SUPPLEMENTARY MATERIAL

A. Validation of ANOVA conditions

ANOVA analysis requires normality of data and homogeneity of variance for each group. Q-Q plots of the residuals are good visual indicators for the quality of the ANOVA model fit. Simply, the data points should be in agreement with the Q-Q line. Fig. S1 shows acceptable agreement of the residuals obtained after N-way ANOVA analysis for our experiments. Plots show the residuals only for the DICE measure, Q-Q plots for TPR and FPR measures are similar (not shown). Fig. S1 shows the residuals for all the ANOVA experiments that are presented in section 3.4. The homogeneity of variance requirement for each group is typically shown using the Barlett test, Fligner-Killeen test or Levene's test. The condition is required for datasets with unequal sample sizes for each group. Since we ran only a single experiment for each of our parameter combinations, our dataset does not contain multiple observations for individual groups that we can show the homogeneity of variance.

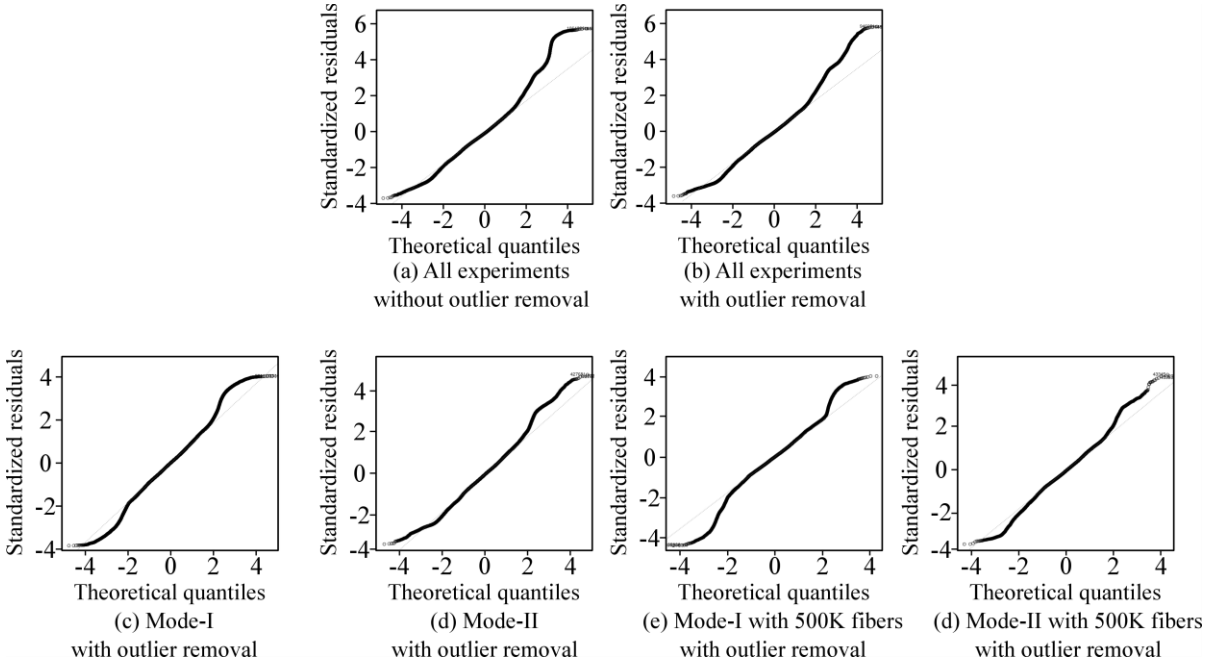

Fig. S1 Normality of residuals is shown using Q-Q plots. Acceptable agreement of residuals with Q-Q lines are observed for all the ANOVA experiments (a) shows the Q-Q plot for the ANOVA analysis used for outlier removal (b-d) show the Q-Q plots obtained for the ANOVA analysis results shown in Figure 6

