

Supplementary Materials for “Non-Negative Data-Driven Mapping of Structural Connections with Application to the Neonatal Brain”

E. Thompson¹, A.R. Mohammadi-Nejad^{1,2}, E.C. Robinson³, J.L.R. Andersson⁴, S. Jbabdi⁴, M.F. Glasser^{5,6}, M. Bastiani^{1,4}, S.N. Sotiropoulos^{1,2,4}

¹*Sir Peter Mansfield Imaging Centre, School of Medicine, University of Nottingham, Nottingham, United Kingdom*

²*Nottingham National Institute of Health Research (NIHR) Biomedical Research Centre, Queen's Medical Centre, University of Nottingham, United Kingdom*

³*School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom*

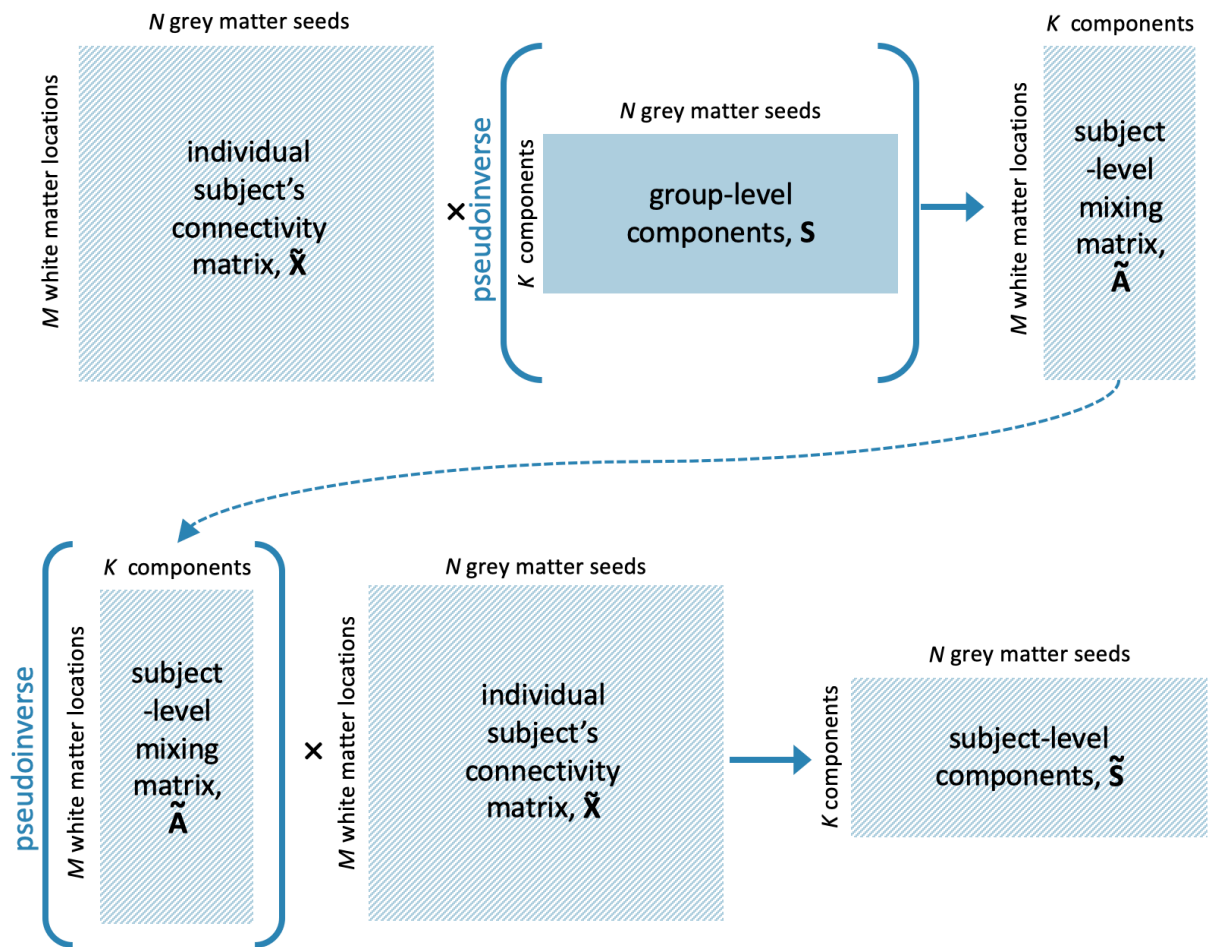
⁴*Wellcome Centre for Integrative Neuroimaging - FMRIB, University of Oxford, Oxford, United Kingdom*

⁵*Department of Neuroscience, Washington University School of Medicine, Saint Louis, USA*

⁶*Department of Radiology, Washington University School of Medicine, Saint Louis, USA*

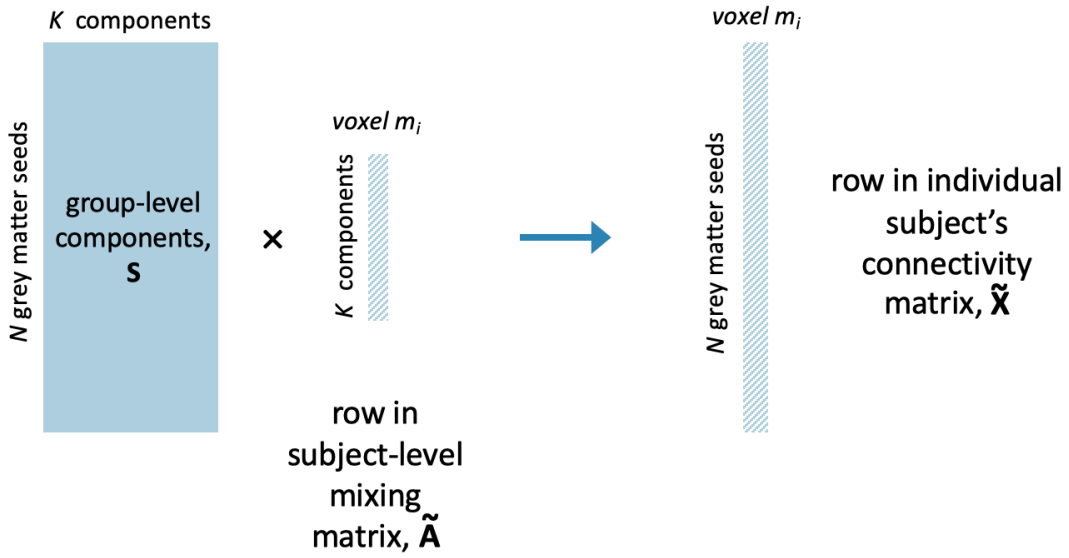
Corresponding author: Stamatios N. Sotiropoulos

