

MEETING ABSTRACTS

Open Access



Abstracts from the Human Genome Meeting 2018

Yokohama, Japan, 12-15 March 2018

Published: 09 March 2018

A1

TCGA whole-exome sequencing data reveals mutational parameters in distinguishing potential cancer drivers

Daniel Wai-Hung Ho^{1,2} and Irene Oi-Lin Ng^{1,2}

¹Department of Pathology, The University of Hong Kong, Hong Kong, Hong Kong; ²State Key Laboratory for Liver Research, The University of Hong Kong, Hong Kong, Hong Kong
Human Genomics 2018, **12**(Suppl 1):A1

Background

With huge amount of genome-wide mutational data generated by cancer genomic sequencing studies, distinguishing cancer drivers from the vast majority of passengers is important. Existing cancer driver prediction methods capture specific mutational aspects in discriminating potential cancer drivers. We explore the possibility of alternative way in doing the task.

Methods

We noted mutational parameters (functional mutation ratio, mutation frequency and sample mutation recurrence) vary differently among mutant genes of different sizes. This led us to develop our novel algorithm (Mutant Gene Ranker - MuGeR), incorporating the comparison of multiple mutational parameters of target gene against the corresponding background derived from a specific subset of genes using a sliding window approach, to estimate the likelihood of target genes for being potential cancer drivers. We applied our MuGeR algorithm on the The Cancer Genome Atlas (TCGA) datasets.

Results

Empirical data on the TCGA datasets and comparison with the prioritization results of 4 other existing tools (MuSiC, MuSig, TUSON explorer and DOTS-Finder) suggested satisfactory performance of our MuGeR algorithm. More importantly, we demonstrated the existence of specific pattern for mutational parameters across cancers.

Conclusions

Empirical data verified the usefulness of our MuGeR algorithm in identifying potential cancer drivers. Moreover, our in-depth appraisal of TCGA liver hepatocellular carcinoma datasets further highlighted the frequent mutational dysregulation of ubiquitin-related proteasomal degradation in driving hepatocarcinogenesis.

A2

Maximizing the information extraction of RNA-seq data for achieving personalized medicine

Tyler Weirick and Shizuka Uchida

Cardiovascular Innovation Institute, University of Louisville, Louisville, KY 40202, United States

Human Genomics 2018, **12**(Suppl 1):A2

Through the efforts of large-scale sequencing projects around the world (e.g., ENCODE, FANTOM), we have learned that 90% of human genome is transcribed as RNA, yet only a minor portion of these RNAs encode for proteins. Similarly, the processes affecting RNA and functions of RNA have become increasingly complex. Given that RNA

is an intermediate between the genomic DNA and proteins, in principle, RNA sequencing (RNA-seq) should be able to reveal the information about the genome (e.g., mutations, including single nucleotide polymorphism (SNPs)) and the final products (i.e., proteins). Although RNA should serve as exact copies of genomic DNA, in reality, RNA can be modified by a variety of enzymes, which results in the addition of 5'-cap and poly A tails as well as various forms of RNA modifications (e.g., RNA editing). To study these multiple processes simultaneously, we created a computational pipeline to maximize the extraction of information from RNA-seq data. In this pipeline, not only do we measure expression levels of RNA (i.e., genes and isoforms), we further detect A-to-I RNA editing sites and clusters of editing or "editing islands" using our previously introduced tool RNAEditor (John, *Brief Bioinform*, 2016; Stellos, *Nat Med*, 2016) as well as circular RNAs (circRNAs) (Boeckel, *Circ Res*, 2015; Militello, *Brief Bioinform*, 2017), which arise from the backsplicing of exons and/or introns. Our pipeline is implemented in the Snakemake as workflow management system that confers numerous benefits, such as easy parallelization on HPC-clusters and cloud computing environments, advanced error detection, and automatically deleting intermediate files. Through the detailed analysis for A-to-I RNA editing events that result in adenosine (A) to inosine (I) conversion in RNA but not in genomic DNA, which can be identified as guanine (G) as replacement of A in RNA-seq sequencing reads, it is possible to correctly infer the actual mutations in the human genome, which differs from the reference genome used as template for RNA-seq data analysis. These mutations should represent the individuality of humans. In this meeting, we would like to share our computational pipeline for the further analysis of RNA-seq data and integrating the additional extracted information for use in personalized medicine, especially related to cardiovascular disease ("Cardiovascular Personalized and Precision Medicine").

A3

Versatile instrument for single cell analysis and complex tissue microdissection

Stanislav L. Karsten, Zhongcai Ma, Lili C. Kudo

NeuroInDx, Inc., 20725 S Western Ave, Ste 100, Torrance, CA 90501, USA
Human Genomics 2018, **12**(Suppl 1):A3

Collection of specific cells or subanatomical regions from complex heterogeneous tissues is a prerequisite step for understanding complex molecular mechanisms underlying health and disease. Single cell analysis (SCA) has become an essential part of cutting edge biomedical research. There are commercially available single cell collection and tissue microdissection technologies including laser based systems, and cell sorting instruments. However, these platforms are typically sample specific, complex and expensive making integration within standard lab workflows difficult. Moreover, part of SCA is the investigation of single cell adhesion properties that is a key in substrate mediated cell behavior essential for understanding cellular properties in health and disease. There are approaches for single cell

adhesion measurements including atomic force microscopy, optical tweezers, and micropipette/capillary aspiration. Unfortunately, none of the existing instruments provide concurrent single cell acquisition and adhesion force measurement prior to its preparation for the downstream analysis (e.g. NGS). Based on our vacuum impulse based cell and tissue acquisition technology (UP8797644) we have developed a universal platform for single cell acquisition and analysis. Developed system collects individual cells from any adherent cultures grown in standard cell culture dishes in as small as 15 nl volume, compatible with downstream SCA and NGS. The system may be used with a wide range of inverted microscopes, and the cells of interest can be identified based on morphology, location or labeling, including fluorescence. Here, individual cells were collected from human neuroblastoma SH-SY5Y, CHO, 3T3 and neural progenitor cell cultures. Collected single cells were dispensed immediately into individual wells for clonal expansion. Clonal expansion for 7 days of these single cells revealed minimal effect on cellular viability (up to 99% when compared to dilution controls). Moreover, trypan blue assay demonstrated survival rates similar with the re-cultivation studies. In addition, we extended the use of our technology for single cell adhesion strength measurement. Adhesion strength for individual cells from several cell cultures was measured using a sensor incorporated in the collection assembly. Measurement of cell adhesion force was performed based on the capillary aspiration techniques reported earlier. Both manual and automatic algorithms in the instrument's software were developed for rapid collection and deposition of target cells and tissue regions. The benefits of the proposed technology include cost-efficiency, simple operation, complete workflow from single cell isolation to adhesion force measurement, compatibility with a wide range of inverted microscopes and use of standard plates and culture dishes.

Keywords: adhesion, single cell analysis, tissue, microdissection, archival tissue, region of interest, acquisition, dispensing

A4

Mutational mechanisms revealed by a comprehensive set of high-quality microdeletions in an ancestry-diverse cohort of 1,314 deep sequenced family trios

Wendy S.W. Wong¹, Pankaj Kumar², John E. Niederhuber^{1,3}, Anindya Dutta²

¹Inova Translational Medicine Institute, Inova Health System, Falls Church, VA, USA; ²Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA; ³Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
Human Genomics 2018, **12(Suppl 1)**:A4

While copy number variants (CNVs) in human chromosomes have been under active research, relatively fewer studies have focused on medium-size variants such as microdeletions (chromosomal deletions in the 100-1000 bp range), and even less attention has been given to germline *de novo* microdeletions. The microdeletions of intermediate length (between 50 and 1,000 bp) are often termed as the twilight zone and are detected with much less sensitivity. Previous studies have shown that microdeletions contribute to a significant number of diseases, but their mechanism of formation is largely unknown. The whole genome sequences of the family trio cohort from Inova Translational Medicine Institute's (ITMI) Childhood Longitudinal Cohort Study provides a unique opportunity to identify pre-existing and *de novo* microdeletions in the human population. By leveraging the state-of-the-art SV detection and genotyping algorithms, namely, Delly, Manta, SV2 and svaba, as well as family relationship in our data, we've created a comprehensive set of common and rare high-quality microdeletions in our cohort. Our analysis demonstrates that the common microdeletions (minor allele frequency >0.01) can be used to identify a subject's ancestry at continental level using Principal Component Analysis (PCA). Next, we identified genomic factors that influence the density of common and rare microdeletions by running a non-overlapping window of size 1 MB. We found that

microdeletions are enriched in high GC and simple repeats regions. Interestingly, as opposed to duplications, the microdeletions mutation rates are higher in late replication regions. Next, we show that the smaller microdeletions (<300bp) prefer microhomology sizes between 2 and 3 bp (>30%) around the breakpoint, suggesting nonhomologous end joining (NHEJ); whereas the larger microdeletions prefer microhomology length of 4bp (21%), followed by 1bp (19%). Finally, the GC richness and microhomologies underlying the junctions of the microdeletions are consistent with the possibility that the microdeletions were generated by mechanisms that also produce extrachromosomal circles of DNA called microDNAs. In summary, the curated set of high quality microdeletions will enhance studies in understanding their mutational mechanism, as well as to understand their disease association and functional impact.

A5

Genome-wide association study for resistant hypertension in a Japanese population

Keiko Yamazaki^{1,2}, Yasuo Takahashi¹, Yukako Katsura³, Yayoi Nishida³, Yoichiro Kamatani^{4,5}, Michiaki Kubo⁶, Satoshi Asai^{1,3}

¹Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University School of Medicine, Tokyo, Japan; ²Laboratory for Genotyping Development, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; ³Division of Pharmacology, Nihon University School of Medicine, Tokyo, Japan; ⁴Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; ⁵Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Sakyo-ku, Kyoto, Japan; ⁶RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Correspondence: Keiko Yamazaki
Human Genomics 2018, **12(Suppl 1)**:A5

Background

Resistant hypertension is defined as uncontrolled blood pressure despite the use of 3 antihypertensive agents or use of 4 antihypertensive agents regardless of blood pressure control. By epidemiological reports in Western countries the prevalence of resistant hypertension is about 15% of all hypertensives. It is expected that genetic factors may play a greater role in resistant hypertensive patients than in the general hypertensives. However, there was unclear about resistant hypertension, whereas genome-wide association (GWAS) studies for hypertension identified over 100 susceptible loci.

Materials & methods

We used genome-wide variant data of 25,450 patients with hypertension obtained from the BioBank Japan project. These subjects were belonging to two groups; group (1) was resistant hypertension requiring at least 4 antihypertensive medications of different classes drug classes to achieve blood pressure control (N=2,723) and group (2) was controlled hypertensives with only 1 antihypertensive medication (N=21,470). After applying stringent quality control for samples, we imputed using 1000 Genome Projects Phase 3 data as reference and examined association of about 800 millions SNPs with accuracy (Rsq) > 0.9 and MAF > 0.05 of both cases and controls.

Results

Our GWAS found 28 suggestive association ($P < 1 \times 10^{-6}$) on chromosome 1, 2, 3, 8, 11, 17, 18 and 21. The genes located around these candidate loci were involving with pathways; ACE inhibitor pathway, regulation of insulin secretion and HSF1-dependent transactivation, neurexins and neurologins. Continuously, we would attempt sex-stratified study to these candidate loci.

Conclusions

We performed GWAS using 25,450 patients with hypertension and identified 28 suggestive signals. Our findings will lead to clarify pathogenesis of resistant hypertension.

Keywords;

resistant hypertension
genome-wide association study (GWAS)
susceptibility

A6**A cell-based translocation assay system identifies active fractions from plant that rescue skeletal dysplasia phenotypes in an achondroplasia mouse model**Yi-Ching Lee¹, Yun-Wen Lin^{1,2} and Yuan-Tsong Chen^{3,4}¹Institute of Cellular and Organismic Biology, Academia Sinica, Taipei, 11529, Taiwan; ²Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, 1490, Taiwan; ³Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529, Taiwan; ⁴Department of Pediatrics, Duke University Medical Center, Durham, NC, 27710, USA**Correspondence:** Yi-Ching Lee (yiching@gate.sinica.edu.tw)

Human Genomics 2018, 12(Suppl 1):A6

Activating mutations of fibroblast growth factor receptor 3 (FGFR3) cause various skeletal dysplasias and are also associated with certain cancers. Because there are no known specific pharmaceutical inhibitors of FGFR3, we established a cell-based protein translocation assay system that can monitor FGFR3 activity and be used for high throughput screening of complex mixtures. With this system we identified ethanol extract from a plant as a FGFR3 inhibitor and performed bioassay-guided fractionation to identify potent active fractions. The functionality of extract and active fractions were validated *in vitro* in FGFR3-activated primary multiple myeloma cells. The therapeutic efficacy and safety of the active fractions were further assessed in *FGFR3^{ACH}* mice, an achondroplasia mouse model. Oral administration significantly improved growth and dwarfism-related clinical features of the *FGFR3^{ACH}* mice. Our results demonstrate the applicability of this discovery approach. The identified plant extracts and active fractions hold therapeutic potential for the treatment of FGFR3-activated skeletal dysplasias and cancers.

A7**A Mendelian randomization phenome-wide association study of thyroid function in 337,199 individuals from UK Biobank**Silvia Ravera¹, Nancy Carrasco¹, Joel Gelernter^{2,3}, Renato Polimanti²¹Department of Cellular and Molecular Physiology, Yale School of Medicine, New Haven, CT, 06510, USA; ²Department of Psychiatry, Yale School of Medicine and VA CT Healthcare Center, West Haven, CT, 06516, USA; ³Departments of Genetics and Neuroscience, Yale University School of Medicine, New Haven, CT, 06510, USA**Correspondence:** Renato Polimanti (renato.polimanti@yale.edu)

Human Genomics 2018, 12(Suppl 1):A7

Background

Thyroid function plays a key role in the regulation of a wide range of biological processes in human body. Circulating levels of the thyroid-stimulating hormone (TSH) and free thyroxine (FT4) are used to assess thyroid function and were investigated by genome-wide association studies that identified risk alleles that partially explain their inter-individual variation. To dissect the causal role of thyroid function in human phenome, we conducted a Mendelian randomization study of TSH and FT4 levels.

Materials and methods

We applied a two-sample Mendelian randomization based on risk alleles for TSH and FT4 levels in euthyroid (healthy) subjects and genome-wide data regarding 2,419 traits assessed in up to 337,199 individuals from UK Biobank. Multiple MR methods were tested to verify the reliability of the results and MR Egger regression intercept was considered to verify the validity of the genetic instruments. False discovery rate (FDR, $q < 0.05$) was applied to correct the results for multiple testing accounting the number of traits tested and the numbers of MR methods applied.

Results

After FDR multiple testing correction, sixteen traits appear to be affected by genetically-determined TSH levels. The strongest results were observed with respect to thyroid-related traits, e.g. hypothyroidism ($\beta = 0.02$, $p = 2.39 \times 10^{-17}$), non-toxic goiter ($\beta = -0.002$,

$p = 9.45 \times 10^{-14}$), and levothyroxine medication use ($\beta = 0.017$, $p = 2.41 \times 10^{-13}$). TSH causal effect was also present with respect to several anthropometric traits, such as arm bioelectrical impedance ($\beta = -0.029$, $p = 2.22 \times 10^{-13}$), standing height ($\beta = -0.023$, $p = 2.76 \times 10^{-10}$), and height at age 10 ($\beta = -0.028$, $p = 1.74 \times 10^{-9}$). Additional significant results were observed for: pulse rate ($\beta = -0.037$, $p = 2.84 \times 10^{-8}$), female infertility ($\beta = 0.001$, $p = 4.91 \times 10^{-6}$), and hearing aid use ($\beta = -0.005$, $p = 7.10 \times 10^{-5}$). No result survived multiple testing correction in the MR analysis based on FT4 genetic instrument.

