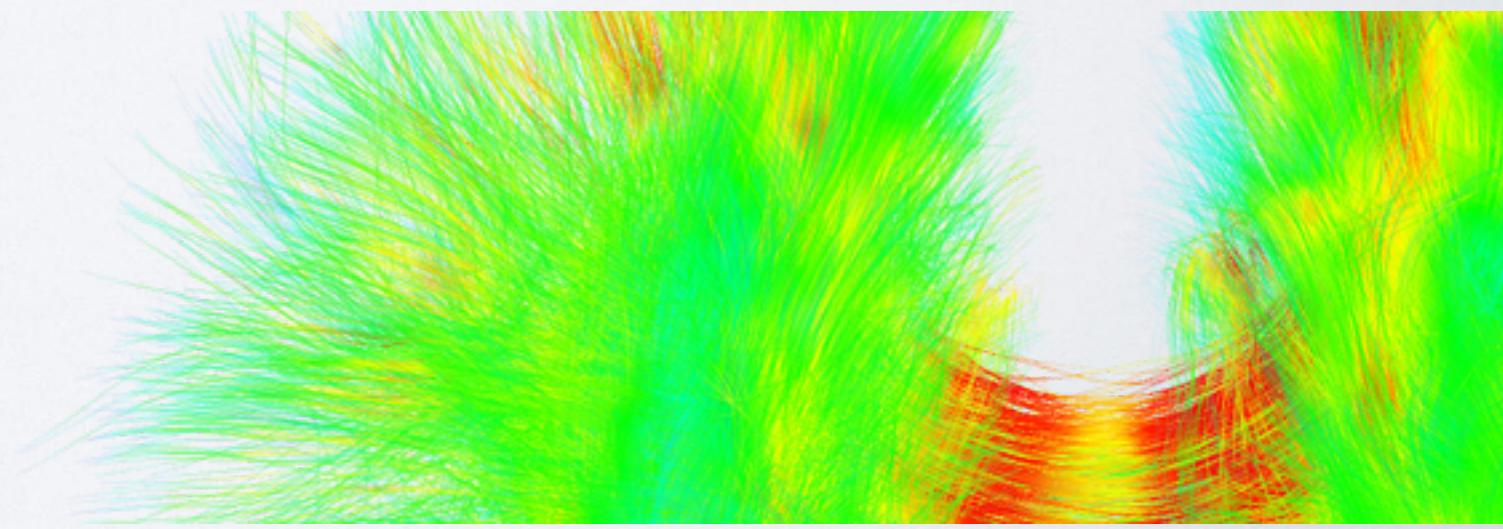


TRACTOGRAPHY CHALLENGE 2015

Peter Neher, Klaus Maier-Hein, Bram Stieltjes
Tim Dyrby, Alessandro Daducci, Emmanuel Caruyer,
Jean-Christophe Houde, Maxime Descoteaux



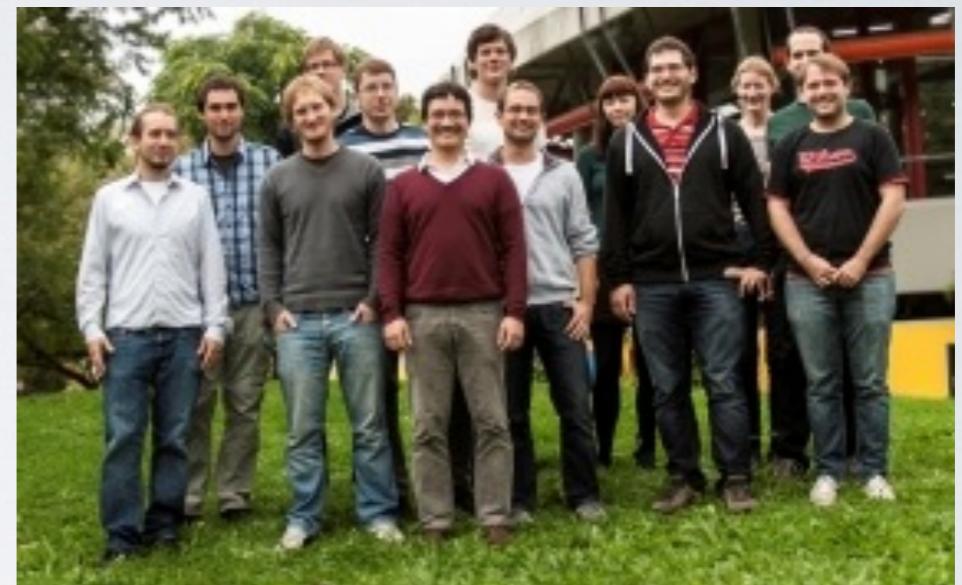
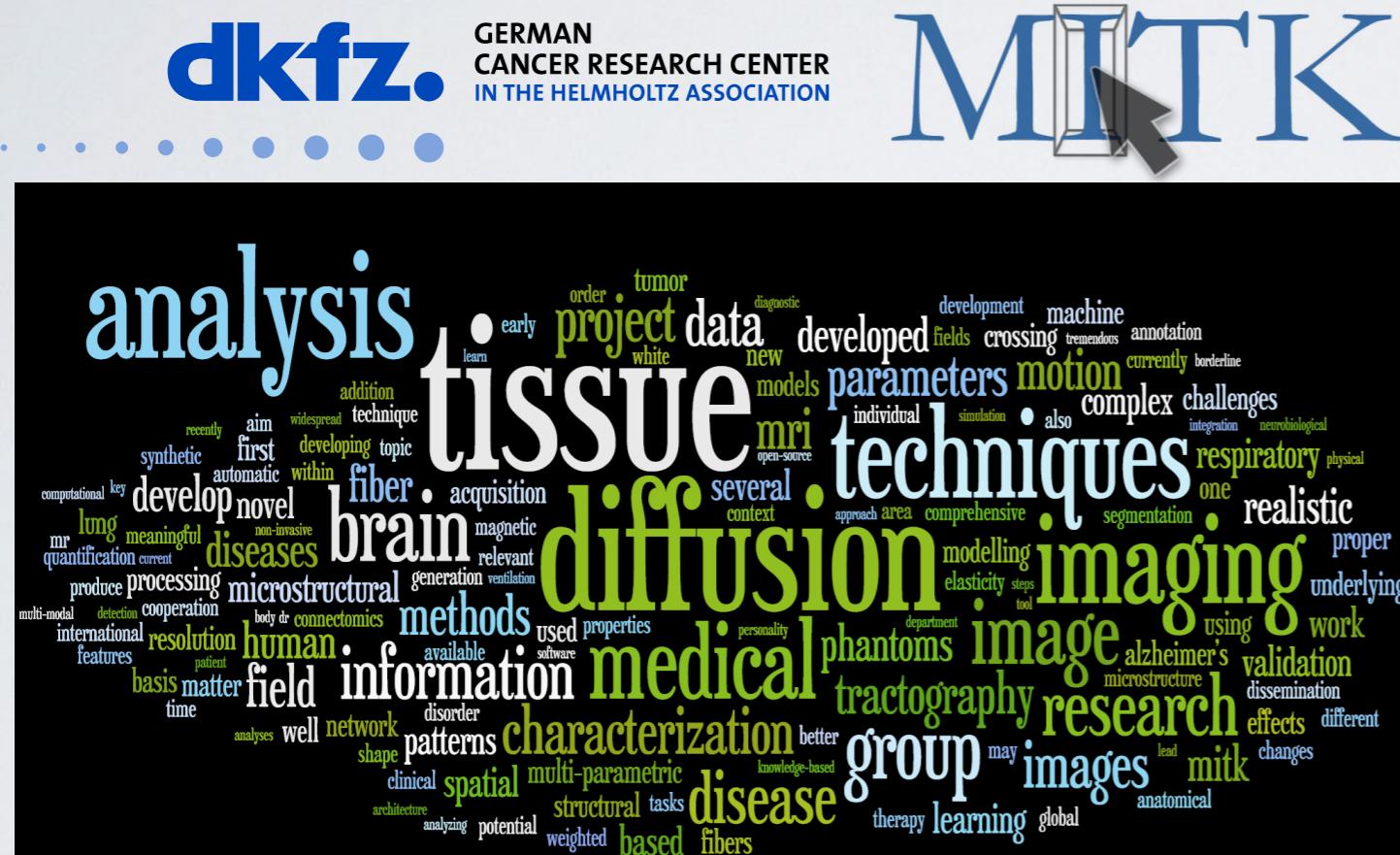
TRACTOGRAPHY SUBMISSIONS

- 20 teams, 96 submissions
- > 50 GigaBytes
- > 20 million streamlines

THANKS TO ALL THE TEAMS AND PARTICIPANTS!!!

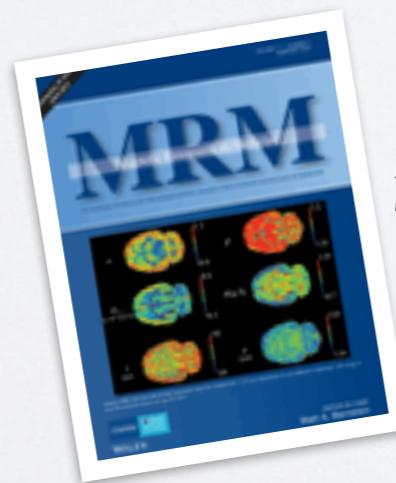
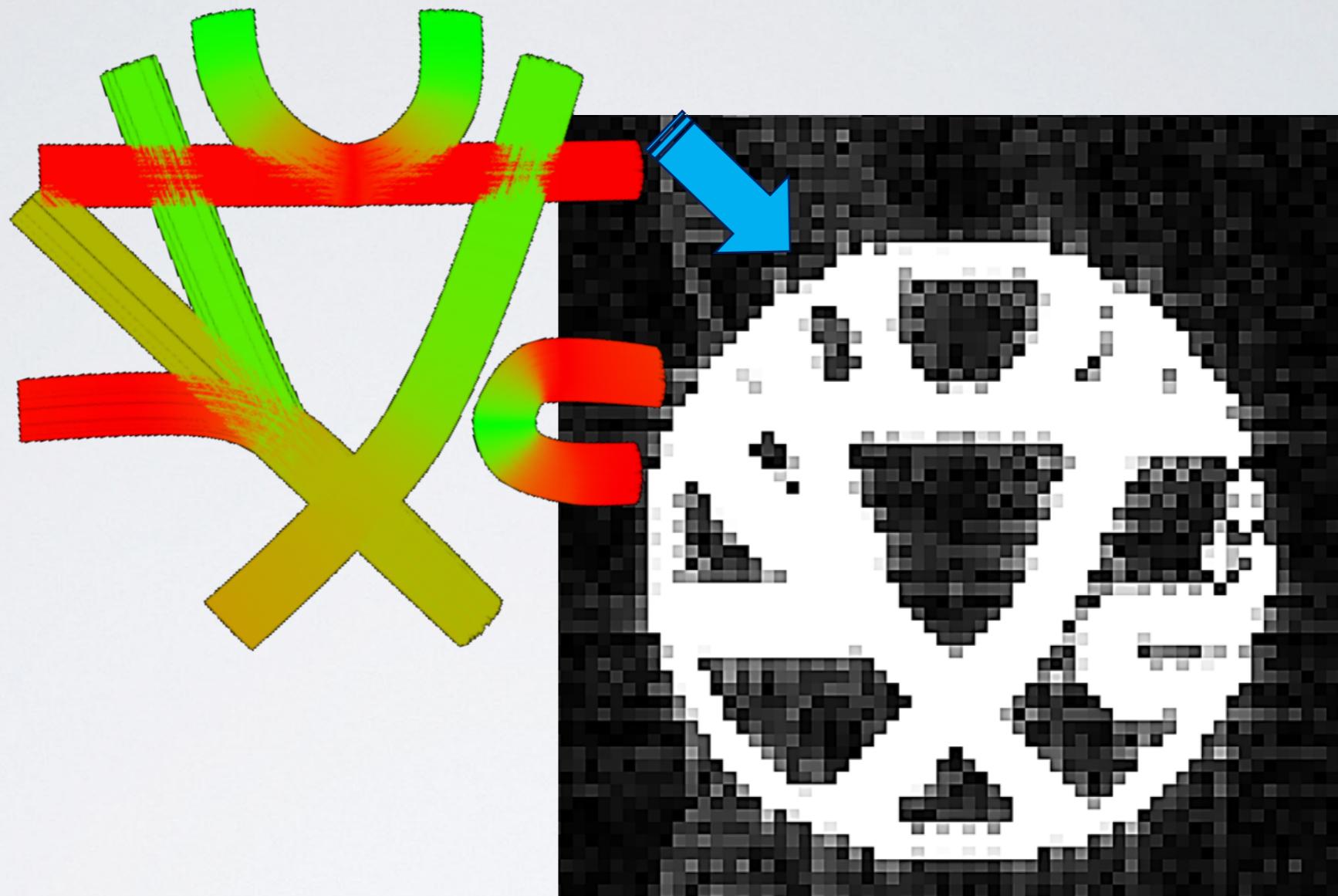
- Jidan Zhong, Maxime Chamberland, Frank Yeh, Ying-Chia Lin, Qing Ji, David Qixiang Chen, Yuanjing Feng, Samuel Deslauriers-Gauthier, François Rheault, Providi Lab + Chantal Tax, Samuel St-Jean, Micheal Paquette, Omar Ocegueda, Gabriel Girard, Arnaud Boré, Alessia Sarica, Ali R. Khan, David Romascano, Muhamed Barakovic, H. Ertan Cetingul, Fabrizio Pizzagalli, Flavio Dell'Acqua et al.
- All teams will receive their results and output tracts
- All results will be on the tractometer.org website
- All ground truth data will be distributed also

