1257F

An analysis pipeline for detecting copy number variations with a low false discovery rate in microarray data. *D.-A. Clevert[1,2], A. Mitterecker[1], A. Mayr[1], G. Klambauer[1], M. Tuefferd[3], A. De Bondt[3], W. Talloen[3], H. Göhlmann[3], S. Hochreiter[1].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Department of Nephrology and Internal Intensive Care, Charité University Medicine, Berlin, Germany; 3) Johnson & Johnson Pharmaceutical Research & Development, a Division of Janssen Pharmaceutica, Beerse, Belgium.

Motivation: A low false discovery rate (FDR) at the detection of copy-number aberrations (CNAs) in microarray data ensures sufficient detection power and prevents failures in CNA-disease association studies. A high FDR means many falsely discovered aberrations, which are not associated with the disease, though correction for multiple testing must take them into account. Thus, a high FDR not only decreases the discovery power of studies but also the significance level of the remaining discoveries after correction for multiple testing. **Methods:** We obtain a low FDR at the detection of CNAs in microarray data by a probabilistic latent variable model, called "cn.FARMS". The model is optimized by Bayesian maximum a posteriori approach, where a Laplace prior prefers models, which represent the null hypothesis of observing a constant copy number 2 for all samples. The posterior can only deviate from this prior by strong (deviation from copy number 2 intensities) and consistent signals in the data, which hints at a CNA - the alternative hypothesis. The information gain of the posterior over the prior gives the informative/non-informative (I/NI) call that serves as a filter for CNA candidate regions. I/NI call filtering reduces the FDR, because a region with a large I/NI call is unlikely to be a falsely detected CNA, which would neither have strong nor consistent measurements. It can be shown that the I/NI call filter applied to null hypotheses of the association study is independent of the test statistic which in turn guarantees that a type I error rate control by correction for multiple testing is still possible after filtering. I/NI-calls perform well for the usually rare CNAs that are seen at few samples only, where variance-based filtering approaches fail. **Results:** cn.FARMS clearly outperformed prevalent methods for CNA detection with respect to sensitivity and especially with respect to FDR on different HapMap benchmark data sets. **Availability:** The software cn.FARMS is publicly available as an R package at Bioconductor and at http://www.bioinf.jku.at/software/cnfarms/cnfarms.html.

1258F

Comparison of different reference genes used for qPCR-based CNV quantification. *N. Fang, A. Missel, C. Beckmann, U. Deutsch.* QIAGEN, Hilden, Germany.

Copy Number Variant (CNVs), the change of the DNA copy number in the genome, has been recently shown to be a widely-spread phenomenon that affects about 10-20% of the human genome. The occurrence of the CNVs has been associated with various diseases such as autism, autoimmune disorders, and cancer. The most commonly used molecular biology tools for discovery of CNVs are array and next-generation sequencing (NGS). These two high-throughput methods can discover multiple potential CNVs, which normally need to be validated with an independent method. Once validated, the confirmed CNVs can also be examined in a large number of samples to identify the statistically significant association of the CNV and phenotype. Quantitative PCR (qPCR), with its ease of use, sensitivity, and scalability, is often the method of choice for CNV validation and association studies. Relative quantification principle is used for this application: first, a reference gene, whose copy number is presumed to be constant in different genomes, has to be defined. The copy number of the genes of interest (GOIs) is then calculated based on the Ct difference of GOI and reference gene among different samples. Since the consistent copy number of the reference gene is essential for the qPCR-based CNV quantification, we evaluated the reliability of commonly-used single copy reference genes such as RNaseP, as well as other candidates. Our results suggest that, compared to single copy genes, stable multi-copy regions can serve as a more sensitive and reliable CNV quantification reference. We also demonstrate more reliable CNV quantification with the REST software, which takes different qPCR efficiency into consideration and performs statistical analysis of the qPCR data.

1259F

CNVs detection from targeted sequencing of genes associated with congenital heart defects. *Y. Lai[1], A. Postma[5], T. Rahman[2], J. Laros[1], Y. Ariyurek[1], S. Sperling[3], S. Klaassen[4], J. Goodship[2], P. 't Hoen[1].* 1) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Institute of Human Genetics, Newcastle University, Central Parkway, Newcastle upon Tyne, UK; 3) Group Cardiovascular Genetics, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Max Delbrück Center for Molecular Medicine, Berlin, Germany; 5) Heart Failure Research Center, Academic Medical Center Amsterdam, Netherlands.