Conclusion

We provided novel data regarding the consequences of inter-individual variability of thyroid function. In particular, genetically-determined TSH levels appears to be involved in causal mechanisms of a wide range of phenotypic traits in agreement with the recognized pervasive regulatory action of thyroid in human body.

Acknowledgements

Part of this research has been conducted using the UK Biobank Resource (application reference no. 19664). This study was supported by the Brain & Behavior Research Foundation, the American Foundation for Suicide Prevention, and the Simons Foundation Autism Research Initiative.

A8**Analysis of gene expression profiles of apical papilla tissues, stem cells from apical papilla and cell sheet to identify the key modulators in MSCs niche**Shu Diao^{1,2}, Zhipeng Fan¹¹Laboratory of Molecular Signaling and Stem Cells Therapy, Beijing Key Laboratory of Tooth Regeneration and Function Reconstruction, Capital Medical University School of Stomatology, Beijing 100050, China;²Department of Pediatric dentistry, Capital Medical University School of Stomatology, Beijing 100050, China**Correspondence:** Zhipeng Fan (zpfan@ccmu.edu.cn)

Human Genomics 2018, 12(Suppl 1):A8

Background: In tooth development and regeneration research, a key variable for successful tissue regeneration and engineering is the environment in which cells and tissues grow. The apical papilla is essential for tooth development, stem cells from apical papilla can contribute to the formation of dentin/bone-like tissues and represent good cell sources for dental tissue regeneration. However, limited by the current research techniques, the micro environment can not be simulated when stem cells isolation and culture *in vitro*. Disruption of the micro environment may lead to failure of tooth regeneration. In this study, we focus on the differential gene expression between SCAP and apical papilla tissues in attempt to identify the genes that are crucial for inducing SCAP.

Materials and methods: Gene chip analysis on cultured SCAP and apical papilla tissue: with patients' informed consent, SCAP and apical papilla tissue were collected separately from third molar of five female patients aged 18-22 years old. SCAP were isolated and cultured as previously described. Cell and tissue samples were subjected to Trizol to extract RNA for genechip analysis, the results were further verified by Real-time RT-PCR.

Results: SCAP was discrete distribution, stretched after 4 days. The shape of the cells were spindle-shaped with oval and polygonal. Gene chip data reveals that 2325 genes that were differently expressed, genes-S1004A, FOXM1, FGF5 were up-regulated, whereas genes-CXCL14, IGF2, BMP6 were down-regulated. The genechip results for these six genes were further verified by RT-PCR analysis.

Conclusion: Genes were differently expressed between apical papilla tissues and SCAP suggesting the importance of micro environment for SCAP to proliferate and differentiate. Our data may be an indication of possible artificial micro environment that is useful for tooth regeneration.

Key words: Stem cells from apical papilla; Niche; Differentiation potential; Gene expression

A9**Analysis of retrotransposition in neurodegenerative disorders**

Giovanni Pascarella¹, Kosuke Hashimoto¹, Annika Busch¹, Anton Kratz¹, Shigeo Murayama², Shinji Saiki³, Stefano Gustincich⁴ and Piero Carninci¹
¹Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Saitama, Japan; ²Department of Neuropathology, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan; ³Department of Neurology, Juntendo University School of Medicine, Tokyo, Japan; ⁴Area of Neuroscience, Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

Correspondence: Giovanni Pascarella
Human Genomics 2018, **12(Suppl 1):A9**

Nearly 50% of the human genome is composed by Transposable Elements (TEs), discrete units of DNA capable of moving within the genomic environment. DNA transposons duplicate through a cut-and-paste mechanism and have been inactivated during evolution, while retrotransposons duplicate through RNA intermediates that are reverse-transcribed by a self-encoded variant of RT and then inserted at new genomic locations. The human genome harbors at least three families of retrotransposons that are currently capable of transposition: LINE-1 (L1), Alu and SVA elements, all belonging to the subclass of non-LTR retrotransposons and accounting for approximately one-third of the human genome. Recent published data shows that the impact of active retrotransposons on the shape and the integrity of the human genome is far deeper than previously considered. With an estimated rate of one new insertion every 10-100 live births, active retrotransposons are able to influence genetic diversity and to cause diseases, as supported by several cases of de novo insertions involved in the outbreak of cancers and other genetic disorders. Yet, clinical studies of insertions involved in human diseases are usually focused on a very small and local genomic scale. Somatic retrotransposition has been shown to occur in the normal human brain at higher rates compared to other tissues, and de novo insertions can happen at different frequencies in different brain regions. We have developed an oligonucleotide based liquid-phase capture protocol in order to specifically capture and sequence young, active LINE-1 and AluY retrotransposons in the human genome. We have used this approach to study the brain retrotransposome in samples obtained from donors affected by idiopathic Alzheimer's disease (AD) and Parkinson's disease (PD). Our preliminary results show significant differential content of L1 and somatic retrotransposition in several brain regions of AD and PD patients versus controls.

A10**High accuracy HLA allele imputation using Han Chinese reference panel**

Jiangshan J. Shen¹, Yujun Sheng², Chao Yang², Yongfei Wang¹, Mengbiao Guo¹, Yu Lung Lau¹, Wanling Yang¹, Xuejun Zhang²
¹Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, SAR China; ²Department of Dermatology, No. 1 Hospital and Key Laboratory of Dermatology, Ministry of Education, Anhui Medical University, Hefei, China
Human Genomics 2018, **12(Suppl 1):A10**

The human leukocyte antigen (HLA) region is a genetically diverse region intimately involved in a variety of immune related functions and known to be associated with disease predisposition. Due to the ubiquity of GWAS genotyping data, the expense of HLA allele typing and interest in the association between HLA region and disease, statistical imputation of HLA alleles from genotypes is becoming indispensable. HLA imputation relies on the use of a reference panel where the association between the genotype and HLA alleles is known, and a model is built on capturing the relationship between the two. Recent work has shown the importance of having population specific reference panels for HLA allele imputation. In this study, we use the HIBAG HLA imputation framework to build Han Chinese

specific reference panels for 8 HLA loci, including 3 loci in HLA class I and 5 loci in HLA class II. Using Han Chinese genotype data, we show that using the largest population specific HLA dataset to date as a reference panel leads to increased accuracy in predicting HLA alleles for imputation and increased number of HLA alleles that predicted. We compare our reference panel to existing HIBAG Pan Asian and multiethnic reference panels, and show that using a Han Chinese specific reference panel significantly improved on the existing HIBAG panels by increasing the call rate of the HLA alleles, improving the number of HLA alleles imputed, and increasing the confidence in the imputed HLA alleles. We also provide a web interface where the user can input their genotype data and impute HLA alleles using Han Chinese, Pan Asian or multiethnic reference panels. The authors hope the study demonstrates the importance of using population specific reference panels and increasing reference panel sample size for HLA imputation. Furthermore, it is hoped that the ease of use in HLA imputation will allow more HLA association studies to be conducted and provide greater insight into the fine mapping of HLA alleles in GWA studies.

A11**Proteobacteria profile in blood microbiota by concordance of 16S rRNA metataxa analysis and WGS human reference unmapped reads**

Yejun Yun, Han-Na Kim, Eunju Lee, Hyung-Lae Kim
 Department of Biochemistry, Ewha Medical Research Institute, School of Medicine, Ewha Womans University, Seoul, South Korea
Correspondence: Hyung-Lae Kim (hyung@ewha.ac.kr)
Human Genomics 2018, **12(Suppl 1):A11**

Background

Blood has largely been considered free of microorganisms. Recent analysis in humans, however, detected bacterial species in blood from healthy and disease individuals, suggesting the existence of a resident blood microbiota regardless of sepsis. Of these bacteria, Proteobacteria phylum was consistently reported to be dominated in the blood. Especially lipopolysaccharides (LPS) which mainly Proteobacteria contains is related to the chronic inflammation correlated with obesity and other metabolic chronic diseases. Therefore, we focused on this phylum to clarify the bacterial component in the blood.

Materials and methods

The study was conducted with blood DNA of 240 healthy individuals. We performed a whole genome sequencing (WGS) of 30x coverage as well as a 16S rRNA gene amplicon sequencing to analyze the blood microbiota. About 99% of whole genome sequencing were mapped to the human reference genome, which remaining unmapped reads are then assigned to taxonomic labels by Kraken pipeline. DADA2 pipeline of QIIME2 was used for 16S rRNA data analysis.

Results

The Proteobacteria was the most abundant phylum in both 16S rRNA metataxa and WGS unmapped reads, around 80%. We observed differences between 16S metataxa and WGS in the relative abundances of bacterial classes, families, and genera after extraction of only Proteobacteria phylum. Gammaproteobacteria were major component (84.5%) in WGS unmapped result, while Alphaproteobacteria (44.9%) and Betaproteobacteria (30.9%) were abundant in 16S metataxa result.

Conclusions

The blood microbiota, even when present in very low abundance, may be implicated in important physiologic role. However, to quantify and characterize the microbiome in human blood is limited because blood contain mostly human DNA and a low abundance of microbial DNA. So, further investigation will be needed to find the concordance between 16S rRNA amplicon sequencing and WGS unmapped reads.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1A6A3A11932719 and NRF-2017R1D1A1B03035501). It was also supported by GSDC and KREONET in KISTI.

A12**Comparison of taxa-matching algorithms for association analysis of the gut microbiome**

Han-Na Kim, Yeojun Yun, Eun-Ju Lee, Hyung-Lae Kim
Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, South Korea

Correspondence: Hyung-Lae Kim
Human Genomics 2018, **12(Suppl 1):A12**

Background

Recent advances in genomic sequencing technology have been applied to the human microbial environment to provide a new understanding of the role of the microbiome in human health. However, several studies have shown that the methodology rather than the biological variation is responsible for the observed sample composition and distribution. As new data analysis pipelines are introduced, it is essential to evaluate their capabilities in specific applications. The object of this study was to assess whether the same biological conclusions regarding microbiome composition could be reached using different bioinformatics pipelines.

Materials and methods

Using the personality data from a cross-sectional study of human gut microbiota, 672 samples were included in the study. The gut microbiome was analyzed by 16S rRNA gene sequencing using Illumina MiSeq. We compared bioinformatics pipelines commonly used for amplicon data analysis: QIIME1 (open reference OTU picking and closed reference OTU picking), UPARSE, and DADA2. Assignment of taxonomy was performed on each table using both the DADA2 taxonomy classification method and the QIIME taxonomy classification method. Generalized linear models implemented in multivariate association with linear models (MaAsLin) packages of RStudio were used for quantitative and qualitative analysis of each personality domain. We used the zero-inflated regression model on a gaussian distribution for zero-inflated microbiota data.

Results

Overall the taxa profiles were comparable between pipeline, however, average relative abundance of specific taxa varied depending on bioinformatics analysis. Not all of our results were not consistently significant in the all algorithm. The reason why the results among the algorithms were different might be that there is a fundamental conceptual difference between what OTU picking methods do and what dada2 is doing. Nonetheless, the association of Neuroticism with the Gammaproteobacteria and the *Haemophilus spp.*, and Conscientiousness with the Proteobacteria were significant in at least three algorithms. The others also showed the suggestive results (p-value <0.05 and q-value <0.25).

Conclusions

This study suggested that, specifically for our cross-sectional study, the same biological conclusions could be drawn from data regardless of the taxa-matching algorithms.

Acknowledgements

This research was supported by the National Research Foundation of Korea funded by the Ministry of Education (NRF-2016R1A6A3A11932719 and NRF-2017R1D1A1B03035501) and it was also supported by GSDC and KREONET in KISTI.

A13**Blood microbial profiles correlated with age**

Eun-Ju Lee, Han-Na Kim, Yeojun Yun, Hyung-Lae Kim
¹Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, South Korea

Correspondence: Hyung-Lae Kim
Human Genomics 2018, **12(Suppl 1):A13**

Background

Blood has generally considered a sterile environment; however, microbiome has been consistently detected in blood in healthy individuals. The healthy individuals harbor a rich microbiota in their blood, it raises the question of the role of this microbiota and its impact on the risk associated with several diseases. To examine how blood microbiomes differ according to the age, here we characterize bacterial species in 37 blood samples.

Materials and methods

The study was conducted with blood DNA of 37 individuals from 5 families. The cohort consisted healthy population from 20s to 80s and included mono- and dizygotic twins. We performed a whole genome sequencing (WGS) of 30X 10 samples, 60X 17 samples, 90X 10 samples to analyze the blood microbiome. About 99% of whole genome sequencing were mapped to the human reference genome (hg38). The remaining unmapped reads are then assigned to taxonomic labels using Brackens (<http://ccb.jhu.edu/software/bracken>; v1.0.0).

Results

In all samples, the Proteobacteria was the most abundant phylum (more than 90%) also, Gammaproteobacteria were major component (on average 57.6%) within this phylum. Proteobacteria exhibit the most variation relative to all phylum across samples. The elderly older aged groups (over 60 years old) showed highly correlation with the distribution of Gammaproteobacteria ($\rho=0.57$, p-value=0.0002) compared to younger adults (under age 40) and middle-aged group (aged 40-60).

Conclusions

This study suggested that blood microbial profiles, especially proteobacteria, may have strong correlation with age like the composition of human gut microbiota changes with age.

Acknowledgements

This research was supported by the National Research Foundation of Korea funded by the Ministry of Education (NRF-2016R1A6A3A11932719 and NRF-2017R1D1A1B03035501) and it was also supported by GSDC and KREONET in KISTI.

A14**Molecular Biological Elucidation of *in vitro* Spermatogenesis by Transcriptome Analysis**

Takeru Abe^{1,3}, Hajime Nishimura³, Kazuaki Kojima¹, Hiroyuki Sanjo², Takuya Sato¹, Harukazu Suzuki³, Takehiko Ogawa^{1,2} and Takahiro Suzuki^{1,3}

¹Laboratory of Biopharmaceutical and Regenerative Sciences, Institute of Molecular Medicine and Life Science, Yokohama City University Association of Medical Science, Yokohama, Kanagawa, 236-0004, Japan; ²Department of Urology, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, 236-0004, Japan; ³Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama City, Japan

Correspondence: Takahiro Suzuki (takahiro.suzuki.aa@riken.jp)
Human Genomics 2018, **12(Suppl 1):A14**

Objectives

About a half of infertility is caused by male side factors. Of that, about 90% is related to spermatogenesis dysfunctions which lead to oligozoospermia or azoospermia. However, about 60% of them is classified as idiopathic, and no effective treatment has yet been established. Therefore, the effective culture system for sperm production has been expected for research purpose as well as for possible therapeutic options. Our group has succeeded in producing fertile sperm from newborn mouse testis *in vitro* with an organ culture method. However, its efficiency remains low compared to that taking place *in vivo*, in spite of various efforts to improve the culture condition. Now, we consider that it is important and possible to reveal the molecular dynamics underlying the complexity of spermatogenesis. Here, we examined the difference in gene expression profile between *in vivo* and *in vitro* testis, and tried to clarify the phenomena occurring specifically in the cultured testis.

Materials and Methods

We used *Acr-Gfp* transgenic mice expressing GFP in germ cells at mid-pachytene stage of prophase meiosis onward as the experimental animal. We cultured testis tissue of the mice at 7day postpartum (dpp) for two about weeks. These cultured testis samples along with testis tissues directly taken from mice at 7, 14 and 21dpp, as *in vivo* samples, were analyzed by FACS after cell dissociation or applied for microarray analysis.

Results

The number of GFP positive cells, an indicator of spermatogenesis progression, was much fewer in testis tissues cultured for 2 weeks than those in the testes of 21dpp, the *in vivo* counterpart, when analyzed by FACS. The principal component analysis using the

microarray data showed that the gene expression profile of the cultured testis was quite similar to that of 14dpp rather than to that of 21dpp *in vivo* testis, suggesting that the *in vitro* spermatogenesis was retarded or mostly arrested at around 14dpp state. Through differential gene expression analysis between the *in vitro* and 14dpp *in vivo*, we found that the genes relating to innate immune system were significantly upregulated in the *in vitro* samples.