Medical Image Computing (MIC)



Peter Neher, Jonas Cordes, Christoph Kolb, Jan Hering, Tobias Norajitra, Christian Weber, Romy Henze, Michael Götz, Caspar Jonas Goch, Jens Petersen, Bram Stieltjes, Jens Kleesiek, Vincent Reuter, Klaus H. Maier-Hein

Simulation of realistic diffusion-weighted images



Neher PF, Laun FB, Stieltjes B, Maier-Hein KH. **Fiberfox**: facilitating the creation of realistic white matter software phantoms. Magn Reson Med. 2014 Nov;72(5):1460-70.



TRACTOMETER RESULTS

[**Jean-Christophe Houde**, Marc-Alexandre Côté, Alessandro Daducci, Eleftherios Garyfallidis, Maxime Descoteaux]

VALID CONNECTIONS (VC)

Connects 2 connected regions

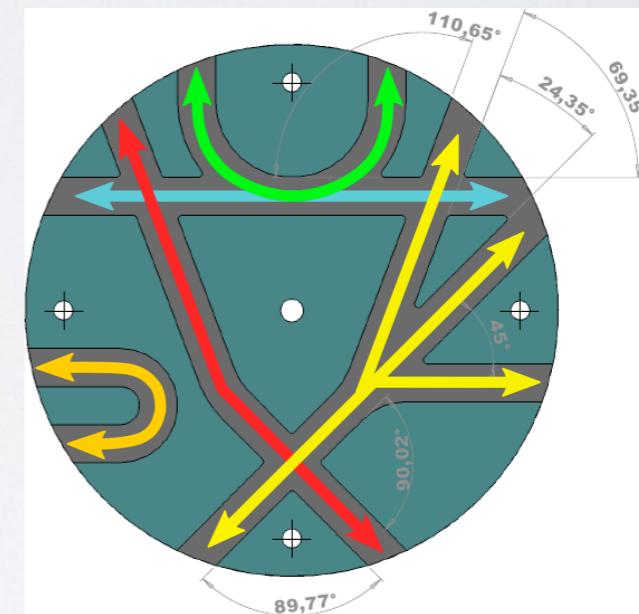
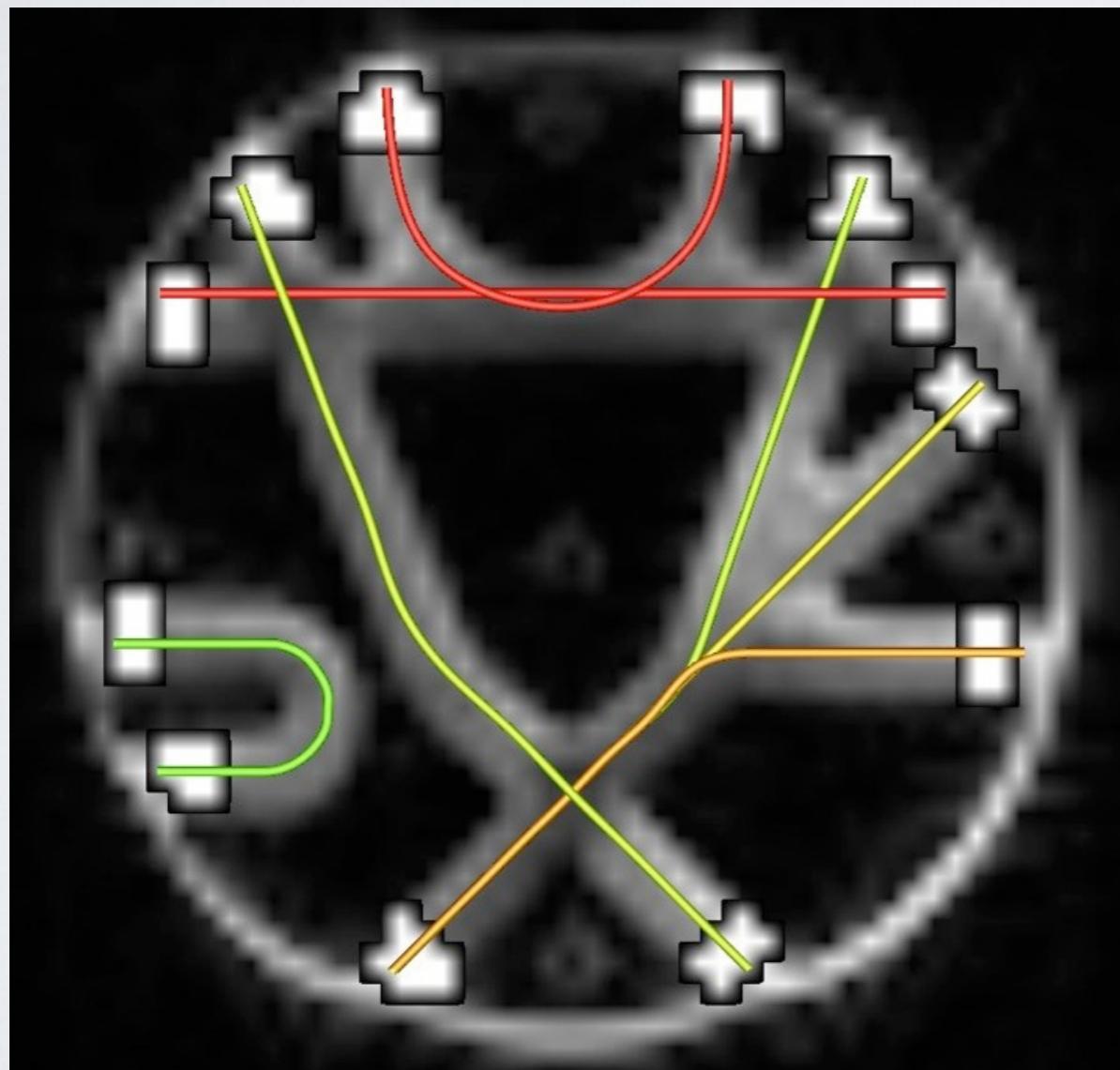
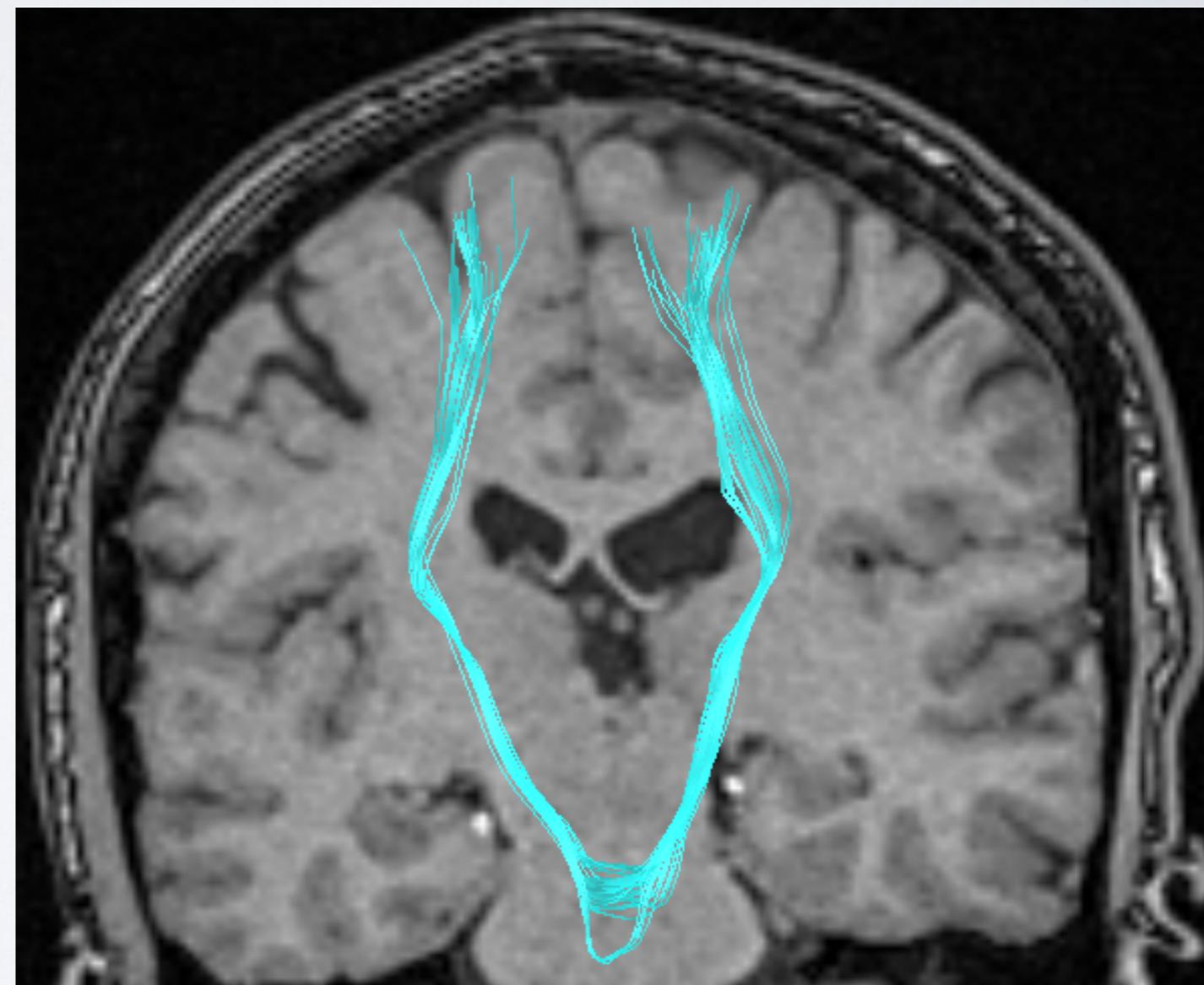
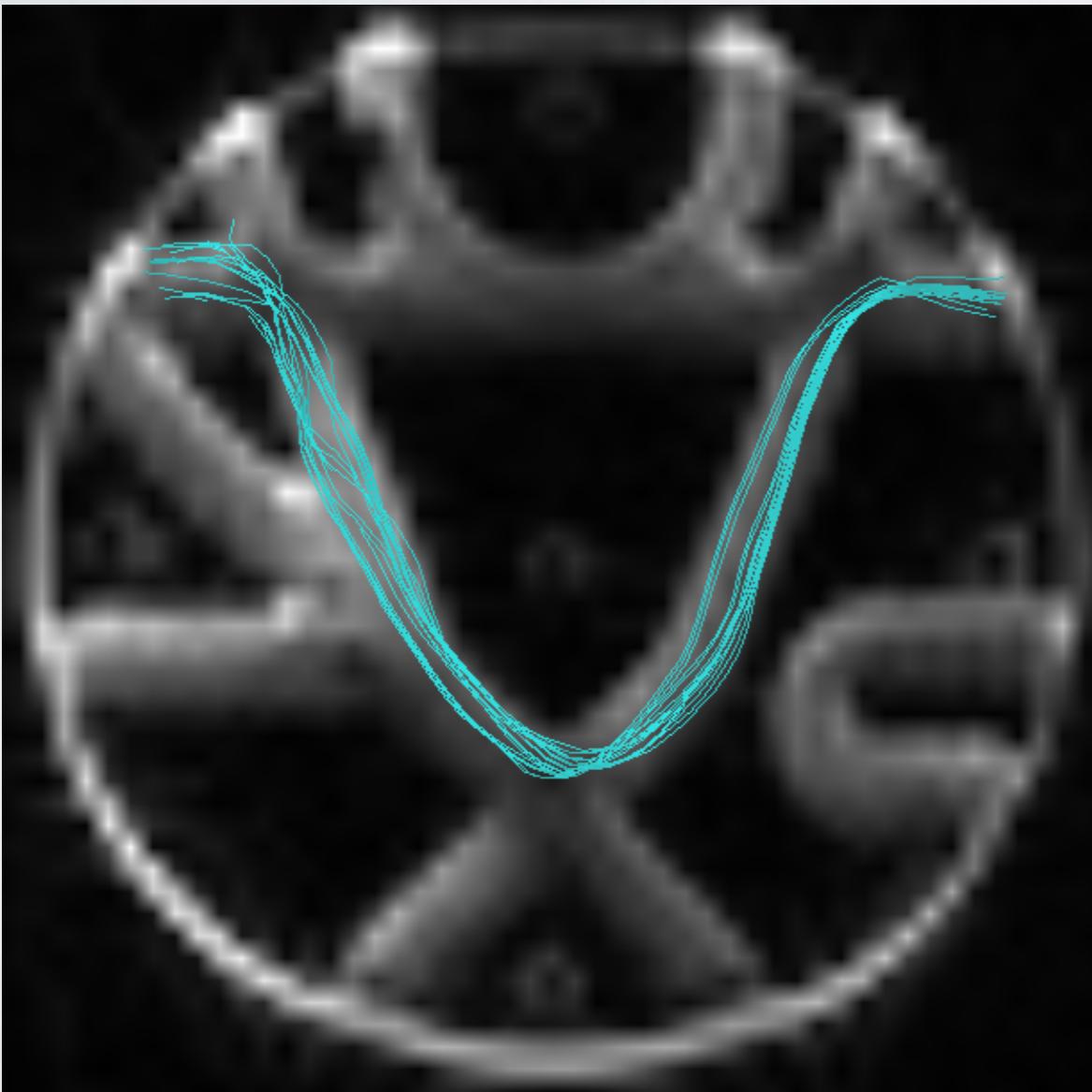


