

Supplementary materials

Pyramidal tract - Internal Capsule (IC)

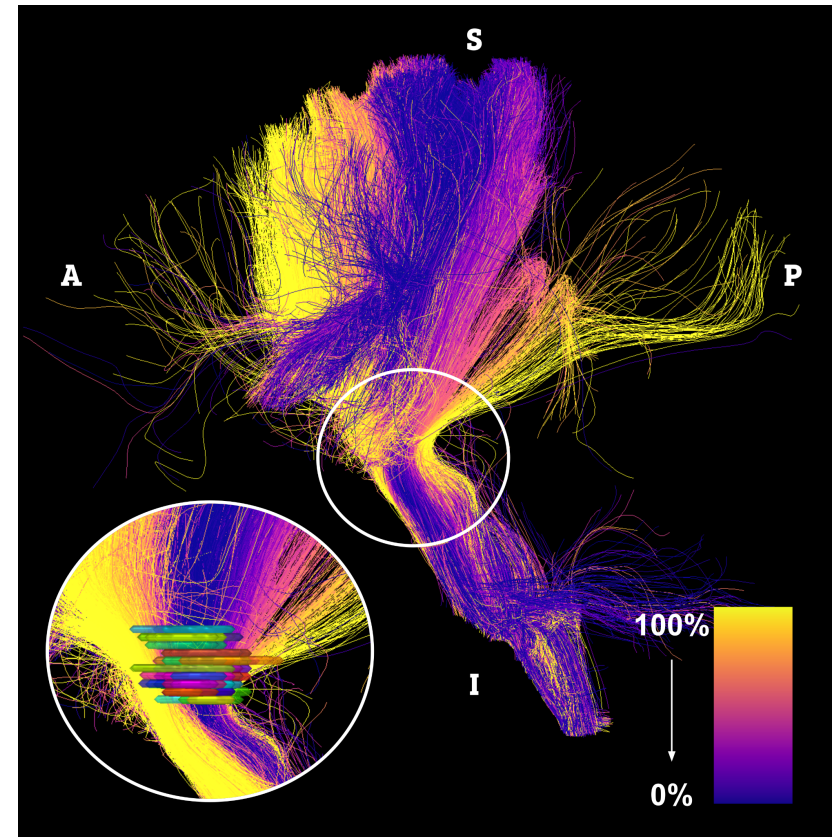
The internal capsule (IC) region of interest specifically targets the genu, which is the point of passage for the motor and somatosensory that originate from the cortex. This specific region acts as a stem in which the pathway of interest must go through.

Raters were expected to identify the ROI using an approximate location (in the I/S axis) from which they could locate the thalamus. Using the barycenter of the thalamus, they had to find a hypersignal in the b0 image (due to high myelination on top of a blue region on the RGB).

The identification of the thalamus and its approximate barycenter will lead to variability in the I/S axis. When the IC region itself is found, the genu of the IC can be difficult to identify due to the poor contrast of the b0 and the diffuse blue region on the RGB. This leads to variability in L/R and A/P directions as well as variation in the size (radius) of the ROI to draw to encompass the IC.

Deviations from the true IC genu lead to a decrease in density of streamlines and often a loss of fanning in the pre/post-central (medial regions). Large deviation results (positioning the ROI in the anterior or posterior limb of the IC) can result in segmenting a different pathway.

The difficulty to identify the I/S axis did not lead to major variability, but the L/R and A/P axis. Increasing the recommended size of the ROI could reduce variability. Instead of trying to exactly identify the genu of the IC, identify the IC as a whole and relying on other regions (already part of the protocol) to constrain the pathways would improve reproducibility.



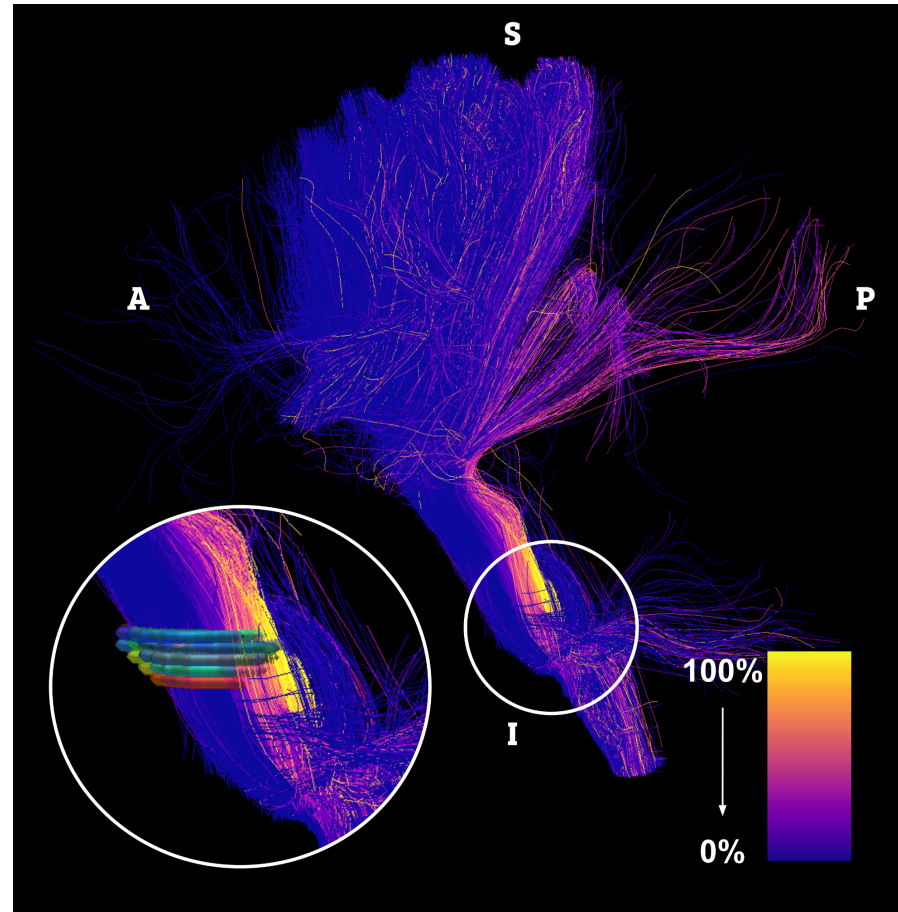
Pyramidal tract - Midbrain (MB)

This ROI is used to discriminate the PYT (motor pathway) located in the anterior part of the midbrain from the lemniscus tract (sensory pathway) located in the posterior part of the midbrain.

An approximate position (in the I/S axis) was recommended as a starting point in the brainstem, then the exact position of this ROI was located using the RGB color map and hypersignal in blue.

Since this ROI is located in between the IC and MO, there was relatively small variability in the I/S axis and the most anterior blue hypersignal was clearly distinguishable. The ROI had to encompass the blue hypersignal, this variable was clear to determine. Raters located and drew the region with high consistency.

Variation in position and size had very small effects on the resulting bundles. Overall, the spatial extent of the bundle is mostly unchanged by the small disagreements between raters' ROI.



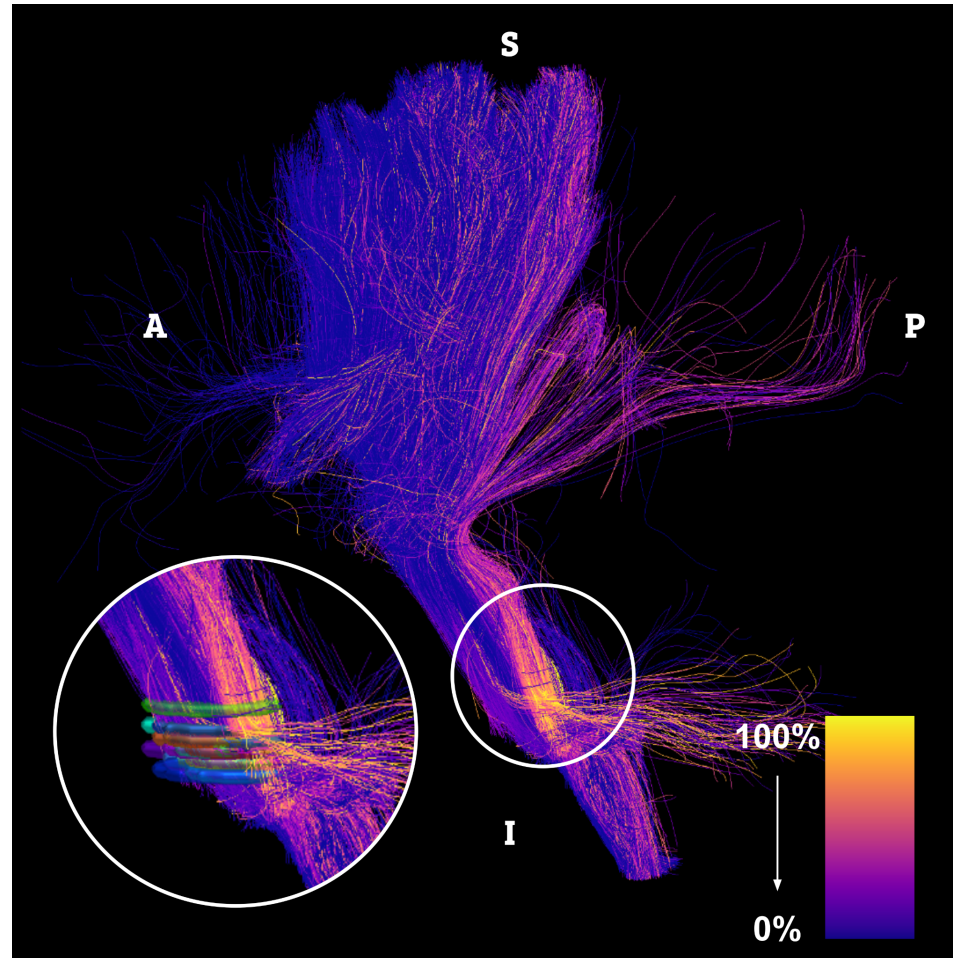
Pyramidal tract - Medulla Oblongata (MO)

This ROI is used to select fibers entering in (or going into) the medulla while not passing through the cerebellum. This ensures the selected pathways do not end in the pons and continue down the brainstem.