Conclusions

Our results revealed that *in vitro* spermatogenesis delayed or arrested at the point around 14dpp *in vivo*. Furthermore, it was suggested that the innate immune system is compromising the efficiency of *in vitro* spermatogenesis. The data obtained in this study will be useful for developing new culture conditions that can support more efficient spermatogenesis.

A15

Disease gene discovery in amyotrophic lateral sclerosis using innovative next generation sequencing and genetic linkage strategies

Emily P. McCann¹, Jennifer A. Fifta¹, Kelly L. Williams¹, Natalie Twine^{1,2}, Dennis Bauer², Dominic B. Rowe¹, Ian P. Blair¹

¹Centre for MND Research, Macquarie University, Department of Biomedical Sciences, Macquarie University, Sydney, New South Wales, Australia; ²Health & Biosecurity Flagship, Commonwealth Scientific and Industrial Research Organization, Sydney, New South Wales, Australia

Correspondence: Emily P. McCann (emily.mccann@hdr.mq.edu.au)

Human Genomics 2018, **12(Suppl 1)**:A15

Background

Amyotrophic lateral sclerosis (ALS; also known as motor neuron disease, MND; or Lou Gehrig's disease) is an ultimately fatal, genetically heterogeneous neurodegenerative disease. Approximately 10% of cases are hereditary (familial; FALS). While gene mutations are the only proven cause of disease, one third of ALS families carry an unidentified mutation. These families are often small, and exhibit incomplete penetrance, both factors inhibiting traditional disease gene mapping.

Materials and methods

We employed innovative approaches using next generation sequencing (NGS) and custom bioinformatics to identify novel genetic contributors to FALS. A FALS family were recruited to the Macquarie University Neurodegenerative Disease Biobank under informed written consent, as approved by Macquarie University. Genome and exome sequencing was performed for two affected individuals and one obligate carrier. After standard bioinformatics processing, a custom bioinformatics pipeline was applied to both datasets. Shared variant analysis was performed, and the NGS reads for the resultant candidates visualised, to ensure correct genotype calling and appropriate annotation of multi-allelic variants. Subsequent filtering removed non-exonic variants and common genetic variants present in ethnically matched control cohorts. SNP-based linkage analysis was also performed in these individuals and eleven additional at-risk family members, using a parametric model and liability classes. Both NGS datasets were then sub-set to only those genomic regions with positive LOD scores, and shared variant analysis was repeated as above, with the addition of potential regulatory variants being retained.

Results

Genome and exome sequencing identified approximately 8,000,000 and 185,000 variants across the individuals of this family, respectively. Custom bioinformatic analysis revealed 12 and 15 novel coding variants, equating to a total 27 candidates. However, when overlaid with linkage results, none of these variants were found within suggestive linkage regions, and are therefore less likely to cause ALS. In total, 40 positive linkage peaks were observed across the genome. As such, the above shared variant analysis pipeline was repeated, however, only the suggestive linkage regions were considered, and regulatory region variants were included in addition to exonic variants, resulting in 12 and one candidate(s) respectively, with the one exome candidate also being present in the genome candidate list.

Conclusions

It is hoped that Sanger sequencing validation and *in silico* evaluation of the identified candidate variants will reveal a single mutation causing ALS in this family. By elucidating the remaining genetic

contributors to ALS we hope to enhance our understanding of disease and inspire downstream studies, particularly therapeutic development.

A16

Optimized homology directed repair for treatment of inherited retinal diseases using the CRISPR/Cas9 system

Brian Rossmiller, Takeshi Iwata

Molecular and Cellular Biology Division, National Institute of Sensory Organs, Tokyo Medical Center, National Hospital Organization, 2-5-1 Higashigaoka, Meguro, 152-8902. Tokyo, Japan

Human Genomics 2018, **12(Suppl 1)**:A16

Nearly 1.5 million people, worldwide, are affected by hereditary vision impairment each year. For many of these, there is little to no effective treatment. Advancements in CRISPR/Cas mediated site specific gene editing provide a new and powerful tool for disease treatment but relies on inefficient homology directed recombination (HDR). The purpose of this project is to treat two models of inherited retinal degeneration using CRISPR/Cas9 mediated HDR. Prior to treatment, we will assess optimal conditions for use of the CRISPR/Cas9 system in the retina. These include age of injection, delivery vehicle (nanoparticle or adeno-associated virus (AAV)), and method of increasing homologous integration of gene replacement (ligase IV inhibitors and neurotrophic factors).

An array of non-viral delivery methods were screened first utilizing transfection of 2.5 micrograms of plasmid containing cytomegalovirus promoter driven GFP in 661W cells and secondly in B6 mice. Mice were injected with the same plasmid at either postnatal day 5 or 15. 2 weeks post injection, retinas were dissociated and measured via FACS. Assessment of sgRNA cleavage efficiency was performed using a novel dual luciferase approach with a molar ratio of 1:4 target to H1-sgRNA expression vector. Luciferase activity was measured 48 hours post-transfection. Each vector will be assessed in two animal models of inherited retinal degeneration. First, a mouse model of retinitis pigmentosa, Rho (I307N), kindly provided by Dr. Nishina of Jackson laboratory. The second model is of normal tension glaucoma, OTPN (E50K) and produced in the Iwata lab.

Here, we have confirmed transfection, of CMV-GFP plasmid, by nanoparticles, into 661W cells by GFP expression. These results currently show p5 injections by either one of two commercially available nanoparticles, TransIT-202 and TransIT-X2 from Mirus®, to be optimal for DNA delivery. The sgRNA cleavage showed H1 driven sgRNA to result in the highest knockdown, 77% reduction in OPTN. Whereas RNA mediated knockdown resulted in a 50% and 63% knockdown of OPTN and RHO respectively.

Our current nanoparticles have shown significant transfection in tissue culture. However, we will continue to screen several additional mixtures to improve efficiency. This is coupled with a significant reduction in target gene expression by both DNA and RNA derived sgRNA in both RHO and OPTN models *in vitro*.

A17

Whole exome sequencing of 14 schizophrenia multiplex families in Japan

Miho Toyama¹, Yuto Takasaki¹, Branko Aleksic¹, Tomoo Ogi², Norio Ozaki¹

¹Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Japan; ²Department of Genetics, Nagoya University Research Institute of Environmental Medicine, Nagoya, Japan

Correspondence: Branko Aleksic (branko@med.nagoya-u.ac.jp)

Human Genomics 2018, **12(Suppl 1)**:A17

Background

Schizophrenia is a psychiatric disease which has a relatively high heritability estimated up to around 80% [1]. Regarding the onset risk, genetic studies focused on schizophrenia suggested that rare variants might have large effect size [2] while common variants detected by genome-wide association studies had relatively small effect size. Exploring disease-associated rare variants has become more feasible than before by utilizing Whole Exome Sequencing (WES), however, it is still challenging to extract candidates efficiently from a large number of

mutations detected. Thus, most of those studies have been conducted focusing on *de novo* genetic variants because of its interpretability. Nevertheless, it is also essential to investigate shared variants among patients in order to elucidate association between transmitted variants and schizophrenia. To achieve this purpose, WES in multiplex families would be a prospective solution to identify disease-associated rare variants.

Materials and Methods

We defined a family which has more than one patient with schizophrenia as a multiplex family. In order to explore the shared genetic risk factors within the patients, we preformed WES study using 29 patients with schizophrenia, one patient with obsessive compulsive disorder and ten healthy controls from 14 schizophrenia multiplex families in Japan. We focused on rare single nucleotide variants (SNVs) with allele frequency (AF) $\leq 1\%$ among databases of Japanese, Eastern Asian and total population, considering difference of AF among them. SNVs shared only within patients as well as *de novo* SNVs in highly intolerant genes with percentile residual variation intolerance score [3] $\geq 25\%$ were selected as the final results.

Results

We identified 209 transmitted mutations including two homozygous SNVs in *CNTN6* and *MAOB* genes, and three *de novo* SNVs in *CACNA1C*, *ODC1* and *BRD4* genes. Out of them, eight genes including *CNTN6* and *CACNA1C* have been repeatedly reported to have rare variants associated with developmental disabilities and psychiatric disorders (Table 1).

Conclusions

Based on WES analysis of 14 schizophrenia multiplex families, we identified SNVs in genes that have been reported in the psychiatric field. Our results suggested that these genes might be strongly related to pathogenicity of common symptoms among those disorders. Now we are carrying on bioinformatics approaches for further interpretation of these results.

Ethics Approval

This study was approved by the Ethics Committees of Nagoya University Graduate School of Medicine.

References

1. Cannon TD, Kaprio J, Lonnqvist J, Huttunen M, Koskenvuo M: The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. *Archives of general psychiatry* 1998, 55(1):67-74.
2. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landen M, Moran JL, Purcell SM, Sklar P, Sullivan PF et al: Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci* 2016, 19(11):1433-1441.
3. Petrovski S, Wang QL, Heinzen EL, Allen AS, Goldstein DB: Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *Plos Genet* 2013, 9(8).
4. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL: A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *Jama Psychiatry* 2016, 73(3):275-283.

Table 1 (abstract A17) Eight genes registered in Developmental Brain Disorder Genes Database [4]. Tier shows strength of evidence, where 1 expresses the strongest evidence with three or more *de novo* pathogenic variants and 4 expresses the weakest with no *de novo* variant reported. For more detail of definition see the reference.

Gene	Inheritance	Mutation type	Genomic position	Amino-acid change	DBD Tier
<i>CSMD1</i>	transmitted	nonsynonymous	8:2857629	p.P2685L	2
<i>SETBP1</i>	transmitted	nonsynonymous	18:42643545	p.Q1558L	1
<i>MYT1L</i>	transmitted	non-frameshift del	2:1946788	p.155_157del	1
<i>KAT6A</i>	transmitted	nonsynonymous	8:41798389	p.I1004V	1
<i>ANK2</i>	transmitted	nonsynonymous	4:114278014	p.R2747H	1
<i>TRIO</i>	transmitted	stopgain	5:14489082	p.W11X	1
<i>CNTN6</i>	transmitted, homozygous	nonsynonymous	3:1363404	p.S278A	4
<i>CACNA1C</i>	<i>de novo</i>	nonsynonymous	12:2224447	p.A36V	4

A18

Co-evolving residue network analysis of proliferating cell nuclear antigen PCNA indicates few mutations among residues of high betweenness

Subinoy Biswas¹, Christian Schönbach^{1,2}

¹International Research Center for Medical Sciences, Kumamoto University, Kumamoto, Japan; ²Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan

Correspondence: Subinoy Biswas (subinoy@kumamoto-u.ac.jp)

Human Genomics 2018, 12(Suppl 1):A18

Background

PCNA a core protein in the DNA replication process also interacts with multiple proteins involved in DNA repair, recombination and cell cycle regulation. Little is known how mutations in human PCNA and its interacting partners affect protein interactions and phenotypic outcomes, particularly in cancers. The objectives of the this study are to investigate i) the evolutionary constraints in residue-residue interactions by network topology, ii) the mutational impact of coevolving residue on PCNA and its partner proteins at structural and functional levels, and iii) the relationships of hub residues in context of cancer-associated mutations.

Materials and Methods

Eukaryotic orthologous protein sequences of PCNA (38 organisms) and its partner proteins from DNA repair pathway were collected from Ensemble and assembled to common sets. Coevolving residues of PCNA and nine coevolving residues sets from three partners (FEN1, RFC1, RFC3-4 and POLD1-4) were determined using CAPS2 server using correlation value $r \geq 0.6$, $p < 0.01$ subjected to bootstrapping. These sets of coevolving residues were matched against mutations reported in COSMIC database and PolyPhen/SIFT predicted deleterious SNPs of dBSNP, and used for creating intra and inter coevolving residue networks of PCNA and its interacting partners with igraph R package.

Results

High betweenness residues (HBR) in intra and inter coevolving networks were identified and analyzed. These critical residues hold the coevolving networks together. HBR and other node residues were compared with SNPs of COSMIC database. Among 31 intra and 35 unique inter PCNA coevolving residues we found Q125 of the inter-domain connecting loop (Fig. 1a) strongly coevolved (total degree 128) with residues of RFC1 (117), POLD2 (9), FEN1 (4) and RFC3 (2), and ranked among the top five HBR inter coevolving networks (Fig. 1b and c). The occurrence of Q125 in PCNA intra coevolving residue network implies critical functional and structural roles. Among PCNA coevolving residue only Q125H (COSM5777338) and I88V (COSM21734) matched to mutations reported in cancer samples and PNCA A252T (COSM4712855) was found in the neighborhood of intra coevolving residues 248, 249 and 250. The impact of these mutations on PCNA interactions and their roles in cancer development if any remains to be tested experimentally.

Conclusion

Preliminary results of the PCNA coevolving residue network analysis suggest that residues of high betweenness and degree seem to be selected against disease-associated mutations.

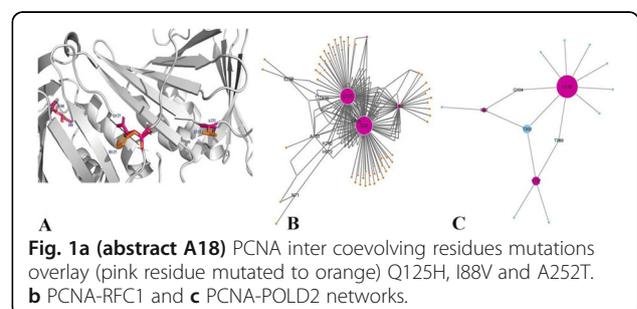


Fig. 1a (abstract A18) PCNA inter coevolving residues mutations overlay (pink residue mutated to orange) Q125H, I88V and A252T. **b** PCNA-RFC1 and **c** PCNA-POLD2 networks.

A19**Investigation the anti-carcinogenic effect of curcumin analog (PAC) by modulating DNA damage signaling pathway gene expression in Human colon cancer cell line**

Abdullah Al Amri¹, Abdelilah Aboussekhra², Mohammad Alanazi¹ and Abdelhabib Semlali¹

¹Genome Research Chair, Department of Biochemistry, College of Science King Saud University, Riyadh, Kingdom of Saudi Arabia; ²Head Cancer Biology & Experimental Therapeutics Section, King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia

Human Genomics 2018, **12**(Suppl 1):A19

Background: Colorectal cancer is one of the most common malignancies encountered in the world and it is the third among males and the fourth most common among females worldwide and the second most common cause of cancer-related mortality. Genetic factors common, with the development of colon cancer, as well as exposure to environmental factors. Although current chemotherapeutics as well as modern targeted therapy-approaches have been developed against colon cancer, the response rates to these agents remain low; the magnitude of tumor regression is variable and transient. *De novo* and acquired resistance to chemotherapeutics and targeted therapies and the toxicity to normal cells are the major causes of treatment failure. Therefore, it is necessary to search for new and better treatments for colon cancer. To these end, attention could be drawn toward phytochemicals derived from folk medicine due to their safety. Curcumin and their analogues due to their anti-cancer effect and their safety could be used as an efficient agent in cancer therapy. Research evidence suggests that curcumin, affects a variety of biological pathways involved in apoptosis, and the proliferation of the tumor, and induces apoptosis in various cell lines such as human prostate, lung cancer through multiple apoptotic pathways. However, curcumin analogues such as 3,5-Bis (4-hydroxy-3-methoxybenzylidene)-N-methyl-4-piperidone (PAC) have been explored to improve curcumin's efficacy in chemoprevention and therapeutics of cancer by enhancing its low systemic bioavailability. In the present study, we attempt to explore the effect of the Curcumin analogs (PAC) on DNA repair signal pathway in human colon cancer cells.