Illustration with the *FiberCup* dataset

MICCAI Challenge 2009. [Fillard, Descoteaux et al Neurolimage 2010]

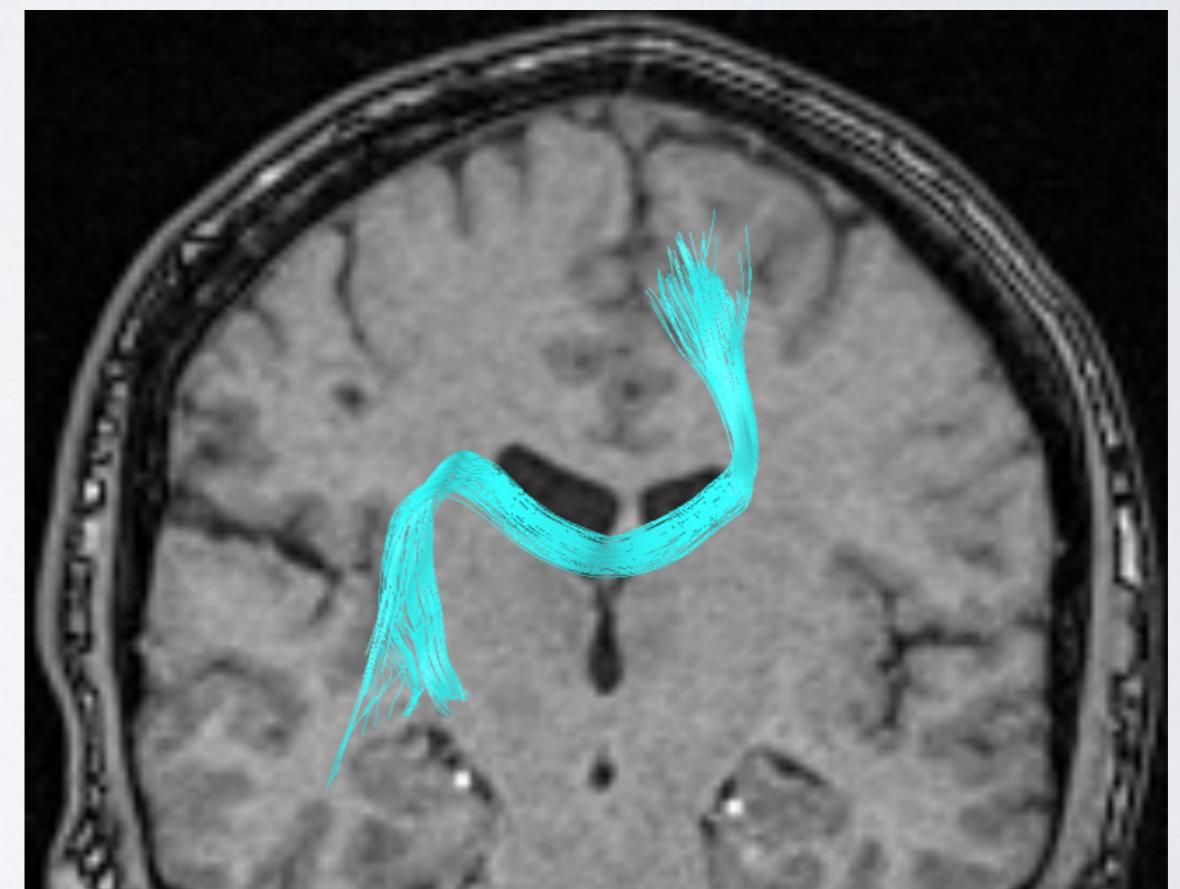
VALID CONNECTION WRONG PATH



Connects 2 valid regions through a wrong path

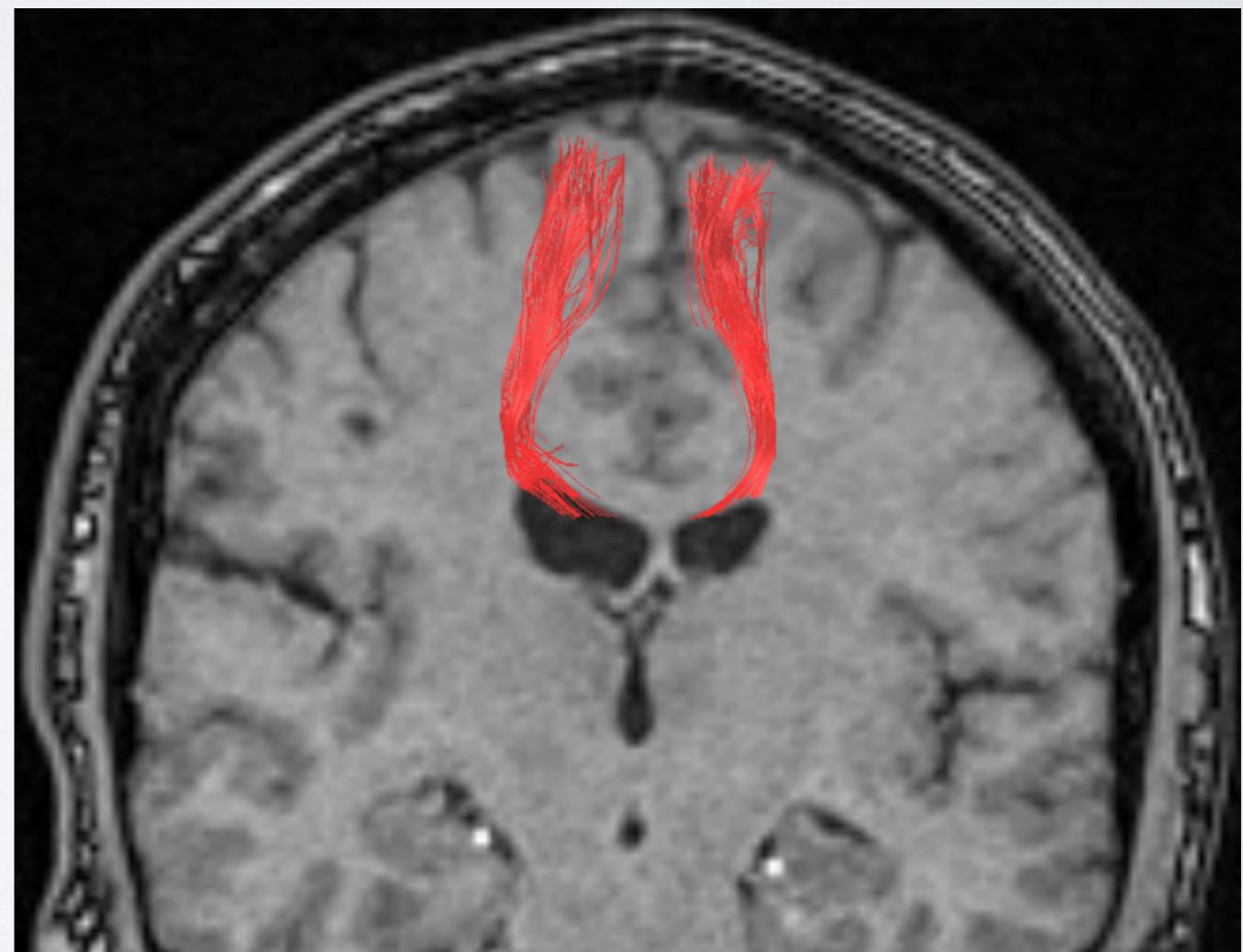
INVALID CONNECTIONS

Connects 2 regions that should not be connected



NO CONNECTIONS

Does not connect 2 end regions
(stops prematurely in white matter or ventricles)



