-image shift induces beam tilt
- Beam tilt creates coma effects
  - Correction via an applied beam tilt creating equivalent coma in the opposite direction
- Implement hardware coma correction since Summer of 2018

# INCREASING THROUGHPUT: MORE BEAM-IMAGE SHIFT AND FEWER STAGE MOVEMENTS

80s/movie
Up to 1000 /day



1 image per stage movement
No beam tilt

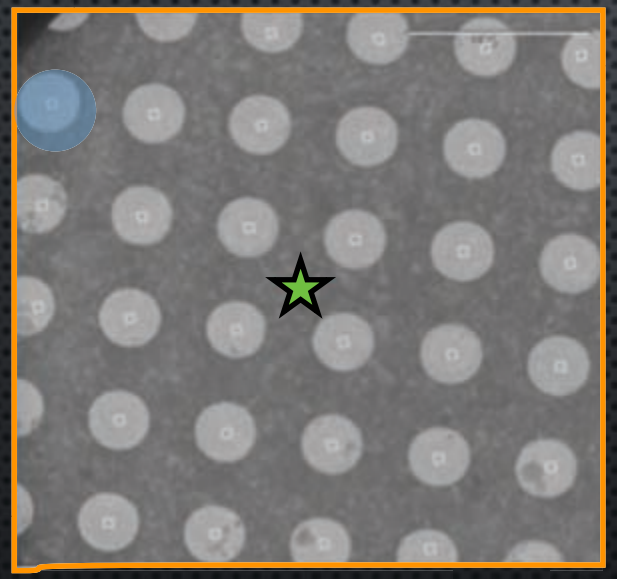
45 s/movie
Up to 2,000 /day

5 images per stage movement
Beam tilt 0.5 mrad

**Overhead**

30s stage movement and settling
30s drift check and focus
20s K2 40-frame movie to save

# INCREASING THROUGHPUT: MORE BEAM-IMAGE SHIFT AND FEWER STAGE MOVEMENTS



30 images per stage movement
Beam tilt 2 mrad

35s/movie
Up to 2500/day

# INCREASING THROUGHPUT: MORE BEAM-IMAGE SHIFT AND FEWER STAGE MOVEMENTS

80 images per stage movement
Beam tilt 3 mrad

22s/movie
Up to 4000/day


# OPTIMIZING DATA COLLECTION: SLACK-LEGINON INTEGRATION

# WARNING SYSTEM

- Leginon is connected to the slack messaging system
- Alerts staff to errors
- About 90% of problems can be solved remotely

**Leginon** APP 9:02 AM
Krios2 Leginon has been idle for 30.0 minutes

**Leginon** APP 9:19 AM
Krios2 K1 Presets Manager Error: Move to target failed: unable to set instrument

Krios2 K1 Presets Manager Error: COMError in _setStagePosition: (-2147024891, 'Access is denied.', (None, None, None, 0, None))

**anchi2c** 9:25 AM
What is wrong?

**laura_yen** 9:25 AM
not sure, i'm troubleshooting now

i think some sq or hl target is too close to edge of stage

**anchi2c** 9:27 AM
What is the value ?

**Friday, April 26th**

**Leginon** APP 12:12 AM
Krios3 BlackStripeDetector Error: error: Black stripe detected on stripe 2

12:13 AM Krios3 BlackStripeDetector Error: error: Black stripe detected on stripe 2

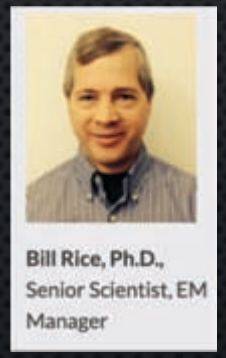
# SUMMARY

- AT THE NYSBC WE USE LEGINON AND APPION TO COLLECT AND PROCESS DATA ON-THE-FLY
- INTEGRATION OF LEGINON + APPION HELPS US COLLECT MORE USEFUL DATA
- DATA COLLECTION OPTIMIZATION
  - TONS OF METADATA AND LIVE FEEDBACK
  - HARDWARE BEAM TILT CORRECTION FOR LARGE IMAGE SHIFT
  - LEGINON-SLACK INTEGRATION