Methods: The effect of PAC on cell viability of human colon cancer cell lines, was assessed by using MTT colorimetric assay. Real-time PCR was used to study pro-apoptotic and anti-apoptotic genes expression on the mRNA level. The effect of PAC on DNA repair signaling pathways was examined by real-time PCR using RT2 Profiler PCR Array and confirmed by quantitative-PCR (Q-PCR).

Results: Our results indicate that curcumin analog (PAC) inhibit the colon cancer proliferation, increase the apoptosis and up-regulation of 11 genes involved in DNA damage signaling pathway.

Conclusions: This study has demonstrated that PAC may be viable as an anti-cancer agent through the modulation of apoptosis, cell survival and DNA repair system in colon cancer cells.

Keywords: Colon cancer, Curcumin analogue, PCR array, DNA damage.

A20**The consequences of promoter birth and death for human genetic and phenotypic variation**

Robert S Young, Martin S Taylor
MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK

Correspondence: Robert S Young
Human Genomics 2018, **12**(Suppl 1):A20

Promoters are the site at which gene regulatory signals are integrated and the site of transcription initiation during gene expression. Gene expression changes are thought to underlie much phenotypic variation between species and individuals, and may be responsible for individual variation in disease susceptibility and prognosis. We have previously identified that the complete birth and death of promoter sequence has been a common occurrence since the divergence of human and mouse [1], using Cap Analysis of Gene Expression (CAGE) data [2].

We are now integrating these evolutionary records with human variation data to assess the likely phenotypic role of these evolutionary volatile but transcriptionally active elements. Human-specific promoters that show sequence turnover (inserted or deleted promoter sequence) are more likely to contain human variants that accompany sequence gain or loss (insertions, deletions, copy number variants) than variants which preserve the nucleotide content (single nucleotide polymorphisms). Promoters whose sequence has been inserted along the human lineage more frequently overlap these polymorphic sites than those whose sequence was deleted down the mouse lineage since the divergence with human (11/14 tests, significant by permutation testing, nominal $p \leq 0.05$). Subsequent analyses of the derived allele frequency of these variants will reveal selective pressures which these evolutionary volatile promoters have recently been experiencing. Analysis of gene expression data from the GTEx and GEUVADIS consortia has revealed a specific enrichment of expression quantitative trait loci (eQTLs) within those promoters that show sequence turnover between human and mouse but not those which show functional turnover (where the underlying sequence is conserved). Despite this, promoters of all volatile evolutionary histories harbour variants which reduce gene expression relative to those contained within promoters which are conserved between human and mouse. We are currently analysing other sets of QTLs, such as those for transcription-factor binding, to determine the precise mechanism(s) by which human-specific promoters may function to suppress gene expression.

By analysing the transcriptomic and potential phenotypic consequences of evolutionary volatile promoters we hope to better understand the effect of these common evolutionary events in subsequently driving biological variation within the human population.

References

1. Young, R.S., et al., *The frequent evolutionary birth and death of functional promoters in mouse and human*. *Genome Research*, 2015. **25**: p. 1546-1557.
2. Forrest, A.R., et al., *A promoter-level mammalian expression atlas*. *Nature*, 2014. **507**(7493): p. 462-70.

A21**A novel homozygous nonsense mutation in *MLPH* causing Griscelli syndrome in a Pakistani family**

Naveed Wasif^{1,2}, Asia Parveen¹, Fareeha Ashraf¹, Ehtsham Ali¹, Guntram Borck²

¹Institute of Molecular Biology and Biotechnology (IMBB), Center for Research in Molecular Medicine (CRIIMM), The University of Lahore, Lahore, Pakistan; ²Institute of Human Genetics, University of Ulm, Ulm, Germany

Correspondence: Guntram Borck (guntram.borck@uni-ulm.de)
Human Genomics 2018, **12**(Suppl 1):A21

Background

Griscelli syndrome (GS) is a rare autosomal recessive disorder characterized by hypopigmentation, manifesting as silver-gray hair, presence of large clusters of pigment in the hair shaft, and the occurrence of either a primary neurological impairment or a severe immune disorder. Three different genetic forms; GS1 with an exclusive neurological, GS2 with immunological phenotypes and GS3 without any association of primary neurological or immunological defect, have been reported in literature. Mutations in *MYO5A*, *RAB27A* and *MLPH* cause GS1, GS2 and GS3 respectively.

Materials and Methods

A consanguineous Pakistani family with two affected members manifesting hypopigmentation of hair and skin was recruited for the study. Exome sequencing combined with Sanger sequencing was used to search for a genetic cause of GS phenotype in this family.

Results

Exome sequencing of one affected individual revealed a novel homozygous nonsense variant (c.70C>T; p.Arg24*) in *MLPH*. This variant lies in a 5.4Mb region of homozygosity. Sanger sequencing of two affected and two unaffected siblings as well as of the unaffected

parents revealed that this variant co-segregates in the family with the GS phenotype. The variant was rarely reported in the Genome Aggregation Database (gnomAD) with an allele frequency of 0.00002525 (where the number of homozygote alleles is 0, the number of heterozygote alleles is 7 and the total allele number is 277,248). The variant was also absent in 511 in-house exomes of individuals with distinct diseases.

Conclusion

In the present study, we have identified a novel nonsense mutation (p.Arg24*) in *MLPH* causing GS in a Pakistani family. *MLPH* is a primary connection between *RAB27A* and *MYO5A* (*RAB27A-MLPH-MYO5A*) for regulating the melanosome transport in human melanocytes. Loss of function of this tripartite complex due to mutations in *MYO5A*, *RAB27A*, or *MLPH* leads to GS.

Acknowledgements

We wish to indicate our deep gratitude to the family for their participation in this study. The work was financially supported by Alexander von Humboldt Foundation, Bonn, Germany.

A22

Understanding the genesis and oncogenic consequences of tandem duplicator phenotypes in human cancers

Francesca Menghi¹, Mayuko Furuta¹, Floris Barthel¹, Vinod Yadav¹, Ming Tang², Bo Ji³, Gregory Carter³, Jos Jonkers⁴, Roeland Verhaak¹, Ralph Scully⁵ and Edison T. Liu³

¹The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA;

²MD Anderson Cancer Center, Houston, TX, USA; ³The Jackson

Laboratory, Bar Harbor, ME, USA; ⁴Netherlands Cancer Institute, Amsterdam, Netherlands; ⁵Beth Israel Deaconess Medical Center, Boston, MA, USA

Correspondence: Edison T. Liu (Ed.Liu@jax.org)

Human Genomics 2018, 12(Suppl 1):A22

Genomic instability is a universal phenomenon across different types of cancers, however, its underlying mechanisms are diverse and not well understood. Previously, we and others have described an enrichment of head-to-tail somatic segmental tandem duplications (TDs), a configuration known as the tandem duplicator phenotype (TDP), in a subgroup of breast and ovarian cancers. Here, we performed a pan-cancer meta-analysis of 2,717 human cancer genomes using publicly available datasets, to determine the incidence and molecular features of the TDP across different tumor types. In total, the TDP is found in ~14% of cancers and it is remarkably prevalent in triple negative breast cancer (50%), ovarian carcinoma (55%), and uterine carcinoma (45%). Detailed analysis of the TD span size in TDPs showed that TDP tumors could be subgrouped based on their TD span size profiles, which follows either a modal or a bimodal distribution with peak values corresponding to only three major discrete span size intervals: class 1 (~11 Kb), class 2 (231 Kb) and class 3 (1.7 Mb), none of which are found in non-TDP tumors. We determined commonly altered genes or pathways for each class of TDs and identified that TDPs with a prevalence of class 1 TDs are significantly associated with conjoint *TP53* mutations and *BRCA1* deficiency across different tumor types while TDPs with predominant class 2 and class 3 TDs are associated with CCNE1 pathway activation and *CRKRS/CDK12* gene disruptions. As a proof of our hypothesis, mammary tumors from tissue-specific knock out mice of *BRCA1* and *TP53* showed TDP with dominant class 1 TDs suggesting that the conjoint abrogation of *BRCA1* and *TP53* genes are the drivers for the induction of the short-span TDP configuration. In addition, we describe how heterogeneous combinations of tumor suppressors and chromatin topologically associating domains are disrupted, and oncogene and tissue-specific super enhancers are duplicated as a consequence of TD formation in TDP tumors.

In summary, our work unifies observations around a specific cancer genomic signature, TDP, and determine the factors driving each class of TDs and their mechanisms and how the consequences of the TDs generate oncogenic diversity that allows for further formation of genetic subgroups. The implications of TDP-based stratification for therapy response, by means of *in vitro* genomic manipulation, patient-derived xenograft pre-clinical trials as well as retrospective analysis of clinical specimens, are currently under investigation.

A23

Mutation in Leber's congenital amaurosis causing gene, *cct2*, evokes retinal hypoplasia in zebrafish

Yuriko Minegishi^{1,2}, Naoki Nakaya², Stanislav Tomarev², Takeshi Iwata¹

¹Tokyo Medical Center, National Hospital Organization, Tokyo, 152-8902, Japan; ²National Eye Institute, National Institutes of Health, Bethesda, MD, USA

Human Genomics 2018, 12(Suppl 1):A23

Background

The Leber's congenital amaurosis (LCA) is one of the most severe genetic retinal dystrophies. The compound heterozygous mutations in the β subunit of chaperonin containing TCP-1 (CCT), encoded by the gene *CCT2*, were identified as LCA causative from the whole exome sequencing. The patients carrying these compound heterozygous mutations exhibit a hypoplastic retina with typical LCA manifestations. The physiological role of *CCT2* in the retina remain unclear. To investigate possible functions of *CCT2* that may be associated with the onset of LCA, we established a zebrafish model with *cct2* mutation.

Materials and Methods

A zebrafish line with *cct2* mutation was established by CRISPR-Cas9. The phenotype was analyzed by microscopy, immunostaining, TUNEL and EdU/PH3 double labeling assays, and Western blotting at distinct larval stages. Phenotypic rescue was conducted by injection of *CCT2* RNA. *Chi-square* test was used to validate the Mendelian inheritance and two-tailed Student's t-test was used for statistical analysis.

Results

The zebrafish carrying the mutation in *cct β* protein with the L394H replacement and deletion of 7 amino acid residues from 395-401 (L394H-7del) was chosen to be analyzed. The homozygous of L394H-7del mutation exhibited a small eye phenotype at 2 days post-fertilization (dpf). The frequency of this phenotype was not significantly deviated from the Mendelian inheritance as judged from 5 independent mating (total 970 embryos; average incidence 25.3%; *p*-value of *Chi-Square* test 0.74). Injection of RNA encoding wild-type human CCT β rescued the small eye phenotype and also rescued RGC development and dramatically reduced cell death at 3 dpf while injection of RNA encoding LCA-causing mutant (T400P) or mock injections failed to rescue the phenotype. The differentiation of retinal ganglion cells (RGC) was attenuated, cell cycle was arrested, and the cell death occurred in the neural retina of the homozygous mutant at 2 dpf. The retinal structure was grossly disturbed, and it was difficult to distinguish the photoreceptor layer after 3 dpf. These results indicate that *cct β* plays essential role in retinal cell cycle and RGC development.

Conclusions

Dysfunction of *cct β* leads to the hypoplastic retina that resembles retinal pathology of human *CCT2*-LCA patients. Injection of *CCT2* mRNA dramatically rescued the hypoplastic retina. It is necessary to establish the mammalian model of *CCT2* to validate the efficacy of gene therapy in future.

A24

Inferring transcriptional regulatory networks controlling direct conversion of induced neurons using Convert-seq

Joachim Luginbühl¹, Tsukasa Kouno¹, Divva Sivaraman¹, Charles Plessy², Jay W Shin¹

¹RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Cell Conversion Technology Unit, Yokohama, Japan;

²RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Genomics Miniaturization Technology Unit, Yokohama, Japan

Human Genomics 2018, 12(Suppl 1):A24

The direct conversion of cell fates is controlled by hierarchical transcriptional regulatory networks (TRNs) that induce remarkable alterations of cellular and transcriptome states. The identification of key transcription factors from myriad of candidate genes within these networks, however, poses a major research challenge. Here we present Convert-seq, combining single-cell RNA sequencing (RNAseq) and ectopic gene expression with a new procedure to discriminate sequencing reads derived from exogenous and endogenous transcripts. We demonstrate Convert-seq by associating hundreds of

single cells at multiple time-points during direct conversion of human fibroblasts to induced neurons (iN) with exogenous and endogenous transcriptional signatures. Convert-seq simultaneously identified TRNs that modulate the emergence of parallel and divergent developmental trajectories during iN conversion and predicted combinatorial interactions of exogenous transcription factors controlling iN subtype specification. Validation of iN subtypes generated by novel combinations of exogenous transcription factors establish Convert-seq as a broadly applicable workflow to rapidly identify key transcription factors and TRNs orchestrating the direct conversion of virtually any cell type.

A25

Variant of FXBO40 gene is associated with general cognitive performance and its domains in the normal elderly

Vadim Stepanov^{1,2}, Anna Bocharova¹, Kseniya Vagaitseva^{1,2}, Andrey Marusin¹, Valentina Markova³, Larisa Minaycheva^{1,3}, Oksana Makeeva^{1,3}
¹Institute of Medical Genetics, Tomsk National Medical Research Center, 634050 Tomsk, Russian Federation; ²Tomsk State University, 634050 Tomsk, Russian Federation; ³Nebbiolo Center for Clinical Trials, 634009 Tomsk, Russian Federation

Human Genomics 2018, **12**(Suppl 1):A25

Cognitive performance is an important endophenotype for various neurodegenerative and neuropsychiatric diseases. In the present study genetic variant in F-box 40 protein (FBXO40) gene, previously reported to be genome-wide significant for Alzheimer's diseases, was examined for association with the various domains of cognitive abilities in normal elderly from the Russian population. Rs3772130 in FBXO40 were genotyped by PCR and MALDI-TOF mass spectrometry in a sample of 720 normal elderly subjects tested for cognitive performance using the Montreal Cognitive Assessment (MoCA). MoCA measures several cognitive domains: memory, attention, naming, visuospatial/executive, language, abstraction, subtraction and orientation domains. MoCA scores ranged between 0 and 30 points, and higher scores indicate better cognitive function. Association of genetic variability with the total MoCA scores and separate domains was estimated by generalized linear model (GLM) using age, gender, body mass index and educational level as independent variables; by non-parametric analysis of variance and by frequency comparison between upper and lower quartiles of MoCA distribution. Significantly higher frequency 'A' allele and 'AA' genotype of rs3772130 in FBXO40 gene were found in a subsample of individuals with MoCA score less than 20 comparing to fourth quartile's subsample (MoCA > 25). Odds ratio values for lower MoCA score associated with allele 'A' and genotype 'AA' of rs3772130 were 1.55 (95% CI 1.03 - 2.33, $p = 0.026$) and 1.86 (95% CI 1.12 - 3.10, $p = 0.011$), respectively. GLM analysis validated by non-parametric tests indicates that genetic variability of FBXO40 is associated with the general cognitive performance assessed by MoCA ($p=0.010$), as well as with 4 out of 8 domains of cognitive functions: memory ($p=0.020$), language ($p=0.003$), visuospatial abilities ($p=0.000$), and subtraction ($p=0.028$). The data suggests that genetic variability in FXBO40 locus associated with neurodegenerative diseases is also contributed to normal variability in cognitive performance in the elderly. This work was funded by the Russian Science Foundation (project No. 16-15-00020).