OVERLAP %



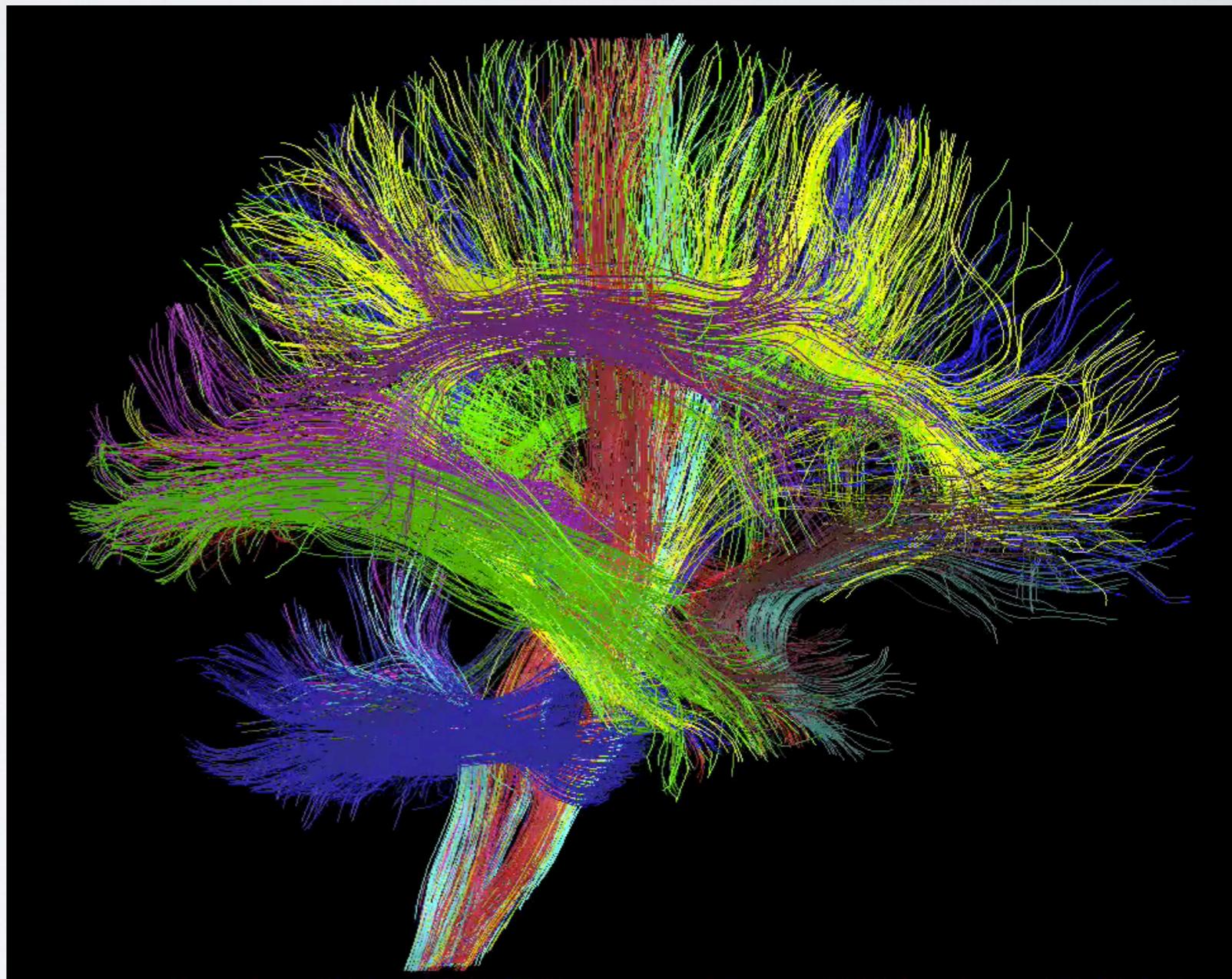
Submitted streamlines
Ground truth streamlines
(Example on the CST)



Overlapping voxels
Non-overlapping voxels

GROUND TRUTH FIBERS

- Perfect results
- 25/25 Valid bundles
- 100 % Valid connections
- 0 % Invalid connection
- 0 % No connection



OVERVIEW OF SUBMISSIONS

anat priors

MITK

1x1x1 denoising

6

THRESHOLD

DSI studio

super resolution

MRtrix

DIPY

motion

GLOBAL

STREAMLINE

MAXIMA

A
EED
IS

WHITE MATTER

MAXTHETA

FIBER LENGTH

PROBABILISTIC MASK DETERMINI

distortion correction

clustering

spikes/outliers

ESL

STEP SIZE

COMMIT

2x2x2

HARD

**LOCAL
ATION**

VISUALISATION

ExploreDT

PRE-PROCESSING

- **Motion-correction** - 17/20 teams
 - Rotating bvecs accordingly, **7 teams**
- **Distortion corrections** - 20/20 teams
- **Denoising** - 14/20 teams
- **Outlier - spike detections** - 6 / 20 teams
- **Super-resolution** - 1 / 20 team



LOCAL RECONSTRUCTION

- Diffusion tensor (**DTI**) - 6 teams
- **HARDI** reconstruction - 18/20 teams
 - Almost all fiber orientation distribution (FOD)-like MRtrix, ExploreDTI, StarTrack, DIPY
 - 2 multi-tensor
 - 1 generalized q-space imaging
 - 2 in-house ODFs

TRACTOGRAPHY

- **Deterministic** - 15 teams
- **Probabilistic** - 8 teams
- **Global** - 3 teams
- **Merge** (union) of det, prob, global - 2 teams

SPACE FOR TRACKING

- **2x2x2** - Diffusion space - 9 teams
- **1x1x1** - T1 space - 11 teams
- **Using anatomical priors** - 11 / 20 teams

TRACTOGRAPHY POST-PROCESSING

- **SIFT** - 2 teams
- **COMMIT** - 1 team
- **Clustering** - 4 teams
- **Fiber compression** - 1 team
- **Short fiber elimination** - 7 teams
- **Manual dissection and editing** - 2 team

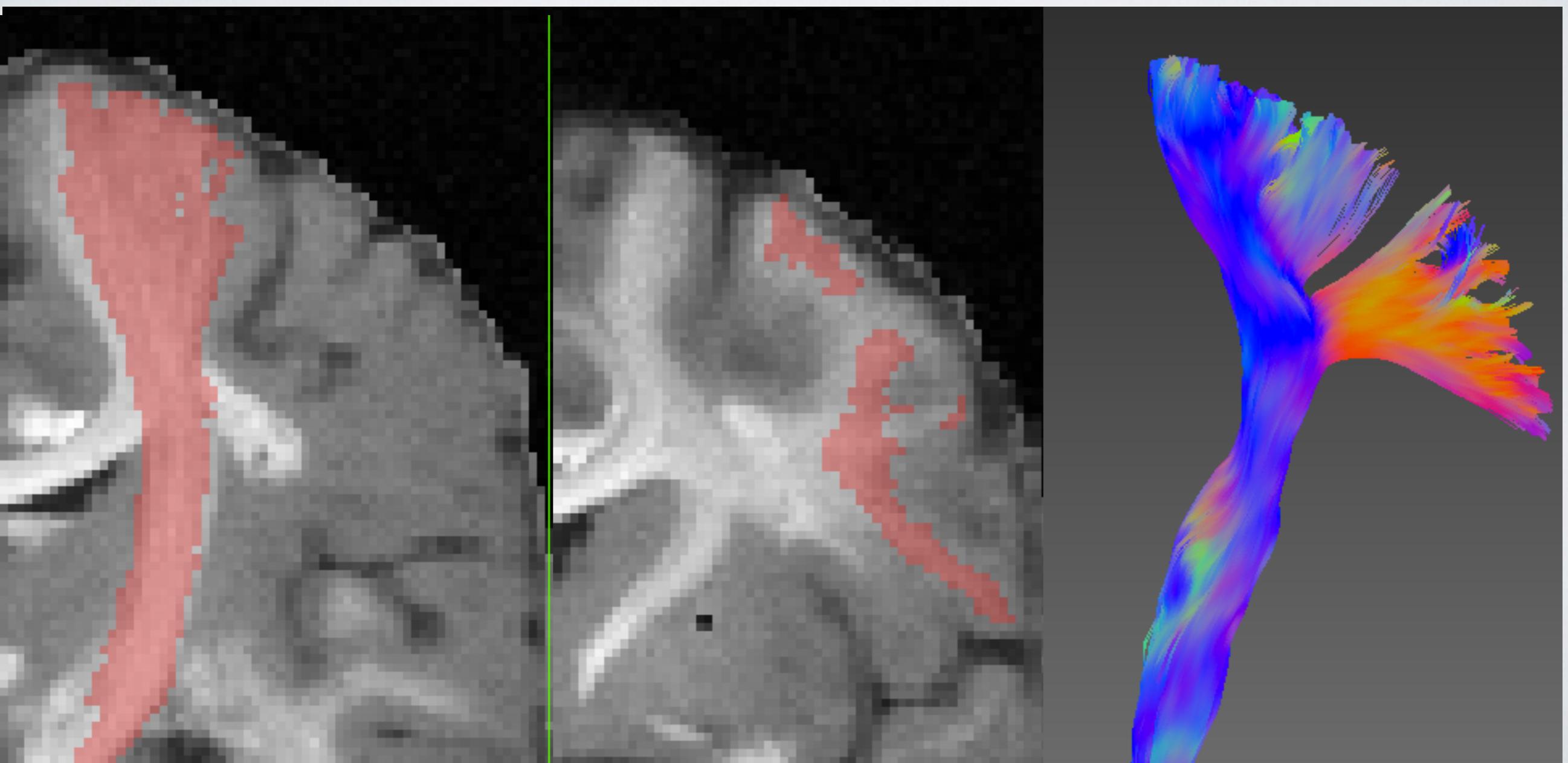
TRENDS

CHALLENGING ENDPOINTS

- Submitted tractograms are **rarely (almost never) voxel-perfectly** aligned with ground truth
- => HUGE number of Invalid Connections & No Connections

EVALUATION

- Challenging ground truth endpoints and bundle mask
- 25 valid bundles, 2450 potential invalid bundles



CHALLENGING ENDPOINTS

- Submitted tractograms are **rarely (~ never)** **voxel-perfectly** aligned with ground truth