B. Overall overlap variation due to tractography parameters and anatomical constraints

To illustrate the impact of each individual parameter on tractography, we plotted in Fig. S2 the variability of overlapping measures with respect to the parameters: step, curvature, and cut-off using injections I1 and I6 on the two subjects M1 and M2. In each plot of Fig. S2 that corresponds to a specific parameter as indicated on top of each column, every bar represents the mean value and standard deviation of the overlapping measures obtained by fixing this parameter to the current value while varying other two parameters. Because there is a clear trend that higher streamline numbers produce converging results, we report here only the results computed using the largest number of streamlines that is 500K. The parameter values for each bar are shown on the bottom of the second row. Results from Mode-I analysis were shown here to illustrate the overall effect of the tracking parameters. (Mode-II results are similar and not shown.) Fig. S2 shows that DICE values typically have less variance than the TPR and FPR measures. The large variability in TPR and FPR indicate that some levels of optimization are necessary to obtain good tractography results. Despite the large standard deviations, however, the overall trends are similar to the ones observed in Figure 5 and Figure 6. Varying the cut-off parameter introduces the largest effect on TPR and FPR. With the increase of the cut-off value, we also observe that the standard deviations of TPR and FPR decrease. This clearly shows the constraining effect imposed by this parameter. For the curvature parameter, its effect on tractography performance is not as pronounced below the value of 59 μm. With the increase of the curvature parameter, we also observe a decrease of the variability in performance measures, which reflects also the strong constraints imposed by this parameter. On the other hand, changing the step size does not significantly alter the standard deviations of overlapping measures even though both TPR and FPR grow with the increase of the step size. These observations suggest that cut-off and curvature introduce more variability than the step size parameter although it also affects the performance. From Fig. S2 we also observe that compared to tractography parameters, changes in injection location and subject introduce more substantial performance differences.

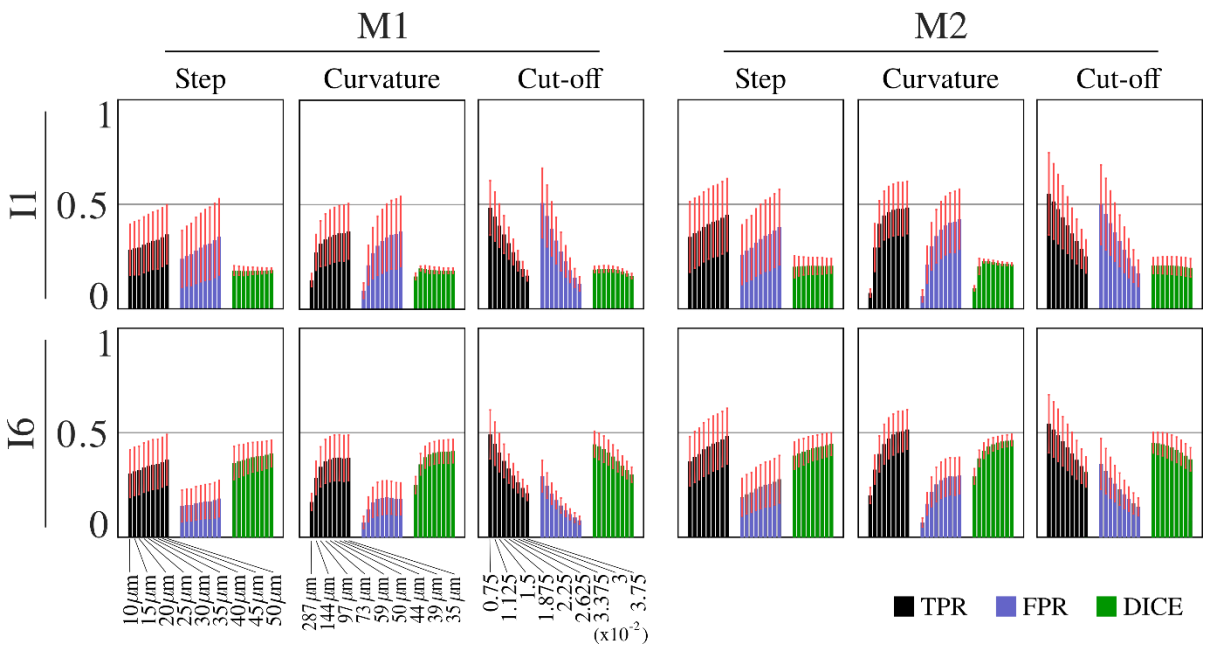

Fig. S2 Overall performance variability due to tractography parameters for Mode-I analysis. Bar plots show mean values with standard deviations as error bars. We show the results for injections I1 and I6 using subjects M1 and M2. Despite large standard deviations in all results, we observe gradual changes in mean performance with respect to variations in tractography parameters. Compared to tractography parameters, injection and subject introduce more drastic changes in performance.