(stamatios.sotiropoulos@nottingham.ac.uk)

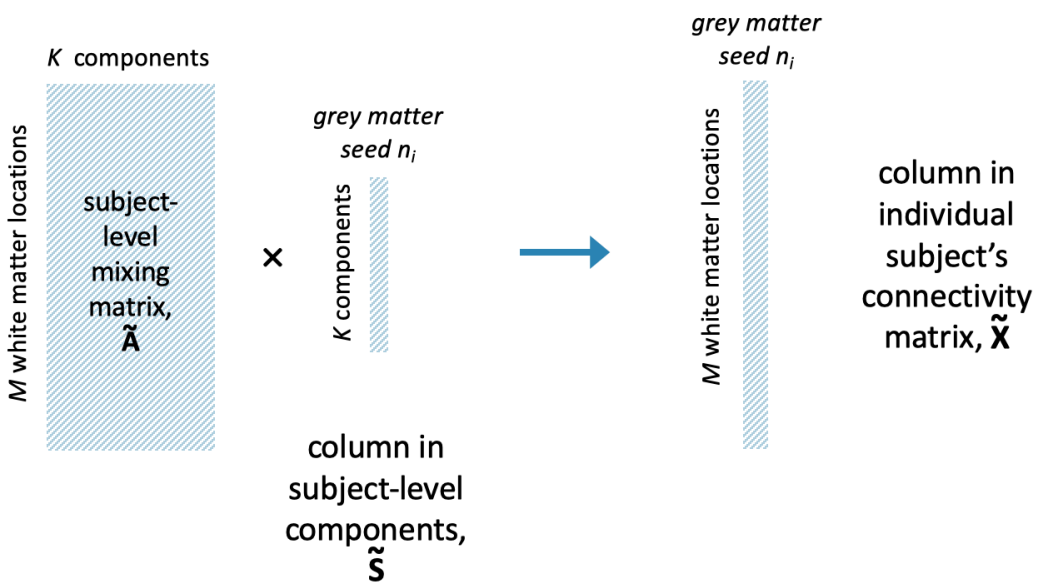


Supplementary Figure 1. Schematic illustration of the dual regression step used for ICA, that generates subject-level representations of the group-level components. The group-level grey matter components are first regressed onto the subject's connectivity matrix to obtain the subject-level representations of the components in white matter. We then use the pseudoinverse of this mixing matrix to obtain the subject-level grey matter components.

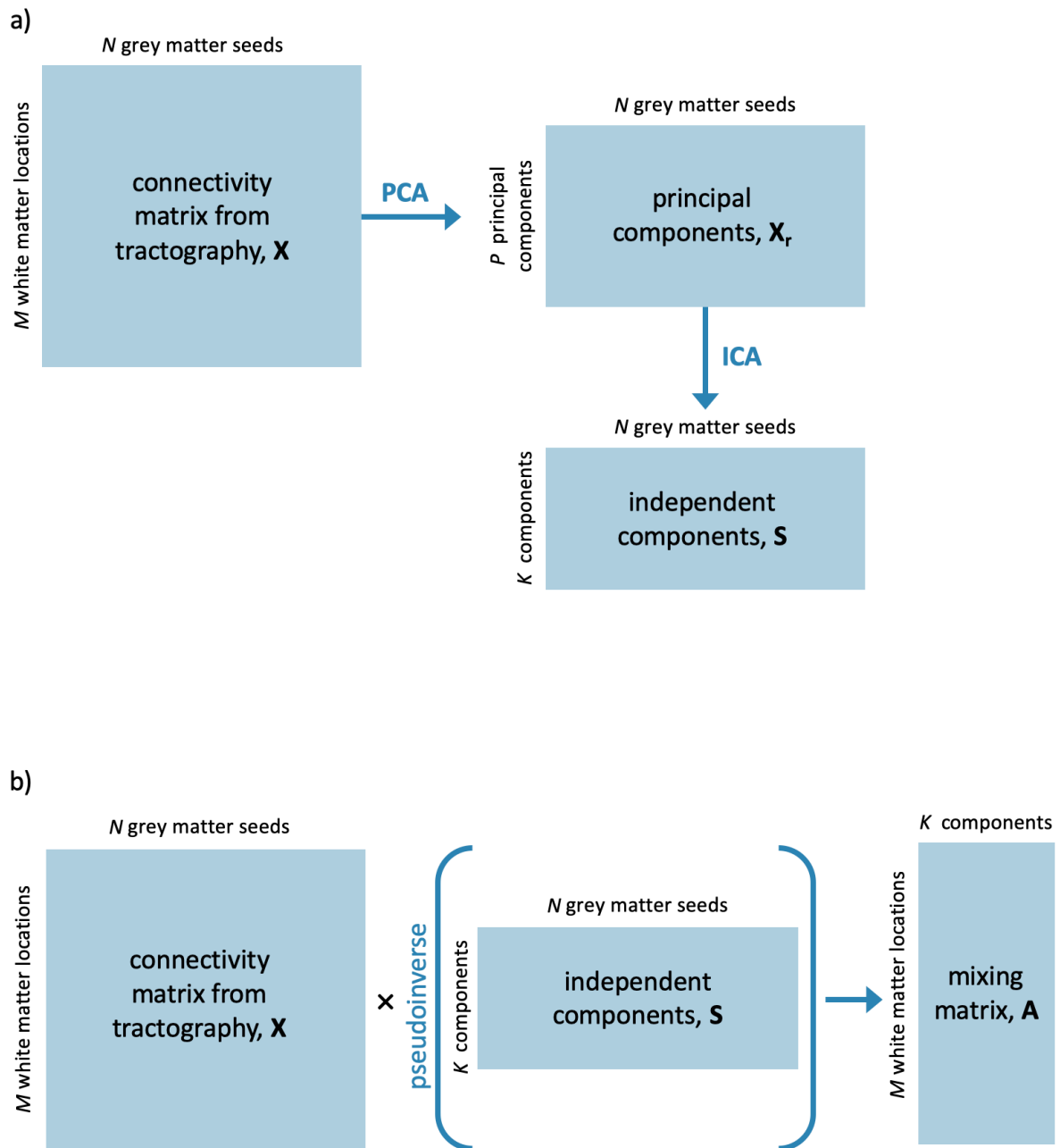
For each voxel m_i in M target voxels:



For each seed n_i in N grey matter locations:



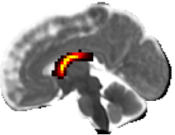

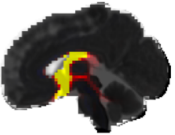

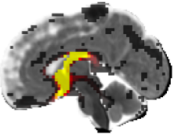
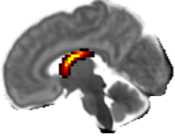
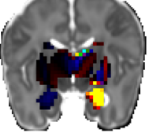
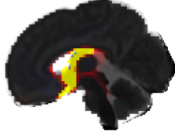

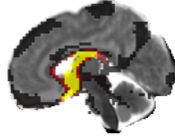

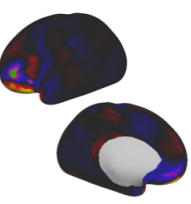
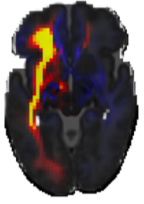
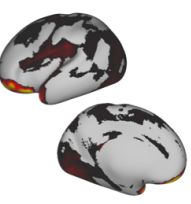
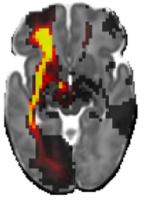
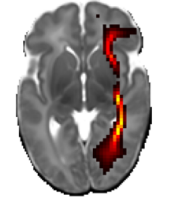
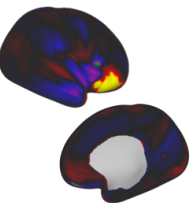
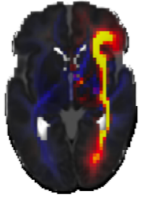
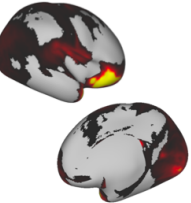
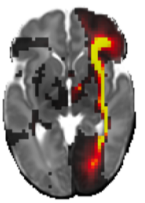

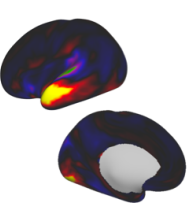
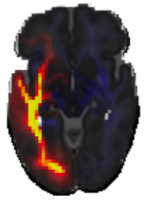
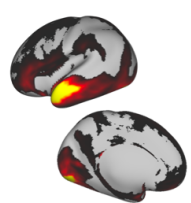
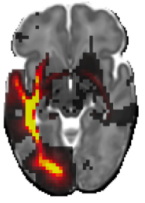

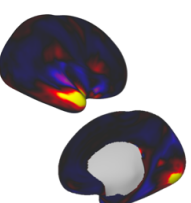
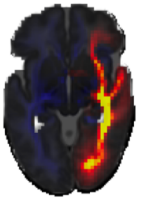
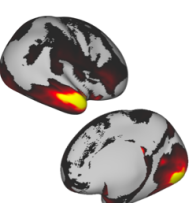

Supplementary Figure 2. Schematic illustration of the non-negative dual regression step used for NMF, that generates subject-level representations of the group-level components. We first use NNLS to solve the top equation, using the group-level grey matter components and the subject's connectivity matrix to solve for each row in the subject-level mixing matrix. We then use this mixing matrix to find the subject-level grey matter components by the same method.



