

## Statistical issues for haplotype analysis in Neuroimaging

Speaker: Tom Nichols, Ph.D.

University of Michigan, Department of Biostatistics

Imaging data is inherently multivariate. As haplotypes are indeed more powerful than single SNP's (Single Nucleotide Polymorphism) for detecting functional variance, the use of haplotype counts to predict/model brain function is possible through haplotype regression methods. Typically, multiple haplotypes are consistent with the SNP data, however ambiguity can occur when one have 2 or more heterozygotic SNP's in the sample. One way of dealing with ambiguity is the use of expected counts of haplotypes, via the Error-in-variables Model (EM), on which haplotype counts are not integers, but fractions. Although EM theory brings the problem that predictors are only *estimates of true predictors*, it also implies that an influence of haplotyping errors on BOLD residuals seems implausible (Independence assumption). In this sense, Dr. Nichols has been able to apply the Haplotype Trend Regression (HTR) model to test the association between statistically inferred haplotypes and particular traits. He has analyzed fMRI data using the HRT model by defining contrasts of interest for each subject (first-level analysis) and regress those values on the estimated haplotype frequencies (second-level analysis). According to Dr. Nichols, second-level analysis allows for two kinds of interrogations: 1) **estimate the average effect**, which tells where the non-genetic specific effects are, identifies regions of the whole brain with *any* response, and pay the price for multiple comparisons. And 2) **estimate haplotype-specific effects** (F-contrast/tests, t-contrast/tests), considering only those regions on where he got specific findings. Of the several ways to deal with the spatial multiple testing problem; Familywise Error Rate (FWER), False Discovery Rate (FDR), Bonferroni test, Random Field Theory (RFT) and Permutations, he emphasized the merits of the last two in terms of adaptability and less computational intensity (RFT) and exchangeability (Permutations). In conjunction with Dr. Meyer-Lindenberg, he has analyzed the effects of COMT haplotypes in working memory, using these methods to control for false positives. Finally he proposed "deep thoughts" on the function vs. structural effects dichotomy. On one side, there are the systematic variations on Gray Matter (GM) density by genotype/haplotype. On the other side, responses to cognitive challenges may depend on the haplotype too. But in practice we probably never get a pure effect, he said. Thus, he raised the question: "Does this matter? What if I report a functional effect when in actual fact it's purely a structural effect?" His two possible answers were: a) No it doesn't matter because finding any sort of genetic variation is useful. Furthermore, these differences might as well be arbitrary at some level (is dendritic plasticity a functional or structural effect?) and b) Yes it does matter because we are not getting the full story. He proposed dealing with this ambiguity with a two-step approach, on which structural data is analyzed first, the GM data is modeled with the genetic data, and those results are interpreted. Then, that GM information should be included as a nuisance variable in the subsequent analysis of the functional data. And when the haplotype effects on the BOLD signal are interpreted, "you know that you already discounted your anatomical differences."