Supporting Information

Krienen et al. 10.1073/pnas.1510903113

Local–Distant Connectivity Metric

Sepulcre et al. (13) used graph theoretic analysis of resting-state functional connectivity data in humans to demonstrate heterogeneity in the extent to which cortical regions were coupled to preferentially local or preferentially long-range cortical locations. Unimodal sensory cortical regions display preferentially local connectivity, whereas higher-order heteromodal association regions distributed across the cortex have preferentially distant connectivity.

Fig. 7*A* shows the relative preference for distant versus local connectivity across the human cortical surface from Sepulcre et al. (13). Specifically, degree centrality quantifies the number of links or edges connected to a node. Sepulcre et al. binned the degree centrality metric according to spatial criteria: immediate local neighborhoods were defined as a 14-mm radius around the target voxel. Local degree was computed by counting for each voxel the number of voxels above a threshold (Pearson's $r > 0.25$) inside and outside the 14-mm sphere to compute a relative measure of local–distant connected voxels. Degree connectivity maps were z-scored separately for local and distant degree. Thus, this plot reflects the relatively greater locally or distantly connected links for each cortical region.

Fig. 7*C* shows the correlation between group-averaged HSE gene profiles for all cortical components (Fig. 1*A*) and the absolute difference in their distant–local degree connectivity score. The negative correlation ($\rho = -0.38$; $P < 0.001$) indicates that transcriptional profile similarity tends to decrease with more divergent distant–local degree connectivity between pairs of brain regions.

Permutation Tests

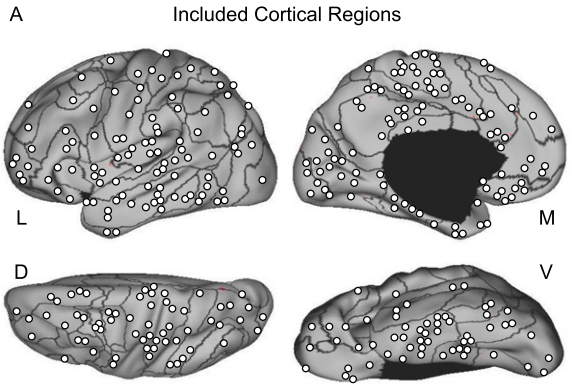
Two permutation analyses were conducted. The first analysis asks whether the elements in the HSE 17 × 17 correlation matrix are significantly different from zero. The second asks whether the elements in the HSE 17 × 17 correlation matrix are significantly greater (positive or negative) than the corresponding correlations obtained from an alternative gene set.

To determine which edges are significantly different from zero in the HSE correlation matrix, the following steps are taken: (*i*) For each subject, average each gene's expression across all samples falling within a given network. The result is a 19 × 17 (average expression of each gene × network) matrix for each subject. That is, each of the 17 networks is associated with a single vector that describes average gene expression for each of the 19 genes. (*ii*) Compute resulting correlation matrix for each subject, average across subjects. This results in a 17 × 17 group-averaged correlation matrix. The diagonal is not meaningful (autocorrelation), leaving 136 (= 17 × 16/2) unique edges. (*iii*) Repeat (*i*) and (*ii*) for 10,000 iterations, each time permuting gene labels between pairs of networks in step (*i*) before computing the correlation matrix for each subject. For each permutation, the permutation order is consistent across subjects.

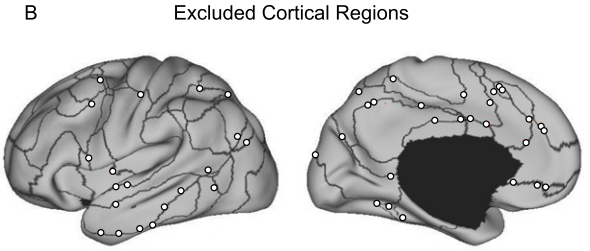
The result is that 103 of 136 edges of the nonpermuted HSE correlation matrix are significantly different from 0 (FDR $q < 0.05$) (Fig. S3*A*).