C. Overall impact of injection sites, mouse brains, and tractography parameters

To gain a more complete understanding about the interaction of injection sites, subjects, and tractography parameters, we present in Fig. S3 and Fig. S4 the overlapping measures from all injection sites and mouse brains as we perturb the tractography parameters with respect to a reference point. Results for Mode-I and Mode-II analyses are shown in Fig. S3 and Fig. S4 respectively. In each row of the figures, we show the distribution of the TPR, FPR, and DICE values from all injection sites and mouse brains as we vary one of the parameters with respect to a reference point. Step, curvature, and cut-off values of the reference points are selected as the middle value of all 9 possible choices listed in Table 2. The number of streamlines in the reference point is set as the maximum value 500K.

Compared with the large variability across different injection sites, we can see overall the performance of tractography experiments across different mouse brains are relatively stable for a specific injection site. As each of the tractography parameter varies the overall trend of how the TPR and FPR changes accordingly are consistent across injections and mouse brains. However, the trends in DICE depend on both the mouse brain and the injection site. For the same subject, the DICE coefficient can have an increasing trend for one of the injections yet it can have a decreasing trend for another injection; or, for the same injection, DICE coefficient can have an increasing trend for one of the mouse brain and decreasing trend for another. The large variability in the DICE coefficient for different injections and subject indicates that finding an optimal parameter combination that maximizes DICE coefficient is a challenging problem.