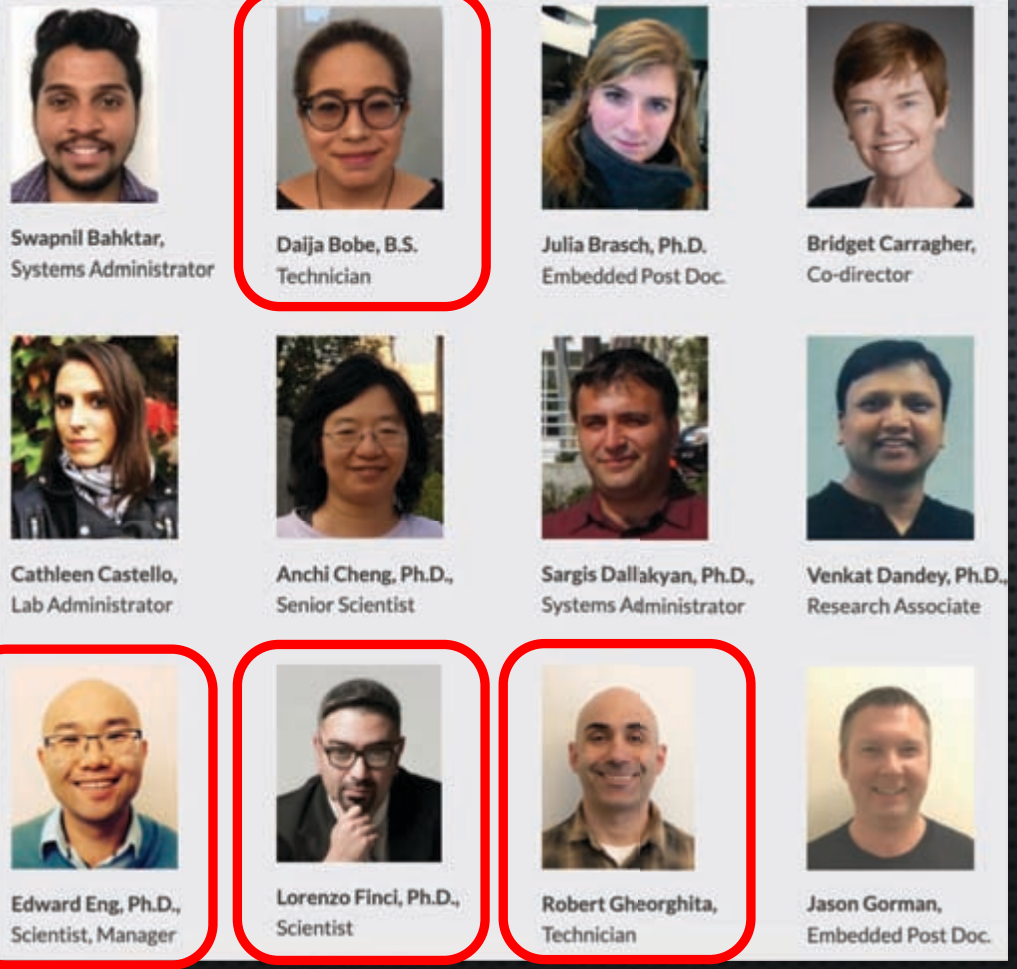
# ACKNOWLEDGMENTS: EMG OPS TEAM

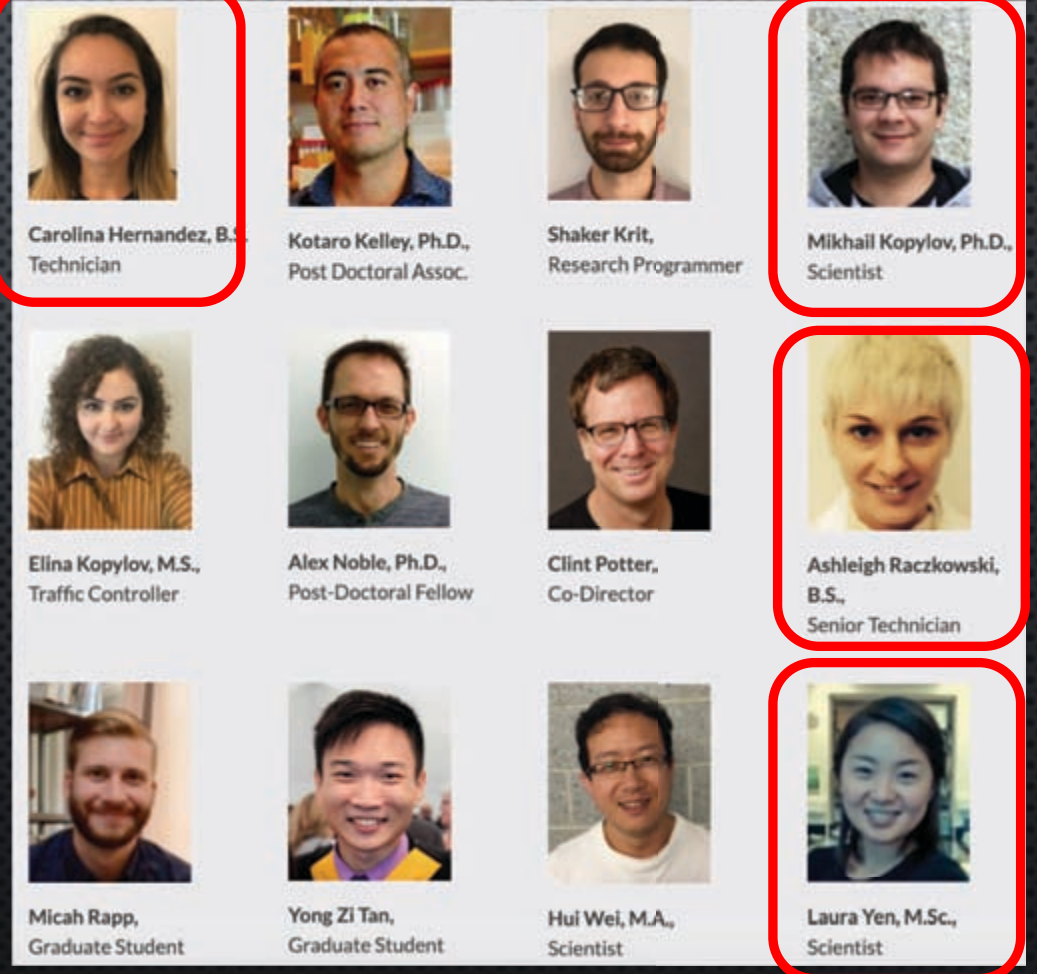
**Grants from**
- Simons Foundation (SF349247)