To determine which edges are significantly greater (positive or negative) in the HSE set vs. alternative gene sets, the following steps are taken: (*i*) For each subject, compute the difference between the 98 × 98 region (omitting regions that are not sampled for that subject) HSE set correlation matrix and a given alternative set. (Here we use the 98 × 98 region matrix, where each region is an individual component of the 17-network parcellation instead of first averaging within networks so that we can compare within-network correlations.) Average correlations for regions falling within the same network and across subjects to obtain a group-averaged difference matrix of correlations within and across networks. (*ii*) For 10,000 iterations, randomly exchange genes between the HSE gene set and the alternative gene set. For example, pool the 19 HSE genes with the 381 Rodent Connectivity genes and randomly split into new sets of 19 and 381 genes. Repeat step (*i*) to obtain a null distribution.

Fig. S3*B* shows the edges that are significantly greater (either positive or negative, FDR $q < 0.05$) in the HSE correlation matrix compared with alternative gene sets.



Network	# Included Samples	Network	# Included Samples
	48		65
	75		45
	168		33
	69		22
	52		86
	20		108
	98		44
	33		64
	112		



	# Excluded Samples	% Subcortical Exclusions
H0351.1009	121	0.96
H0351.1012	106	0.95
H0351.1015	88	0.97
H0351.1016	107	0.94
H0351.2001	156	0.94
H0351.2002	101	0.94

Fig. S1. Locations of included and excluded brain samples. (*A*) Approximate locations of brain samples on a Caret surface representation. Regions were included if centroid MNI coordinates as well as neighboring coordinates had the same network assignment. (*B*) This excluded subcortical structures and structures that lay on the borders between networks.

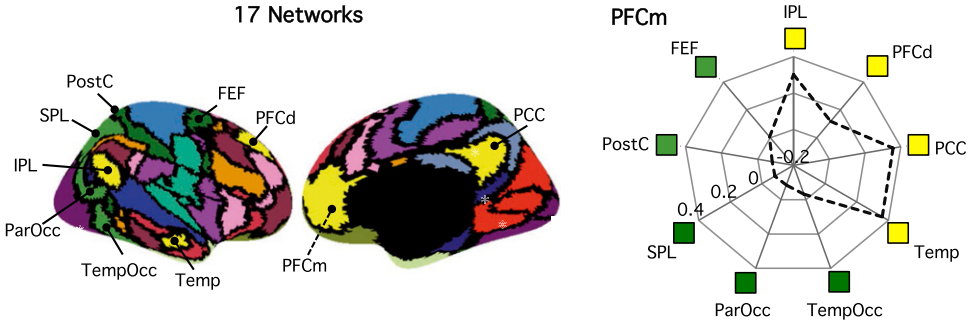


Fig. S2. Transcriptional similarity within the default network. Polar plot showing transcriptional profile correlations between a medial prefrontal cortex region within the default network and other default network regions, as well as dorsal attention network regions. FEF, frontal eye fields; IPL, inferior parietal lobule; ParOcc, occipitoparietal cortex; PCC, posterior cingulate cortex; PFCd, dorsal prefrontal cortex; PFCm, medial prefrontal cortex; PostC, postcentral gyrus; Temp, temporal cortex; SPL, superior parietal lobule; TempOcc, occipitotemporal cortex.