An approximate position (in the I/S axis) was recommended as a starting point in the brainstem. To position the ROI in the axial plane the RGB was used to locate where the white matter of the medulla oblongata (rather in blue) separate from the white matter of the cerebellum. Then, the ROI is centered on the blue signal located in the anterior part of the Medulla Oblongata.

Due to the clear landmark and approximate starting position, the positioning of the ROI was consistent. Only small effects can be observed on the resulting bundles. Overall, the spatial extent of the bundle is mostly unchanged by the small disagreements between raters' ROI.

However, there is a potential algorithmic pitfall for this region (which did not occur in our study). Tractography often abruptly ends just after the medulla (due to partial volume effect and a bend in the spinal cord). This means that while tens of thousands of streamlines can reach a ROI positioned just above the medulla, only a few thousands will likely reach a ROI positioned just below it. Anatomically speaking, the pathway is expected to go beyond the medulla, but there is a balance between tractography reconstruction and anatomical accuracy. Positioning this ROI too low will result in a major loss of density, which will affect the spatial extent and likely the lateral fanning of the PYT.



Pyramidal tract - Mid-sagittal plane

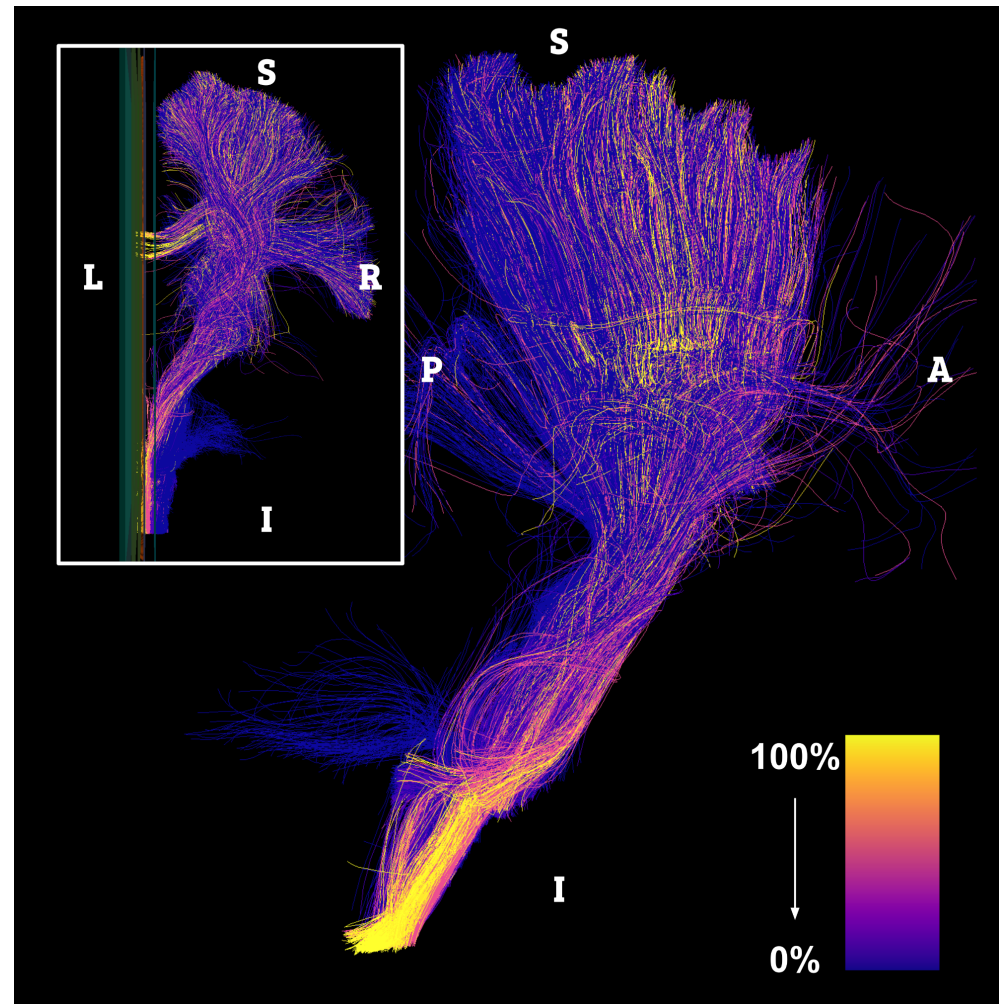
This region aims to prevent spurious streamlines from switching paths into the corpus callosum (outliers) and into the other hemisphere.

Raters were expected to identify the interhemispheric fissure and select the entire plane as a ROI.

There was little potential confusion for this region, the observed variability was only in the L/R axis. However, by selecting the whole plane if raters selected a slice too lateral it is possible to cut the brainstem. In the current dataset, this kind of error was rarely observed and only had a limited impact.

To avoid this potential source of variability, it would be wise to use a hand-drawn ROI instead of using an entire plane. In the case of non-AC-PC aligned datasets, this small modification could avoid a major loss of streamlines by accident.

As shown in the figure, the variability comes mainly from the brainstem and only affects a few commissural outliers. While useful to remove undesired outliers, the ROI creates variability in an unexpected region.



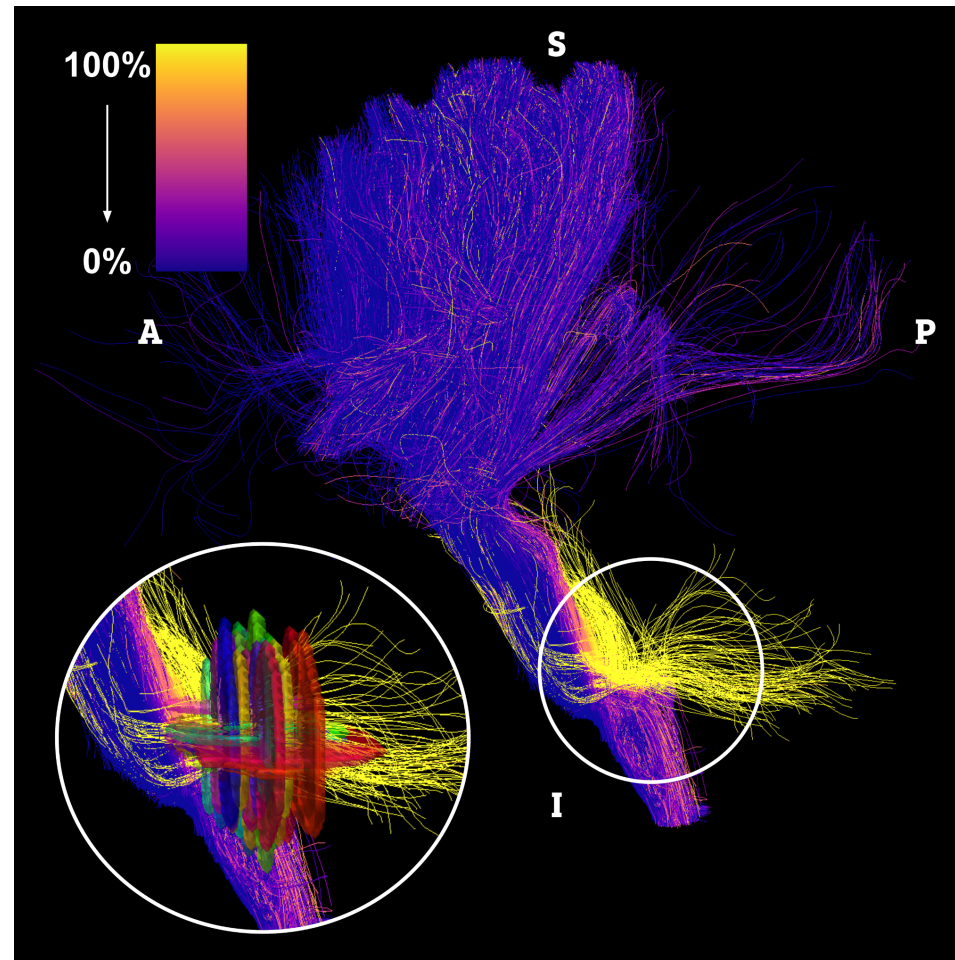
Pyramidal tract - Cerebellum cleaning

This region aims to prevent spurious streamlines from entering the cerebellum and avoiding contamination from the lemniscus tract (sensory pathway) located in the posterior part of the midbrain.

