1639M

Exploring the genome-wide roles of transcription factors and their complexes in chromosome interaction. *MJ. Li[1,2], LY. Wang[1], PC. Sham[3], MQ. Zhang[4], JS. Liu[2], JW. Wang[1].* 1) Biochemistry, The University of Hong Kong, Hong Kong; 2) Statistics Dept, Harvard University, Boston, MA, USA; 3) Psychiatry Dept, The University of Hong Kong, Hong Kong; 4) Molecular and Cell Biology Dept,, The University of Texas at Dallas, Dallas, TX, USA.

The tight regulation of genes in different cells is governed by temporal and spatial biological signals. It is very important to pinpoint the pattern of transcription factors (TFs) and their complexes in looping interactions and to detect TF complexes as well as the underlying cis-regulatory modules (CRMs) in different human cell types. Existing studies on analysis of TFs and their complexes were only performed at one dimension and not at genome-wide scale. Recently, the unbiased chromosome conformation capture, Hi-C, can detect the genome-wide chromatin interactions, but has restrictions on its resolution due to the variable cell-to-cell chromosome structures and inadequate sequencing depth. In this study, we provide a comprehensive analysis on TFs regulatory pattern within chromosome looping by combining Hi-C and ENCODE ChIP-Seq data from three human cell types (GM12878, H1-hESC and K562). We first devised a strategy to map ChIP-Seq peaks of each TF to a normalized 10kb Hi-C contact matrix and construct an interaction matrix for each participant TF. We observed tight correlation for TFs participant activities in high resolution chromosome looping between biological replicates, which indicate the TF activities is more stable than local DNA interactions. To check the enrichment of different chromatin marks and genomic features in the interaction region of each participant TF, we performed enrichment test on several histone modifications marks (H3K4me1, H3K4me2, H3K4me3, H3K27me3, H3K36me3, H3K9ac, H3K27ac), CAGE, DHSs, DNA methylation (RRBS) and conservation signals. We found chromatin markers and genomic features of each TF is highly correlated between replications. We also observed that grouped TFs shared consistent patterns of chromatin marks and genomic features, which indicate their similar roles in gene regulation. Furthermore, we developed a method to detect the candidate TF complexes by reducing the dimension of huge sparse interaction matrix for each participant TF followed by community clustering. We successfully detected several known complexes mediating long range enhancer-promoter interaction in different cell types. Using the combinatory motifs scanning, we can predict the genome-wide CRMs for specific TF complex with known motifs. Finally, by correlating gene expression profiles of TF and its targets , we can pinpoint the role of detected TF complexes (active or repressive) in controlling cell type specific gene regulation.

1640T

Comparing blood and brain gene expression networks in Huntington's Disease by semantic analysis. *E. Mina, P.A.C. 't Hoen, W. van Roon-Mom, M. Roos.* Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Huntington's Disease (HD) is a neurodegenerative disease with the most prominent pathology in the brain. However, human brain tissue is not easily accessible for molecular biology studies and it cannot be isolated from living patients. The widespread pathology of HD indicates that the phenotype is not limited to brain dysfunction. Several studies have shown that understanding alterations in peripheral tissues could be valuable for monitoring disease progression. Considering that transcriptome technologies have successfully been used for biomarker discovery and the study of physiological and pathophysiological mechanisms, it is evident that studying transcriptional changes in peripheral tissue can provide new insights that can lead to the development of new therapies and markers to monitor disease progression. In addition, novel sequencing technologies and the wide availability of data and published articles, create opportunities for novel, more effective ways of data exploitation. Our methodology identifies and prioritizes disease signatures from blood that are also associated with abnormalities in brain, in a robust way by a combination of large scale network analysis, integration of heterogeneous datasets, and incorporation of prior knowledge mined from literature. Our preliminary data show that modules in gene co-expression networks are significantly more reproducible than individual gene markers between blood and brain. We additionally link significant HD signatures to biological and pharmacological knowledge sources to prioritize biomarkers based on their potential as drug targets using a 'Linked Data' approach. This allowed us to identify drug targets by querying across four different data sources, including drugbank and pathway data to target pathways that are impaired in HD. For instance, we found Imatinib, associated with the gene ABL1 that targets autophagy and is known to delay production of beta-amyloid. Moreover, in the HD context, we prioritize drug targets according to their structural or functional similarity to known targets that can pass the blood-brain barrier. In conclusion, our approach is a powerful method to identify genes that are potentially involved in physiological and pathophysiological processes in HD. Using the 'Linked Data' approach, our results can be easily extended and will become available to the scientific community as a resource for further investigation for understanding disease pathology in HD.