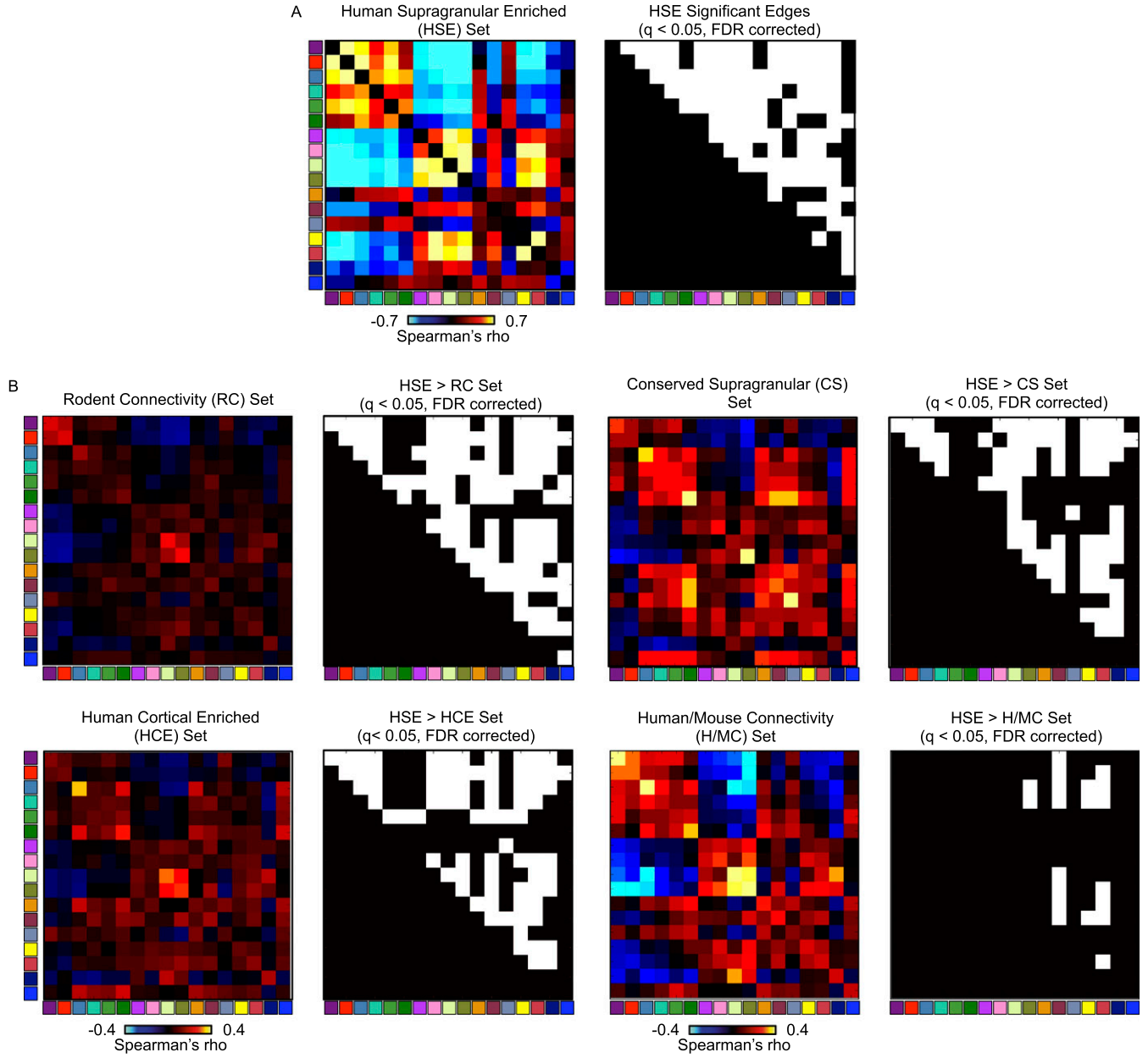


Fig. S3. Significant edges in the HSE set correlation matrix and in the difference in correlation strengths between HSE and alternative gene sets. For each heatmap plot, network-averaged correlation coefficients for the 17 networks are color coded as in Fig. 1 *A* and *C*. For each binary plot, a thresholded matrix shows network pairs in the correlation matrix that are significant at FDR-corrected $q < 0.05$ in white. (*A*) Significant edges in the HSE correlation matrix are determined by permuting gene labels between pairs of networks. Average expression for each of the 19 genes is computed across all brain samples falling within a given network before the correlation is computed. (*B*) Significantly different edges between the HSE set and comparison sets are determined by permuting genes between the two sets for each brain sample and averaging correlations for regions that fall within the same network. See text in *Supporting Information* for additional details.