This is the only '2D' ROI of the protocol (two intersecting disks). The raters were asked to draw a large (axial) disk behind the MO-ROI on top of a blue signal on the RGB (to block the posterior part of the midbrain). Then, a second (coronal) disk had to be drawn to cut off the entrance of the cerebellar peduncle without affecting the streamlines reaching down to the spinal cord.

While the intent of the ROI is intuitive, such an exclusion region in the brainstem can be risky. If too large or placed too low, the coronal disk is at risk of intersecting the lower brainstem. This would severely impact the density and spatial extent of the bundles. Similarly, if the axial disk is too large or placed too low, it comes near the anterior portion of the pathway and is a risk of preventing the streamlines from reaching the lower brainstem.

For the description, interpretation and placement of the ROI, using two independent disks instead of presenting it as a single ROI made of two intersecting disks would likely decrease variability. It would at least reduce confusion, raters feedback showed that this ROI was perceived as difficult. However, as seen in the figure, the ROI is a low source of variability and the impact of raters' variation only impacts the desired area (cerebellum). Despite its potentially high risk, the ROI was well executed by most raters.



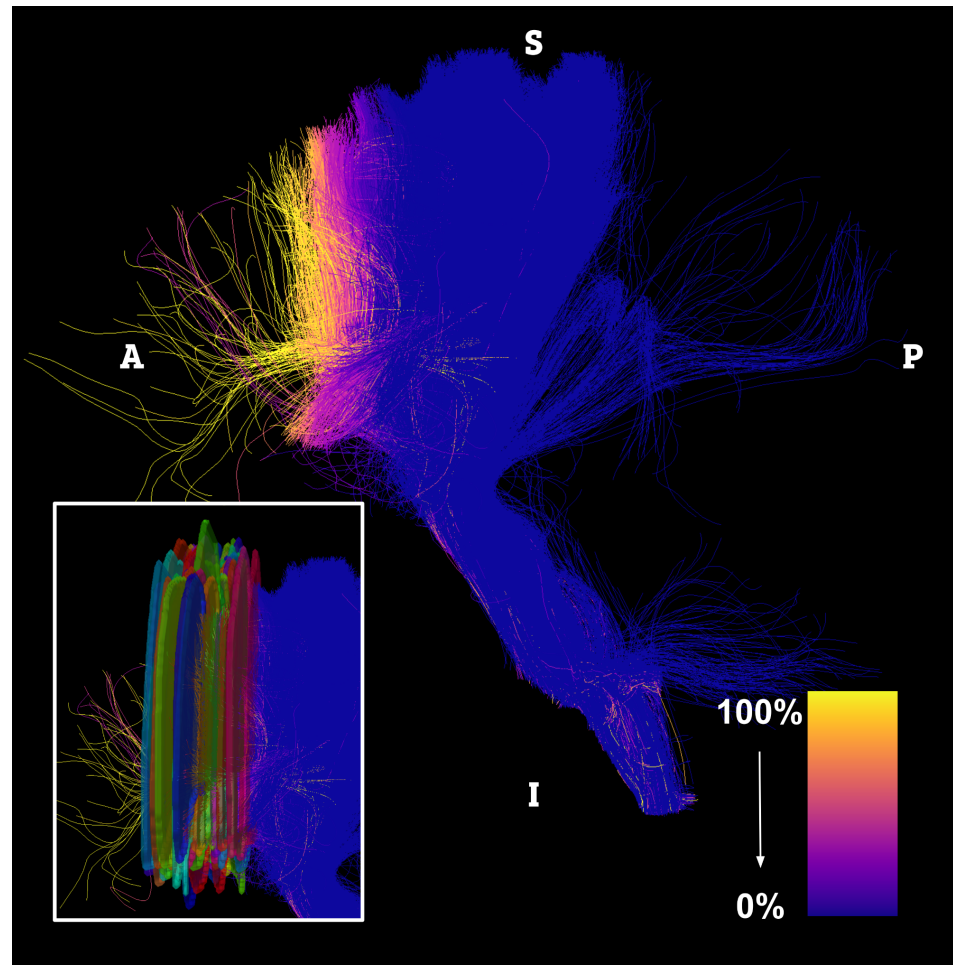
Pyramidal tract - Precentral gyrus

This region aims to prevent streamlines from reaching into the frontal lobe; it acts as an anterior limit to the bundle when reaching the cortex. The ROI is a large planar drawing blocking any streamlines going too much forward.

Since the precentral gyrus has a complex orientation, it would have been difficult for raters to accurately identify the whole gyrus. Instead, the raters had to identify the most anterior limit of the precentral gyrus by first locating where Sylvian fissure starts and then scrolling medially until the lateral sulcus divides both the frontal lobe and temporal lobe. Finally, raters had to draw the planar ROI where the precentral gyrus generally runs in the I/S axis.

The precentral ROI shows little variability in the L/R and I/S axis. However, major variation in the A/P axis can be seen in the figure, despite raters locating the ROIs across more than 20mm distance only small impacts can be observed. Variations in size and shape have a low impact on the final bundles.

As an exclusion ROI, adding an extra margin of error (moving it more anteriorly) would increase reproducibility. However, this ROI is used by the corpus callosum (as an exclusion) and arcuate fasciculus (as an inclusion) and thus this solution may not be optimal for the other bundles.



Pyramidal tract - Postcentral gyrus

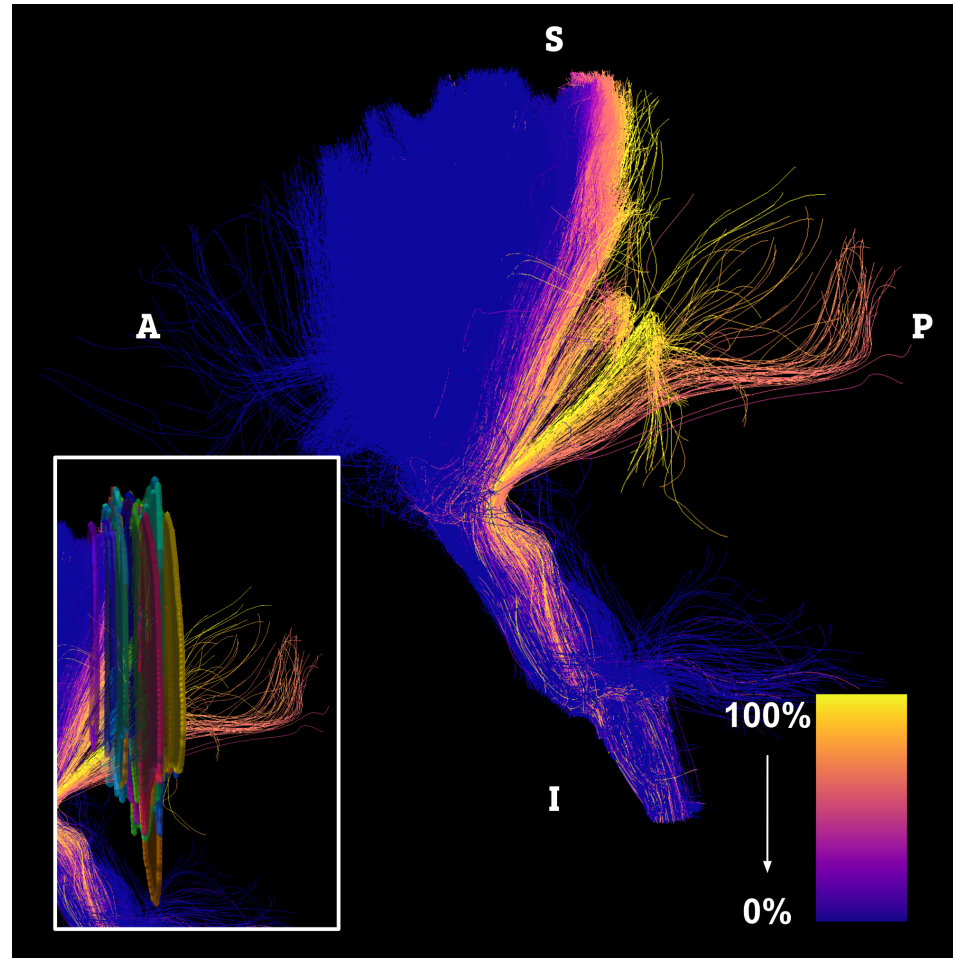
This region aims to prevent streamlines from reaching into the parietal/occipital lobe; it acts as a posterior limit to the bundle when reaching the cortex. The ROI is a large planar drawing blocking any streamlines going too much backward.