- NYSTAR
- NIH GM103310
- NIH OD019994

Now at NYU!



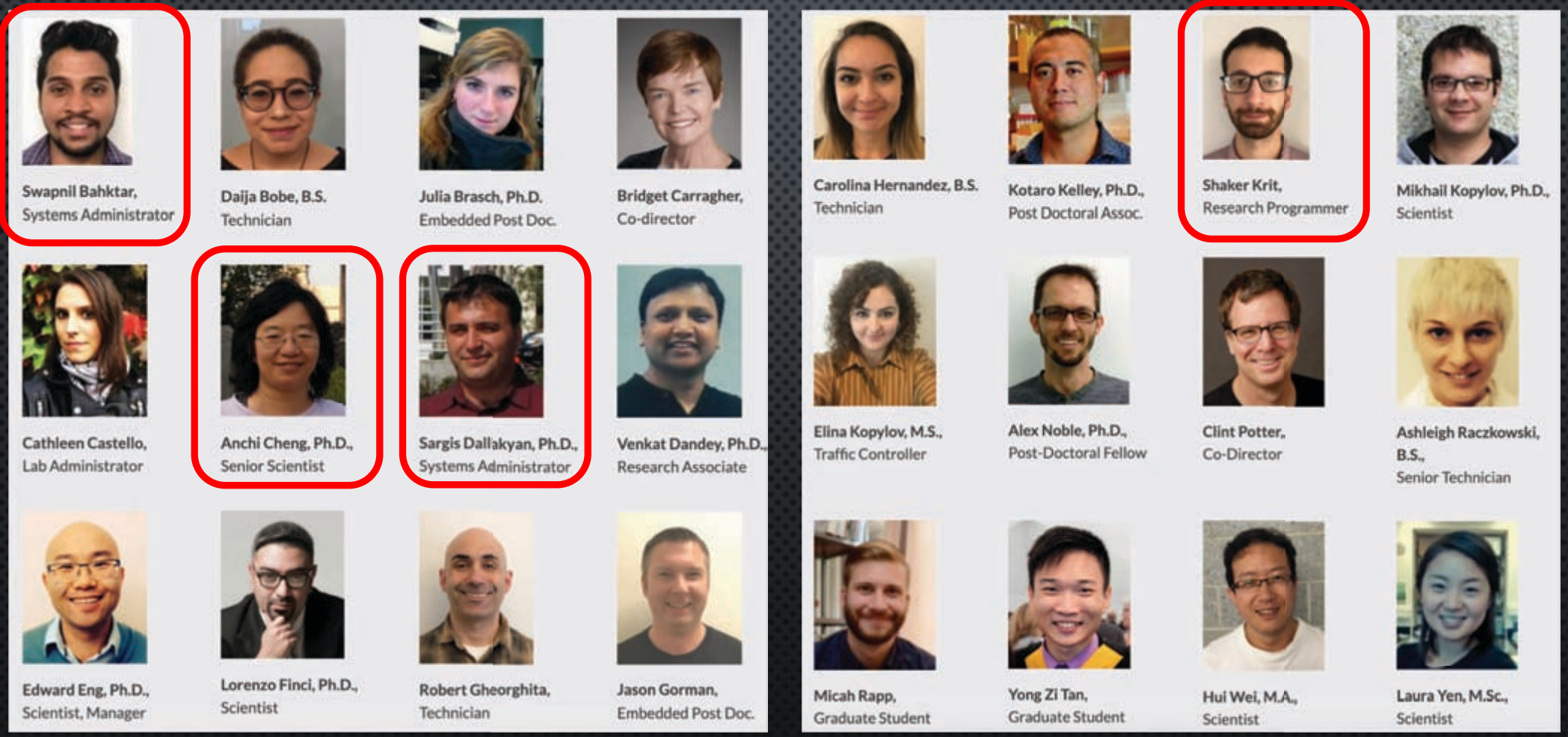
Bill Rice, Ph.D., Senior Scientist, EM Manager