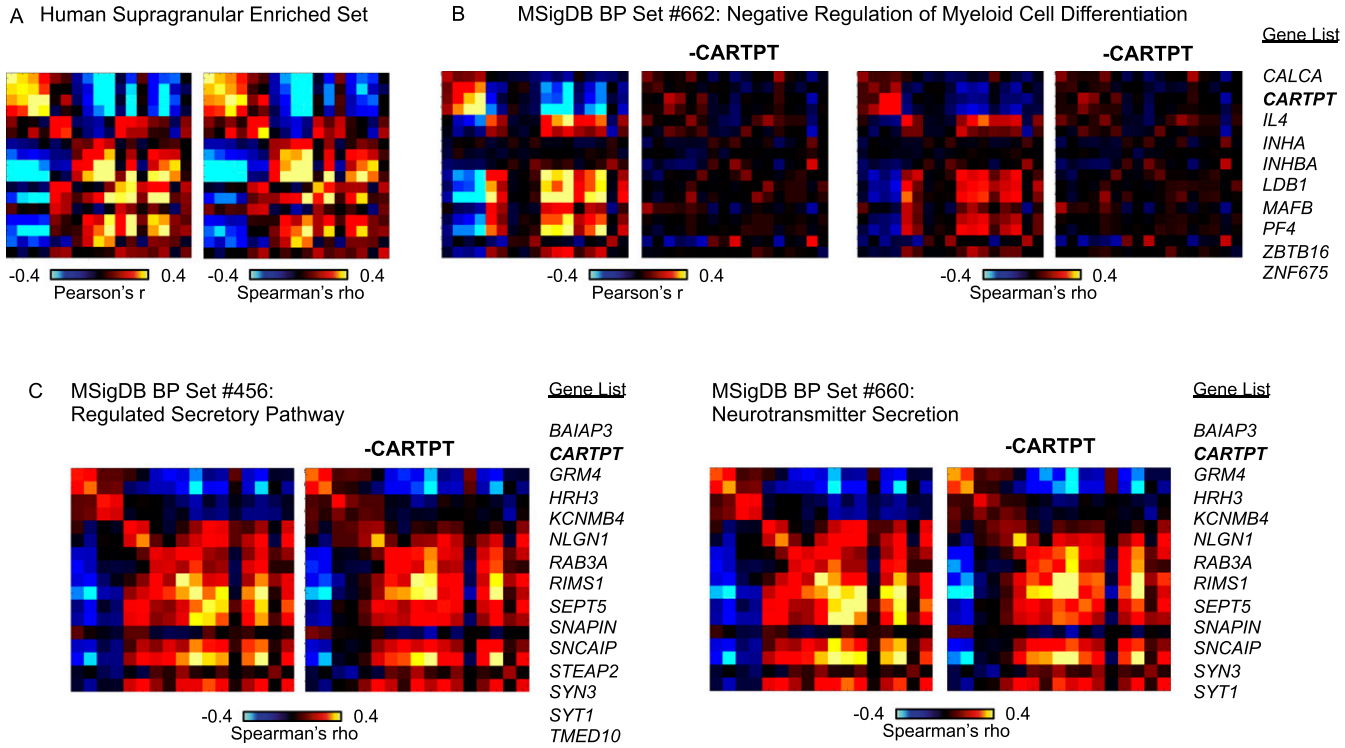


Fig. S4. Spearman's ρ is robust against outliers. (*A*) Network-averaged correlation coefficients for the 17 networks for the HSE genes using either Pearson's product-moment correlation or rank-signed Spearman's ρ. Resulting matrices are comparable using the two metrics. (*B*) Correlation matrices for one of 13 MSigDB gene sets that had large connected components using the NBS metric (see main text). Although weaker than the HSE set, each set showed strong within-network similarity as well as anti-correlations between sensory and association/paralimbic networks. Inspection of these sets revealed one gene in common: *CARTPT*. For 11 of 13 gene sets, *CARTPT* appeared to be an outlier, as removing it from these sets substantially diminished the within- and across-network correlations using Pearson's *r*. Using Spearman's ρ attenuates the influence of such outliers. (*C*, *Left*) Network-averaged correlation coefficients for MSigDB Biological Process Set #456 (Regulated Secretory Pathway), averaged into the 17 networks. (*Right*) MSigDB Biological Process Set #660 (Neurotransmitter Secretion). The two sets are highly overlapping; the Neurotransmitter Secretion set is a subset of 13 of 15 genes contained in the Regulated Secretory Pathway set. Both sets contain *CARTPT*. Unlike other MSigDB gene sets (e.g., *B*), the correlation structure persists when this gene is removed.