Copy number variations (CNVs) in the genome are an important source of genetic variability and underlie many disease phenotypes. If a gene residing in a copy number variable region is dosage sensitive, this can affect its gene expression level. For instance when one has only one functional copy of a gene, which may affect the abundance of the protein to support its normal function. Previous studies showed that numerous genes encode transcription factors play an important role in regulating heart formation. We hypothesize that haploinsufficiency of genes related with heart development is the reason for various severity of congenital heart defects. Haploinsufficiency of genes can be due to copy number change of dosage-sensitive genes or unmasking of recessive alleles by single base substitutions or short indels of the functional copy. Here we focus on the aspect of copy-number alterations within the disease cohort. Most current CNVs detection methods are designed for whole genome resequencing data and are able to identify large (>10 kb) or small (10 bp) indels (insertions or deletions). The objective is to discover medium sized deletions from exome sequencing data. Within the EU-sponsored HeartRepair project, we targeted exons of ~400 genes associated with congenital heart defects and sequence the targeted regions to high depth for ~160 patient samples. From this data, we compiled a matrix of reads that can be aligned uniquely to the targeted regions in samples passing minimum coverage criteria (>80% of targeted regions with coverage >20X) and normalized the matrix to the total number of reads on targets. We calculated the coefficient of variation on the depth of coverage for each targeted region to assess variability among samples, assuming that copy number polymorphic regions will show higher dispersion of the coverage across samples. ChrX can be seen as a copy number polymorphism between male and female samples. We used the 119 targeted regions on ChrX as bench mark for calling copy number variable regions and could clearly detect the higher coefficient of variation in the complete data set compared to data sets containing females or males only. 117 copy number polymorphic regions were detected in the 4200 targeted autosomal regions, of which two are overlapping with known CNVs in DGV database. We further assess the validity of the detected copy number polymorphic regions by checking the proportion of unbalanced heterozygous calls.

1260F

Complex genomic aberrations can be the cause of variable phenotypes of 22q11.2 deletion or duplication syndrome. *D. Li, M. Buch, M. Tekin, YS. Fan.* University of Miami, Miller School of Medicine, Miami, FL.

It is well known that the phenotypes of patients with DiGeorge syndrome can be extremely variable from near normal to severe developmental disabilities including congenital heart defects and risk of psychological problems such as schizophrenia. The causes of lack of genotype-phenotype correlation are not well understood. Prior to the clinical use of array CGH, deletion or duplication in the 22q11.2 region was detected by FISH, and therefore changes in the genome other than the 22q11.2 region was barely known. Array CGH studies have made it possible to reveal the imbalances genome-wide, and the findings of copy number changes other than the 22q11.2 microdeletion or duplication may explain the complexity of the genotype-phenotype correlations in these patients. By array CGH, we have detected copy number changes in the 22q11.2 region in 16 (11 deletions; 5 duplications) of 1292 cases referred for intellectual disabilities and/or congenital anomalies. We observed complex genomic imbalances in 2 of the 16 cases. The first case was a baby female of 15 days of age with tetrology of Fallot revealed by ultrasound. Array CGH showed a 2.47 Mb deletion in the 22q11.2 DiGeorge region and a 1.56 Mb deletion in the Xp22.31 region involving multiple genes including the STS gene which causes X-linked recessive ichthyosis when deleted or mutated. The second case was a baby boy of 15 days of age also with an abnormal ultrasound showing complex congenital heart defects. In this baby, array CGH detected multiple pathogenic copy number alterations, including a 2.84Mb duplication in the 22q11.2 DiGeorge region, a 605kb duplication in the 15q13.3 region involving the CHRNA7 gene, as well as a 209 kb deletion in 16p13.2 region involving the A2BP1 gene. Copy number changes of CHRNA are associated with mental retardation, autism and schizophrenia. Deletion of A2BP1 has been reported in patients with mental retardation, autism and seizure. Our observations have provided evidence that complex genomic changes are not rare in DiGeorge patients and they may contribute to the extremely variable phenotypes in this disease. Also, the additional changes such as duplication of CHRNA and deletion of A2BP1 may add additional risk for psychological diseases in the patient. Continuing follow-up and collection of detailed clinical information in these 2 patients may help our understanding on the complicated phenotypes of the DiGeorge or 22q11.2 microdeletion/duplication syndrome.