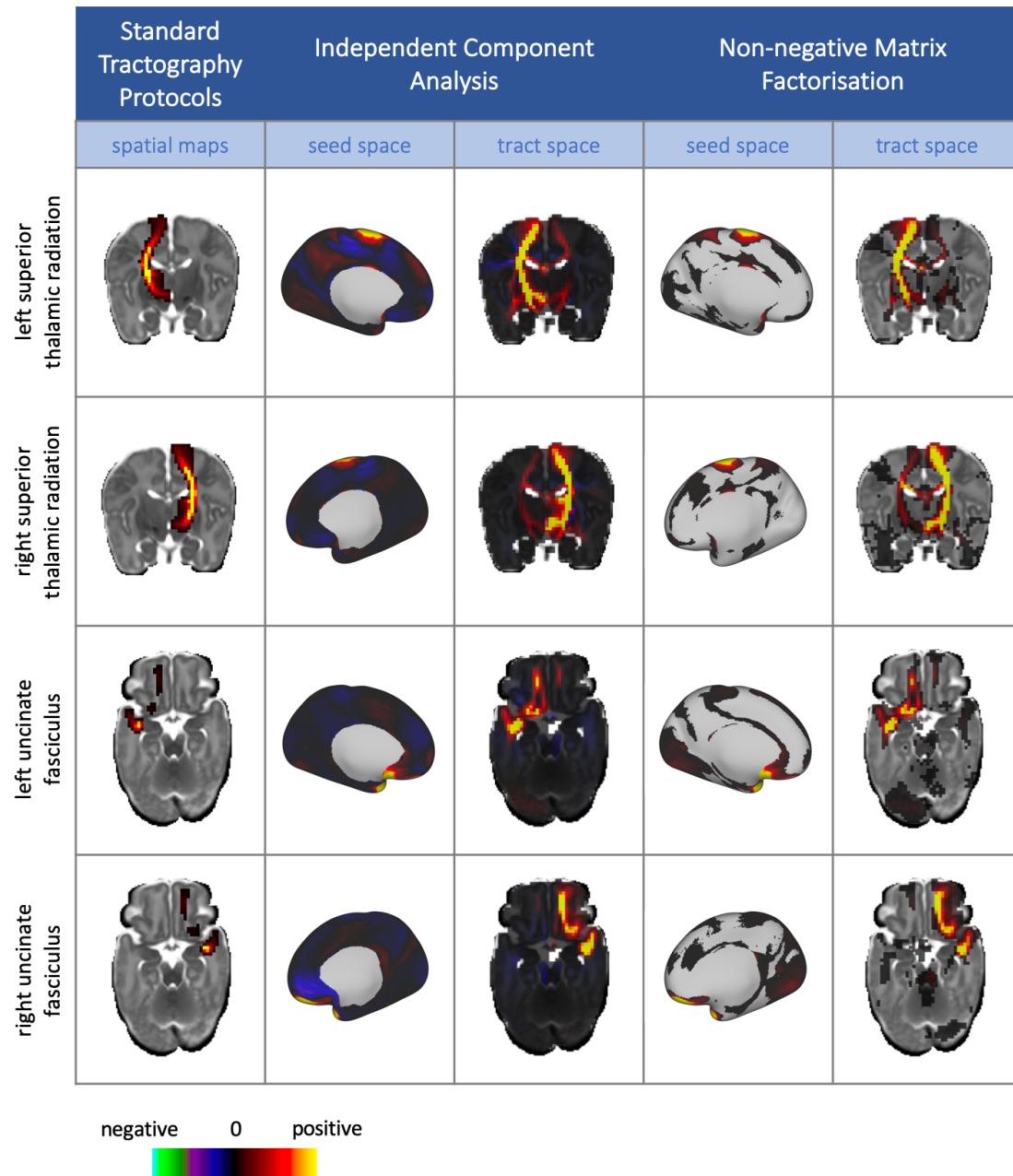
Supplementary Figure 3. Illustration of a back projection step to obtain a white matter mixing matrix after data reduction by PCA. a) Data are first reduced to P principal components, and ICA applied to the data in the reduced subspace. b) Independent components are regressed onto the original data to obtain a mixing matrix in white matter space.

	Standard Tractography Protocols	Independent Component Analysis	Non-negative Matrix Factorisation	
	spatial maps	seed space	seed space	tract space
left acoustic radiation				
right acoustic radiation				
left anterior thalamic radiation				
right anterior thalamic radiation				
left cingulum				
right cingulum				

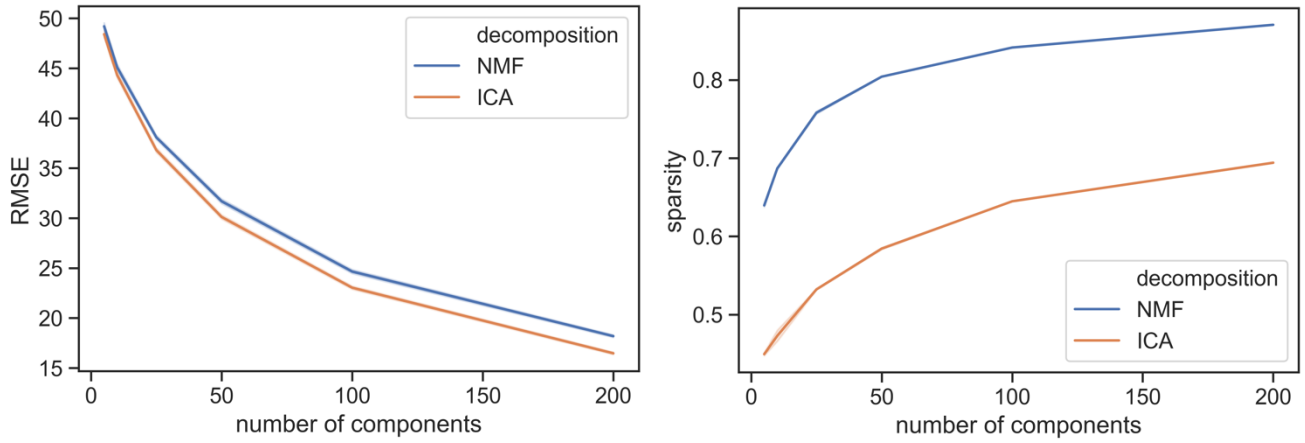
	Standard Tractography Protocols	Independent Component Analysis	Non-negative Matrix Factorisation		
	spatial maps	seed space	tract space	seed space	tract space
left cingulum (hippocampal part)					
right cingulum (hippocampal part)					
left cortico-spinal tract					
right cortico-spinal tract					
forceps major					
forceps minor					