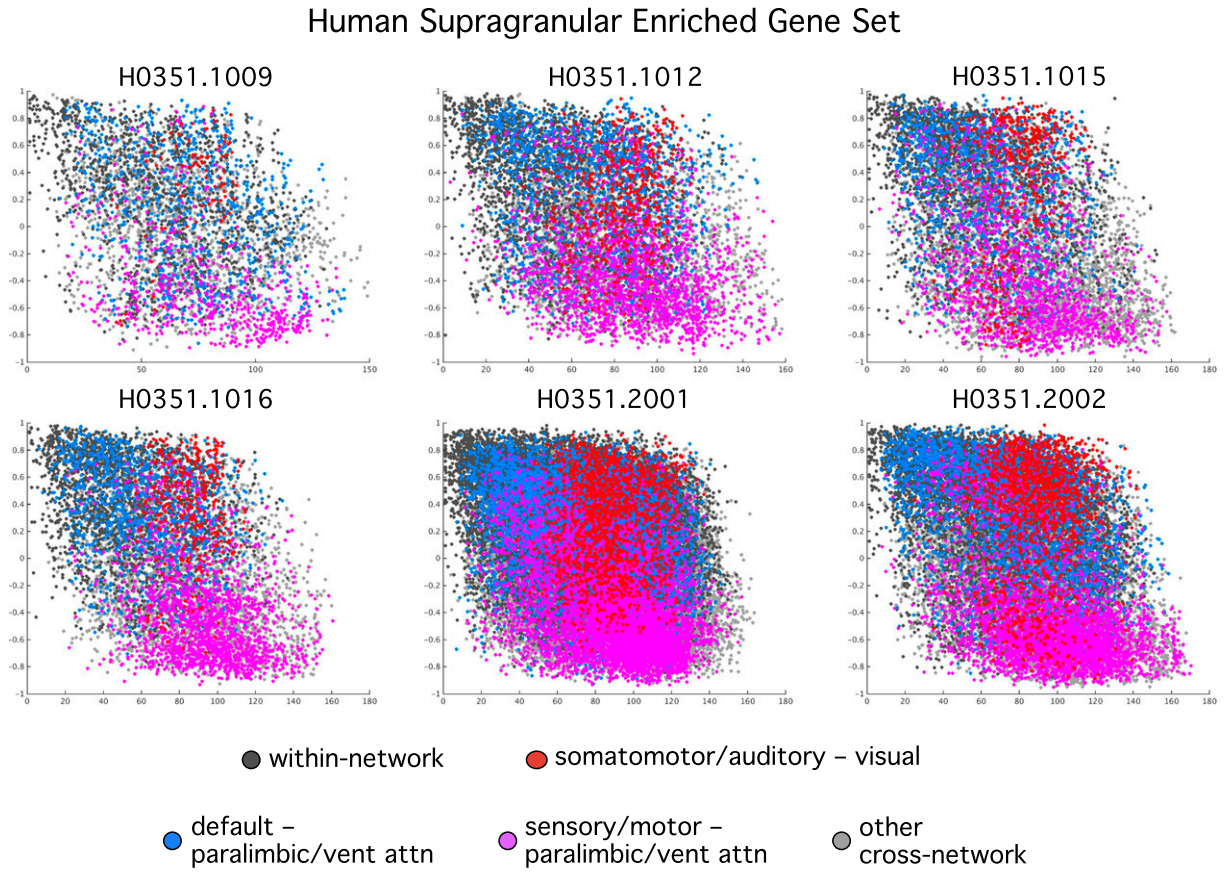


Fig. S5. HSE gene set transcriptional profiles are a function of spatial proximity as well as network identity and cortical type. Transcriptional profile correlations are plotted against Euclidean distance measurements for pairs of brain samples. Dark gray points are within-network pairs. Notice that dark gray points tend to group at the top of the graph, meaning that they have higher correlations even at long distances. Red points are visual network to somato/motor network pairs. Note that they tend to have high correlations despite long distances. Blue points are paralimbic or ventral attention to default network pairs. Note that they tend to have high correlations at both close and long distances. Magenta points are paralimbic or ventral attention to visual network or somato/motor network pairs. Note they tend to have low correlations at all distances. Light gray points are all other combinations. Note that they tend to follow the overall negative correlation and are particularly evident at the long-range, low-correlation corner of the graph.