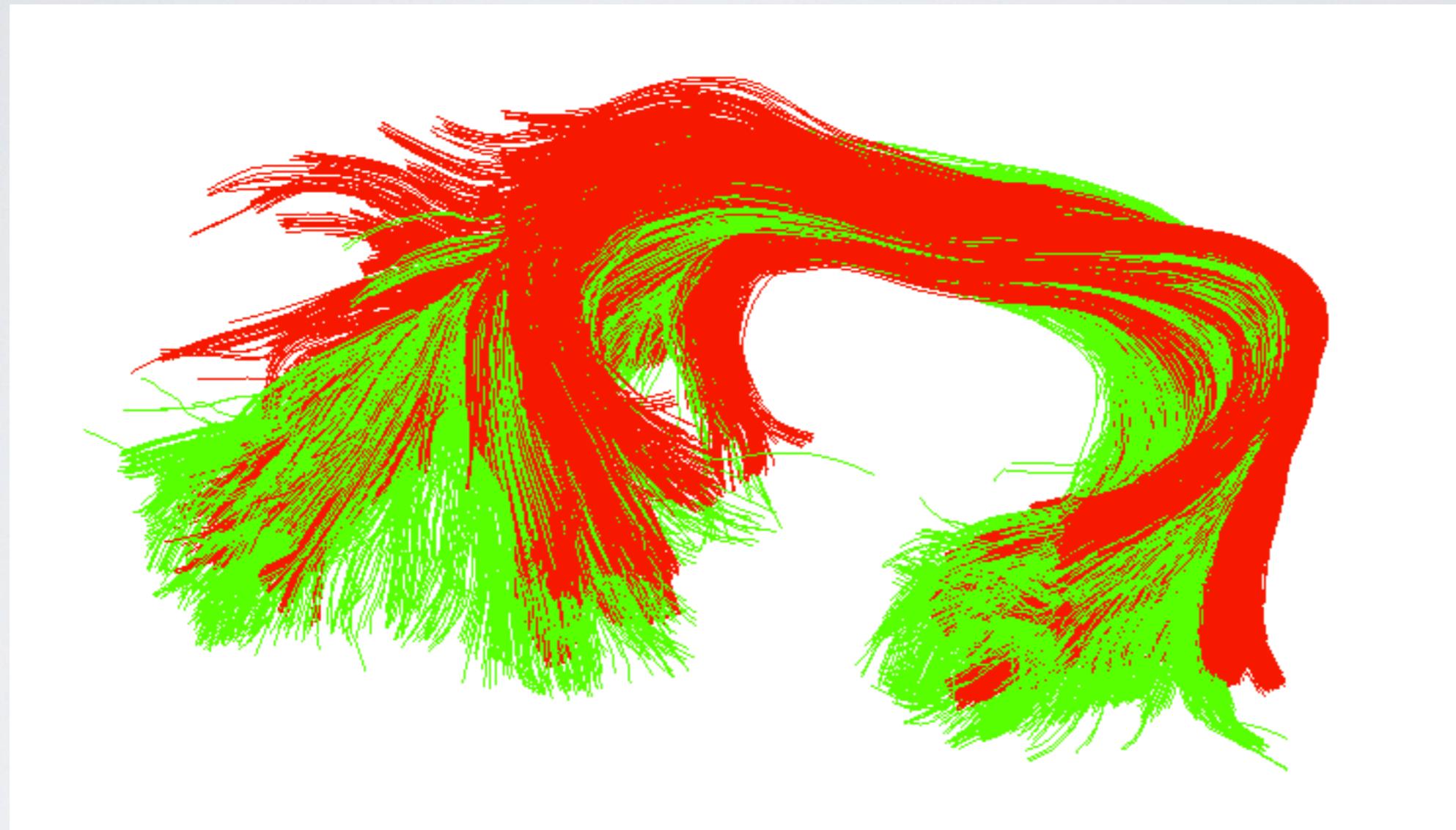
=> HUGE number of Invalid Connections & No Connections

- At best,
 - < 19 valid bundles - mean 13 +- 3 VB
 - < **7% valid connections, mean 3% +- 2%**
 - > **90% invalid and no connections, mean 376 +- 150 IB**
 - > 300 invalid bundles
 - ~ 5% valid connections but wrong path

CHALLENGING ENDPOINTS

- Submitted tractograms are **never voxel-perfectly** aligned with ground truth
- But, valid bundles seem there
 - Just not at the exact right place
 - Original evaluation scheme is too aggressive

UNCINATE EXAMPLE



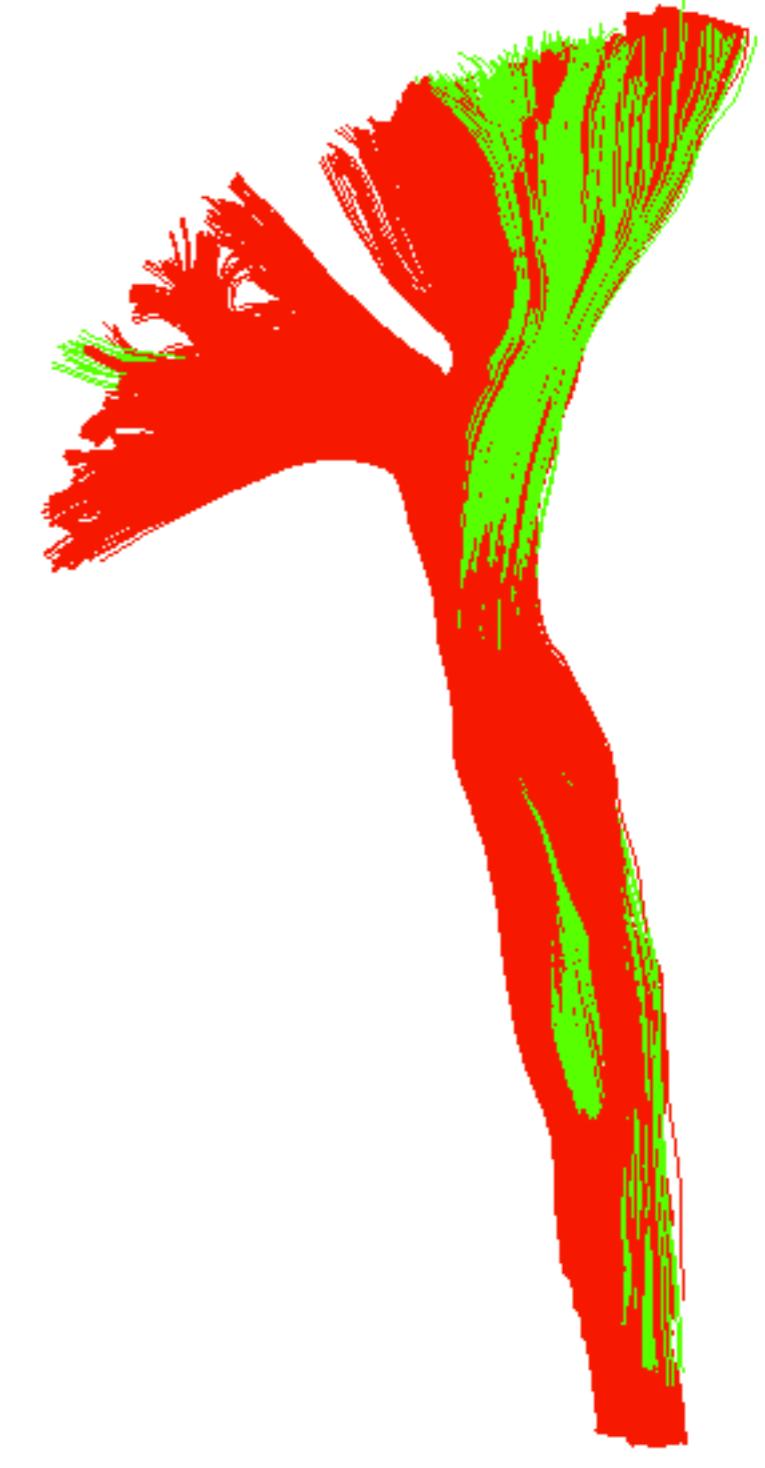
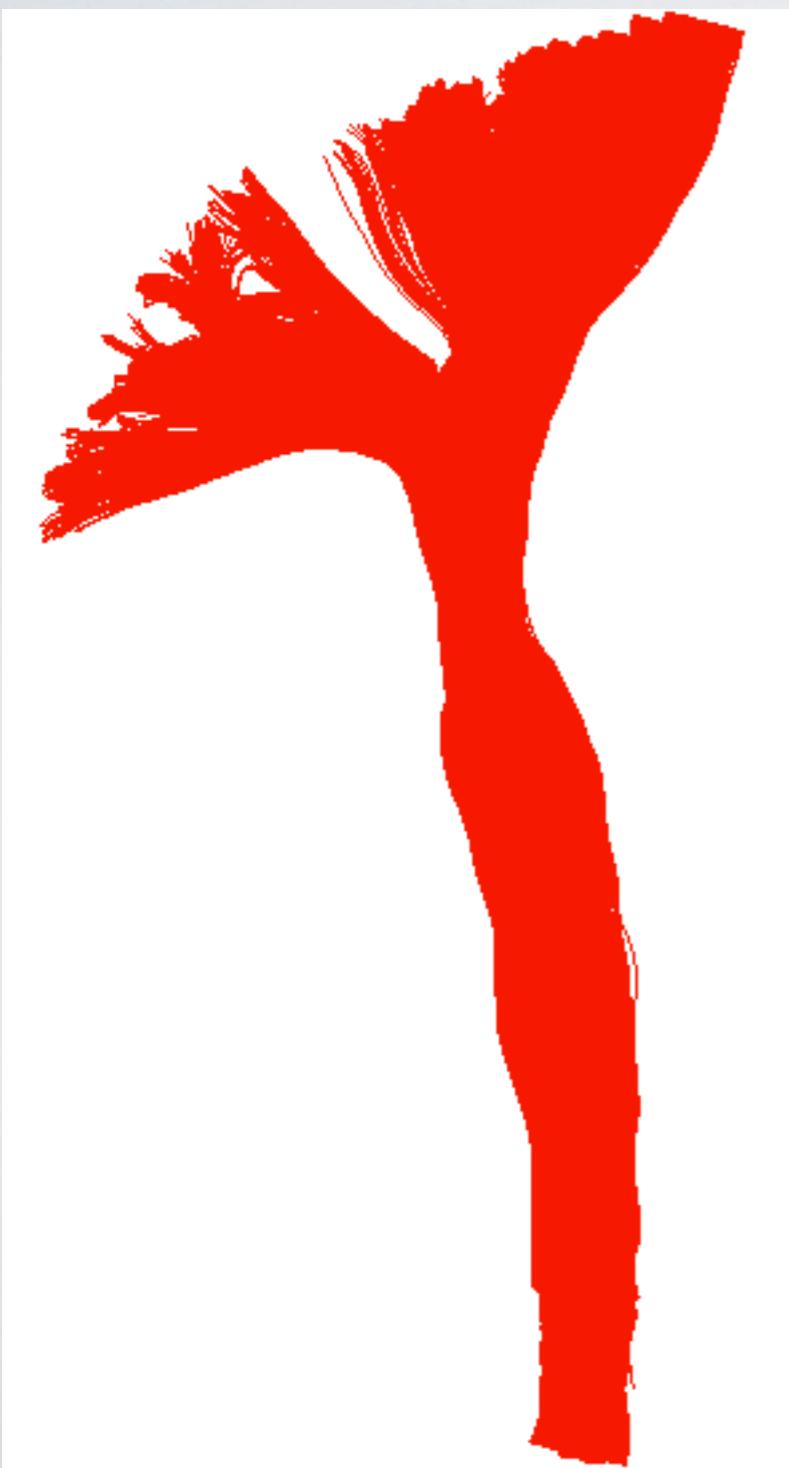
Submitted streamlines
Ground truth streamlines

