Title:

Commentary: BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. S*cience* 348, 1241-4.

Spiro P.Pantazatos[1], Xinyi Li[2]
Departments of Psychiatry[1] and Biomedical Informatics[2], Columbia University, New York, NY

Supplementary Discussion

Although not reported in the original article (Richiardi et al., 2015), the authors claim that SF remains significant after a linear regression-based distance correction is applied and only positive connections are included (personal communication). However, there are two problems with this: 1) The assumption that tissue-tissue correlation strength various linearly with distance is too strong. A plot of the tissue-tissue correlations vs. distance shows that the best-fit curve is steep for short edges and less steep at around 20mm: after adjusting for the best-fit line there will still be a distance bias. No matter what model is adjusted for, the correction will not be as optimal as simply removing proximal connections. 2) Applying an arbitrary cutoff of zero for connections contributing to the SF is not well justified (this applies to the main analyses as well). What, biologically, distinguishes a correlation of 0.1 vs -0.1 other than i.e. noise in the expression vector that nudges the correlation across zero? Furthermore, after regression, about half of the connections (that were included in the original main analyses) will be negative due to mean centering and omitted in the new analysis, making the cutoff of zero even more arbitrary.

It is likely that the optimization approach in Richiardi et al. used to derive the 136 consensus genes (i.e. multiplying each gene's expression by 10 and recalculating the strength fraction) identified genes with both high local spatial autocorrelation and variability across the cortex. This is consistent with the observation that >75% of these 136 consensus genes are in the

top 10% of genes found to have consistently high region-to-region variability (so called differentially stable, DS genes) across the cortex identified in Hawrylyzc et al. (Hawrylycz et al., 2015). Furthermore, GO functions related to potassium channels (featured prominently in Richiardi et al. Supplemental Table S3) were most over-represented among high-DS genes ($P < 1.70 \times 10^{-12}$) in Hawrylyzc et al. (Hawrylycz et al., 2015). Given that genes high in DS (i.e. consistent region-to-region variability), *irrespective of belonging to resting state functional networks*, are more likely to be involved in brain functioning (Hawrylycz et al., 2015), this could account for the enrichment (p=0.006) of SNPs associated with functional network SF observed in the IMAGEN portion of the Richiardi et al. analyses.

Figure 2 in Richiardi et al. is misleading, and does not constitute evidence for "definite differences in functional connectivity strength mostly within the functional networks themselves." Given that the authors used a post-hoc, biased approach to generate the loosely thresholded functional connectivity difference matrices and maps, it is unclear whether comparable results could be generated when applying their scoring procedure to 136 genes randomly selected from the background set or from the top 10% of genes showing variability across the cortex (i.e. cortical DS genes reported in (Hawrylycz et al., 2015)). Finally, the results from mouse tractography data (p=0.011 Mantel correlation, Figure 3 in Richiardi et al.) does not make any adjustment for spatial proximity, and is likely also confounded by spatial proximity.

The Richiardi et al. study is an important step towards identifying genes whose spatial pattern of cortical expression relate to distributed functional networks consistently observed in resting state fMRI. However, we are not quite there yet. Further work will be required to adequately control for the confounding effects of spatial proximity. While here distances were computed in 3D MNI space, computing distances in flattened cortical surface (2D) space would

make distance measurements more accurate. While distance correction using 3D Euclidean distance is suboptimal compared to 2D Euclidean, it is optimal compared to using region labels. 2D Euclidean distance would be more accurate than 3D distance, and while the slope and shape of the curve in Fig 1C might change, SF would still fall monotonically as short-range edges are removed and be no greater than the null distribution at around 24-32 mm.

Future studies to relate gene expression with resting-state functional networks will require valid and more appropriate null distributions, and could benefit from "non-parametric" approaches to correct for distance (i.e. calculating outcome measures such as within-network SF across distance bins to directly visualize distance effects etc.)

References:

1. Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL, et al. Canonical genetic signatures of the adult human brain. Nat. Neurosci. 2015;18(12):1832-44.