		Standard Tractography Protocols		Independent Component Analysis		Non-negative Matrix Factorisation	
		spatial maps	seed space	tract space	seed space	tract space	tract space
left fornix							
							
left inferior fronto-occipital fasciculus							
							
left inferior longitudinal fasciculus							
							

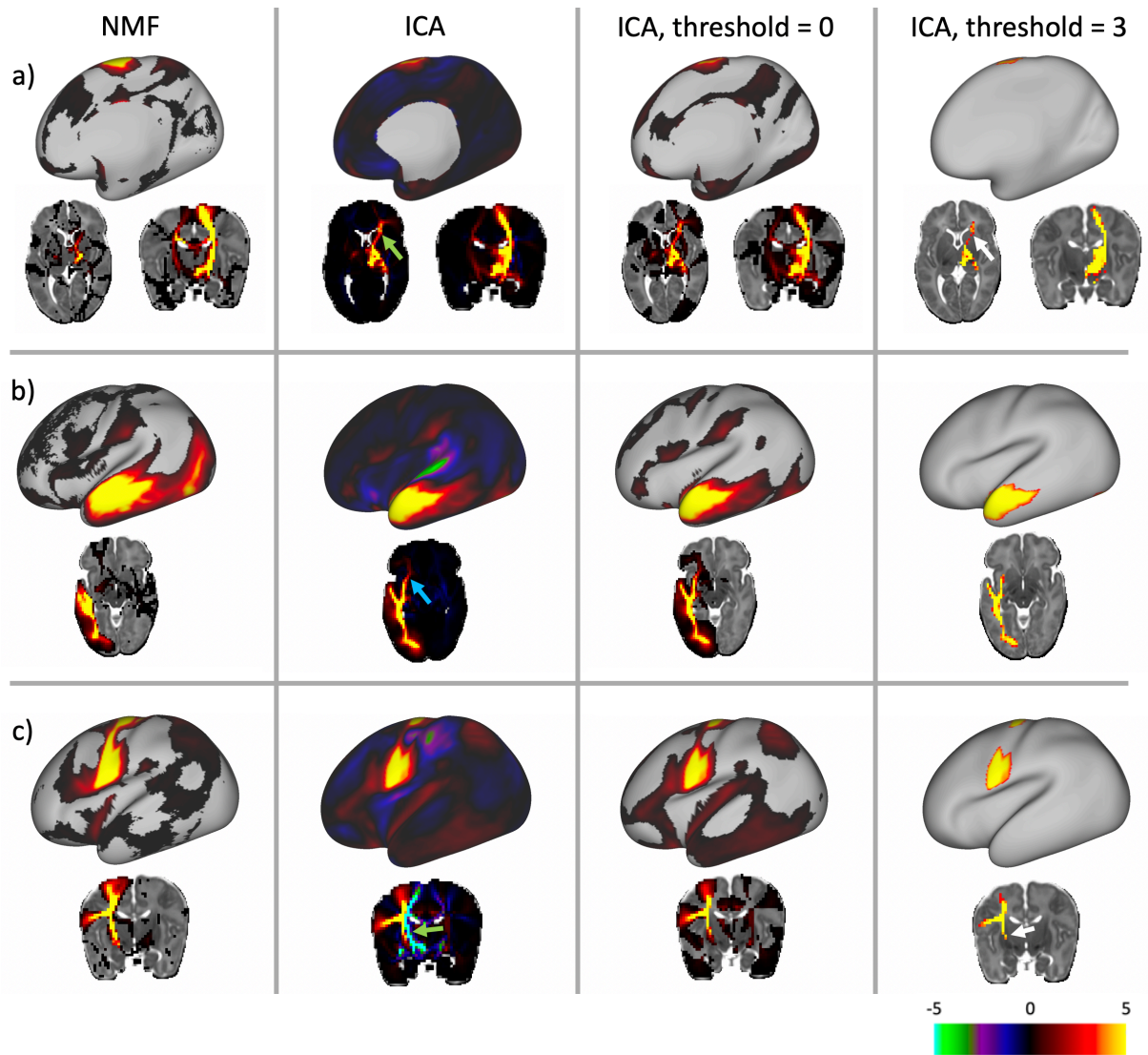
		Standard Tractography Protocols		Independent Component Analysis		Non-negative Matrix Factorisation	
		spatial maps	seed space	tract space	seed space	tract space	tract space
left medial lemniscus							
right medial lemniscus							
left posterior thalamic radiation							
right posterior thalamic radiation							
left superior longitudinal fasciculus							
right superior longitudinal fasciculus							



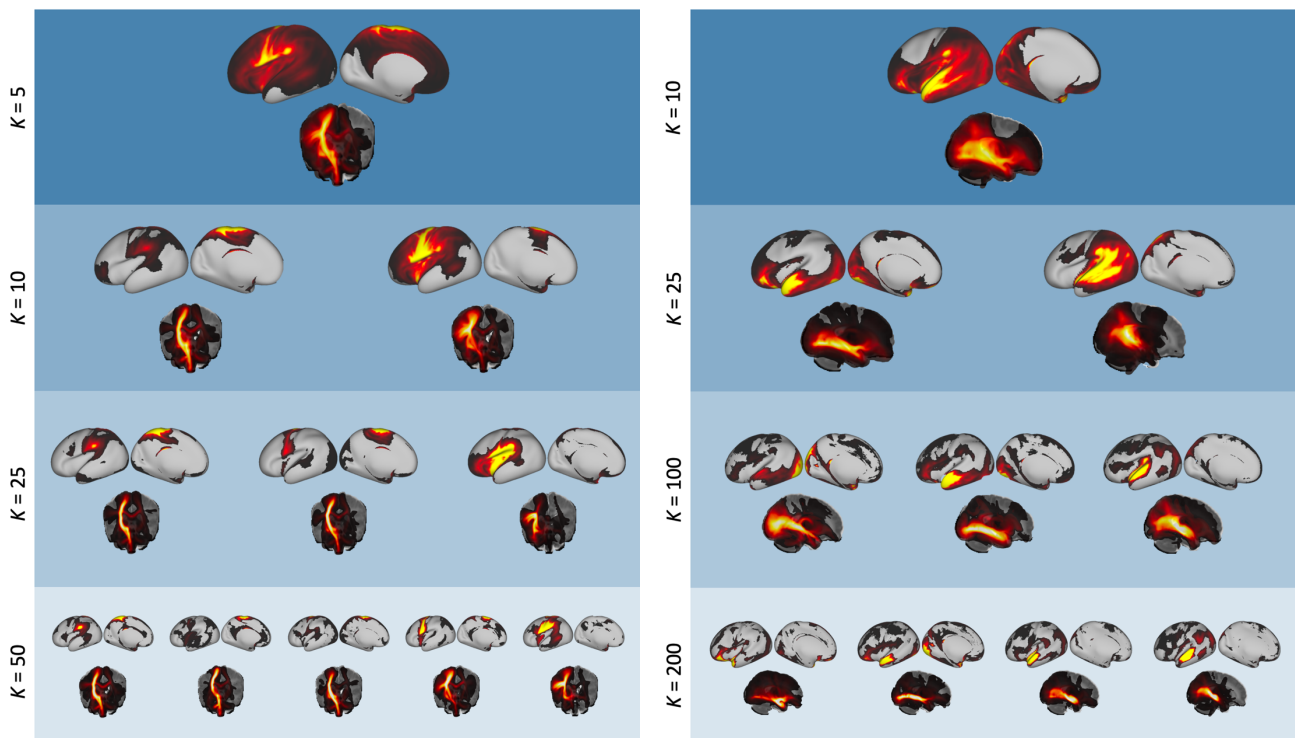
Supplementary Figure 4. The full set of 28 tracts from the standardised protocols (Bastiani et al., 2019), alongside there corresponding components from ICA and NMF. Data-driven components are unthresholded to enable the comparison between the negative values in the ICA components and the sparse, non-negative representations from NMF, whereas the maps from standard protocols are lower thresholded at 0.001 for clearer visualisation of the tract. All tractography and data-driven results are taken from split 1 of the split-half analysis.