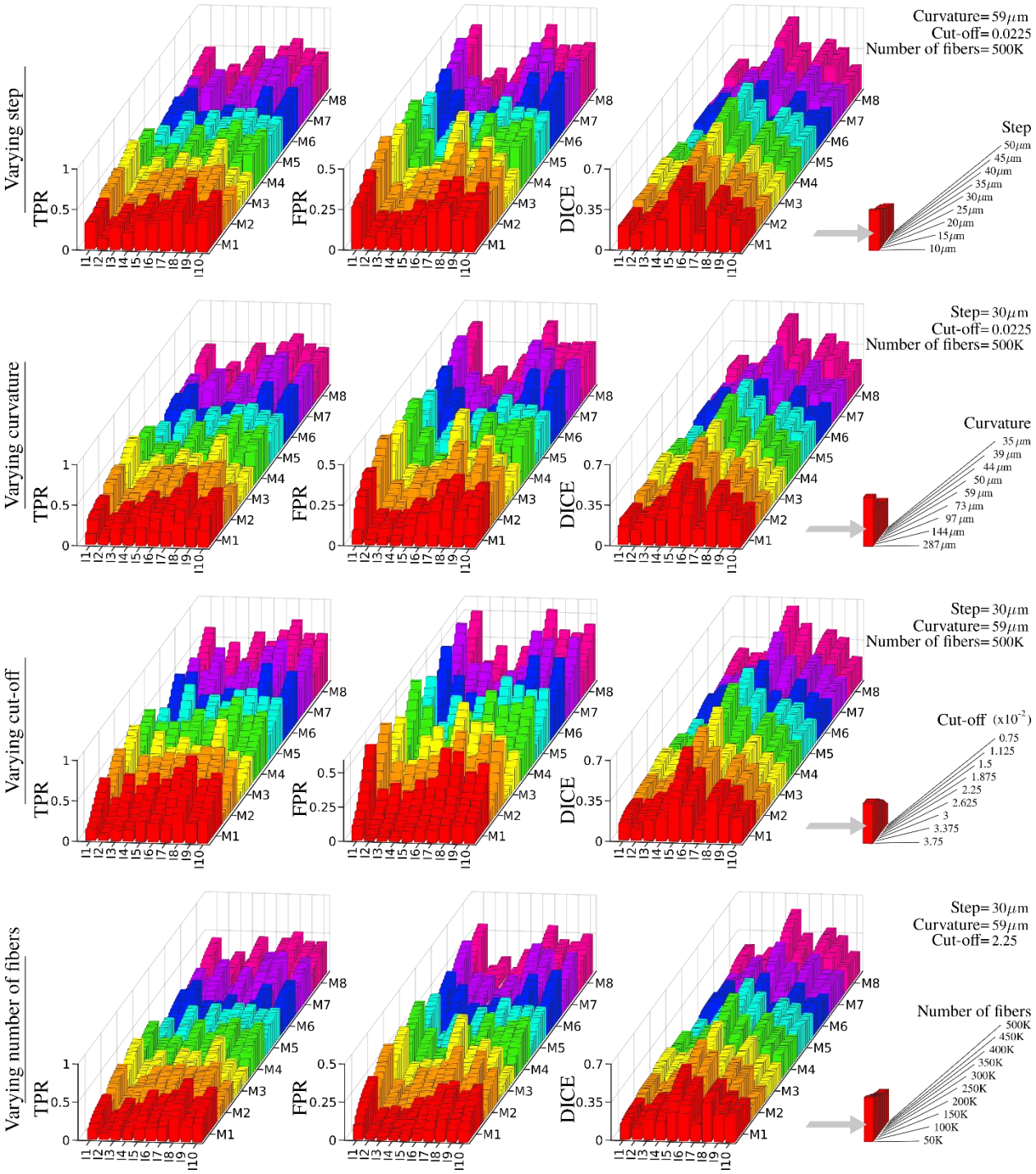

Fig. S3 Impact of injection sites, mouse brains, and tractography parameters on the performance of tractography results for Mode-I. In each row, only one parameter is varied around a reference point shown on the top right corner of each row. From the first to the fourth rows, we plot the distribution of TPR, FPR, and DICE across injection and mouse brains as we vary the step, curvature, cut-off, and number of streamlines. The detailed values of each parameter are shown on the right side of each row.