Similar to the precentral gyrus, requesting an accurate segmentation of the whole structure would be difficult. To avoid this a single planar ROI had to be positioned at the most posterior portion of the postcentral. Raters were asked to identify the Rolando sulcus, known as the deepest (most medial termination) sulcus seen on the axial slice and that includes the omega shape. Then, the postcentral gyrus was easy to identify (one gyrus back) and from this position, they had to draw the ROI.

The postcentral ROI shows little variability in the L/R and I/S axis. However, major variation in the A/P axis can be seen in the figure (similar to the precentral, around an average 20mm distance). Only small impacts can be observed, variations leading to major differences are rare.

Just like adding an extra margin of error (moving it more posteriorly) could increase reproducibility for the precentral ROI, this could work for the postcentral ROI.

However, this ROI is used by the corpus callosum (as an exclusion) and arcuate fasciculus (as an inclusion) too. In the first version of Tractostorm, it was observed that some raters drew this region too large (reaching as far down as the brainstem) and led to streamlines being blocked from the lower brainstem. Likely due to the description and demonstration from the online course, this error was avoided by the rater. This shows how the information is shown in a didactic course (even small details) leads to an increase in reproducibility.

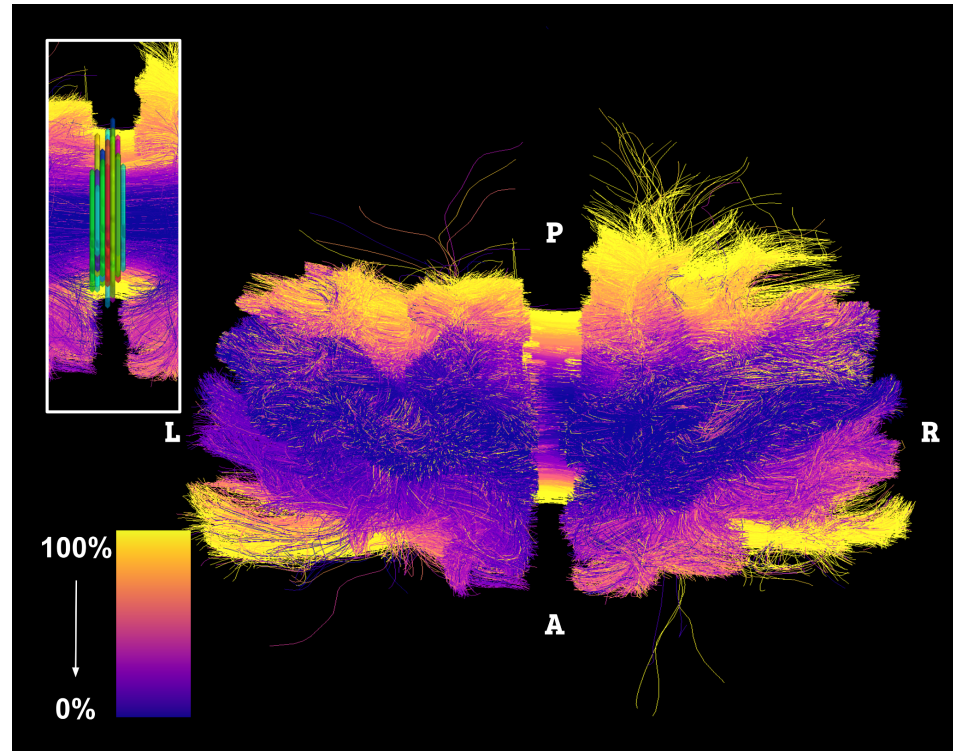


Corpus callosum - Central corpus callosum

This ROI aims to approximate the mid-anterior, central and mid-posterior regions of the corpus callosum. This was made to target streamlines that fan into the precentral and postcentral gyri.

Raters were requested to identify the small rectangular region on the mid-sagittal slices, in between the precentral and postcentral ROI drawn previously. The region was expected to be 20-30mm long rectangle with minimal curvature in the corpus callosum, excluding the genu and the splenium)

The intended 'core' of the bundle was adequately segmented by most raters. Variations in the L/R axis had very little effect. The A/P axis position of this ROI is first approximated by the 'center' between the precentral and postcentral gyri regions, this leads to a bias. Raters that estimated these ROIs too forward or too backward will not only constrain the bundle too little (either anterior or posterior), but will also likely estimate this ROI in the wrong location. Variations in the A/P axis resulted in a bundle that overflowed anteriorly or posteriorly. It was also observed that missed voxels in the I/S axis due to the region being too small resulted in hundreds of streamlines being missed.



For the I/S axis variation, the easiest solution is to enlarge the region to make sure the whole thickness of the corpus callosum is captured. As for the A/P axis variation, the solution is much more complex. First, the ROI could also be enlarged in this axis, but to constrain the bundle a good segmentation of the precentral and postcentral is needed. This is a careful balance, by enlarging this region of interest, more of the 'responsibility' is on the precentral and postcentral regions. As seen in the figure, variations have a large visual impact. However, the obtained bundle generally conserved a stable density, but the spatial extent varied drastically.

Corpus callosum - Lower axial limit

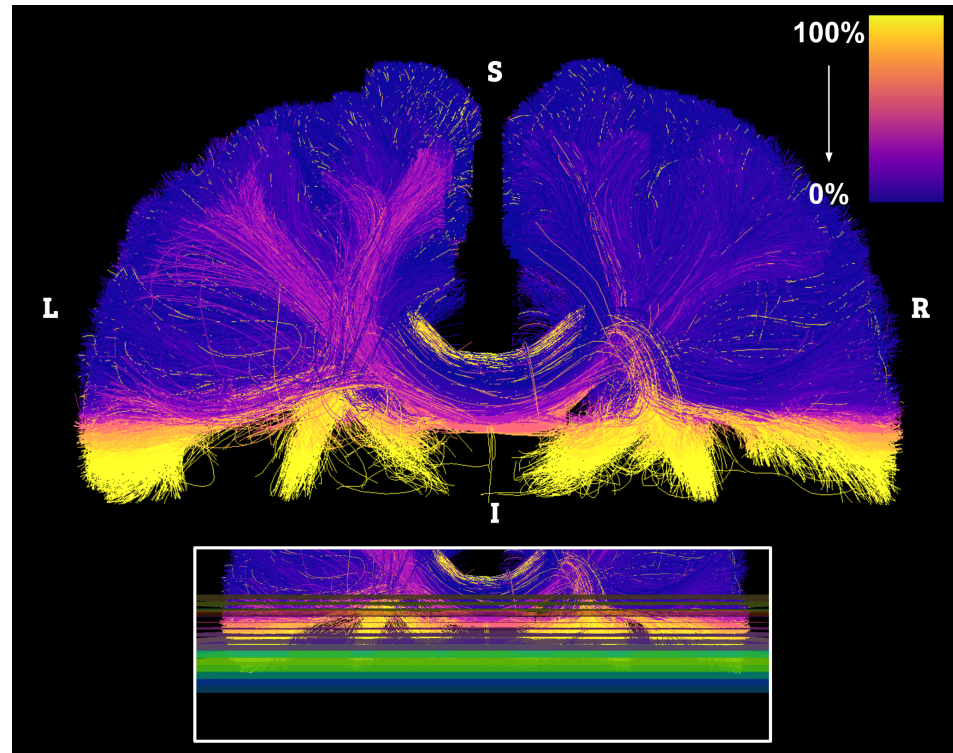
This region aims to prevent spurious streamlines from switching paths into the pyramidal tract (outliers) and from reaching into the deep nuclei.

Raters were expected to identify the lower axial limit using another ROI as a landmark (body of the corpus callosum). The plane is supposed to be located below the ventricle or the Sylvian fissure, whichever is lower.

Depending on the coronal slice selected to evaluate the position of that limit, a lot of variability occurs in the I/S axis. When placed “too high” the ROI would intersect fanning in the lateral portion of the corpus callosum.

To avoid this potential source of variability, it would be wise to use a hand-drawn ROI instead of using an entire plane. By limiting the ROI to only the region immediately below the ventricles, this would preserve lateral fanning. In the case of non-AC-PC aligned datasets, this small modification could avoid a major loss of streamlines by accident.

As shown in the figure, despite high variability in the I/S axis only a small portion of the corpus callosum is impacted. The ROI could also be drawn lower by adding a few millimeters as a margin of error. This would still eliminate the outliers switching into the pyramidal tract while reducing the variability observed at the cortex.