A26

Exploring the roles for repressive and quiescent chromatin regions in 3D organization of the human genome

Yichao Cai¹, Yan Ping Loh², Jia Qi Tng², Mei Chee Lim², Deepak Babu², Omer An², Henry Yang², Greg Tucker-Kellogg¹, Melissa Jane Fullwood^{2,3,4,5}
¹Department of Biological Science, Faculty of Science, National University of Singapore, Singapore, Singapore; ²Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore; ³School of Biological Sciences, Nanyang Technological University, Singapore, Singapore; ⁴Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore; ⁵Yale-NUS Liberal Arts College, Singapore, Singapore

Correspondence: Yan Ping Loh

Human Genomics 2018, **12**(Suppl 1):A26

Chromatin are decorated by various epigenetics marks and are packaged into different levels of structure, to fit into the cell nucleus. Repressive chromatin, normally marked by H3K27me3, and quiescent (unmarked) chromatin are generally considered to be silenced. Here, we hypothesize that repressive and quiescent chromatin may be important in maintaining the 3D genome organization of the human genome.

First, we classified the state of chromatin interactions from genome-wide Hi-C or ChIA-PET data, using ChIP-seq data of H3K27me3 and H3K27ac. While different chromatin interaction profiles can be seen at specific regions across different cell lines including GM12878, a normal lymphoblastoid cell line, K562, a chronic myelogenous leukemia cell line and MCF-7, a breast cancer cell line, all showed numerous repressive and quiescent chromatin interactions. Several repressive and quiescent regions connect to important gene promoters, including MYC.

Exploring the functions of repressive chromatin interactions, we first found that repressive chromatin interactions are associated with transcriptional silencing, as compared to active chromatin interactions. Next, we performed circular chromosome conformation capture (4C) on estrogen-induced MCF-7 breast cancer cells, focusing on several known estrogen-regulated genes. These cells did not show significant changes in chromatin interaction profile despite upregulation of estrogen-regulated genes, suggesting that some of these chromatin interactions may not be important in gene regulation but may instead play a role in structural maintenance. Looking at RNA-seq data of these cell lines, we found that the expression level of genes involved in chromatin interactions with repressive regions showed more dependency on the state of the genes' promoters rather than the distal interacting regions. Taken together, our results suggest a non-transcriptional, structural maintenance role for this class of chromatin interactions.

A27

Handling uncertainties and biases in the reference genome with EAGLE: Explicit Alternative Genome Likelihood Evaluator

Tony Kuo^{1,4}, Martin C Frith^{1,2,3}, Jun Sese^{1,4} and Paul Horton^{1,2}

¹Artificial Intelligence Research Center, AIST, 2-3-26 Aomi, Koto-ku. 135-0064. Tokyo. Japan; ²Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo. 5-1-5 Kashiwanoha, Kashiwa. 277-8562. Chiba, Japan; ³AIST-Waseda CBB-D-OIL. 3-4-1 Ookubo, Shinjuku-ku. 169-8555. Tokyo. Japan; ⁴AIST-Tokyo Tech RWBC-OIL. 2-12-1 Okayama, Meguro-ku. 152-8550. Tokyo. Japan

Human Genomics 2018, **12**(Suppl 1):A27

In genome-wide studies, a fundamental practice is to study a given genome relative to its differences to a reference genome. Typically, this involves aligning short read sequencing data to the reference then performing a variant discovery procedure. Some common goals include: discovering polymorphisms in a population, finding somatic mutations in cancer cells, or quantifying allele specific expression in single cells.

However, there exist numerous uncertainties in these tasks due to sequence repetitiveness and reference biases. Here we present EAGLE, a statistical model that explicitly postulates the alternative genome versus the reference genome as hypotheses and calculates the likelihood that the sequencing data supports each hypothesis.

The model considers candidate alternative genomes (i.e. variants) which may represent germ line mutations, mutations in normal vs cancer cells, or allele specific genotypes. EAGLE is free from the details of the "pileup" and handles uncertainties related to ambiguous gaps (indels), multi-mapped reads (paralogs), and spurious mappings due to outside sequences (reference bias). These models can aid in improving specificity in variant discovery tasks.

Unlike most other methods, EAGLE considers the read sequence in its entirety and can evaluate phased variants. As well, reads can be unambiguously determined to be from the reference or the alternative genome in the case of allele specific expression based on its likelihood.

All together, EAGLE is useful in rigorously evaluating the degree to which sequencing data supports any given genotype.

A28

Analysis of the evolutionally preserved function of SINEUPs, and its therapeutic features

Hazuki Takahashi, Mariko Chinzei, Kazuhiro R Nitta, Piero Carninci
RIKEN Center for Life Science Technologies, Division of Genomic
Technologies, Yokohama, Japan
Human Genomics 2018, 12(Suppl 1):A28

One of the key conclusion of the Functional Annotation of Mammalian Genome (FANTOM) project is half of transcribed RNAs are constituted by long non-coding RNAs (lncRNAs). Additionally, a large portion of lncRNA sequences are an antisense to protein coding mRNAs, shaping sense-antisense (AS) pairs. Through the FANTOM project, we discovered functional novel AS-lncRNAs, in which an embedded SINE repeat element UP-regulate translation of partially overlapping target sense mRNAs, in mouse genome. We named the AS-lncRNAs as SINEUPs. SINEUPs contain two functional domains. First, the binding domain (BD) is the region of antisense to the target mRNA. Second, SINEB2 represents the effector domain (ED) and enhances translation. Here we show the SINEUPs screening method of a high-throughput, semi-automated imaging system to optimize synthetic SINEUP BD and ED. Using SINEUP-GFP as a model found that tRNA derived SINE from several species including salmon, horse and human show similar effect of translational up-regulation to mouse SINEB2. This finding suggests the SINE function that up-regulate translation of target sense mRNA is evolutionally preserved. Here we also show the therapeutic features of SINEUPs with the BD of mouse hepatocyte nuclear factor 4 alpha (mHnf4a) mRNA. SINEUP-mHnf4a up-regulate translation of mHnf4a mRNA in both mouse hepatocellular carcinoma cells and in mice liver. These results promise SINEUPs are useful tool of gene therapy.

A29

Targeted re-sequencing of genes involved in H3K9 methylation machinery in autism spectrum disorders reveal a loss of function variant in *SUV39H2*

Shabeesh Balan¹, Yoshimi Iwayama¹, Mikiko Fukuda², Atsuko Shirai², Kaoru Kotoshiba², Motoko Maekawa¹, Tomoko Toyota¹, Tetsuo Ohnishi¹, Manabu Toyoshima¹, Yoichi Shinkai², Takeo Yoshikawa¹
¹Laboratory for Molecular Psychiatry, RIKEN- Brain Science Institute, Wakoshi, Saitama, 351-0198, Japan; ²Cellular Memory Laboratory, RIKEN, Wakoshi, Saitama, 351-0198, Japan

Correspondence: Shabeesh Balan (shabeeshbalan@riken.jp)

Human Genomics 2018, 12(Suppl 1):A29

Histone H3 methylation at lysine 9 (H3K9) is a conserved epigenetic signal, mediating heterochromatin formation by tri-methylation, and transcriptional silencing by di-methylation. Concerted interaction of histone lysine methyltransferases and demethylases regulate these processes and the downstream gene expression pathways. Gene knockout mouse models of histone lysine methyltransferases, involved in H3K9 mono- and di-methylation, viz. GLP (*Ehmt1*) and G9a (*Ehmt2*) showed autistic phenotypes and behavioral abnormalities. These evidences have underscored the contribution of the epigenetic component in autism spectrum disorders (ASD) pathogenesis. Notably, loss of function mutations in histone lysine methyltransferases has also been a consistent observation in psychiatric disorders, intellectual disability, and developmental delays. We therefore examined the possible role of genes involved in H3K9 methylation machinery in the etiology of ASD and suggest that rare functional variants in these genes may be associated with ASD. Targeted resequencing of all exonic regions in the genes for histone lysine methyltransferases (*EHMT1*, *EHMT2*, *WIZ*, *SETDB1*, *SUV39H1* and *SUV39H2*) and demethylases (*KDM3A*, *KDM3B* and *PHF8*) was performed using molecular inversion probe (MIP) based- next generation sequencing (NGS) in Japanese ASD patients and controls. The variants detected were prioritized using *in silico* tools based on novelty and functionality, thus unraveling several rare missense variants exclusively in ASD patients.

In vitro functional analysis revealed that the rare missense p.Ala211-Ser located in the Pre-SET domain of *SUV39H2* drastically reduced H3K9 methylation efficiency. Interestingly, *SUV39H1* and *SUV39H2* expressions were found to be downregulated in BA21 and dorsal raphe nucleus, respectively in postmortem ASD brain samples when compared to the controls. Functional evaluation of majority of other variants in the important domains had no impact with respect to their function. Therefore prioritization of variants by *in silico* methods should be carefully pursued, owing to the presence of functionally neutral, rare private variants. To summarize, we have provided evidence for the putative role of *SUV39H2* in ASD pathogenesis.

Keywords: Autism spectrum disorders, Histone lysine methyltransferases, Rare variant, H3K9 methylation

Ethics Approval

All participants gave informed, written consent to join the study, after receiving a full explanation of study protocols and objectives. The human genetics study was approved by the ethics committees of RIKEN and all participating institutes, and was conducted in accordance with the Declaration of Helsinki. The animal experimental procedures were approved by the RIKEN Animal Ethics Committee.

A30

Establishment and activity of the Asian Eye Genetics Consortium (AEGC)

Takeshi Iwata¹, Gyan Prakash², Paul Baird³, S Natarajan⁴
¹National Institute of Sensory Organs, Tokyo Medical Center, National Hospital Organization, 2-5-1 Higashigaoka, Meguroku, Tokyo 152-8902 JAPAN; ²National Eye Institute, National Institutes of Health, Bldg. 31, 6A03, Bethesda, MD, 20892 USA; ³Centre for Eye Research Australia, Ophthalmology, Department of Surgery, University of Melbourne, 32 Gisborne St., East Melbourne, VIC 3002 AUSTRALIA; ⁴Aditya Jyot Eye Hospital, Plot No. 153, Road No. 9, Major Parmeshwaran Road, Opp S.I.W.S. College Gate No. 3, Wadala, Mumbai 400031 INDIA
Human Genomics 2018, 12(Suppl 1):A30

With significant advancement in DNA sequencing and internet data exchange technology, we are experiencing a new era in the field of human genetics. Decades of eye genetic research has shown that genetic variations play significant role in eye diseases. These genetic involvements can be highly penetrant as in Mendelian eye diseases or highly associated as risk factors in common eye diseases. In both cases, patient genome sequence is quickly determined and traced within a family and compared with millions of genome sequences collected around the world and stored in a database. However, most of the information originates from populations of European decent and information on other ethnic groups are limited. The Asian Eye Genetics Consortium (AEGC, <http://asianeyegenetics.org>) was established in 2014 to focus on eye research in Asia, the most populated region of the world where limited data are available on genetic variation in eye diseases has been reported.

AEGC has the following goals and plans:

- Share genetic information in the Asian population to rapidly isolate common disease-associated variants,
- Establish systems for accurate diagnosis and grouping of Asian eye diseases,
- Establish systems for cost effective genetic analysis,
- Develop a research-oriented database to collect, diagnose and catalog eye diseases in Asia,
- Support and foster collaboration among Asian countries for the advancement of research that will provide genetic information in the Asian population,
- Collaborate with other international or regional organizations with similar goals, and
- Organize and hold regional congresses and other educational and scientific activities to promote goals of the consortium

Over one hundred eye researchers from Eye Institutions and Hospitals from more than 20 countries including Australia, China, India, Indonesia, Israel, Japan, Malaysia, Pakistan, Philippines, Saudi Arabia, Singapore, South Korea, Sri Lanka, Taiwan, Thailand, Turkey, UAE, and

USA are now participating. These groups are currently interacting and collaborating to develop programs to share, catalogue and collaboratively work to identify the genetic aspects of eye diseases in Asia. The consortium has brought a collective thinking and ideas from the researchers around the world who have interest in genetic eye research in the Asian region.

A31

Identifying long noncoding RNAs associates in human cells

Saomya Agrawal¹, Jordan Ramilowski¹, Chi-Wai Yip¹, Chung Chau Hon¹, Kayoko Yasuzawa¹, Masayoshi Ito^{1,2}, Takeya Kasukawa¹, Naoto Kondo¹, Ken Yagi¹, Harukazu Suzuki¹, Christopher J.F. Cameron^{3,4}, Josée Dostie³, Jay W. Shin¹, Michiel de Hoon¹, Piero Carninci¹

¹Division of Genomic Technologies (DGT), RIKEN Center for Life Science Technologies, Yokohama, Japan; ²RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan; ³School of Computer Science, McGill University, Montreal, Canada; ⁴Department of Biochemistry, McGill University, Montreal, Canada

Correspondence: Piero Carninci (carninci@riken.jp)

Human Genomics 2018, **12(Suppl 1)**:A31

Background

More than half of the genes in human genome transcribes non-coding RNAs and for the very long time they were thought to be nonfunctional biological "dark matter". However, recent studies have shown that noncoding RNAs especially long noncoding RNAs (lncRNAs) have a key role in various biological processes like transcriptional regulation, enhancer-like functions, differentiation of cells and tissues, epigenetic regulation, cellular senescence, DNA replication, maintenance of the telomeric structure and so on. However, despite their biological significance a comprehensive functional annotation of lncRNAs is still lacking, including the genomic functional elements regulated by each lncRNA. The information about three-dimensional localization of a regulatory element in the nucleus can be used in determining their potential interacting partners or in other words in establishing their functional territory. Therefore, as part of the FANTOM6 project, we have used genomic interaction data and combine them with various regulatory element databases to define the function of the lncRNA in various human cells.

Results

We have generated a high-resolution genomic interaction map using newly produced Hi-C data for somatic and stem cells that were used to identify potential interacting partners of the lncRNAs such as protein-coding genes, other lncRNAs, microRNAs, or enhancers in the human cells. Further, we have also identified common or cell type-specific interacting partners of lncRNAs by comparing interaction map of different cell types. We are currently overlaying various databases for RNA binding proteins, transcription factor binding sites, and specific genomic patterns of lncRNAs to find the potential connecting link between two interacting genomic elements. To further validate the interacting partners, we have knocked-down unbiasedly selected lncRNAs followed by transcriptome profiling using CAGE to assess whether the expression of genes interacting with lncRNAs (as indicated by Hi-C data) was affected by their knockdown. Our analysis reveals that for many lncRNAs interacting genes are indeed enriched among the differentially expressed genes, suggesting that those lncRNAs may have a direct or indirect effect in regulating the expression of the genes.

Conclusions

Our analysis has shown that lncRNA transcribing regions are not isolated genomic regions but frequently interact with both protein-coding genes and other non-coding RNAs in the cells. Further, these interacting partners can be conserved or cell-type specific showing the dynamic nature of the lncRNAs. Overall, we will discuss by how integrating information about genomic contact together with different transcription regulatory elements can be used to predict the potential biological role of lncRNAs.

A32

Selective detection of somatic mutations in circulating tumor DNA from pancreatic cancer patients by next-generation sequencing and a bioinformatic variant filter

Yoji Kukita¹, Ryoji Takada², Kazuyoshi Okawa², Kazuhiro Katayama², Kikuya Kato¹

¹Laboratory of Medical Genomics, Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan, 630-0192;

²Department of Hepatobiliary and Pancreatic Oncology, Osaka International Cancer Institute, Osaka, Japan, 541-8567

Correspondence: Yoji Kukita

Human Genomics 2018, **12(Suppl 1)**:A32

Circulating tumor DNA (ctDNA) is released from dying/dead cancer cells to blood, and an emerging candidate of cancer biomarkers. Although next-generation sequencing is becoming the method of choice, the number of false positives caused by a high read error rate is a major problem to detect ctDNA (somatic mutations with low frequency) from abundant cell-free DNA (cfDNA) in blood. We have developed a high-fidelity target sequencing system, Non-Overlapping Integrated Reads Sequencing System (NOIR-SS), to accurately identify and absolutely quantify individual DNA molecules using barcode sequence tags. To apply that to ctDNA detection in cancer patients, we designed a gene panel that targets hotspots of pancreatic cancer related genes (comprising 2.8 kb of genomic DNA), and devised an assay system by combining NOIR-SS and a bioinformatic variant filter, not only to detect variants but also to remove those unlikely to be tumor-specific: those either absent or occurred at low frequencies in the Catalogue of Somatic Mutations in Cancer database. The performance of the system was evaluated using two independent cohorts (cohort1: 57 pancreatic cancer patients (PCs) and 12 healthy individuals; cohort2: 86 PCs and 20 patients with intraductal papillary mucous neoplasm (IPMN)). Our system was likely to eliminate most non-tumor-specific mutations, and showed comparable sensitivity for detecting tumor-specific mutations to other methods based on digital PCR and deep sequencing.