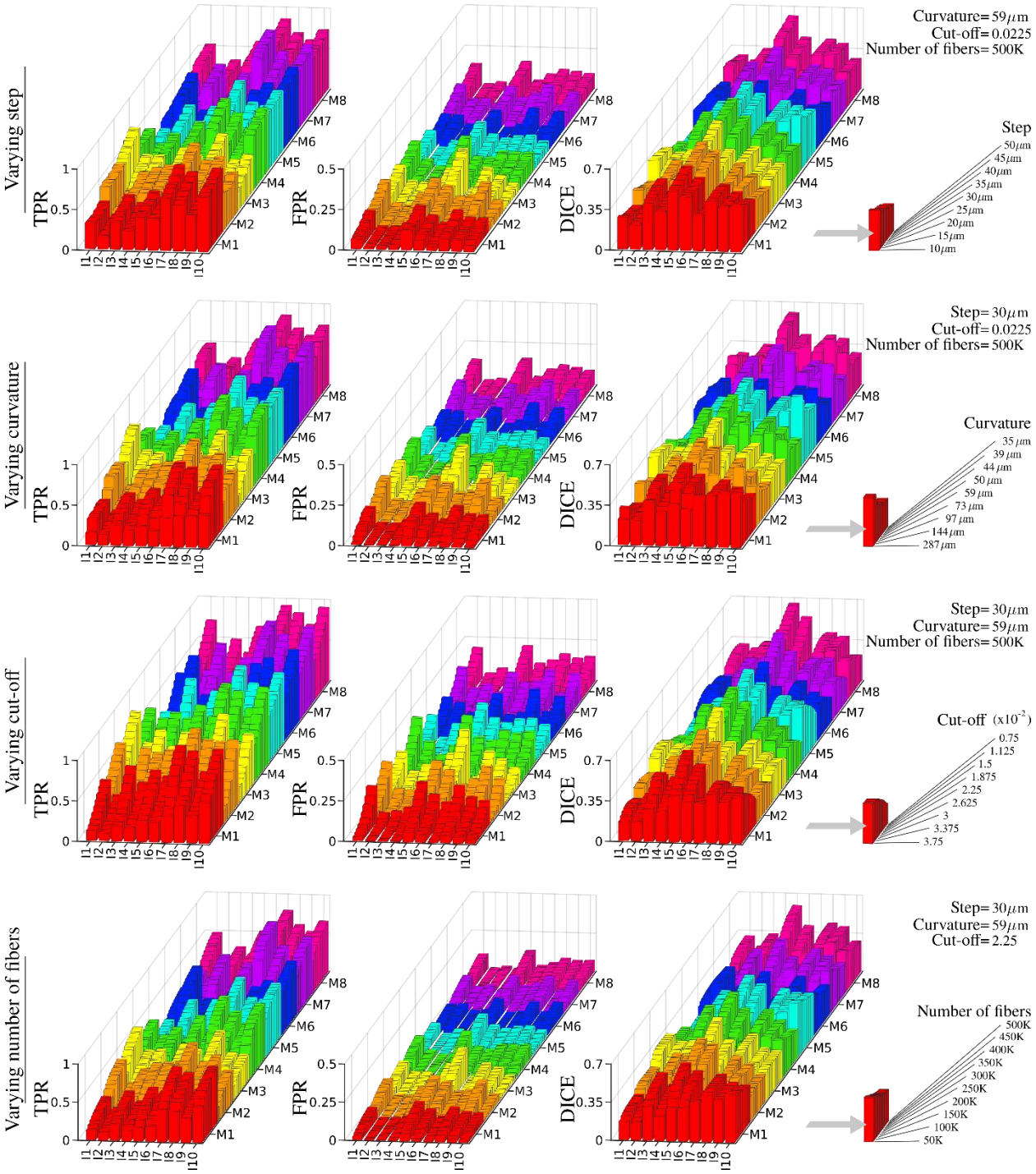

Fig. S4 Impact of injection sites, mouse brains, and tractography parameters on the performance of tractography results for Mode-II. In each row, only one parameter is varied around a reference point shown on the top right corner of each row. From the first to the fourth row, we plot the distribution of TPR, FPR, and DICE across injection and mouse brains as we vary the step, curvature, cut-off, and number of streamlines. The detailed values of each parameter are shown on the right side of each row.

To more clearly illustrate the effect of injection sites versus mouse brains presented in Fig. S3 and Fig. S4, we plotted in Fig. S5 the overall effect of individual injection location or subject on overlapping measures. In each box of Fig. S5, we fixed either the injection site or the mouse brain and calculated the statistics of the corresponding overlapping measure from the data shown in Fig. S3 and Fig. S4. Results from Mode-I and Mode-II analysis are plotted. From the plots shown in Fig. S5, we can see that overlapping measures exhibit more variability with the change of injection sites than mouse brains.

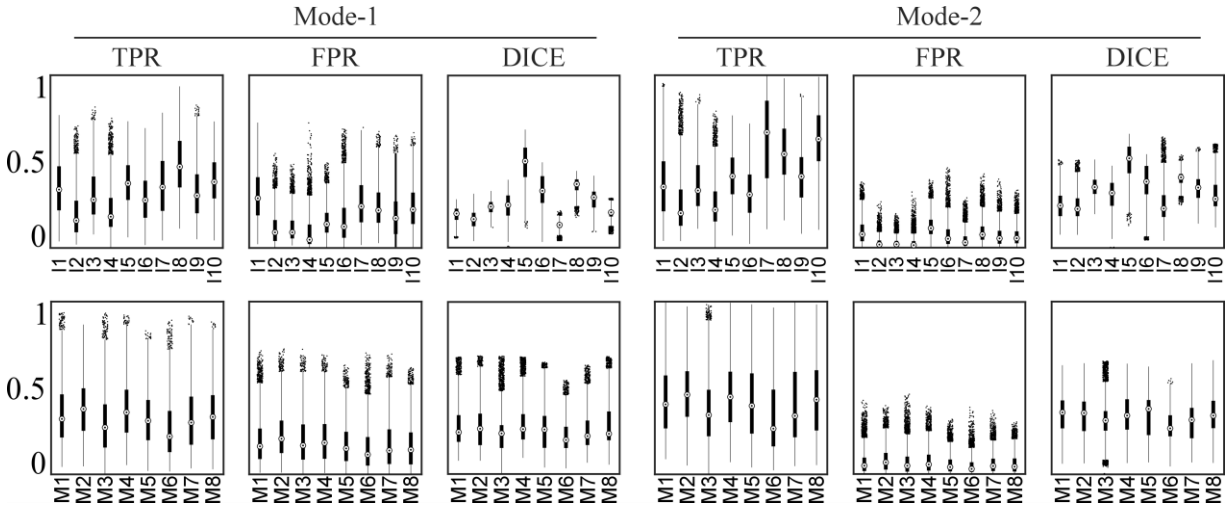

Fig. S5 Box plots show the distributions of TPR, FPR and DICE measured with respect to the injection locations (top row) and subjects (bottom row). In each plot, circles show the median, edges of the boxes indicate 25th and 75th percentiles, and individual points outside the boxes show outliers. The top row shows that for both modes and for all measures, there is a large variability in performance with the change in injection location. The bottom row shows that changes in subject introduces some level of variability but not as pronounced as injection locations.