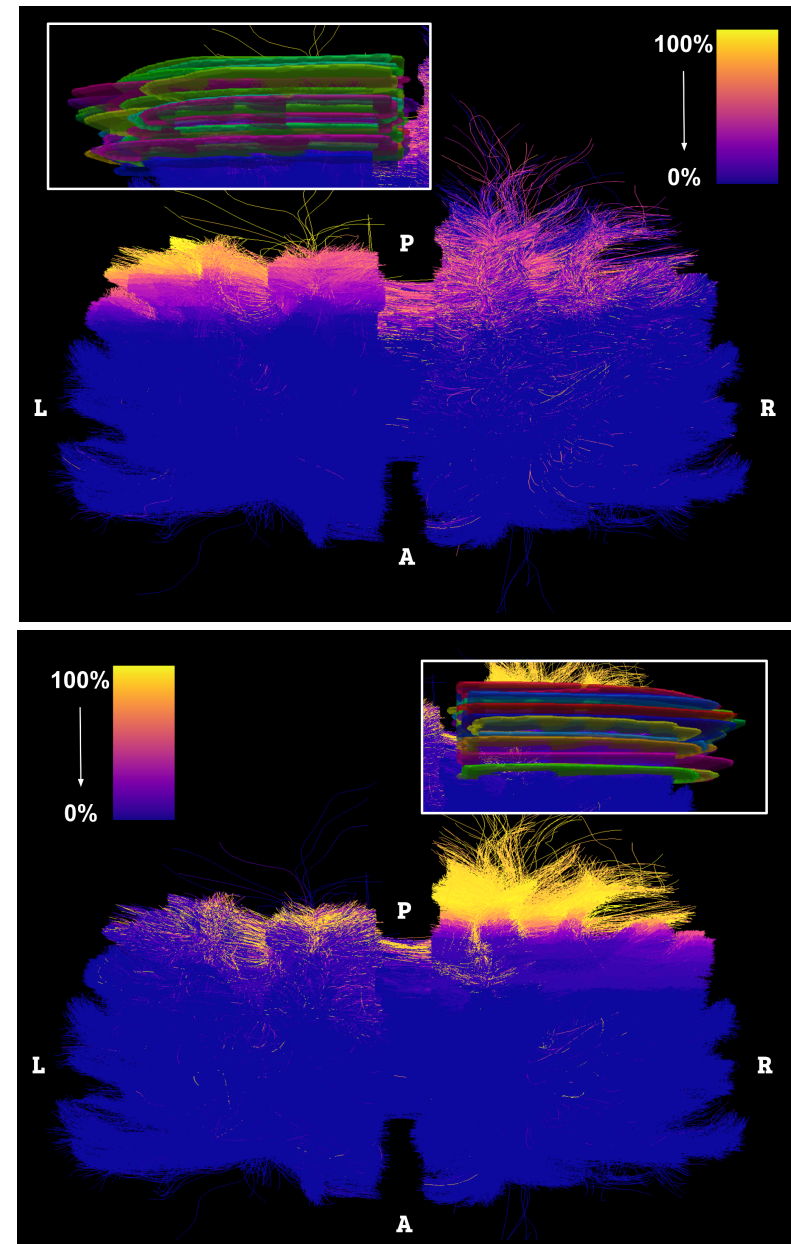
Corpus callosum - Precentral gyrus (L/R)

These regions aim to prevent streamlines from reaching into the frontal lobe; it acts as an anterior limit to the bundle when reaching the cortex. The ROI is a large planar drawing blocking any streamlines going too much forward.

The precentral gyrus ROI on the left is the same as the one used for the pyramidal tract. The goal was to reuse as much as possible to decrease the total duration of the protocol. The ROI on the right is obtained by following the same instructions, but applied to the right hemisphere (for more details, see above).

Similar to the pyramidal tract, variability in the I/S axis causes little variation in the final bundles. Despite raters locating the ROIs across more than 20mm distance, only small impacts can be observed.

The corpus callosum dissection uses the same ROIs (and its counter-hemispheric analog) as an exclusion ROI (just like the pyramidal tract). Adding an extra margin of error (moving it more anteriorly) would increase reproducibility without decreasing the anatomical accuracy of the segmented bundles.



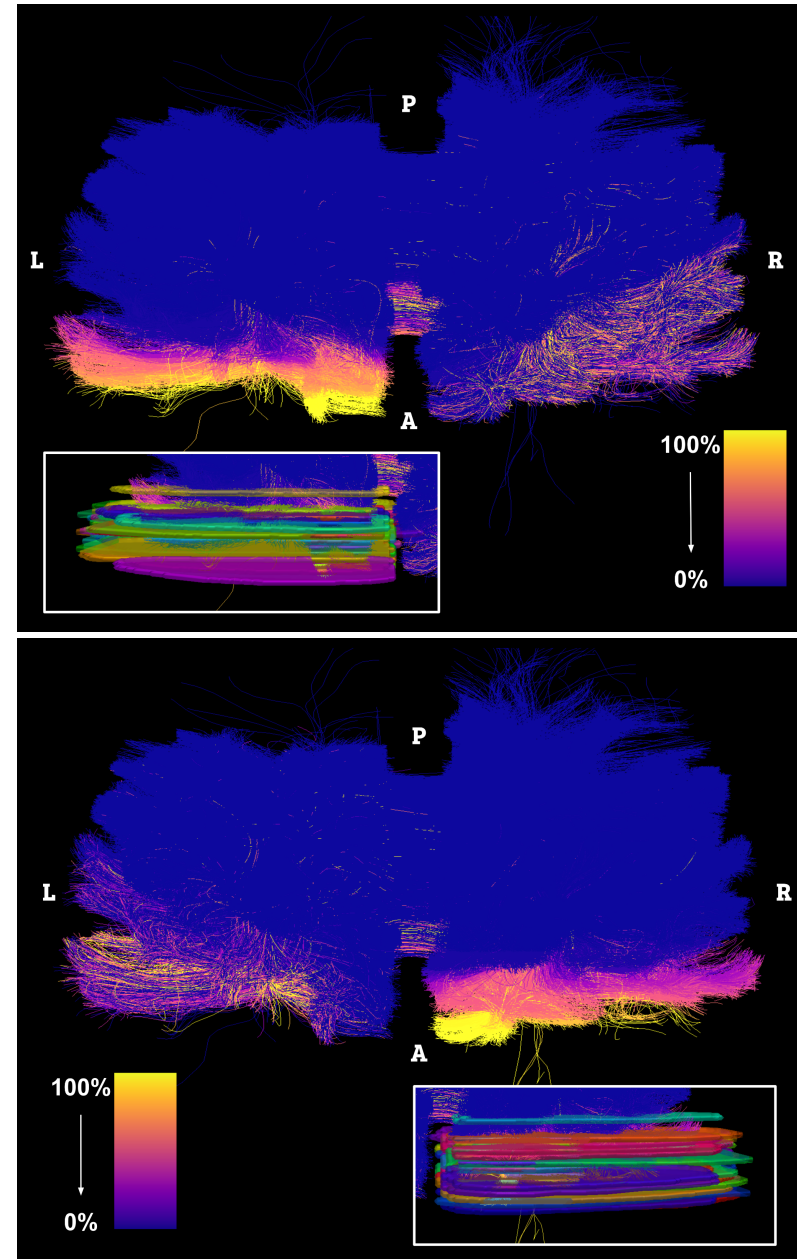
Corpus callosum - Postcentral gyrus

This region aims to prevent streamlines from reaching into the parietal/occipital lobe; it acts as a posterior limit to the bundle when reaching the cortex. The ROI is a large planar drawing blocking any streamlines going too much backward.

The postcentral gyrus ROI on the left is the same as the one used for the pyramidal tract. The goal was to reuse as much as possible to decrease the total duration of the protocol. The ROI on the right is obtained by following the same instructions, but applied to the right hemisphere (for more details, see above).

Similar to the pyramidal tract, variability in the I/S axis causes little variation in the final bundles. Despite raters locating the ROIs across more than 20mm distance, only small impacts can be observed.

The corpus callosum dissection uses the same ROIs (and its counter-hemispheric analog) as an exclusion ROI (just like the pyramidal tract). Adding an extra margin of error (moving it more posteriorly) would increase reproducibility without decreasing the anatomical accuracy of the segmented bundles.