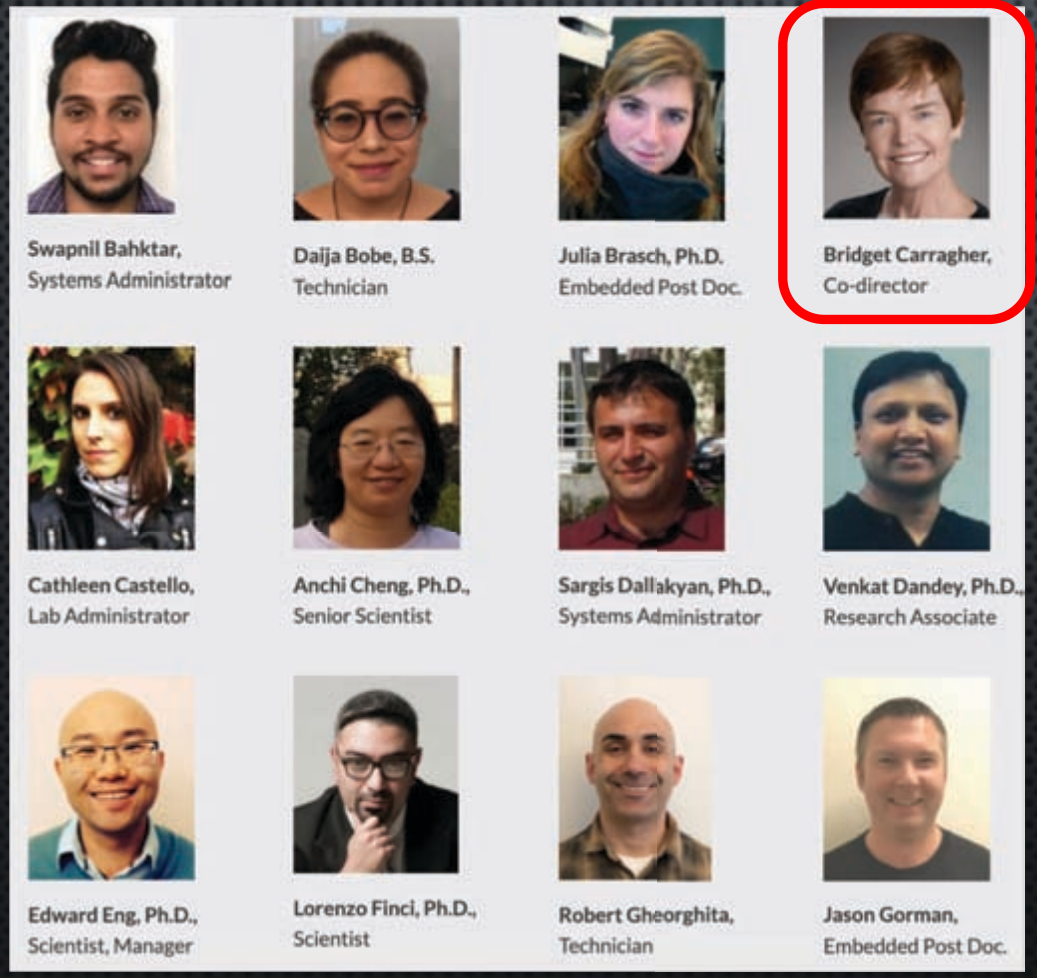
# ACKNOWLEDGMENTS: SEMC IT

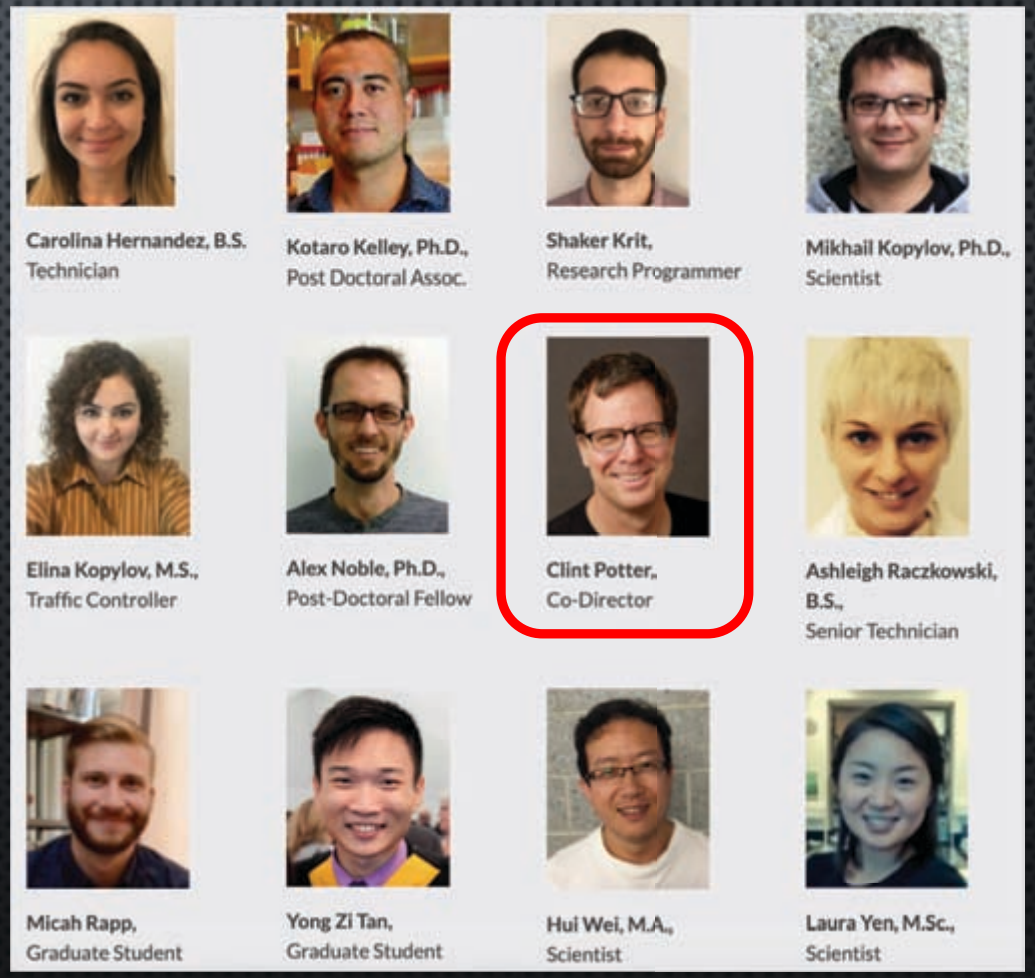
**Grants from**
- Simons Foundation (SF349247)
- NYSTAR
- NIH GM103310
- NIH OD019994



# ACKNOWLEDGMENTS: DIRECTORS

**Grants from**
- Simons Foundation (SF349247)
- NYSTAR
- NIH GM103310
- NIH OD019994



# WE'RE HIRING! JOIN IN ON THE FUN!

## Research Associate and Scientist positions

The New York Structural Biology Center (NYSBC) seeks an experienced electron microscopist to join the staff of the Simons Electron Microscopy Center (SEMC) (http://semc.nysbc.org). The NYSBC is a shared center that supports state-of-the-art research in EM, NMR, and X-ray crystallography. The facilities include nine transmission electron microscopes (including three Titan Krios instruments and three more by the end of the year) and a dual-beam scanning electron microscope equipped with a cryo stage. Projects focus on 3D reconstruction of biological assemblies using techniques including microED, single particle reconstruction, tomography, and preparation of FIB milled lamellae. To assist in these developments, NYSBC seeks an individual with experience in biological electron microscopy and image processing. This individual will carry out experiments in support of collaborative projects with affiliated investigators and will also have opportunities to pursue independent research


# QUESTIONS?