A33

Chromothripsis: evolution of *de novo* small supernumerary marker chromosomes from trisomies

Nehir Kurtas¹, Luciano Xumerle², Massimo Delledonne², Alfredo Brusco³, Krystyna Chrzaowska⁴, Albert Schinzel⁵, Silvana Guerner⁶, Emmanouil Manolakos⁷, Edoardo Errichiello¹, Sabrina Giglio⁸, Thomas Liehr⁹, Orsetta Zuffardi¹

¹Department of Molecular Medicine, University of Pavia, Pavia, Italy;

²Department of Biotechnology, University of Verona, Verona, Italy;

³Department of Medical Sciences, University of Turin, Torino, Italy;

⁴Department of Medical Genetics, The Children's Memorial Health

Institute, Warsaw, Poland; ⁵Institute of Medical Genetics, University of Zurich, Zurich, Switzerland; ⁶Laboratory of Medical Genetics, Fondazione

IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; ⁷Laboratory

of Genetics, Access to genome P.C., Thessaloniki, Greece; ⁸Biomedical

Experimental and Clinical Sciences "Mario Serio", University of Florence,

Firenze, Italy; ⁹Institute of Human Genetics, Jena University Hospital,

Jena, Germany

Correspondence: Nehir Kurtas (nehirkurtas@gmail.com)

Human Genomics 2018, **12(Suppl 1)**:A33

We hypothesize that most *de novo* constitutional non-recurrent small supernumerary marker chromosomes (sSMC) are the remnant of a supernumerary chromosome present in trisomic embryos, which undergoes a chromothripsis event resulting in its massive fragmentation with loss of some portions and disordered reunion of the remaining ones. Based on the mainly maternal origin of trisomies, we also assume that, depending on which of the three homologous undergo chromothripsis, the remaining two are either in maternal hetero-/isodisomy or of biparental origin. To investigate this, we

collected DNA from 20 cases of non-recurrent *de novo* sSMCs, already defined by array-CGH, and from their parents, all in mosaic with a normal cell line. We performed whole genome paired-end sequencing (WGS) in the first 10 cases at 30-40x coverage. We also performed trio analysis by microsatellites spread along the entire chromosome by which the sSMC originated. As a result, in seven cases, namely sSMC2a, sSMC2b, sSMC7a, sSMC8a, sSMC17, sSMC18, sSMC11, markers were constituted by a disordered assembly of two or more segments of their corresponding chromosomes. In one of them, sSMC11, the marker was derived from random fusion and disordered assembly of ten fragments spread along chromosome 11. In three cases, sSMC7b, sSMC1, and sSMC8b, only a single chromosomal region was involved in their construction. In each case the novel junctions were confirmed by breakpoints (bps) PCR amplification and cloning, with a fusion signature showing repair mechanisms such as non-homologous end joining and microhomology mediated processes. Microsatellite and SNPs analysis in the trios indicated a maternal origin of the marker with biparental origin of the related homologous chromosomes in 4 cases, and a paternal origin with maternal hetero/isodisomy of their related chromosomes in the remaining 5 cases. The sSMC8b was of paternal origin whereas the homologous chromosomes 8 were biparental, indicating its postzygotic origin.

Finally, these data demonstrate both a link between numerical and structural anomalies and that early lethal trisomies may leave a dramatic legacy in the postnatal life.

Informed consent had been obtained from all families according to the ethical committee of the Meyer Hospital in Firenze, Italy.

A34

Identification of epigenomic regulatory networks of heart development

Michał Pawlak¹, Katarzyna Kedzierska¹, Maciej Migdał¹, Karim Abu Nahia¹, Hashimoto Kosuke³, Jordan Ramilowski³, Lukasz Bugajski², Aleksandra Marconi¹, Katarzyna Piwocka², Piero Carninci³, Cecilia Winata^{1,4}

¹International Institute of Molecular and Cell Biology in Warsaw, Laboratory of Zebrafish Developmental Genomics, Warsaw, Poland;

²Nencki Institute of Experimental Biology, Laboratory of Cytometry, Warsaw, Poland; ³RIKEN Center for Life Science Technologies, RIKEN Yokohama Institute, Yokohama, Japan; ⁴Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany

Correspondence: Cecilia Winata (cwinata@iimcb.gov.pl)

Human Genomics 2018, **12(Suppl 1)**:A34

Congenital heart disease (CHD) is one of the most common birth defects (1 of 100 live births) and is a major cause of heart diseases. In vast majority, CHD is caused by abnormalities in multilayer molecular processes involving interactions between various signaling pathways which are not fully understood. The lack of understanding on the regulatory network driving heart development often hinders the diagnosis of CHD and development of novel therapeutic strategies.

To profile the genetic and epigenetic regulation of heart development, we employed genomics methodology to profile the transcriptome and chromatin state in wild-type and several zebrafish lines carrying mutations in transcription factors responsible for heart development. Taking the advantage of transgenic zebrafish lines with fluorescently-labeled cardiomyocytes, we isolated specific cell fractions from zebrafish embryos and larvae at different stages of heart development.

We have developed and optimized the strategies of RNA sequencing (RNA-seq) from cardiomyocytes as well as the assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq), providing an in-depth analysis of genetic and epigenetic factors regulating heart development on genome-wide scale. It allowed us to identify gene regulatory networks as well as their downstream regulatory chromatin regions during key stages of heart development.

Finally, we have identified novel candidate genes driving normal cardiac development, the disruption of which may lead to CHD in human.

This work is supported by National Science Centre, Poland, grant no. 2014/13/B/N22/03863 MP is supported by Ministry of Science and

Higher Education, Poland and National Science Centre, Poland, SON-ATA. Program no. 2014/15/D/NZ5/03421.

A35

The development of cardiac pacemaker cells derived from AV node in zebrafish

Karim Abu Nahia¹, Michał Pawlak¹, Agata Sulej¹, Rashid Minhas¹, Łukasz Bugajski², Katarzyna Piwocka², Vladimir Korzh¹, Cecilia L. Winata^{1,3*}

¹International Institute of Molecular and Cell Biology, Warsaw, Poland;

²Laboratory of Cytometry, Nencki Institute of Experimental Biology, Warsaw, Poland;

³Max-Planck Institute of Heart and Lung Research, Bad Nauheim, Germany

Correspondence: Cecilia L. Winata (cwinata@iimcb.gov.pl)

Human Genomics 2018, **12(Suppl 1)**:A35

Cardiovascular diseases are the leading cause of death globally. Among them, heart arrhythmia and conduction disorders are one of the most prevalent. The cardiac conduction system is a key component that coordinates and regulates proper heartbeat. An unaffected, rhythmic heart contraction occurs due to the presence of highly specialized network of pacemaker cells responsible for generation and transmission of electrical impulses throughout the heart. These impulses originate within two major pacemaker sites, the sinoatrial node (SAN) and atrioventricular node (AVN). The AV node is a highly specialized conducting tissue that propagates and slows down the impulse conduction considerably ensuring that the atrium has ejected their blood into the ventricles first before the ventricles contract, which results in unidirectional blood flow. Any disturbances in propagation of these impulses may lead to serious consequences like heart attack, stroke and different types of arrhythmia. However, despite the prevalence of conduction system disorders, current knowledge about its development remains elusive. In our study, we aim to gain an insight and elucidate the molecular mechanism underlying the development and function of cardiac pacemaker cells, especially those that originate from AV node. To achieve our goal, we use transgenic zebrafish lines as a model organism which allows us to easily investigate AVN development by means of genomics. I will present our ongoing analysis of the AVN RNA-seq data which focuses on elucidating the molecular composition of these specialized cells as the first step to understand their development and function. In the long run we hope to establish the zebrafish as a model organism for research on cardiac arrhythmia in humans.

This work is supported by the Foundation for Polish Sciences, First-TEAM grant nr. First TEAM/2016-1/8

A36

Computational analyses of cardiomyocyte regulatory elements using integrative approaches

Maciej Migdał¹, Michał Pawlak¹, Karim Abu Nahia¹, Norbert Dojer², Cecilia Lanny Winata¹

¹International Institute of Molecular and Cell Biology in Warsaw, Laboratory of Zebrafish Developmental Genomics, Warsaw, Poland;

²University of Warsaw, Institute of Informatics, Warsaw, Poland

Correspondence: Maciej Migdał

Human Genomics 2018, **12(Suppl 1)**:A36

One of the most frequent causes of neonatal death is congenital heart defect. While many individual genes and transcription factors crucial to the process of heart formation has been identified, the manner in which they interact and orchestrate gene expression remains mostly unknown. Understanding the complex genetic regulatory network of heart development is necessary to track the causes of congenital heart defect. Here, we study the network of interactions between chromatin landscape and genetic elements that governs gene expression at early stages of heart formation using the zebrafish as a model organism. Based on RNA-seq and ATAC-seq data sets generated in our laboratory, we are using computational methods such as digital genomic footprinting to elucidate the interplay between transcription factors and gene expression. Using this information, we seek to integrate our data sets and construct a model describing the relations between regulatory elements and

their effect on gene expression. Our analyses focus at the core genes required for proper heart formation and the mechanisms regulating their expression. I will present our findings on the relationships between key elements involved in the process of heart development. This work is supported by National Science Centre, Poland, grant no. 2014/13/B/N22/03863

A37

Signalling network map of innate immune response in cancer reveals signatures of cell heterogeneity and polarization in tumor microenvironment

Inna Kuperstein, Maria Kondratova, Urszula Czerwińska, Andrei Zinovyev and Emmanuel Barillot

Institut Curie, PSL Research University, Mines Paris Tech, Inserm, U900, F-75005, Paris, France

Human Genomics 2018, 12(Suppl 1):A37

To describe the balance between components of tumor microenvironment (TME) and to analyse impact of both non-immune and immune cells, we systematically collected information on related molecular mechanisms and represented in a form of comprehensive network maps. The modular map of cancer associated fibroblasts (CAF), the non-immune component of TME, is composed of 681 objects and 585 reactions. The map is covering the main functions of CAFs in tumor among others, interactions with extracellular matrix components, signalling coordinating involvement of CAFs in tumor growth and interactions of CAFs with immune system. There are two types of functional modules on the CAF maps, modules responsible for CAFs activation are associated with pro-tumor activity and modules involved in CAFs inhibition, therefore contributing to anti-tumor activity. In addition, we constructed signalling maps of macrophages, dendritic cells, myeloid-derived suppressor cells, natural killers, neutrophils and mast cells. These cell type-specific maps integrated together and updated by interactions and cross-talks between them and the map of tumor cell, gave rise to a seamless comprehensive meta-map of innate immune response in cancer. The meta-map contains 1466 objects and 1084 reactions and depicts signalling responsible for anti- and pro-tumor activities of innate immunity system as a whole. The meta-map is represented in a geographical-like manner, possessing a hierarchical structure with two functional zones, each divided into meta-modules and smaller sub-modules. The network maps were applied for identification of possible molecular mechanisms regulating CAF and innate immune cells reprogramming in TME in metastatic melanoma. Unsupervised statistical methods were applied for decomposition of single cell RNASeq data of CAF, natural killers and macrophages. Analysis and interpretation of expression patterns in the context of the network maps demonstrated existence of numerous sub-populations within each cell type characterized by different anti- and pro-tumor functional properties. These network-based polarization signatures correlated with the survival status of patients. We concluded that tumor microenvironment may contain a range of CAF and innate immune cells, varying in their polarization status. The fine-tuned balance between the sub-populations will dictate the overall impact of TME onto tumor evolution in each given case.

A38

Single cell RNA sequencing of T cells in Crohn's disease identifies tissue specific drug targets

Uniken Venema, Werna T.¹, Voskuil, Michiel D.¹, Vich Vila, Arnau¹, Graham, Daniel B.^{3,4}, Weersma, Rinse K.¹, Festen, Eleonora A.^{1,2}

¹Department of Gastroenterology and Hepatology, UMCG, Groningen, The Netherlands; ²Department of Genetics, UMCG, Groningen, The Netherlands; ³Broad Institute of MIT and Harvard, Boston, MA, USA;

⁴Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

Human Genomics 2018, 12(Suppl 1):A38

Introduction

Crohn's disease (CD) is a chronic inflammatory disease, that predominantly causes inflammation in the terminal ileum¹. Genome-wide

association studies have identified around 200 CD risk loci². These loci are enriched for genes involved in T cell signaling, highlighting the importance of T cells in CD pathology³. To get true insight in the underlying pathomechanisms it is crucial to study these T cells in the relevant tissue: the intestinal mucosa⁴. Characterization of mucosal T cells in a disease- and tissue-specific context at a single cell resolution has not yet been performed for CD. We aimed to increase our insight in CD pathomechanisms and to identify novel drug targets by characterizing mucosal T cell populations from CD patients using single cell RNA sequencing (scRNA-seq).

Methods

We performed scRNA-seq of 5,292 CD3+ T cells isolated from peripheral blood (PBL) and from ileal mucosal biopsies of CD patients. Biopsies were dissociated and separated into T cells from the epithelium (IEL) and the lamina propria (LPL). scRNA-seq was performed with an adapted SmartSeq2 protocol, using 3'-end library generation and unique molecular identifiers, sequenced on the Illumina NextSeq500.

Results

Unsupervised clustering of our scRNA-seq data identified 5 distinct T cell types. The distribution of cell types differs between IEL, LPL and PBL: while cytotoxic T cells (CTL) dominate the IEL, the blood T cell reservoir is mainly composed of quiescent T cells, and T-helper 17 (Th17) cells are the largest population within the LPL.

Th17 cells and CTLs show the highest proportion of differentially expressed known CD risk genes: such as *CD69* and *FOS* in Th17 cells, and *CX3CR7* and *NKG7* in CTLs. Aligning potential IBD drug targets⁵ to CTL signature genes, we find that Etrolizumab, but not Vedolizumab, specifically appears to target mucosal CTLs. Two potential drug repositioning opportunities for targeting mucosal Th17 cells, are *MSX-122*, an anti-tumor CXCR4-antagonist, and Rivenprost, which targets *PTGER4*, and has been tested in UC patients. (Fig. 1)

Conclusion

We have conducted the first detailed transcriptomic characterization of disease- and tissue-specific effector cells in Crohn's disease. We have profiled 5 distinct T cell types using single cell transcriptome data. We show that CD risk genes are significantly overexpressed in ileal mucosal Th17 and CTLs and provide promising targets for future cell-specific therapies in CD patients.