1641S

Genetic Risk Prediction and Neurobiological Understanding of Alcoholism. *A. Niculescu[1, 12], D. Levey[1], H. Le-Niculescu[1], J. Frank[2], M. Ayalew[1], N. Jain[1], B. Kirlin[1], R. Learman[1], E. Winiger[1], Z. Rodd[1], A. Shekhar[1], N. Schork[3], F. Kiefer[4], N. Wodarz[5], B. Muller-Myhsok[6], N. Dahmen[7], M. Nothen[11], R. Sherva[8], L. Farrer[8], A. Smith[9], H. Kranzler[10], M. Rietschel[2], J. Gelernter[9], GESGA Consortium.* 1) Psychiatry, Indian Univ Sch Medicine, Indianapolis, IN; 2) Central Institute of Mental Health, Mannheim, Germany; 3) Department of Human Biology, The J. Craig Venter Institute, La Jolla, California, USA; 4) Dept. of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University. Germany; 5) Dept. of Psychiatry, University Medical Center Regensburg, Univ. of Regensburg, Germany; 6) Dept. of Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany; 7) Dept. of Psychiatry, Univ. of Mainz, Germany; 8) Boston University School of Medicine, Department of Medicine (Biomedical Genetics); 9) Yale University School of Medicine, Department of Psychiatry, Division of Human Genetics; and VA CT Healthcare Center; 10) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, and Philadelphia VAMC; 11) Dept. of Genomics, Life & Brain Center, Univ. of Bonn; Inst. of Human Genetics, Univ. of Bonn, Germany; 12) Indianapolis VA Medical Center, Indianapolis, Indiana, USA.

We have used a translational convergent functional genomics (CFG) approach to discover genes involved in alcoholism, by gene-level integration of genome-wide association study (GWAS) data from a German alcohol dependence cohort with other genetic and gene expression data, from human and animal model studies, similar to our previous work in bipolar disorder and schizophrenia. A panel of all the nominally significant p-value SNPs in the top candidate genes discovered by CFG (n=135 genes, 713 SNPs) was used to generate a genetic risk prediction score (GRPS), which showed a trend towards significance (p=0.053) in separating alcohol dependent individuals from controls in an independent German test cohort. We then validated and prioritized our top findings from this discovery work, and subsequently tested them in three independent cohorts, from two continents. In order to validate and prioritize the key genes that drive behavior without some of the pleiotropic environmental confounds present in humans, we used a stress-reactive animal model of alcoholism developed by our group, the DBP knock-out mouse. A much smaller panel (n=11 genes, 66 SNPs) of the top CFG-discovered genes for alcoholism, cross-validated and prioritized by this animal model, showed better predictive ability in the independent German test cohort (p= 0.041). The top CFG scoring gene for alcoholism from the initial discovery step, synuclein alpha (SNCA), remained the top gene after the stress-reactive animal model cross-validation. We also tested this small panel of genes in two other independent test cohorts from the United States, one with alcohol dependence (p=0.00012), and one with alcohol abuse (a less severe form of alcoholism) (p=0.0094). SNCA by itself was able to separate alcoholics from controls in the alcohol dependent cohort (p= 0.000013) and the alcohol abuse cohort (p= 0.023). So did 8 other genes from the panel of 11 genes taken individually, albeit to a lesser extent and/or less broadly across cohorts. SNCA, GRM3 and MBP survived strict Bonferroni correction for multiple comparisons. Taken together, these results suggest that our stress-reactive DBP animal model helped to validate and prioritize from the CFG-discovered genes some of the key behaviorally relevant genes for alcoholism. These genes fall into a series of biological pathways involved in signal transduction, transmission of nerve impulse (including myelination), and cocaine addiction.