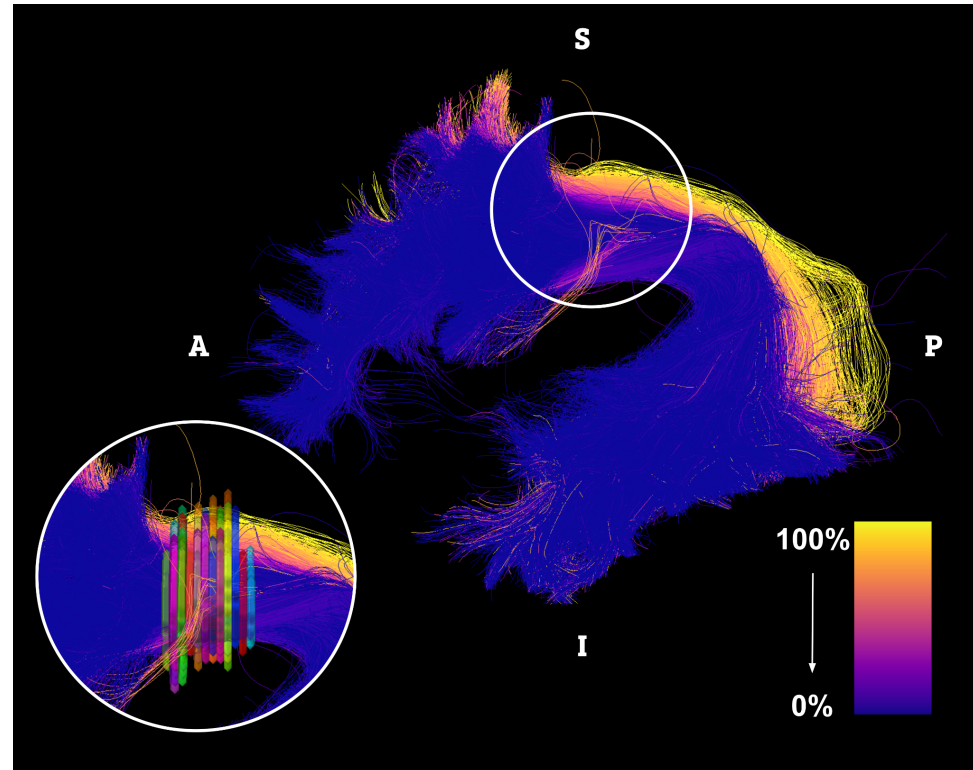
Arcuate fasciculus - Centrum semiovale

This region aims to force streamlines to funnel into the portion of the centrum semiovale close to the precentral and postcentral gyri. It acts as a bottleneck for the bundle before fanning into the frontal lobe. The ROI is a large disk that targets the core/stem of dorsal association pathways.

The approximate location of the ROI was determined by using the center point (in the A/P axis) between the precentral and postcentral ROI delineated earlier. Then, a large triangular green hypersignal on the RGB had to be located (generally below the curve of the corpus callosum pathway). A large disk encompassing the arcuate fasciculus core/stem (15-20mm) had to be positioned at the center of the green hypersignal.

Variations in the I/S axis caused little impact on the final bundles. However, the variations in the L/R and I/S locations as well as the disk radius differences are causing variability at the periphery of the curve into the temporal lobe while the core/stem is correctly selected. This has a low impact on the overall volume and no impact on the distribution of streamline density in the bundle.

For this particular region of interest, increasing the size is likely not a desirable alternative since it aims to target the stem of the pathways and disentangle the arcuate from other association pathways (e.g. superior longitudinal fasciculus). Focusing on better identification of the precentral and postcentral gyri (and consequently the center point in between) would provide a better starting position of the raters to identify the green hypersignal on the RGB. More details are needed to identify the center of the green hypersignal. More specifically, instructions on how to 'follow' the curvature of the corpus callosum pathways to give a superior limit. Since the medial limit of the arcuate fasciculus plane (describe below) is easier to describe, it should be used as a reference (instead of using the centrum semiovale as a reference for medial limit)



Arcuate fasciculus - Medial sagittal limit

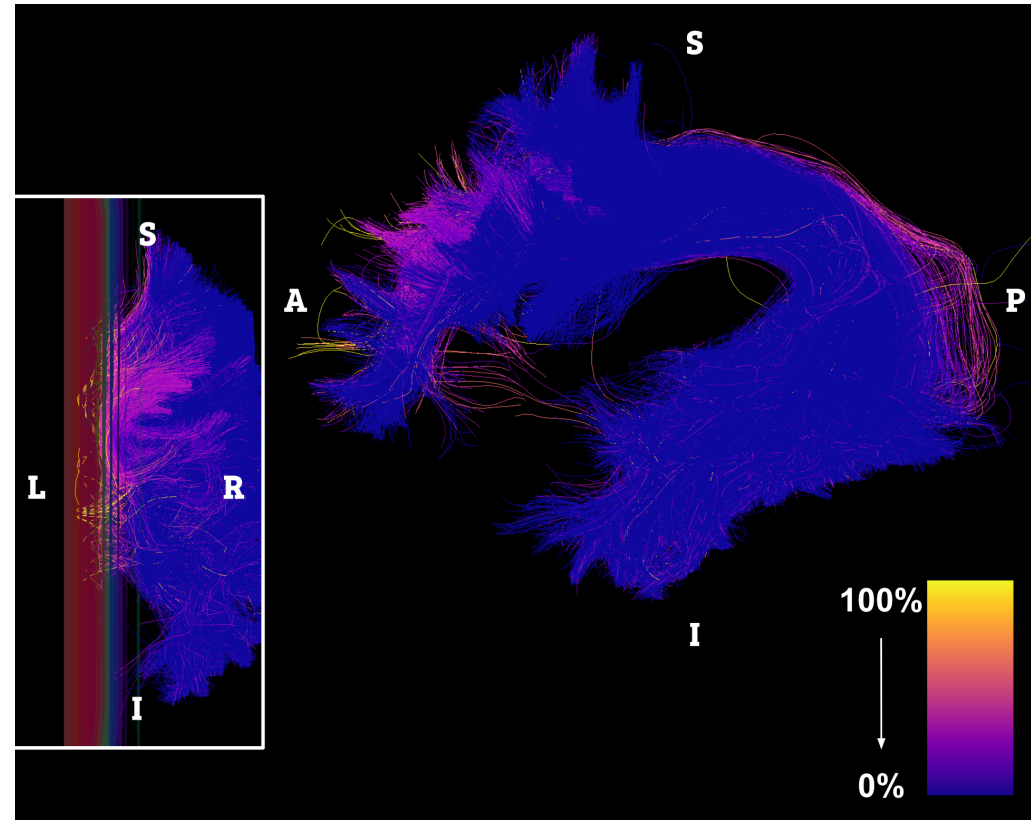
The medial sagittal limit acts as a constraint to avoid streamlines from the superior longitudinal merging into the arcuate fasciculus. This region also prevents spurious streamlines from switching paths into another bundle such as the corpus callosum (outliers)

Raters were expected to identify the ROI by using a previously drawn region (centrum semiovale) as a slice reference and use the outer edge of a green region on the RGB as a limit.

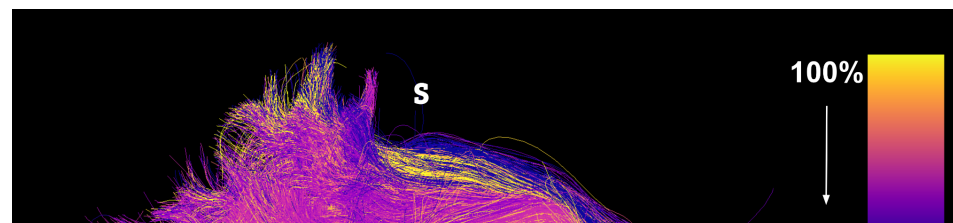
Since the plane selection strongly relies on another ROI, the previous decisions will have repercussions. For example, the choice of a slice to visualize the green signal will impact the outer edge limit. The green signal can mix with another color (e.g. blue with the PYT) and so determining the most medial position of the green signal is not trivial.

The impact of variations in the L/R axis (from the centrum semiovale segmentation) was limited. However, determining with accuracy the end of the green region medially was the main source of variation. Personal interpretation of what is and is not “green” probably was the source of confusion. Despite the variation in the L/R axis, the impact on the final bundles was limited.

Choosing a slice too laterally would lead to a major difference in the bundle as it would intersect the centrum semiovale and effectively cut in half the arcuate fasciculus. Adding an extra few millimeters medially as a safe margin of error would reduce variability while keeping the same purpose.



Arcuate fasciculus - Temporal lobe entrance



This ROI is designed to approximate the entrance to the temporal lobe and ensure that reach into the temporal lobe (and thus achieve the U-shape of the arcuate fasciculus)

The raters were asked to locate a slice where the temporal lobe is clearly distinguishable approximately on the same coronal slice as the postcentral gyrus ROI. Raters were also asked to use the lower axial limit ROI as a landmark to ensure the temporal lobe entrance is strictly below the Sylvian fissure. The ROI was described as a large square/circle ROI of approximately 25-35mm in radius to encompass the temporal lobe.

