Semanti Mukherjee

The significance of our work and the key findings:

A germline Janus kinase 2 (*JAK2*) SNP that increases susceptibility to myeloproliferative neoplasms (MPNs) was the focus of our study published in March 2009 in *Nature Genetics*. MPNs are a group of diseases that involve overproduction of one or more types of differentiated blood cells from the bone marrow. The chronic MPNs polycythemia vera (PV), essential thrombocytosis (ET), and primary myelofibrosis (MF) are commonly associated with the mutation *JAK2* ^{V617F}. There likely are additional genetic events that contribute to the pathogenesis of these phenotypically distinct disorders. To identify germline variation contributing to MPN predisposition and phenotypic pleiotropy, we performed a genome-wide association (GWA) study. GWA studies have been advocated as the method of choice for identifying genetic variants associated with complex diseases. In a GWA study design, cases and controls are ideally matched for ethnicity, age, sex, socio-economic background and other environmental factors. Recently, the concept of using shared controls, such as the blood donor cohort in the Wellcome Trust Case Control Consortium (WTCCC), for GWA study has been shown to be a possible approach. We investigated the idea of using controls from published GWA studies as shared controls for our MPN study. The controls from WTCCC provided us with shared controls. We developed a method to match MPN cases with shared controls based on genetic ancestry. To determine population stratification biases that could be introduced by the shared controls, we carried out principal components analysis (PCA) using EIGENSTRAT. MPN cases and shared controls who cluster on the first two principle components were selected for association test. A *JAK2* SNP rs10974944 was significantly associated with MPN risk after correcting for residual population stratification and multiple testing. Further genetic analysis of 324 patients with PV, ET, or PMF determined that the "G" allele (GG or CG) at rs10974944 was more common in MPN patients than control individuals (odds ratio [OR], 3.1; $P = 4.1 \times 10^{-20}$) and was shown to preferentially acquire the V617F mutation. This illustrates a complex interplay between somatic and germline genetics in cancer. These data indicate a germline variation in *JAK2* locus is an important risk factor for MPN predisposition. The present approach of using shared controls can be employed to identify cancer predisposition loci in other GWA studies. Thus shared controls is cost effective and very attractive in terms of not having to identify multiple series of controls for different GWA studies.