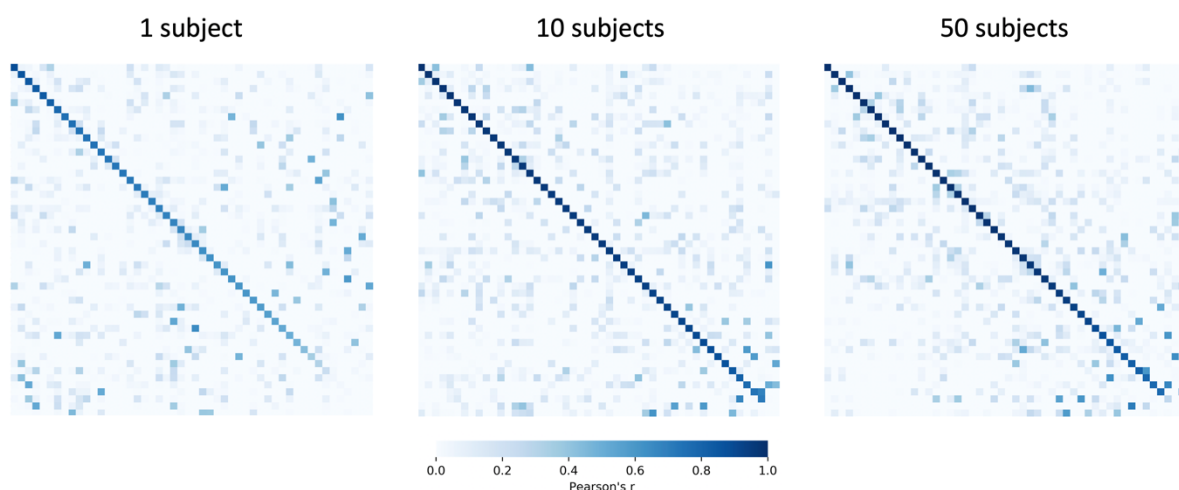
Supplementary Figure 5. a) Reconstruction error at each model order for ICA and NMF. This is the root mean squared error between the connectivity matrix and the dot product of the mixing matrix and the components. b) The sparsity of the derived components, calculated according to equation (2).



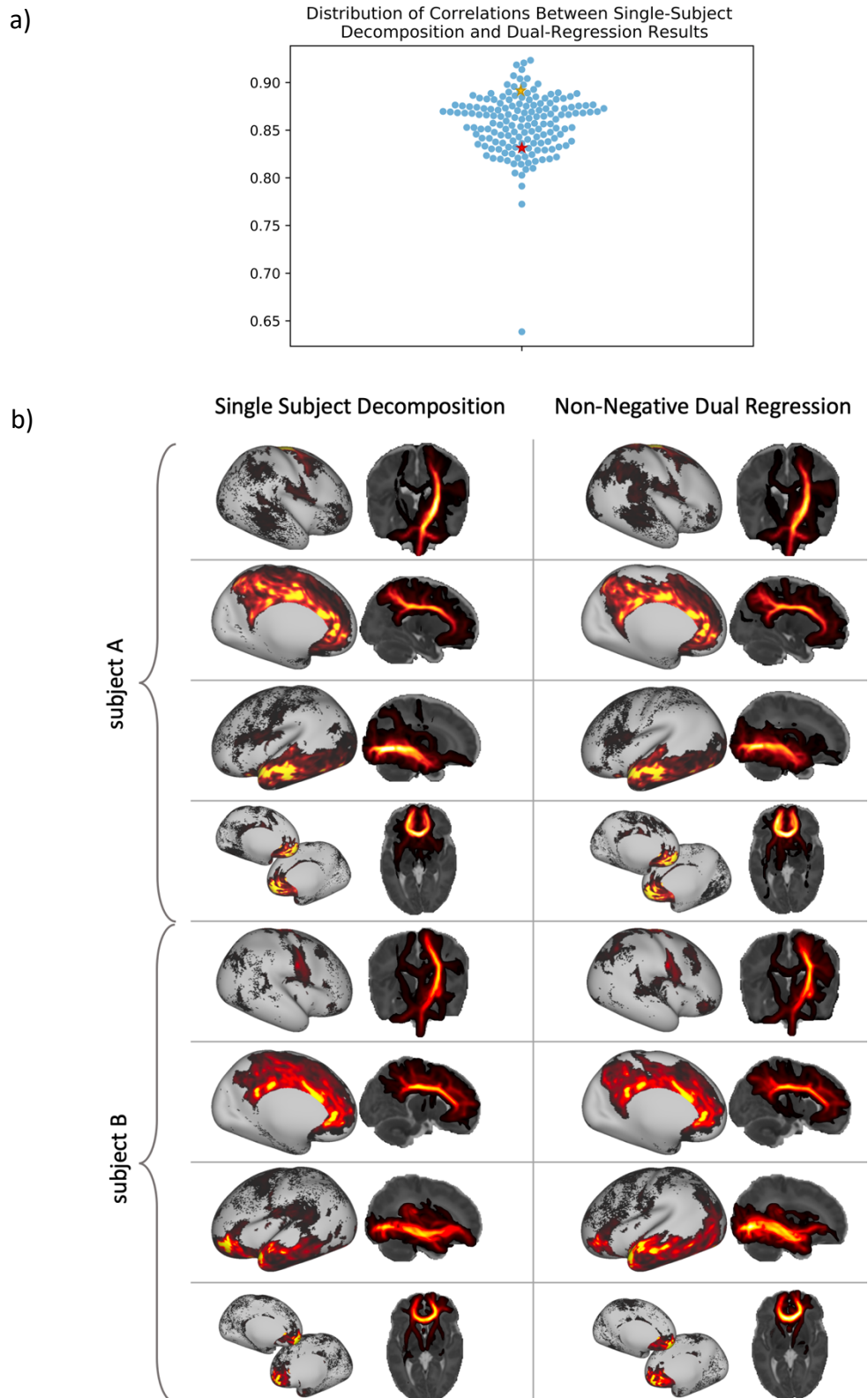
Supplementary Figure 6. Examples of matched pairs of NMF and ICA components from the $K = 100$ decomposition that illustrate differences between the two methods. Components have all been variance normalised for consistent scaling. a) In this example, the right superior thalamic radiation, there is a false positive anterior projection in the ICA component (green arrow) that is not present in the NMF component. This is still present at a high threshold for the ICA maps (white arrow). b) In this component, corresponding to the left inferior longitudinal fasciculus, there is an anterior projection in the ICA component that is not seen in the NMF result (blue arrow). c) The ICA component has strong negative and positive aspects that obscure the main part of the tract (green arrow). High thresholding, such as in the right column, is too conservative and the tract itself is thresholded out (white arrow). These examples further illustrate that the NMF results convey different information than the ICA results, even when the latter are thresholded to only retain positive values (3rd column) or are thresholded to retain the areas with the strongest weights (4th column).