References

1. Baumgart *et al.* (*Lancet*, 2012).
2. Liu *et al.* (*Nat. Genet.*, 2015).
3. Farh *et al.* (*Nature*, 2015)
4. Raine *et al.* (*Gut*, 2015)
5. OpenTarget: <https://www.targetvalidation.org/>

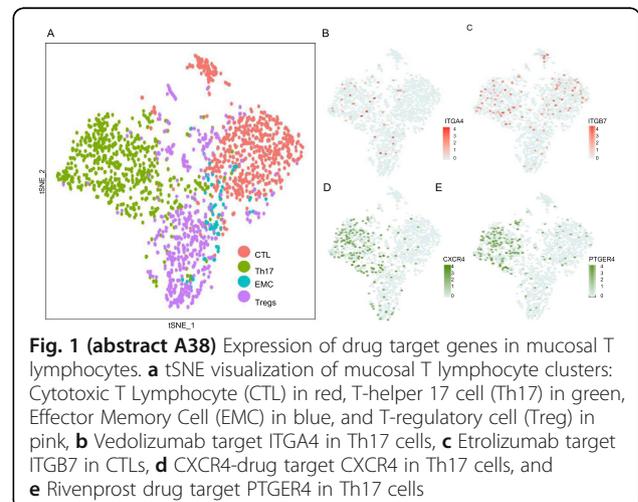


Fig. 1 (abstract A38) Expression of drug target genes in mucosal T lymphocytes. **a** tSNE visualization of mucosal T lymphocyte clusters: Cytotoxic T Lymphocyte (CTL) in red, T-helper 17 cell (Th17) in green, Effector Memory Cell (EMC) in blue, and T-regulatory cell (Treg) in pink, **b** Vedolizumab target ITGA4 in Th17 cells, **c** Etrolizumab target ITGB7 in CTLs, **d** CXCR4-drug target CXCR4 in Th17 cells, and **e** Rivenprost drug target PTGER4 in Th17 cells

A39

Molecular signatures of Alzheimer’s disease in the TOMM40-APOE-APOC1 locus

Alexander M Kulminski, Leonardo Shu, Yury Loika, and Irina Culminskeya
 Biodemography of Aging Research Unit, Social Science Research
 Institute, Duke University, Durham, NC 27708-0408, USA

Correspondence: Alexander M Kulminski
 (Alexander.Kulminski@duke.edu)

Human Genomics 2018, 12(Suppl 1):A39

Background

Although the APOE region is the strongest genetic risk factor for Alzheimer’s diseases (AD), its pathogenic role remains poorly understood. Elucidating genetic predisposition to AD, which represents a subset of age-related diseases characteristic for post-reproductive period, is hampered by the undefined role of evolution in establishing molecular mechanisms of such diseases. This uncertainty is inevitable source of evolution-related (selection-free) genetic heterogeneity in predisposition to AD.

Materials and methods

We examined linkage disequilibrium (LD) structures characterized by seven single nucleotide polymorphisms (SNPs) from TOMM40-APOE-APOC1 locus in 2,661 AD-affected and 16,089 unaffected individuals, who were mainly older than 55 years, from four independent studies and LD structures in 16,089 older unaffected subjects and 5,303 younger individuals aged 20-55 years from two additional studies.

Results

Consistent with the undefined role of evolution in age-related diseases, we found that the LD structures, being heterogeneous, are significantly different in subjects with and without AD, $p < 2.0 \times 10^{-3}$. The pattern of the significant difference represents molecular signature of AD comprised of SNPs from these genes. As this pattern was consistent in all studies, all pair-wise estimates of the LD difference in the pooled sample (except rs8106922 and rs405509) were significant after Bonferroni correction, $p < 2.4 \times 10^{-3}$ (Fig. 1a). There was no significant difference in LD patterns between older unaffected subjects, characterized by exponentially increasing mortality rates, and the younger population, characterized by negligible mortality rates (Fig. 1b).

Conclusions

Significant and highly heterogeneous molecular signature of AD provides evidence on complex polygenetic predisposition to AD in the TOMM40-APOE-APOC1 locus. Significant differences in pair-wise LD in subjects with and without AD indicate SNPs, or their proxies, likely involved in AD pathogenesis. The lack of significant difference in LD patterns between the older and younger subjects indicates that the LD patterns in unaffected subjects are likely evolutionary selected whereas LD patterns in AD-affected subjects are likely the result of LD rearrangement in modern environment. These findings are more consistent with an environment-sensitive complex haplotype than with a single genetic variant origin of AD in this locus.

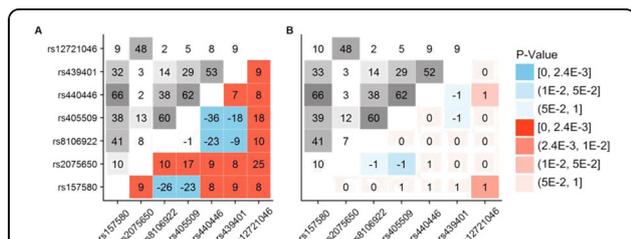


Fig. 1 (abstract A39) a Molecular signature of AD in the pooled sample of older subjects from four studies. Upper-left: LD pattern (r^2 , %) in AD-unaffected subjects. Lower-right: heat map for the difference in LD between AD-affected and AD-unaffected subjects. **b** Upper-left: LD pattern (r^2 , %) in individuals younger than 55 years. Lower-right: heat map for the difference in LD between younger and older AD-unaffected subjects.

A40

Molecular characterization of the SLC transporters in human cancer

Linghua Wang
 Department of Genomic Medicine, The University of Texas MD
 Anderson Cancer Center, Houston, TX, USA

Human Genomics 2018, 12(Suppl 1):A40

Metabolic reprogramming has now been well recognized as one of 10 hallmarks of human cancer. The membrane transporters that enable efficient cellular metabolism and aid in nutrient sensing, are receiving increasing attention in recent years. Membrane transporters can be divided into 3 main classes: ABC transporters, p-type ATPases and the solute carrier family (SLC). The SLC group of transporters includes nearly 400 members organized into 52 families that mediate the transport of a wide variety of substrates across biological membranes. Until recently, little was known about the role of these SLC transporters in the mechanisms of cancer development and progression. Yet, the genomic landscape and clinical relevance of SLC transporters in human cancer have not been investigated systematically. We here performed a comprehensive characterization of 396 SLC transporters by integrating multi-platform data across over 30 tumor types from The Cancer Genome Atlas. We observed frequent dysregulation of SLC transporters across cancer types and identified cancer type- and subtype-specific expression signatures. We identified significant correlations of these SLC transporters with the activity of cancer metabolic pathways including oxidative phosphorylation and glycolysis, and explored their complex regulatory networks. We also examined the association of SLC transporters with known oncogenic pathways including mTOR and AKT pathways and investigated the role of SLC transporters in driving tumor metastasis. We further evaluated the potential clinical relevance of these SLC transporters and identified a bunch of SLC transporters that were significantly associated with patients’ clinical outcomes. Our study provided an in-depth understanding of the role of SLC transporters in cancer metabolic reprogramming, elucidated their mechanisms in cancer development and metastasis, and provided novel biomarkers and potential therapeutic targets that can be further developed for future clinical applications.

A41

Large-scale unsupervised transcriptome annotation reveals candidate small RNAs from unannotated regions

Mickaël Mendez^{1,2}, Michael M Hoffman^{1,2,3}
¹Department of Computer Science, University of Toronto, Toronto, ON, Canada; ²Princess Margaret Cancer Center, Toronto, ON, Canada; ³Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Correspondence: Mickaël Mendez
 Human Genomics 2018, 12(Suppl 1):A41

Background

Recent efforts from the FANTOM5 consortium have shown that specific transcriptional patterns found within intergenic unannotated regions can be associated with different classes of regulatory elements and non-coding RNAs.

However such studies of the transcriptome remain challenging for multiple reasons. First, current methods often rely on pre-existing transcript models and neglect both cell-type specific transcripts and intronic signals. Second, there is a large diversity of transcriptome profiling protocols targeting only specific types of RNAs (polyA⁺ vs polyA⁻, long vs short, etc). As a result, relating the signals obtained from an experiment to others is non-trivial, fastidious and often done case by case by visual inspection of a genome browser view.

Method

We developed a new method based on dynamic Bayesian networks, SegRNA, which brings a new model to Segway semi-automated genome annotation model. SegRNA identifies recurrent combinatorial patterns occurring across libraries of stranded RNA signals and produces one human-understandable annotation per strand. SegRNA model is unsupervised and allows to compare intronic and exonic patterns without relying on pre-existing transcript models. Additionally, SegRNA learns a model of the most likely transitions between all

patterns, allowing us to describe the structure of larger domains made of individual RNA elements.

Results

We applied SegRNA on ENCODE and FANTOM5 transcriptome sequencing libraries from the K562 chronic myeloid leukemia cell line. Each of these libraries was selected to represent the widest possible diversity of RNA including nascent, capped, polyA⁺, polyA⁻, long or short RNAs. We found that some of the patterns re-discovered fundamental gene structures such as promoter, exonic and intronic regions, in an unsupervised fashion. Furthermore, we found that about 10% of genomic signals identified by SegRNA falls within unannotated regions. Specifically, we identified a pattern enriched for known small RNAs which also occurs within intergenic, intronic and exonic regions that have not been characterized yet.

Discussion

SegRNA identified multiple patterns associated with promoter activity and higher gene structures such as gene body and exonic regions from annotated genes. Furthermore, SegRNA's unsupervised annotations allow one to explore intergenic transcriptional stranded patterns occurring across entire genomes, and collections of transcriptome libraries.

A42

PTSD re-experiencing in the Million Veterans Program sample: risk loci and biology

Gelernter J.^{1,2}, Sun N.^{3,4}, Polimanti R.^{1,2}, Pietrzak R.^{1,2}, Lu Q.⁴, Hu Y.⁴, Li B.⁴, Chen Q.^{3,4}, Radhakrishnan K.³, Aslan M.^{3,5}, Cheung K.H.^{3,6}, Li Y.^{3,7}, Rajeevan N.^{3,7}, Sayward F.^{3,7}, Harrington K.^{8,9}, Cho K.^{8,10}, Honerlaw J.⁸, Pyarajan S.^{8,10}, Quaden R.⁸, Gaziano J.M.^{8,10}, Concato J.^{3,5}, Zhao H.^{3,4}, and Stein M.B.^{11,12}

¹Psychiatry Service, VA Connecticut Healthcare System, West Haven, CT, USA; ²Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; ³VA Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, CT, USA; ⁴Department of Biostatistics, Yale University School of Public Health, New Haven CT, USA; ⁵Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; ⁶Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT, USA; ⁷Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, USA; ⁸Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA; ⁹Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA; ¹⁰Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; ¹¹Psychiatry Service, VA San Diego Healthcare System, San Diego, CA, USA; ¹²Department of Psychiatry, University of California San Diego, San Diego, CA, USA

Correspondence: Gelernter J

Human Genomics 2018, **12**(Suppl 1):A42

Background

Posttraumatic stress disorder (PTSD) is a major problem among the veteran population and presents treatment challenges. The US Veterans Affairs (VA) Million Veteran Program (MVP) is building a large medical and genetic information databases, currently >620,000 consented participants. ~350,000 enrollees have genotype information available, linked to VA EHR data and questionnaire responses—the largest current sample for studying PTSD-relevant traits. PTSD symptoms are categorized into 3 major symptom clusters by DSM-IV criteria: intrusive re-experiencing of the trauma, avoidance of trauma-associated stimuli, and alterations in arousal or reactivity.

Materials and methods

We conducted a GWAS on the re-experiencing symptom cluster score based on a sum of 5 items from the PTSD Checklist (recurrent intrusive thoughts/dreams/flashbacks of trauma; emotional or physiological response to reminders of trauma), total score, 5-25. This is the symptom cluster most characteristic of PTSD.

Results

After cleaning, 146,660 European-Americans (EAs) and 19,983 African-Americans (AAs) were retained. In EAs, 8 distinct common-

variant genome-wide-significant (GWS) regions were identified—three with significance >5×10⁻¹⁰. These latter regions map to chrom. 3 – lead SNP rs2777888 (2.1E-11), gene *CAMKV*, same SNP previously implicated in “age at first birth”; chrom. 17 – lead SNP rs2532252 (4.5E-10), closest to *KANSL1* but within a well-known long high-LD region (the site of an inversion common in EAs) that also includes *CRHR1* (corticotropin releasing hormone receptor 1); and chrom. 18 – lead SNP rs2123392 (5.4E-11), at *TCF4*. Other significant associations were observed at *KCNIP4*, *HSD17B11*, *MAD1L1*, and *SRPK2*. *TCF4* and *MAD1L1* have both previously been GWS-associated to schizophrenia and other psychiatric traits. There were no GWS associations in the smaller AA sample, but when EA and AA subjects were meta-analyzed, the lead SNP in the region on chrom. 17 shifted to rs1724409 (and increased significance to 3.6E-11). The chrom. 17 inversion is much less common in AAs than EAs, and this added fine-mapping information. This new lead variant is intronic at *CRHR1*, a very strong functional candidate for PTSD. LD score regression analysis showed polygenic association with many psychiatric and behavioral traits, including “mood swings” (1.07E-56), neuroticism (1.28E-44), and “miserableness” (2.62E-30).

Conclusion

Meta-analyzing results from European- and African-ancestry subjects improved our ability to narrow an associated region in a biologically meaningful way. These results provide new insight into the biology of the most characteristic PTSD symptom cluster in what is the best-powered study undertaken to date.

Ethics Approval

Research involving MVP in general is approved by the VA Central IRB; the current project was also approved by local IRBs in Boston, San Diego, and West Haven (USA).

Acknowledgements

On behalf of the Department of Veterans Affairs Cooperative Studies Program (#575B) and the VA Million Veteran Program.

A43

ChimerDB 3.1 update: a knowledgebase for fusion genes in human cancer

Ye Eun Jang¹, Subin Cho¹, Insu Jang², Inhae Jung¹, Byunguk Lee², Sanghyuk Lee^{1†}

¹Department of Bio-Information Science, Ewha Womans University, Seoul 03760, Rep. of Korea; ²Korean Bioinformation Center (KOBIC), KRIBB, Daejeon 34141, Rep. of Korea

Human Genomics 2018, **12**(Suppl 1):A43

Objectives: Identification and analysis of fusion genes is essential to find prognostic markers and therapeutic targets of cancer. ChimerDB is a knowledgebase for fusion genes encompassing analysis of deep sequencing data and manual curations.

Methods: To build knowledgebase of fusion gene, we curated fusion genes from public resources and reviewed manually. We also utilized an advanced text mining method to build a literature database based on PubMed abstracts. A classifier to identify fusion genes from the candidate sentences were developed using machine learning techniques. Most importantly, we analyzed RNA-Seq data of the TCGA project using two programs of FusionScan and StarFusion in this update.

Results: ChimerDB 3.1 is composed of three modules of ChimerKB, ChimerPub, and ChimerSeq. ChimerKB represents a knowledgebase including 1,041 manually curated fusion genes with the experimental evidences. Notably, 460 fusion breakpoints were catalogued in this update of ChimerDB 3.1. ChimerPub includes 2,767 fusion genes obtained from text mining of PubMed abstracts. ChimerSeq is the archival database of fusion gene candidates computationally identified from RNA-Seq data of the TCGA project. In this update, we expanded the patient samples substantially to include 31 cancer types and 7,775 patients. Based on our benchmark test, we used two reliable and efficient programs of FusionScan and StarFusion.

Conclusion: ChimerDB 3.1 update includes almost twice as many patients reflecting more diverse types of cancer. We also added 460

fusion breakpoints in the ChimerKB via manual curation. ChimerDB 3.1 is available at <http://ercsb.ewha.ac.kr/fusiongene/>.