D. Outlier removal

We performed a post hoc analysis using Tukey's test after the N-way ANOVA analysis that included the complete dataset. Fig. S6 displays differences in the mean levels for each group. From Fig. S6 (a), the first curvature group (287 µm) stands out as outlier. No other clear outlier groups are observed.

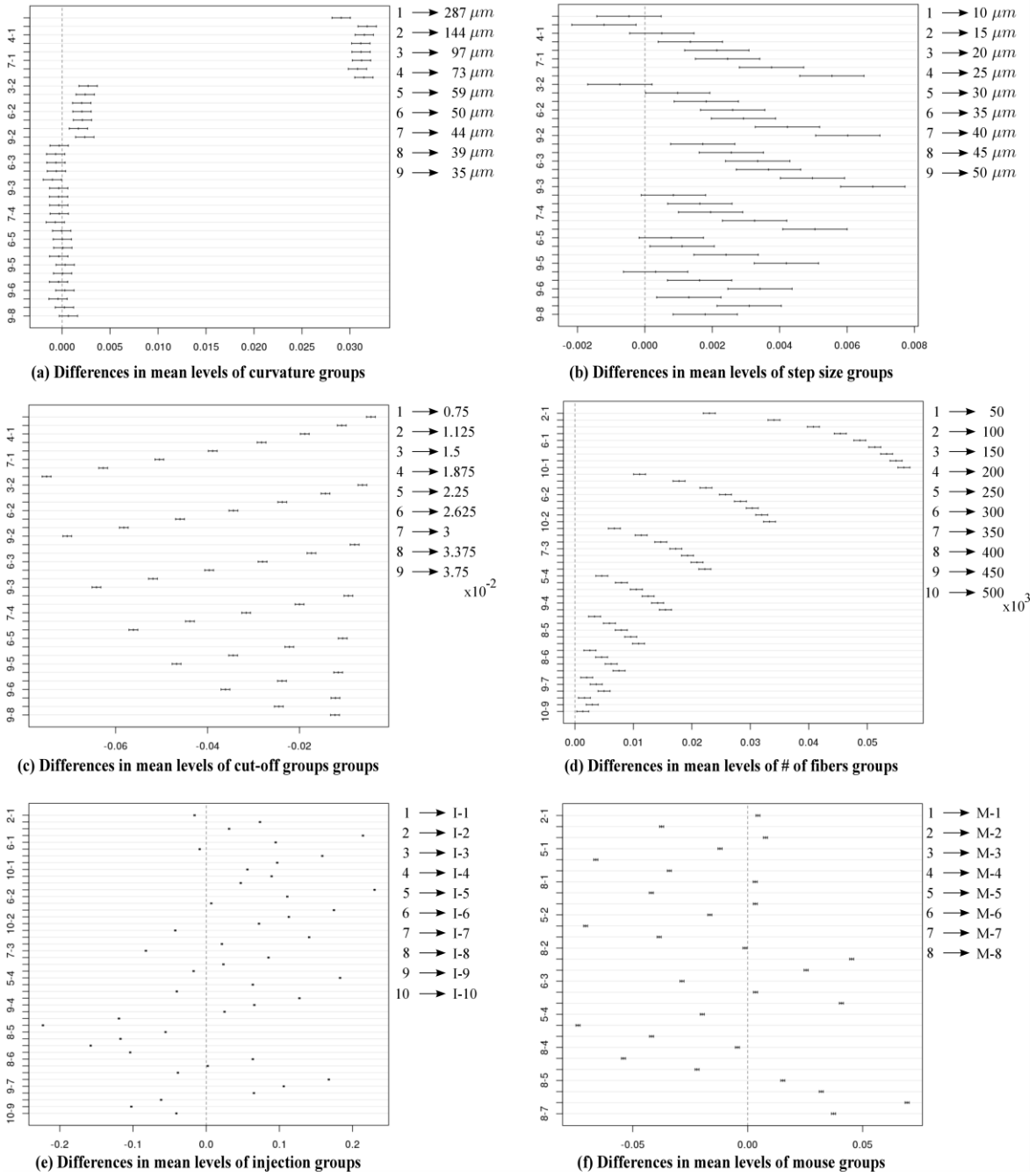

Fig. S6 Post hoc analysis results obtained using Tukey's test (a-f) show differences in mean levels of each group studied in this work. The only clear outlier is the first curvature group shown in (a)