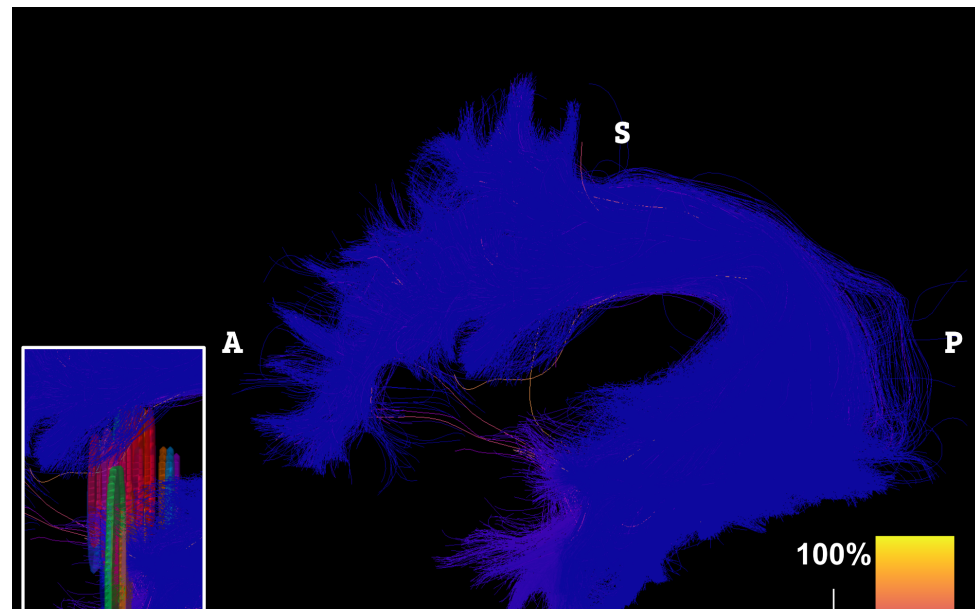
The usage of 2 other ROIs was designed to facilitate rater execution and accelerate the protocol. However, it was observed that the postcentral ROI was highly variable in the A/P axis, sometimes giving the raters a difficult approximate location. This ROI is severely impacted by the variation in the A/P axis. Forcing the streamlines to reach too far back some of the sharp turning streamlines will be excluded, while a ROI position too anteriorly will affect the streamlines terminating at the beginning of the temporal lobe. Most ROI were sufficiently large to correctly encompass the whole desired regions.

Counterintuitively, the best way to improve reproducibility would be to avoid the reference to the postcentral ROI (since it was highly variable in the first place). Identification of the entrance of the temporal lobe could be achieved using a visual description only rather than providing an approximate starting point to the rater. Another, more complex, approach would require a completely new ROI. The high impact of position this ROI (too anteriorly/posteriorly) indicates that locating the stem of the arc (U) of the arcuate fasciculus would be more accurate anatomically and could be more reproducible. This new ROI could be located using the Sylvian fissure to locate the correct axial plane and then identifying a blue hypersignal on the RGB map (when the streamlines switch direction in the curve of the arcuate fasciculus). However, it is difficult to say if nonexperts could reliably find such a region.

Arcuate fasciculus - Temporal stem

This region aims to close off the temporal stem from possible outliers and spurious streamlines. The signal from the uncinate fasciculus could lead streamlines out of the temporal lobe and back into the frontal lobe.

Using the precentral gyrus as a starting point to identify a search region for a strong green hypersignal on the RGB map, a large disk region (20-25mm) had to be drawn. The ROI encompasses the temporal stem while not intersecting the temporal lobe itself or the dorsal pathways



(centrum semiovale), its position is restrained to the insular region.

This region of interest has an exclusion rule and does not affect the final bundles. The spurious streamlines using this passage are extremely rare due to all the rules used to define the arcuate fasciculus. In the current protocol, such streamlines do not exist in sufficient quantity to warrant the use of this “cleaning” ROI.

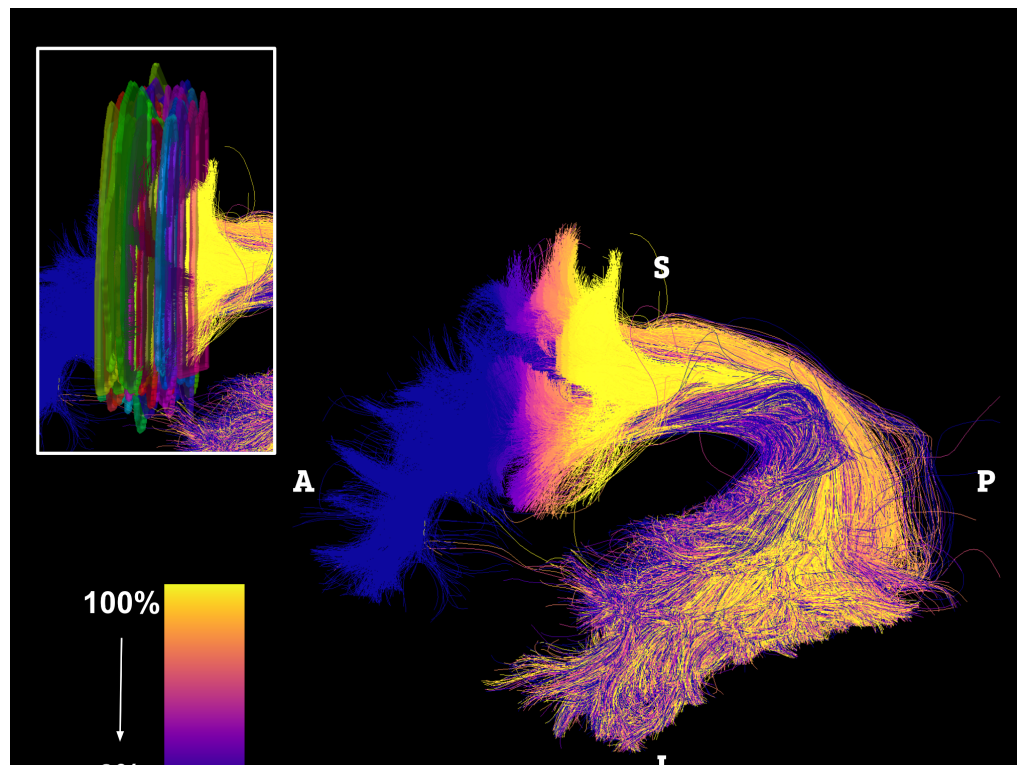
This ROI is a good example of a delineation being a burden on the raters (the variation in size and position indicates it is difficult to locate), not creating any variability and not being useful to the virtual dissection. This ROI should be removed from the protocol as it does not have any positive (or negative) impact, but increases raters' workload.

Arcuate fasciculus - Precentral gyrus

The goal of this region is to force streamlines to reach the frontal lobe to avoid early termination after the centrum semiovale ROI.

The precentral gyrus ROI is the same as the one used for the pyramidal tract. The goal was to reuse as much as possible to decrease the total duration of the protocol.

As opposed to the two other pathways using the precentral gyrus ROI as an exclusion region (low impact), this pathway uses it as an inclusion region (high impact). The large variation in the A/P axis sometimes causes the precentral gyrus to be part of



the final results, but not always. This region leads to the biggest source of spatial and density variability in the protocol (for all bundles). Identification of the precentral gyri planar region is among the hardest ROI of the protocol (also varies from HCP subject to another). This led to the arcuate fasciculus being the pathway with the highest dissection variability.

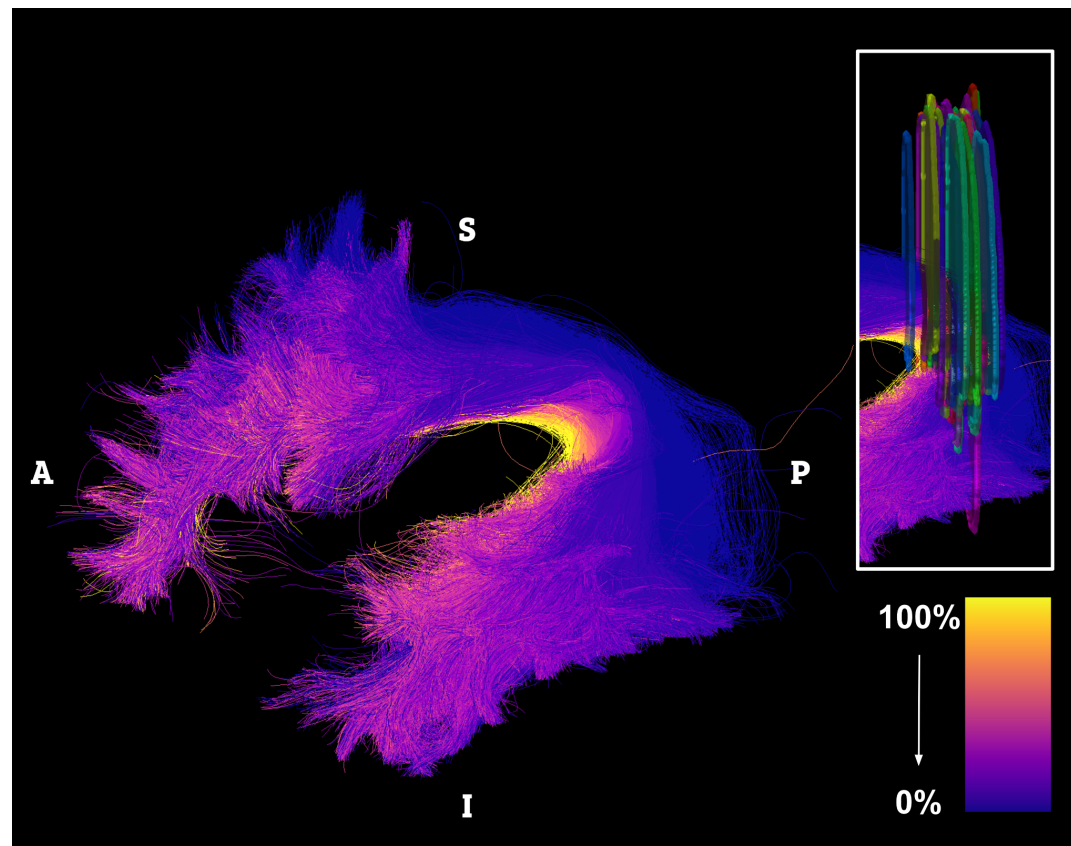
This ROI is a good example of a region of interest needing to be redefined due to anatomical inaccuracies and being a large source of variability. The precentral gyrus planar region was defined as the most anterior point of the gyrus to avoid cutting off the fanning of the corpus callosum and pyramidal tracts (used as exclusion). However, this pathway uses it as an inclusion and so the most posterior point of the precentral gyrus would be needed instead. This leads to two possibilities: 1) Adding a new planar region of interest to the protocol (most posterior portion of the precentral gyrus) or 2) removing this region entirely and moving the centrum semiovale ROI slightly forward to ensure the pathway is effectively reaching into the frontal lobe. Since the identification of landmarks related to the precentral (and postcentral) gyrus seemed to be a difficult task for our raters, we believed option 2) is more appropriate for the protocol. This way the anatomical accuracy is increased from the original protocol, the variability of dissection is decreased and the raters' workload is not increased.