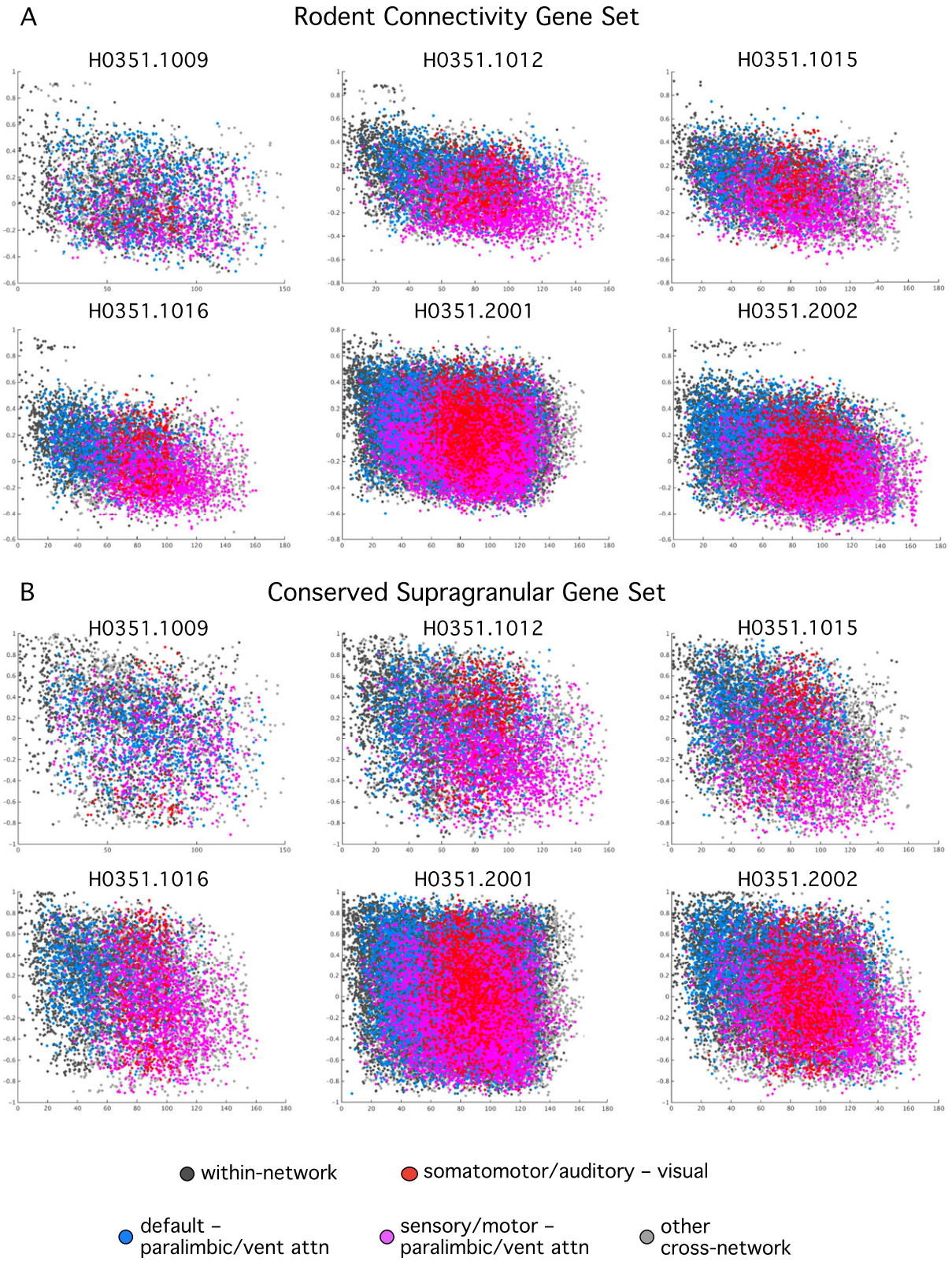


Fig. S6. Effect of spatial proximity and network identity on alternative gene sets. (*A*) Rodent Connectivity gene set. (*B*) Conserved Supragranular gene set. Color-coding follows Fig. S5.