Small shift
Endpoints difference
Streamline outside bundle mask

Submitted streamlines

Ground truth streamlines

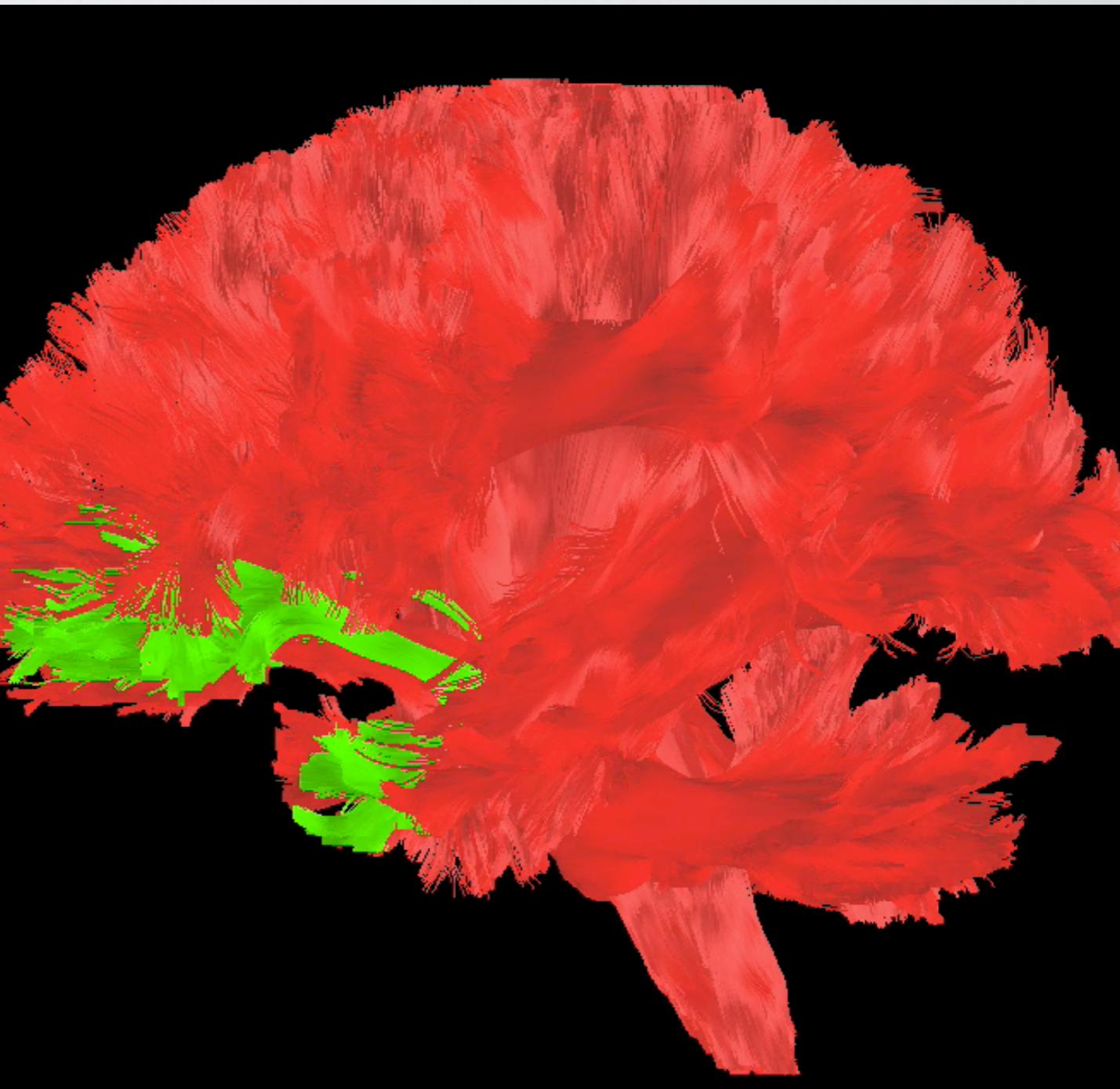
CST EXAMPLE



SO... LET'S FIND THE VALID
BUNDLES IF THEY ARE THERE

Bundle recognition in the space of streamlines

- If the submission has streamlines that are similar to the models:
 - shape (length, geometry)
 - spatial position
 - extract them and flag them as valid connections

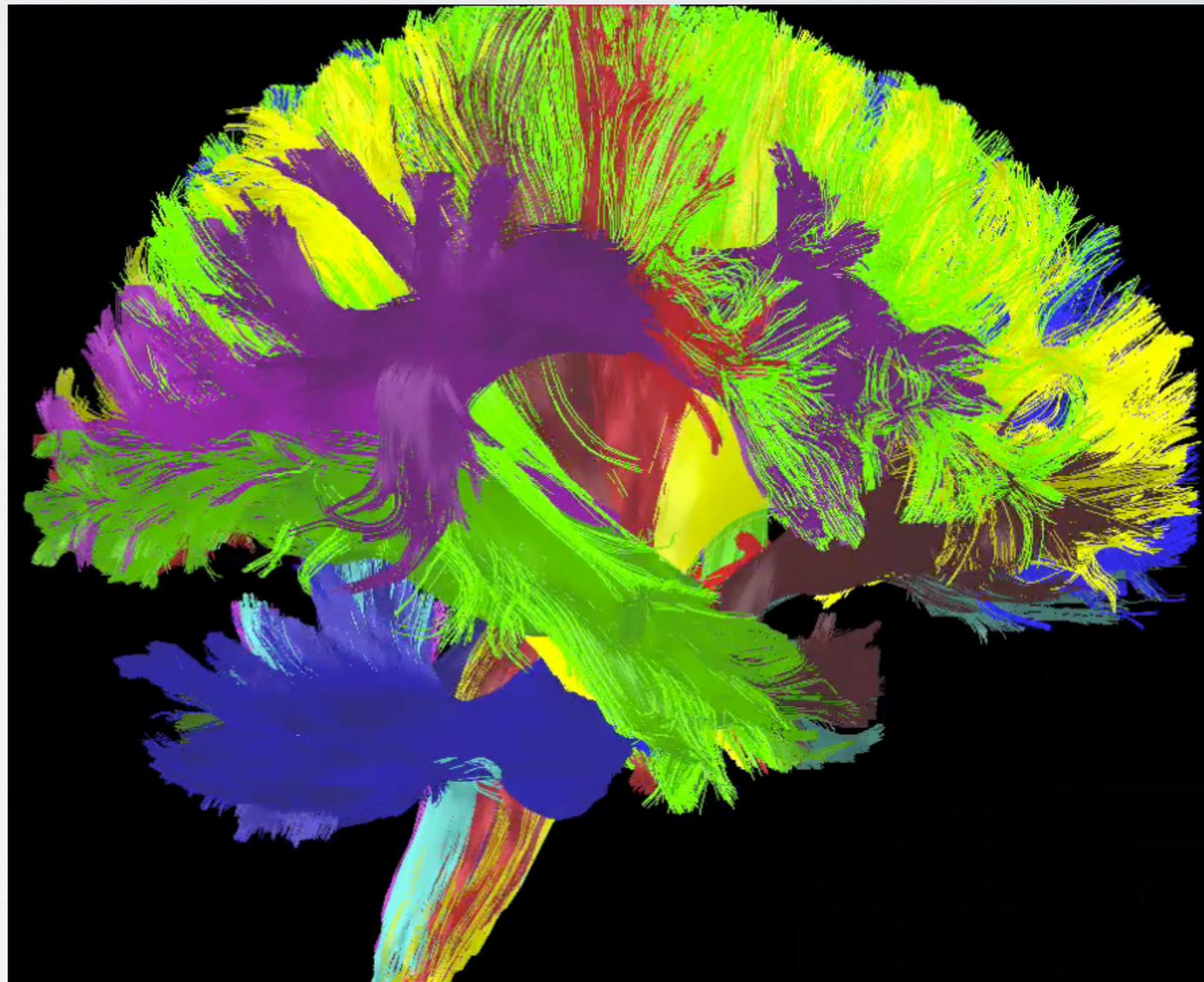


RELAXED EVALUATION

- Auto-extraction of *valid* connections (25 valid bundles)
- Find *invalid* connections (2450 potential invalid bundles)
 - Find *valid* connection but wrong path
- Remaining streamlines are *no* connections

GROUND TRUTH FIBERS

- Perfect results
- 25/25 Valid bundles
- 100 % Valid connections
- 0 % Invalid connection
- 0 % No connection



VALID BUNDLES (VB)

- 1 team got 25/25 valid bundles
3 teams got 24/25 valid bundles
15/20 teams for 23/25 valid bundles
= 60-80% of valid connections
- Average across 96 submissions = 21 valid bundles +- 4 VB
= 54% valid connections +- 24%
= 37% overlap +- 15%
= 280 invalid bundles (20% IC)
- Conclusion:
Most teams find the valid bundles but also find many invalid

MOST DIFFICULT BUNDLES

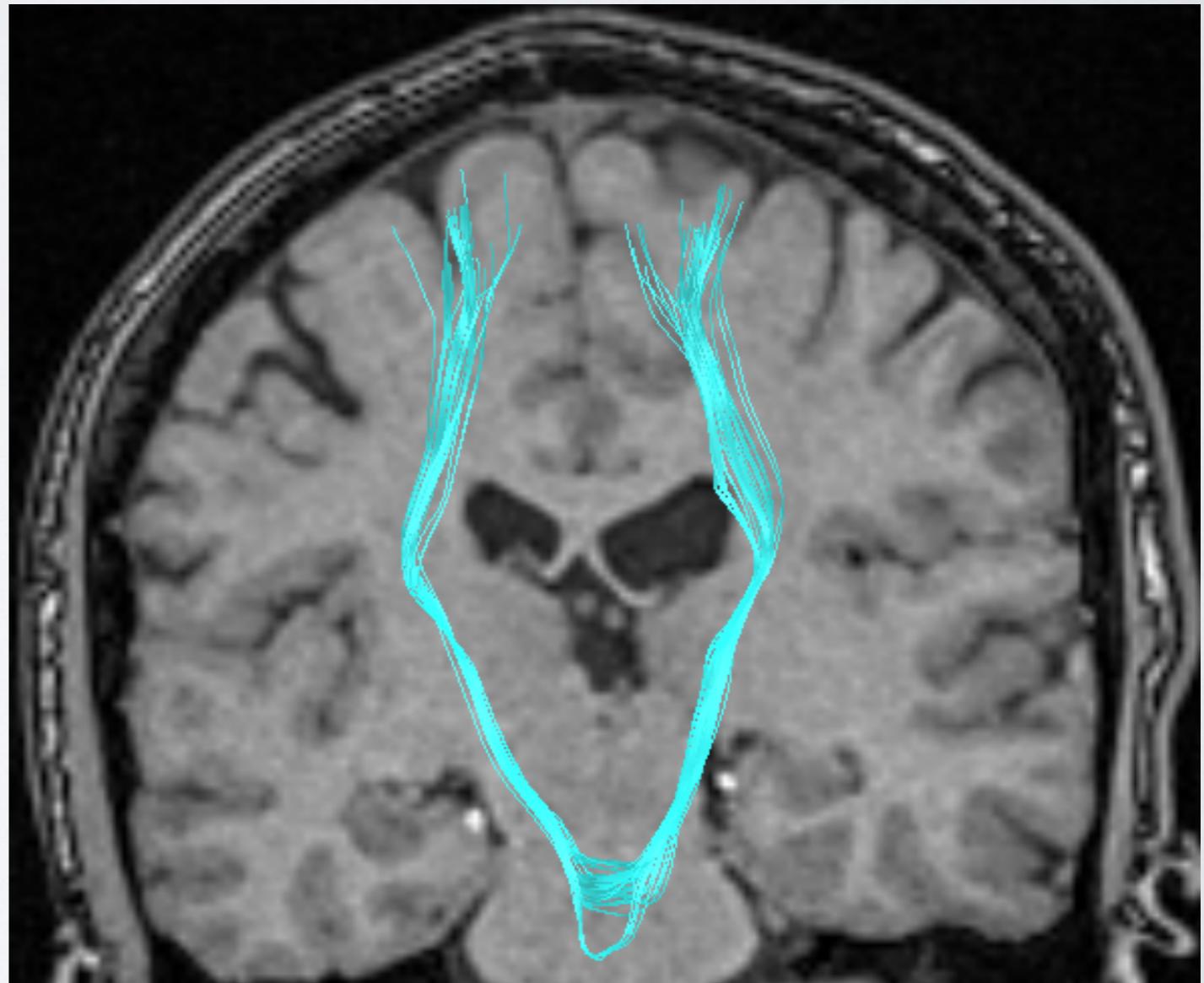
- Anterior commissure (CA) - Posterior commissure (CP)
- 1-2 voxels (1mm space) wide funnel



WHAT ABOUT THE NON-VALID
CONNECTIONS?