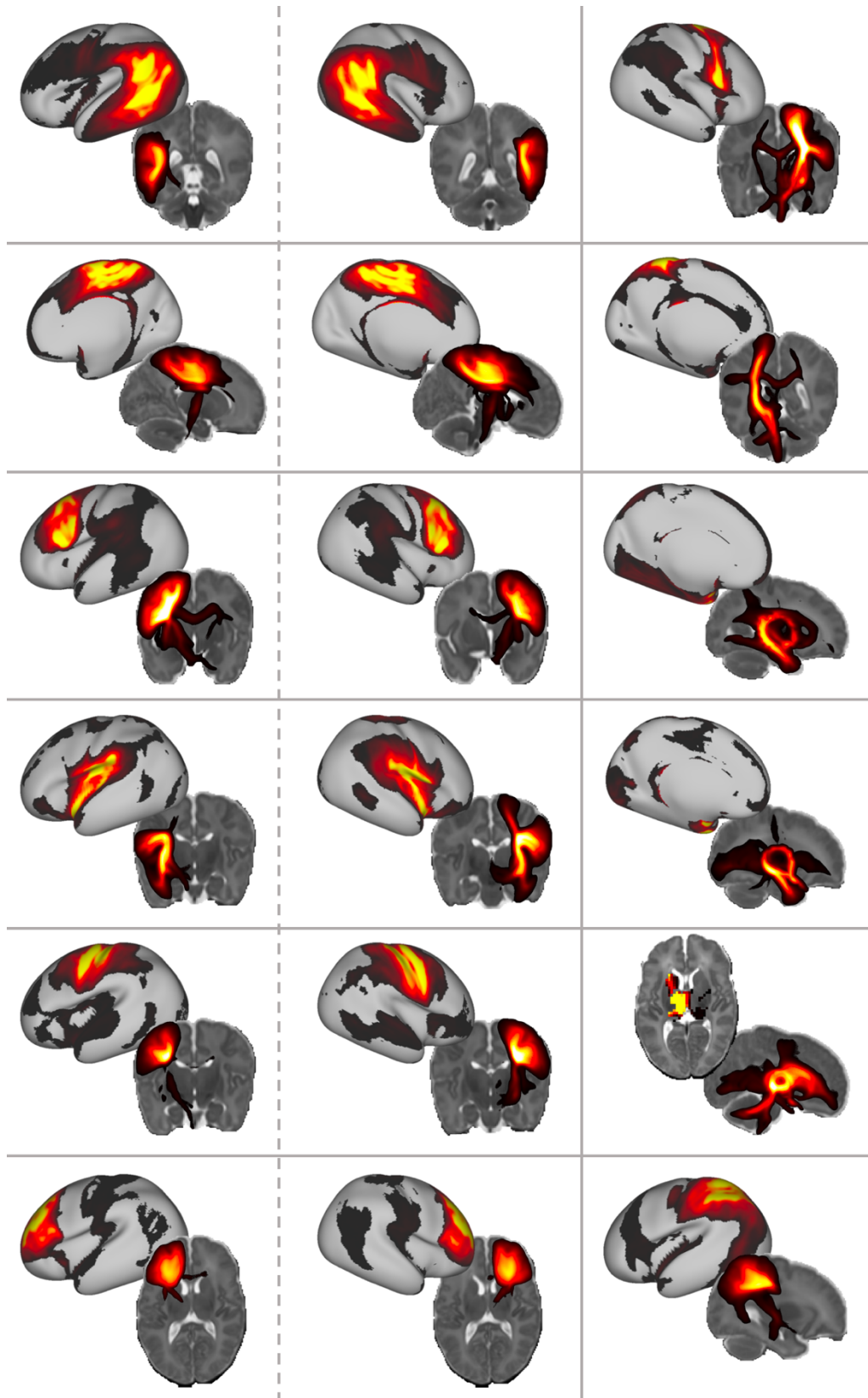
Supplementary Figure 7. Similar components across different model orders, demonstrating the hierarchical nature of the decomposition. Starting with a single component from a lower dimensionality decomposition, we show components from higher model orders that have high spatial correlation with the original component, in tract space ($r > 0.5$). Tract space results are displayed as maximum intensity projections. Left: A component from the $K = 5$ decomposition showing the left cortico-spinal tract, which is split into more localised sub-components for higher K . Right: A component from the $K = 10$ decomposition that includes several different association fibres in the left hemisphere. At $K = 200$, this has been split into the uncinate fasciculus, inferior longitudinal fasciculus and middle longitudinal fasciculus.