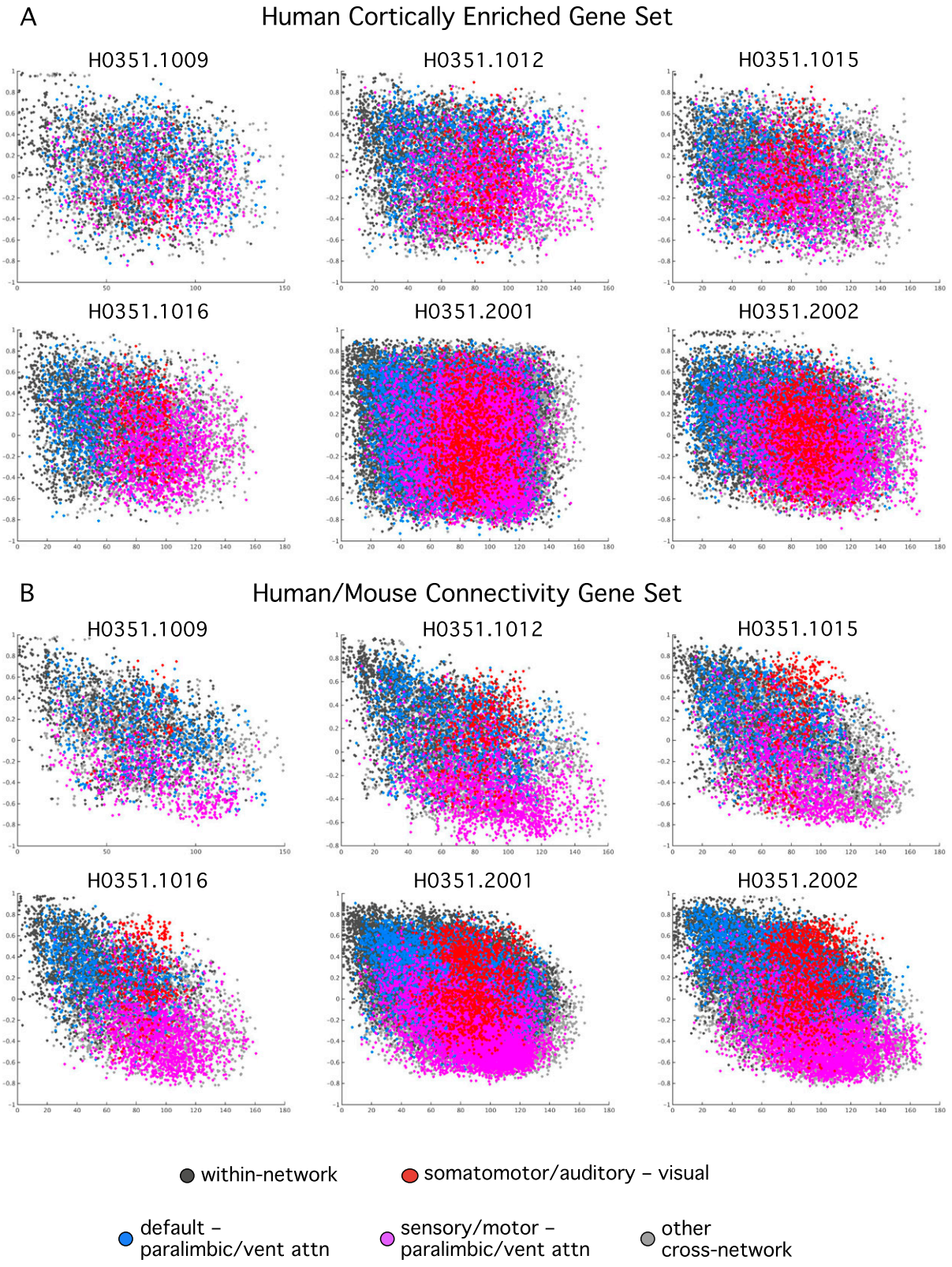


Fig. S7. Effect of spatial proximity and network identity on alternative gene sets. (*A*) Human Cortically Enriched gene set. (*B*) Human/Mouse Connectivity gene set. Color-coding follows Fig. S5.

Other Supporting Information Files

Dataset S1 (XLSX)