VALID CONNECTIONS BUT WRONG PATH

- **Almost none.**
- Less than 1% on average.
- Worst case 2%



INVALID CONNECTIONS / NO CONNECTIONS

- 20-60% of connections are invalid or no connections
- **This is where there are most differences across pipelines**
- About half invalid connections - half no connections
- If submissions specifically tried to remove the no connections,
they are still there! why?
 - **The mask chosen if off. Never perfectly aligned with the ground truth masks (even a dilated version)**

A FEW EXAMPLES AMONG
THE SUBMISSIONS....

VC : valid connections

VB : valid bundles

VCWP : valid connections wrong path

IC : invalid connections

IB : invalid bundles

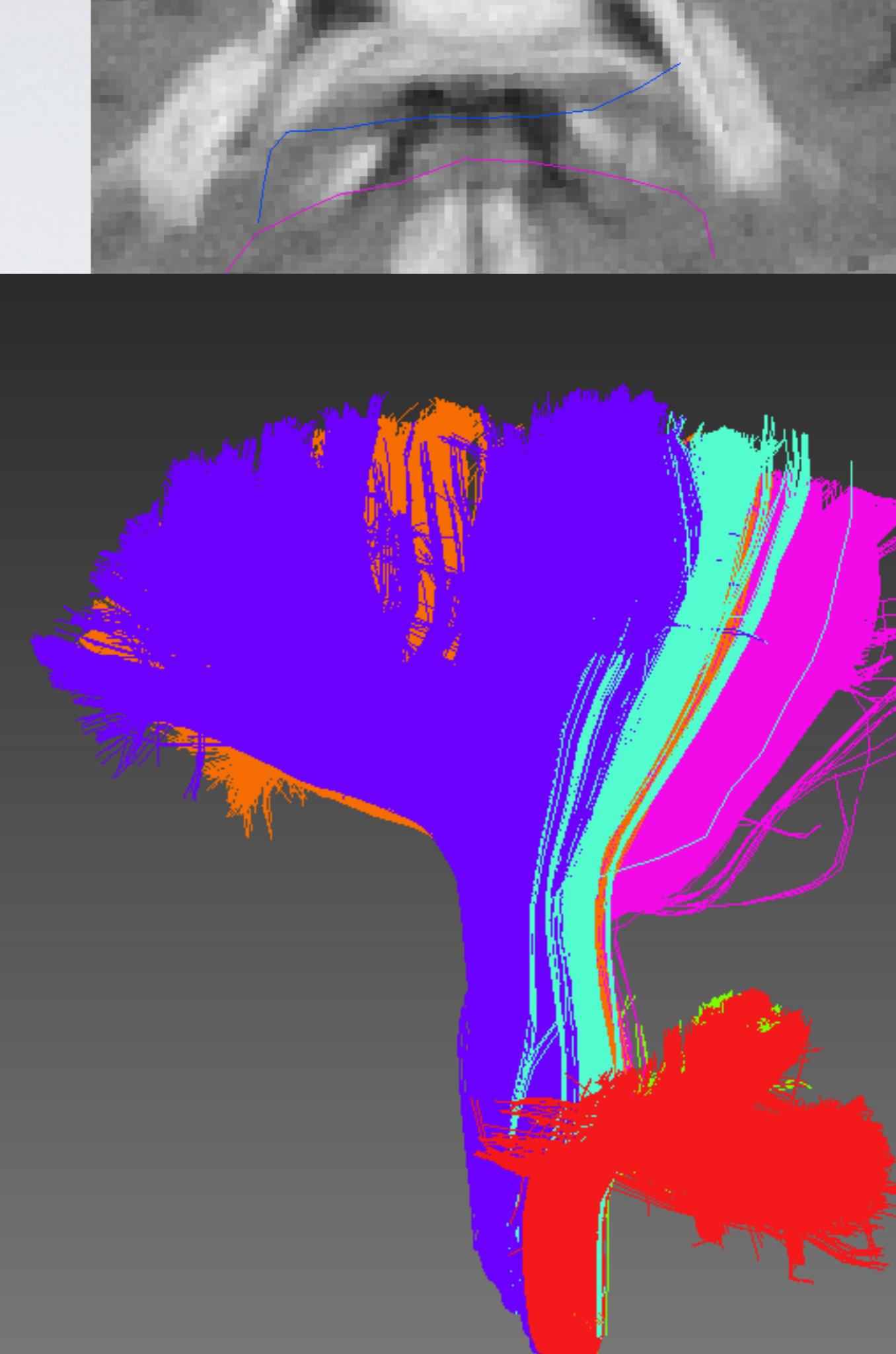
NC : no connections

BEST OF

- **Most number of valid bundles 25/25**
 - 25/25 VB, 53% VC but 1% VCWP, 7% IC, 465 IB, 39% NC
40% overlap +- 20%
- Pipeline:
 - motion, bvec rotation, distortion, denoising, upsampling
 - FODs, $1 \times 1 \times 1$ space
 - wm seeding/masking, 2M fibers, deterministic tractography
 - quickbundle clustering

25/25 VB

- 1-2 streamlines in CA, CP. Luck?
 $< 1\%$ overlap
- Large extent of bundles are found
 - full curving of the cingulum
 - fanning of the projection fibers
- Overlap with ground truth at 40%
- Large number of Invalid Bundles and invalid connections



VC : valid connections

VB : valid bundles

VCWP : valid connections wrong path

IC : invalid connections

IB : invalid bundles

NC : no connections

BEST OF

- **Most number of valid connections 90%**
 - **20/25 VB** and 90% VC but 9% IC, 117 IB, 1% NC,
30% overlap
- Pipeline:
 - motion correction, upsampling
 - ODF from generalized q-sampling, $|x|x|$ space
 - wm seeding/masking, deterministic tractography
 - **clustering 50 large clusters kept, small clusters removed**
 - Least number of no connections

GOOD VALID/INVALID RATIO

- Many valid connections, few invalid, but **weak overlap**, ~30%
- Valid bundles are sparse, not dense, not filled in



Submitted streamlines

Ground truth streamlines

VC : valid connections

VB : valid bundles

VCWP : valid connections wrong path

IC : invalid connections

IB : invalid bundles

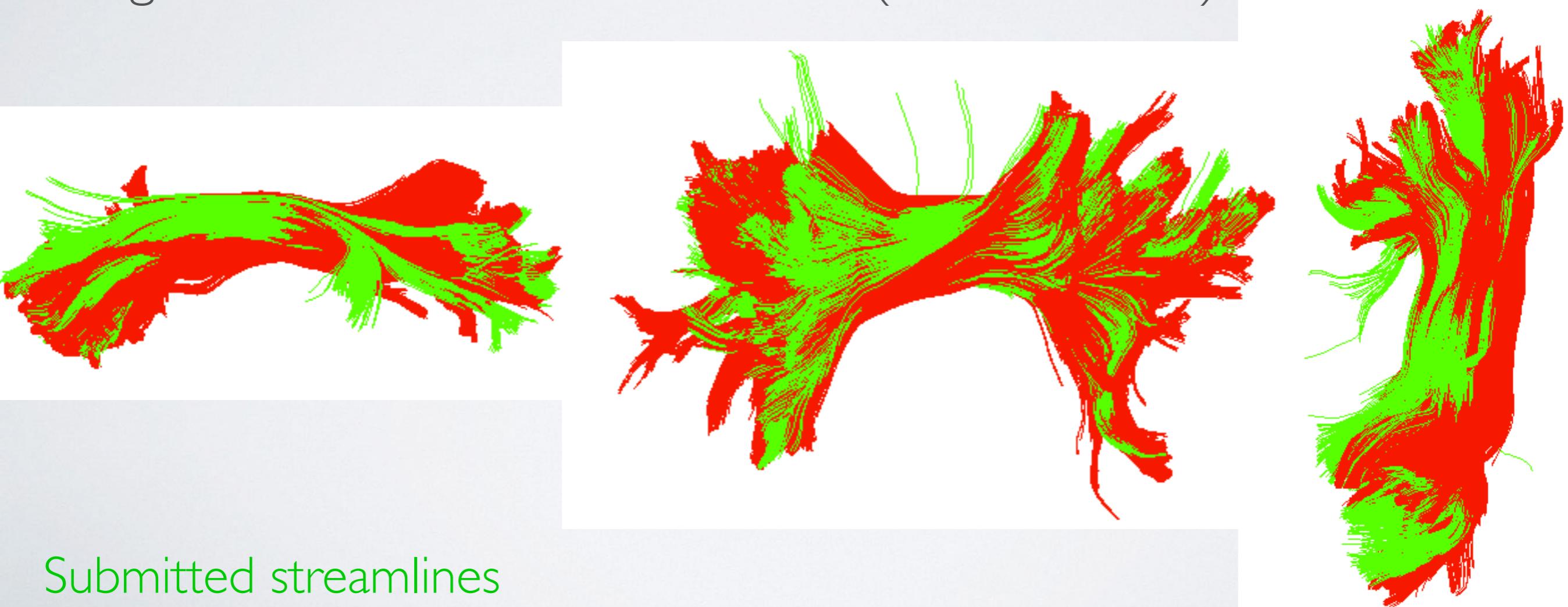
NC : no connections

BEST OF

- **Least number of invalid connections %7**
 - **23/25 VB** and 89% VC but 7% IC, 99 IB, 4% NC
overlap 30% +- 15%
- Pipeline:
 - Denoising LPCA, eddy/topup, motion correction
 - **DTI!**
 - Deterministic tractography
 - **Semi-automatic dissection**
 - **Clean up with anatomy**

LOW INVALID CONNECTIONS PRODUCED BY DTI

- Big association bundles are found (SLF, IOFF, OR)



LOW INVALID CONNECTIONS PRODUCED BY DTI

- Big association bundles are found (SLF, IOFF, OR)
- Most fanning projection bundles are missed (CST, ICP, FCP)
-