Arcuate fasciculus - Postcentral gyrus

The region aims to further enforce streamlines to curve into the temporal lobe while following the desired path.

The postcentral gyrus ROI is the same as the one used for the pyramidal tract. The goal was to reuse as much as possible to decrease the total duration of the protocol.

The postcentral gyrus ROI has a large variability in the A/P axis, but does not seem to lead to increased variability in the final bundles. Only the ROIs positioned too posteriorly affect a small portion of the inner curve of the arcuate fasciculus. When the region is positioned in a more average position, there is no impact on the spatial extent and only minor differences in density. However, it has a bigger effect on the



superior temporal gyrus and a strong effect close to the parahippocampal gyrus.

With the combination of rules for this pathway virtual dissection, the need for this region is minimal. More explicitly, the temporal entry ROI already forces the path into a curve to reach the temporal lobe. A more careful description of the temporal entry ROI would have a bigger impact on the quality of the final pathway.

Since there is a valid anatomical “reason” for this region/rule, it could be kept as part of the arcuate fasciculus dissection. However, more effort is needed to reduce the variability of the postcentral (and precentral for that matter) gyri. In the current state of the protocol, the postcentral gyrus planar ROI potentially introduces more variability to the arcuate fasciculus dissection than it contributes to its anatomical accuracy.

General recommendations (full-text of section 3.2)

Dissection in standardized space. Virtual dissection protocol should always be performed in a standardized space (e.g. MNI or AC-PC aligned). Either performed before the protocol or in the software itself. The reason for this recommendation is simple: Landmarks do not look the same depending on the brain orientation due to slicing. It is more difficult for raters to locate structures when their shape is affected by the orientation of the slices and when the initial approximate position is harder to find. Relative spatial location is easier to interpret (by human or by computer) in MNI/AC-PC align space (medial to, lateral to). This should be done using linear transformation only (affine or rigid) to avoid deformation of landmarks and the reference images should always be from the subjects themselves (not a template).

Clear instruction on the software does help raters. While the description of regions of interest is crucial, a detailed presentation of the software is likely as important. Presenting functionalities, tips to prevent major pitfalls, avoiding crashes and saving time all are important to reduce the mental workload of a rater. By knowing the software limitations and inconveniences, a virtual dissection protocol can be optimized to reduce the burden on raters and thus increase reproducibility and improve the general experience. Having a single example (video or class) does help raters to use the software more efficiently and reduce critical mistakes. Written text is a reference guide, but step-by-step instruction (or better, a video) is a good way to increase reproducibility.

Virtual dissection workflow. Another way to reduce variability is to establish a workflow for the tasks. Organizing the input, the order in which the data should be loaded in the software, requiring a naming convention for all ROIs and BOIs are all solutions to reduce the workload of raters (and likely reduce the stress associated with the tasks). Saving all ROIs and BOIs for the raters to

acknowledge that each ROIs and BOIs is accounted for is a simple, yet effective verification method. This also allows for a quick verification at the end of the protocols (does every dataset have X ROIs and Y BOIs, do they have the right filename, is the file size consistent enough or are there empty files).

Getting in the zone. While the accuracy of a virtual dissection does not significantly change over time, each rater workflow (software manipulation, files structure, details about the protocol, filename convention) needs some time to get up to speed. It is good practice to set time to focus and perform a few virtual dissections at the time and to always complete a dataset before stopping. This not only avoids mistakes, but also reduces the overall time needed to complete large datasets. It was observed that each protocol 'restart' was taking around 5-10 minutes when the actual process of virtual segmentation of a single subject was 20-30 minutes.

Being too strict is worse than being permissive. Using very small ROIs or using too many exclusion regions is not only harder for raters, but it leads to poor results. Using permissive and easier to locate ROIs leads to good results. Combining multiple ROIs will facilitate the process of virtual dissection. A good balance of inclusion and exclusion is a better approach than relying mostly on one or the other, but using too many ROIs can increase the workload of raters.

Using other ROIs as landmarks. This is generally a good practice to facilitate and accelerate the virtual dissection process. By providing a lot of details on specific ROIs, the subsequent regions should become easier to locate as they reference previously drawn ones. However, this is true only if used for ROIs that are highly consistent. A region with high variability will 'transfer' its variability to any other region referencing it (e.g. central centrum semiovale being based on the center position between the precentral and postcentral gyri: Too variable). Any ROIs being used as reference should receive special attention to its description and limit as much as possible its intrinsic variability.

Generating more than one ROI from the same landmarks. Only re-use ROIs if they have the same impact, if not a new ROI may be needed. For example, if a ROI is used twice, once as inclusion and once as exclusion, and they lead to different effects on their respective dissection, it could be more reproducible and easier for raters to create two separate ROIs from the same set of instructions. Once a landmark is located, one copy could be created with a small translation and the other with a bigger radius, for example. This would allow that time to be used efficiently and optimal ROIs are drawn for each use in a different set of rules.

Providing ranges. Each ROIs should have an approximate initial location and an approximate size to accommodate the anatomical variability of subjects. When in a standardized space, a range of slices can be provided to facilitate location. The exact value should be avoided, raters should be instructed to start from an approximate location and scroll until something specific is located.

Locating landmarks. It appears that instructions using the RGB map are easier to follow than the ones using only T1. This is likely because RGB contains hypersignal of 3 main colors, much easier to describe and locate than a general area of 'gray' on the T1. It also appears that descriptions that involve multiple orientations at once are difficult (axial, coronal sagittal). Landmarks that are perpendicular to the slicing orientation should be prioritized to limit switching from one plane to the other (in the software).

Optimal ROIs design. When using the **inclusion rule** for a stem/core, the positioning along the pathway is less important, the focus should be targeted at the perpendicular positioning (perpendicular to a strongly oriented core). A larger diameter is likely better than a smaller one. A small diameter is prone to large effects if badly positioned and this can lead to losing a large number of streamlines. However, larger diameters can lead to spurious streamlines, but this is often solved by using a combination of ROIs and is, therefore, less problematic. When using the **exclusion rule**, bigger is often better as long as you do not accidentally cut in half the pathway somewhere else (e.g. middle sagittal plane to prevent commissural streamlines accidentally blocking the brainstem). For exclusion rules, adding a few millimeters in the opposite direction of what is supposed to be stopped provides a good margin of error to reduce variability. Indeed, an exclusion region placed too close to the landmark (as some raters will likely *cut/exclude* before the landmark and ruin the dissection) is much worse than one being positioned slightly after the landmark.

Increasing complexity. The virtual dissection protocol should start with basic descriptions of the bundles of interests, a summary of the steps that will follow and a general overview of all regions that will be drawn. Afterward, each ROI should start with a description of how it will be used and overall meaning to inform the raters. Then, providing an approximate starting position (using major, easy-to-find structures) is important to reduce uncertainty for raters. The next step is to provide details on how to locate a specific landmark from this previously identified approximate location (e.g. scroll slowly until you locate X, you should not go past Y). Additionally, ROIs' instruction should start with a fast (and less accurate) set of instructions to delineate the region followed by a slower (more accurate) detailed instruction. This way a protocol is not only more intuitive to follow and provides redundant verification of validity to the raters, but the same documents/course can be used for different applications. Finally, adding a verification step so the raters can confirm if they made a mistake or not (e.g. top portion of the ROI should not go above W or go below Z) limits doubtfulness.

Cleaning ROIs. Some protocols attempt to increase the reproducibility of bundles by allowing cleaning ROIs to manually remove spurious streamlines. This operation can inadvertently lead to more variability by leaving too much room for interpretation and potential software manipulation mistakes. If cleaning ROIs are needed (not anatomically meaningful, but simply to clean outliers from tractography limitations), they should be described and justified, follow the previously listed recommendations and all cleaning ROIs should be accounted for. It should not be left to the raters to decide how big, how many, where cleaning ROIs are needed.