Ruprecht-Karls-Universität Heidelberg
Medizinische Fakultät Mannheim
Dissertations-Kurzfassung

Biologically informed risk scoring in schizophrenia based on genome-wide omics data

Autor: Junfang Chen
Institut / Klinik: Zentralinstitut für Seelische Gesundheit Mannheim (ZI)
Doktorvater: Dr. E. Schwarz

Extensive efforts in characterizing the biological architecture of schizophrenia have moved psychiatric research closer towards clinical application. As our understanding of psychiatric illness is slowly shifting towards a conceptualization as dimensional constructs that cut across traditional diagnostic boundaries, opportunities for personalized medicine applications that are afforded by the application of advanced data science methods on the increasingly available, large-scale and multimodal data repositories are starting to be more broadly recognized. A particularly intriguing phenomenon is the discrepancy between the high heritability of schizophrenia and the difficulty in identifying predictive genetic signatures, for which polygenic risk scores of common variants that explain approximately 18% of illness-associated variance remain the gold standard. A substantial body of research points towards two lines of investigation that may lead to a significant advance, resolve at least in part the 'missing heritability' phenomenon, and potentially provide the basis for more predictive, personalized clinical tools.

First, it is paramount to better understand the impact of environmental factors on illness risk and elucidate the biology underlying their impact on altered brain function in schizophrenia. This thesis aims to close a major gap in our understanding of the multivariate, epigenetic landscape associated with schizophrenia, its interaction with polygenic risk and its association with DLPFC-HC connectivity, a well-established and robust neural intermediate phenotype of schizophrenia. As a basis for this, we have developed a novel biologically-informed machine learning framework by incorporating systems-level biological domain knowledge, i.e., gene ontological pathways, entitled 'BioMM' using genome-wide DNA methylation data obtained from whole blood samples. An epigenetic poly-methylation score termed 'PMS' was estimated at the individual level using BioMM, trained and validated using a total of 2230 whole-blood samples and 244 post-mortem brain samples. The pathways contributing most to this PMS were strongly associated with synaptic, neural and immune system-related functions. The identified PMS could be successfully validated in two independent cohorts, demonstrating the robust generalizability of the identified model. Furthermore, the PMS could significantly differentiate patients with schizophrenia from healthy controls when predicted in DLPFC post-mortem brain samples, suggesting that the epigenetic landscape of schizophrenia is to a certain extent shared between the central and peripheral tissues. Importantly, the peripheral PMS was associated with an intermediate neuroimaging phenotype (i.e., DLPFC-HC functional connectivity) in two independent imaging samples under the working memory paradigm. However, we did not find sufficient evidence for a combined genetic and epigenetic effect on brain function by integrating PRS derived from GWAS data, which suggested that DLPFC-HC coupling was predominantly impacted by environmental risk components, rather than polygenic risk of common variants. The epigenetic signature was further not associated with GWAS-derived risk scores implying the observed epigenetic effect did likely not depend on the underlying genetics, and this was further substantiated by investigation of data from unaffected first-degree relatives of patients with SCZ, BD, MDD and autism. In summary, the characterization of PMS through the systems-level integration of multimodal data elucidates the multivariate impact of epigenetic effects on schizophrenia-relevant brain function and its interdependence with genetic illness risk.

Second, the limited predictive value of polygenic risk scores and the difficulty in identifying associations with heritable neural differences found in schizophrenia may be due to the possibility that the manifestation of the functional consequences of genetic risk is modulated by spatio-temporal as well as sex-specific effects. To address this, this thesis identifies sex-differences in the spatio-temporal expression trajectories during human development of genes that showed significant prefrontal co-expression with schizophrenia risk genes during the fetal phase and adolescence, consistent with a core developmental hypothesis of schizophrenia. More specifically, it was found that during these two time-

periods, prefrontal expression was significantly more variable in males compared to females, a finding that could be validated in an independent data source and that was specific for schizophrenia compared to other psychiatric as well as somatic illnesses. Similar to the epigenetic differences described above, the genes underlying the risk-associated gene expression differences were significantly linked to synaptic function. Notably, individual genes with male-specific variability increases were distinct between the fetal phase and adolescence, potentially suggesting different risk associated mechanisms that converge on the shared synaptic involvement of these genes. These results provide substantial support to the hypothesis that the functional consequences of genetic risk show spatiotemporal specificity. Importantly, the temporal specificity was linked to the fetal phase and adolescence, time-periods that are thought to be of predominant importance for the brain-functional consequences of environmental risk exposure. Therefore, the presented results provide the basis for future studies exploring the polygenic risk architecture and its interaction with environmental effects in a multivariate and spatiotemporally stratified manner.

In summary, the work presented in this thesis describes multivariate, multimodal approaches to characterize the (epi-)genetic basis of schizophrenia, explores its association with a well-established neural intermediate phenotype of the illness and investigates the spatio-temporal specificity of schizophrenia-relevant gene expression effects. This work expands our knowledge of the complex biology underlying schizophrenia and provides the basis for the future development of more predictive biological algorithms that may aid in advancing personalized medicine in psychiatry.