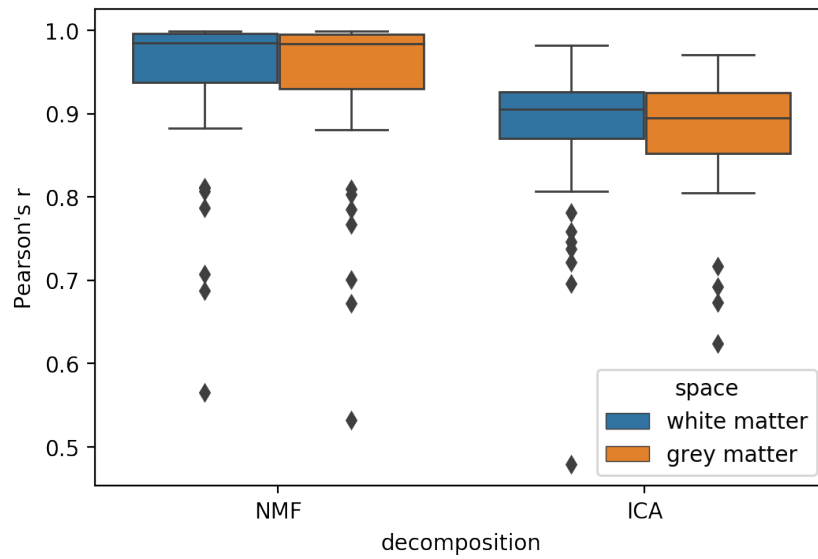
Supplementary Figure 8. Correlation matrices between the white matter $K = 50$ NMF spatial maps from smaller group-level decompositions with the split 1 components from the main analysis. Each entry corresponds to the correlation between a component obtained from a group of L subjects ($L=1,10,50$) and the best-matching component from one of the split-halves of the full cohort (161 subjects). The matrices have been reordered so that matching components lie on the diagonal and are in descending order in terms of correlations. We can see that there is no significant increase in agreement between the 10 subject group with the full cohort and the 50 subject group and the full cohort (average Pearson's r of matching components = 0.89 for both), which indicates that the method is robust even for small group sizes.



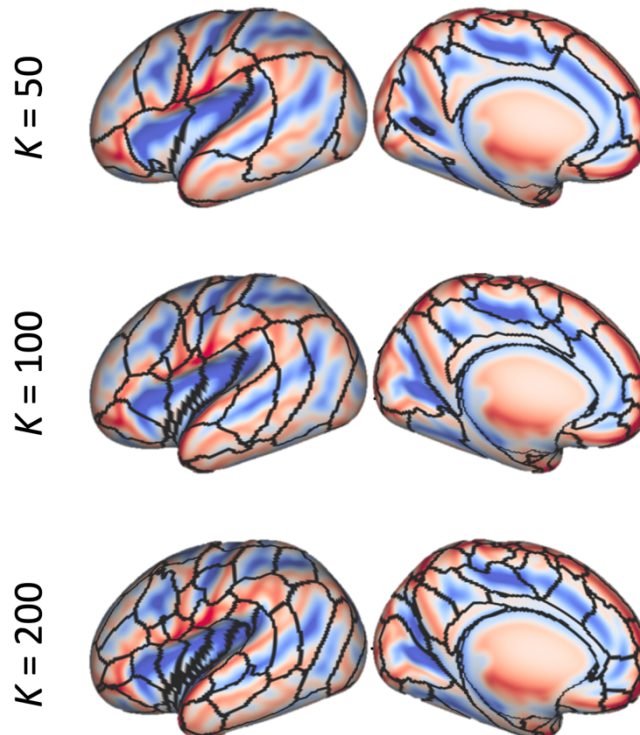
Supplementary Figure 9. We assessed the similarity between $K=50$ subject-level decompositions and non-negative dual-regression results against the group decomposition obtained from the split 1 of the cohort. a) Swarm-plot of the median correlation between matching components for each subject across the $K = 50$ components. b) Selected components from two representative example subjects from the cohort, Subject A (median correlation = 0.89, yellow star), and Subject B (median correlation = 0.83, red star).



Supplementary Figure 10. Unassigned components from the $K = 50$ NMF decomposition. One-to-one matching was performed with the 28 standard tracts, leaving 22 unassigned components. Four of these were callosal fibres, similar to those shown in Figure 7, leaving 18 additional components that we show here. The components in the left and centre columns are bilateral, unpaired components are in the right column.



Supplementary Figure 11. Correlations between decompositions applied to connectivity matrices in both the (seed x white matter) and (white matter x seed) configuration. Highest Pearson's r is plotted for each component or column of the mixing matrix with the equivalent matrix from the transposed decomposition.



Supplementary Figure 12. Our group-level NMF parcellation overlaid on the dHCP's 40-week PMA sulcal depth template (Bozek et al., 2018) (blue for sulcal fundi, red for gyral crowns). There is no consistent pattern for parcellation boundaries to follow sulci or gyri, which indicates that our parcellation is not driven (at least to a large degree) by the gyral bias.