Submitted streamlines

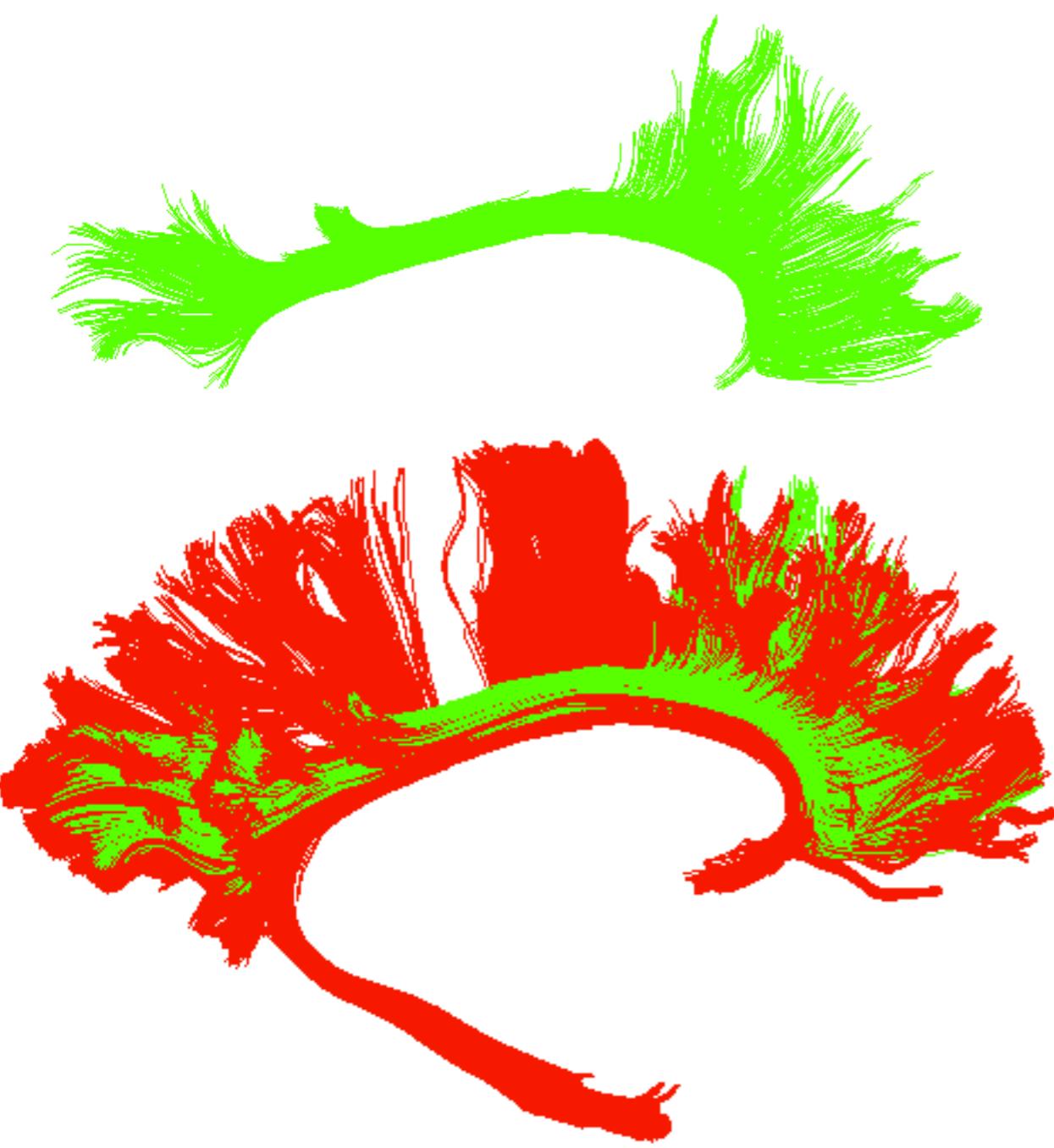
Ground truth streamlines

LOW INVALID CONNECTIONS PRODUCED BY DTI

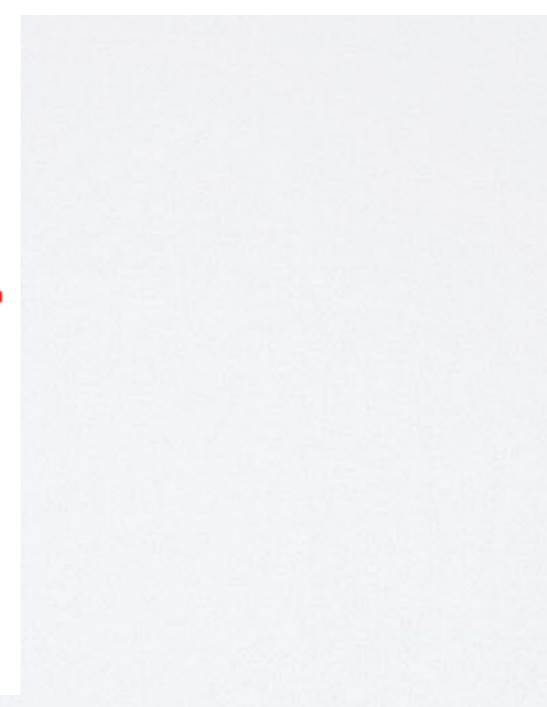
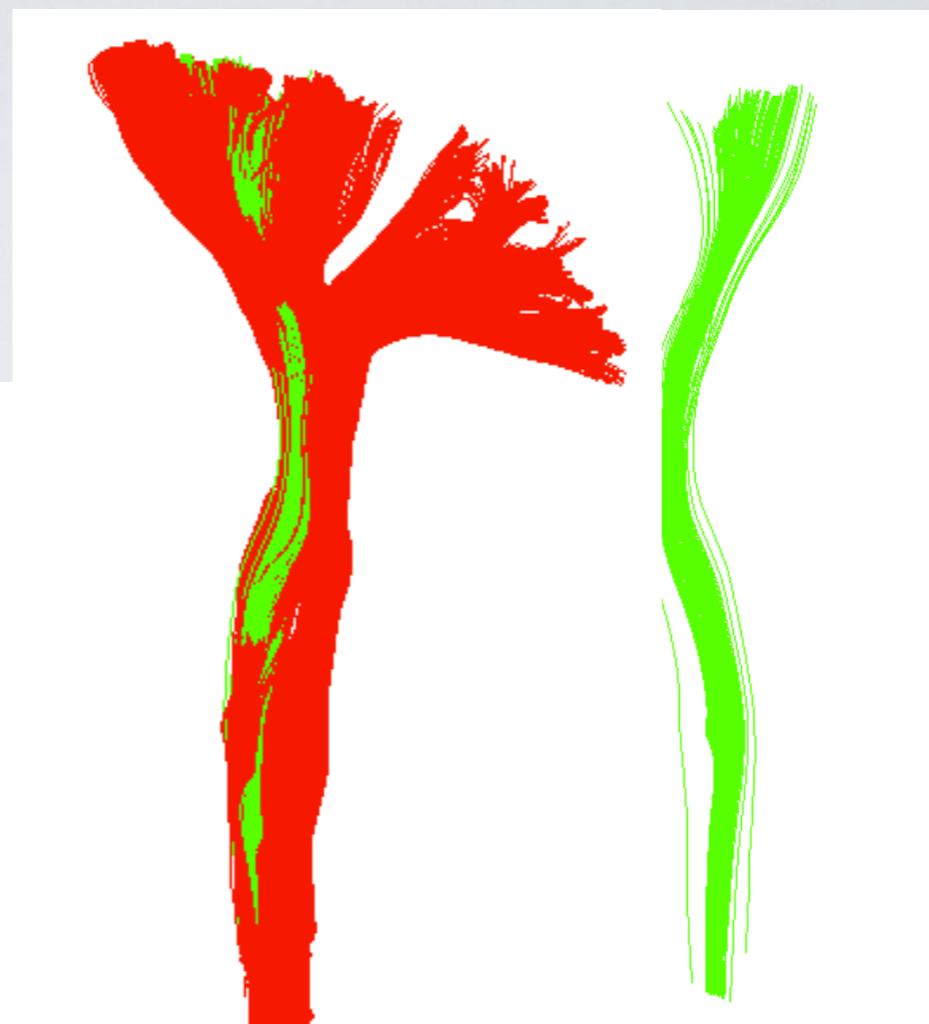
- Big association bundles are found (SLF, IOFF, OR)
- Most fanning projection bundles are missed (CST, ICP, FCP)
- Bundles found tend to be shorter, not have the full extent of the fanning endpoints and also are challenged by high curvature (Cingulum)

Submitted streamlines

Ground truth streamlines



Submitted streamlines
Ground truth streamlines



DTI VS HARDI

- HARDI produces more filled up bundles
 - Both at the endings and inside the bundle
 - Able to get the fanning of bundles
 - High curvature
 - More invalid connections
- DTI produce good Valid/Invalid ratio
 - Bundles less filled in, less curving & fanning

VC : valid connections

VB : valid bundles

VCWP : valid connections wrong path

IC : invalid connections

IB : invalid bundles

NC : no connections

BEST OF

- **Most overlap with ground truth bundles, 84%**
 - **23/25 VB**, 45% VC but 31% IC, **623 IB**, 2% VCWP, 22% NC
overlap **84%** +- 8%
- Pipeline:
 - Unspiking, motion/distortion correction, denoising
 - $I \times I \times I$ space, fiber ODF
 - **Probabilistic tractography**
 - Clean up with TI priors

SPACE 1X1X1 OR 2X2X2?

- $1 \times 1 \times 1$ space finds
 - slightly more valid connections
 - slightly less invalid connections

DETERMINISTIC VS PROBA

- Probabilistic tractography finds
 - Large overlap with ground truth
 - Twice as much invalid connections

NO PRE-PROCESSING

- DTI pipeline (MRtrix)
 - Original evaluation
VB: 13, VC: 4%
IB: 359, IC: 51%
VCWP: 3%, **NC: 42%**
 - Rexaled evaluation
VB: 22 - VC: 34%
IC: 27.54% - IB: 365,
VCWP: 1.75%, **NC: 36.36%**

NO PRE-PROCESSING

- HARDI pipeline (MRtrix)

- Original evaluation

VB: 17, VC: 2%

IB: 590, IC: 37%

VCWP: 4%, **NC: 57%**

- Rexaled evaluation

VB: 23 - VC: 17%

IB: 496, IC: 27%,

VCWP: 3%, **NC: 53%**

VC : valid connections

VB : valid bundles

VCWP : valid connections wrong path

IC : invalid connections

IB : invalid bundles

NC : no connections

BEST OF

- **Least number of invalid bundles 75 / 2450**
 - **14/25 VB** and 87% VC but 8% IC, 75 IB, 5% NC,
20% overlap +- 10%
- Pipeline:
 - denoising, motion, distortion correction
 - outlier rejection, REKINDLE
 - **DTI**
 - post-processing with gray matter from TI

CONCLUSIONS

- Best scores come from the following pipelines:
 1. State-of-the-art pre-processing
 2. Tractography in $|x| \times |y| \times |z|$ space
 3. Post-processing clean up (clustering, $|T|$ priors)
- **Trade-off between valid and invalid:**
 - Highest valid connections and bundles, overlap => HARDI
 - Lowest invalid connections and bundles => DTI

CONCLUSIONS

- Relaxed evaluation because
 - **None of the submissions are voxel-perfectly aligned**
 - **ALL have a shift and mostly in the A-P direction**
- **!!Warning!!**
for connectomics/tractometry, parcellation of TI, counting tracts in ROIs defined from TI

CONCLUSIONS

- More to come... Need to further explore the results, visualize them and compare them...
- At first glance, biggest improvements come from
 - careful pre-processing and artefact correction
 - &
 - careful post-processing and manual dissection

DISCUSSION

We need a **unified tract format** &
more explicit **reference frame** &
a convention

DISCUSSION

- How to improve for the next challenge?
 - Voxel-based evaluation too restrictive
 - Need a cortex-white matter ground truth parcellation
 - More datasets
 - eddy currents, more b-values, directions
 - More bundles (U fibers, Medial forebrain, etc)

DISCUSSION / DEBATE

- What is the bottleneck for tracking algorithms?
 - Distortions?
 - Motion?
 - Noise?
 - Resolution?
 - Number of directions? b-values? multishell?
 - Where do we go from here?

DISCUSSION

- Publications
 - No participation
 - Anonymous
 - Full participation