A44

Functional characterization of long non-coding RNAs in iPSCs using CRISPRi based genetic perturbation studies

Divya Mundackal Sivaraman¹, Yip Chi Wai¹, Youtaro Shibayama¹, Joachim Luginbühl¹, Tsukasa Kouno¹, Chung Hon Chan², Saumya Agarwal², Michiel De Hoon², Piero Carninci³, Jae Woo Shin¹

¹Cell Conversion Technology Team, Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Saitama 351-0198, Japan;

²Genomics Algorithms Development Unit, Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Saitama 351-0198, Japan; ³Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Saitama 351-0198, Japan

Correspondence: Divya Mundackal Sivaraman (divya.sivaraman@riken.jp)

Human Genomics 2018, **12**(Suppl 1):A44

The FANTOM projects have shown that mammalian genome is pervasively transcribed which largely include long non-coding RNAs (lncRNAs) and majority of them with no characterized function. lncRNAs are generally key regulatory elements and few of them are reported to be involved in different cellular processes. However, their functional role is still elusive. Here, we investigated the role of lncRNAs in maintaining pluripotency and neural differentiation in human induced pluripotent stem cells (hiPSCs). Based on the dynamic expression patterns of hiPSC differentiation to neurons, candidate lncRNAs were selected for functional perturbation. We used different knockdown strategies such as anti-sense oligos (ASOs) and clustered regularly interspaced short palindromic repeats interference (CRISPRi) to perturbate lncRNAs in human iPSCs. We evaluated the effect of knockdown by qPCR and single molecule RNA- fluorescence in situ hybridization (FISH). Knocking down of selected lncRNAs showed downregulation of pluripotency marker gene Oct3/4, indicating a role in stem cell maintenance and neuronal differentiation. Analysis of chromosome conformation capture (HiC) data from hiPSCs further underscored direct interaction of selected lncRNA with genes involved in pluripotency network. Mechanistic role of the lncRNA is being queried by RNA Immunoprecipitation sequencing (RIP-seq) and Chromatin Isolation by RNA purification sequencing (ChIRP-seq). Additionally, we are also performing molecular phenotyping of these lncRNA perturbations by gene expression profiling by CAGE to reveal the dynamic regulatory functions involved in the maintenance of pluripotency and neural differentiation propensities. The nature of different perturbation technologies has also been discussed, which exhibits its own strengths and weaknesses.

A45

Functional study of peptidylarginine deiminase type4 as a genetic risk factor for rheumatoid arthritis

Akari Suzuki¹, Yuta Kochi¹, Hirofumi Shoda², Keishi Fujio² and Kazuhiko Yamamoto^{1,2}

¹Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, 1-7-22 Suehiro, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan; ²Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence: Akari Suzuki

Human Genomics 2018, **12**(Suppl 1):A45

Background

Rheumatoid arthritis (RA) is one of the most common autoimmune diseases, which affects approximately 1% of the world population. RA is a chronic systemic inflammatory disease characterized by the

inflammation of synovial joint tissues. Previously, we identified peptidylarginine deiminase type 4 (PADI4) as a susceptibility gene for RA by genome-wide association studies (1, 2). PADI4 is highly expressed in immune cells, such as bone marrow, macrophages, neutrophils, and monocytes. Peptidyl-citrulline is an important molecule in RA because it is a target antigen of anti-citrullinated peptide antibodies (ACPAs), and only PADs (translated protein from PADI genes) can provide peptidyl citrulline via modification of protein substrates. The aim of this study was to evaluate the importance of the PADI genes in the progression of RA.

Materials and Methods

We generated Padi4 knockout (Padi4^{-/-}) DBA1J mice. Padi4^{-/-} DBA1J and wild-type mice were immunized with bovine type II collagen (CII) to develop collagen-induced arthritis (CIA). We compared the incidence and severity score, and performed measurements of expression levels of various inflammatory cytokines and Padi genes in immune cells by real-time TaqMan assay. Also cytokine concentration and CII antibodies in sera were measured by enzyme-linked immunosorbent assay. We investigated PAD4 effect on the transcriptional pattern of macrophage from Padi4^{-/-} mice by microarray.

Results

We demonstrated that the clinical disease score was significantly decreased in Padi4^{-/-} mice and Padi4 expression was induced by CII immunization. In Padi4^{-/-} mice sera, serum anti-type II collagen (CII) IgM, IgG, and inflammatory cytokine levels were also significantly decreased compared with those in wild-type mice sera. Interestingly, Padi2 expression was compensationally induced in CD11b⁺ cells of Padi4^{-/-} mice (3). We also identified Fus gene as a significant gene between WT and Padi4^{-/-} mice by independent two microarray tests.

Conclusions

On the basis of these studies, it appears that Padi4 enhances collagen-initiated inflammatory responses. Our results revealed that PAD4 affected on expression of various cytokines and also controlled Padi genes and Fus gene.

Acknowledgments

We are grateful to all members of the Laboratory for Autoimmune Diseases for their expert technical assistance

References

1. Suzuki, A, R. Yamada, X. Chang, S. Tokuhiko, T. Sawada, M. Suzuki, M. Nagasaki, M. Nakayama-Hamada, R. Kawaida, M. Ono, M. Ohtsuki, H. Furukawa, S. Yoshino, M. Yukioka, S. Tohma, T. Matsubara, S. Wakitani, R. Teshima, Y. Nishioka, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, Y. Nakamura, and K. Yamamoto. 2003. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 2003; 34:395-402.
2. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A et al. Rheumatoid arthritis contributes to biology and drug discovery, et al. *Nature*. 2014;506:376-81.
3. Suzuki A, Kochi Y, Shoda H, Seri Y, Fujio K, Sawada T, Yamada R and Yamamoto K. Decreased Severity of Experimental Autoimmune Arthritis in Peptidylarginine Deiminase Type 4 Knockout Mice. *BMC Musculoskeletal Disorders*, 2016; 17; 205.

A46

CRISPR/Cas9 genome editing to create mouse models for human deafness

Mor Borodnik-Cohen¹, Nada Danial-Faran^{1,2}, Ofer Yizhar-Barnea¹, Yael Noy¹, Tal Koffler¹, Stavit Shalev², Rebecca Haffner³, Karen B. Avraham¹

¹Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; ²Genetics Institute, Ha'Emek Medical Center, Afula, Israel and Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; ³Department of Veterinary Resources, Weizmann Institute of Science, Rehovot, Israel

Correspondence: Karen B. Avraham

Human Genomics 2018, **12**(Suppl 1):A46

Pathogenic variants in almost one hundred genes lead to hereditary hearing loss in humans worldwide. Due to the inaccessibility of the human inner ear, and the lack of an inner ear cell line to study the pathogenesis of deafness, the mouse has been a critical model for deciphering the mechanisms of hearing loss. Using the CRISPR/Cas9 genome editing tool, we have created several mouse models representing pathogenic variants associated with deafness in the Israeli Jewish and Palestinian Arab populations. These include *Atoh1*, a transcription factor essential for the development of cerebellar neurons and generation of inner ear hair cells, and *Slc25a21*, a gene that encodes the 2-oxoadipate mitochondrial carrier (ODC) that transports C5-C7 oxodocarboxylates across inner mitochondrial membranes. The non-syndromic hearing loss in the families ranges from congenital to childhood onset, and in severity from severe to profound. We designed sgRNAs directing the Cas9 to cut at the mutation sites, and oligonucleotides containing the mutation, in order to facilitate homologous recombination. These were injected into mouse zygotes, along with Cas9 RNA. Both heterozygote and homozygote founder mice were obtained and mated further to create lines for further experiments. Auditory brainstem response (ABR) to evaluate hearing, behavioral tests to determine vestibular function, and scanning electron microscopy (SEM) of inner ears to examine the morphology of the hair cells were performed. If the hearing loss found in each family will be mirrored in the CRISPR/Cas9 mice, it will verify these variants as the cause of deafness. Moreover, these mice will provide a relevant tool for studying the pathogenesis and mechanisms of deafness for each gene. Finally, the Crispr/Cas9 editing tool may further be used to rescue hearing loss in mutant mice.

A47

Functional annotation of very long intergenic (vlinc) RNAs using systems biology approaches reveals strong association with cancer

Huifen Cao, Fan Gao, Lorena Salazar Garcia, DingDing Cai, Huifang Wang, Ye Cai, Xueer Han, Jinling Tang, Philipp Kapranov
Institute of Genomics, Huaqiao University, 668 Jimei Road, Xiamen, China 361021

Correspondence: Philipp Kapranov (philippk08@hotmail.com)
Human Genomics 2018, **12(Suppl 1):A47**

Despite growing appreciation of the importance of long non-coding RNAs (lncRNAs) in the progression and metastasis of many malignancies, knowledge of cancer-related lncRNAs remains limited. We have recently discovered a widespread class of vlincRNAs in human cells. These transcripts with minimal length of 50kb and often reaching 100's of kb's represent over 2000 different loci and cover at least 10% of human genome. A subset of these transcripts regulated by endogenous retroviral (ERV) promoters appear to function in early embryonic development and cancer. Despite being a prominent class of human transcripts, most vlincRNAs remain functionally uncharacterized. Here, we present a robust functional annotation system that combines experimental and computational methods and aims to globally annotate all vlincRNA in human genome. Firstly, we integrated multiple RNAseq expression datasets from public sources and our in-house database to identify the expression patterns of vlincRNAs. Secondly, based on the co-expression network among vlincRNAs and known-genes, multiple functional enrichment tools were integrated to annotate the vlincRNAs. Finally, the system is constantly modified and improved by experimental validation to build a robust annotation platform for vlincRNAs. One of the major conclusions was that a very high proportion of vlincRNAs, not only the ones regulated by ERVs, participate in cancer-related pathways. This annotation system will significantly contribute to our understanding of molecular mechanisms of

lncRNAs functioning and provide an important resource for new potential prognostic biomarkers and therapeutic targets in the field of cancer.

A48

Genomic characterization of biliary tract cancers identifies their driver genes, cell-of-origin, and predisposing mutations

Hidewaki Nakagawa¹, Christopher Wardell¹, Masashi Fujita¹, Toru Yamada², Michele Simbolo³, Matteo Fassan³, Rosa Karlic⁴, Paz Polak⁵, Masakazu Yamamoto⁶, Hiroki Yamaue⁷, Kazuaki Chayama⁸, Satoru Miyano⁹, Gad Getz², Aldo Scarpa³, Satoshi Hirano², and Toru Nakamura²
¹Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan; ²Department of Gastroenterological Surgery II, Hokkaido University Graduate School of Medicine, Sapporo, Japan; ³Department of Pathology and Diagnostics, University of Verona, Verona, Italy; ⁴Department of Molecular Biology, Division of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia; ⁵Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; ⁶Department of Gastroenterological Surgery, Tokyo Women's Medical University, Tokyo, Japan; ⁷Second Department of Surgery, Wakayama Medical University, Wakayama, Japan; ⁸Department of Medicine & Molecular Science, Hiroshima University School of Medicine, Hiroshima, Japan; ⁹Laboratory of DNA Information Analysis, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Human Genomics 2018, **12(Suppl 1):A48**

Background & Aims: Biliary tract cancer (BTC) or cholangiocarcinoma is a rare cancer worldwide, but prevalent in some areas including Japan, where a specific risk factor of environmental exposure is involved in BTC development such as chronic inflammation and chemical exposure. The genetic features of BTC remain poorly understood and the molecular profiles of BTCs are as heterogeneous as their pathology and biology, making large sample sizes necessary for comprehensive analysis and understanding of its molecular carcinogenesis and clinical associations. We performed large scale genome sequencing analyses on BTCs to investigate their somatic and germline driver events and characterize their genomic landscape.

Methods: We analyzed 412 BTC samples from Japanese and Italian populations, with 107 whole exome sequencing (WES), 39 whole genome sequencing (WGS), and targeted sequencing of a further 266 samples. The BTC subtypes were 136 intrahepatic cholangiocarcinomas (ICCs), 101 distal cholangiocarcinomas (DCCs), 109 peri-hilar types (PHCs), and 66 gallbladder or cystic duct cancers (GBCs/CDCs). We identified somatic alterations and searched for driver genes in BTCs, and found pathogenic germline variants of cancer-predisposing genes. We predicted cell-of-origin for BTCs by combining somatic mutation patterns and epigenetic features.

Results: We identified 32 significantly and commonly mutated genes including *TP53*, *KRAS*, *SMAD4*, *NF1*, *ARID1A*, *PBRM1*, and *ATR*, and *FGFR2* fusion was detected in 7% of ICCs. Strong negative effects on overall survival were observed in patients harboring mutations in *ARID1A* ($P=0.0011$) and *KRAS* ($P=0.0042$). Smokers had significantly worse prognosis ($P=4 \times 10^{-3}$). A novel focal deletion of *MUC17* at *7q22.1* affected patient prognosis ($P=3 \times 10^{-4}$), which was confirmed by immunohistochemical analysis on the tissue microarrays. The cell-of-origin (COO) of a cancer can be determined by comparing the genomic distribution of mutations to the chromatin organization of specific cell types, as mutations are more likely to occur in open, transcriptionally active chromatin. COO predictions using WGS data of BTCs and liver cancers and 424 epigenetic features by the Epigenome Roadmap suggest hepatocyte-origin of hepatitis-related ICCs. Deleterious germline mutations of cancer-predisposing genes such

as *BRCA1*, *BRCA2*, *RAD51D*, *MLH1*, or *MSH2* were detected in 11% of BTC patients, indicating that BTC is a tumor related with Lynch syndrome and HBOC.

Conclusions: BTCs have distinct genetic features including somatic events and germline predisposition. These findings could be useful to establish treatment and diagnostic strategies for BTCs based on genetic information.

A49

Improved prediction of functional annotation of genes

Rezvan Ehsani¹, Finn Drabløs²

¹University of Zabol, Zabol, Iran; ²NTNU – Norwegian University of Science and Technology, Trondheim, Norway
Human Genomics 2018, **12**(Suppl 1):A49

A large number of genes do not have useful annotation, meaning that we do have sufficiently reliable information about which processes the gene products may be involved in. This is in particular the case for non-coding RNA (ncRNA) genes, like the long ncRNA genes (lncRNA), where most of the known genes are without any functional annotation. It is therefore of great interest to be able to predict function of un-annotated genes.

One possible approach is to use “guilt by association”, i.e., to identify well-annotated genes that seem to be involved in similar processes as a given un-annotated gene. This can for example be co-expression, indicating potential co-regulation. It is then possible to predict function of the un-annotated gene using information from the well-annotated co-expressed genes. This has been tested in previous implementations, but with somewhat limited success. In particular, the lncRNAs are challenging because they can have a very cell type-specific expression pattern, which may bias co-expression analyses.

Here we present an improved approach for function prediction. We use several measures for estimating co-expression, including statistical (Pearson, Spearman) and geometrical ones (Sobolev, Riemannian). This gives improved identification of true co-expression. We then use an enrichment analysis to identify enriched GO-terms in the co-expressed gene set, and use this to predict GO terms for the un-annotated gene.

We have tested this approach on subsets of well-annotated protein coding genes, using a leave-one-out procedure. For each gene the GO terms were predicted (without using the known terms), and the fit between predicted and known GO terms was measured using semantic similarity measures, in particular GOSemSim and TopoCSim. This showed good correlation between predicted and known GO terms, in particular for terms related to biological process (BP). The procedure was also tested on lncRNAs, in particular a set of five well-described lncRNAs tested in previous publications. The predicted GO-terms showed good correspondence with published functional descriptions of these lncRNAs.

This shows that it is possible to predict the function of both protein-coding and ncRNA genes, given a reliable set of expression data.

A50

Integrative genome and epigenome analysis of gene regulatory regions in human lung cancer cells isolated by laser-micro-dissection

Masafumi Muratani¹, Shingo Sakashita², Naomi Kaneko¹, Masayuki Noguchi²

¹Department of Genome Biology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, 305-8575, Japan; ²Department of Diagnostic Pathology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, 305-8575, Japan

Correspondence: Masafumi Muratani (muratani@md.tsukuba.ac.jp)
Human Genomics 2018, **12**(Suppl 1):A50

Understanding of genome regulation in native human cancer cells is severely limited due to availability of fresh tumor sample and tissue heterogeneity. Previously, we analyzed gastric cancer tissue samples and discovered cryptic promoter activation events which derive expression of important developmental genes [1]. It indicated that ectopic expression of lineage-specific factors could contribute to epigenetic reprogramming of normal cells in cancer development.

However, it was unclear if such cancer associated epigenetic changes occurred in cancer cells or other cell types in bulk tissue sample.

In this study, we developed a workflow for integrative chromatin immunoprecipitation-sequencing (ChIP-seq) and RNA-sequencing (RNA-seq) analysis of clinical tissue samples in combination with isolation of specific cell population by laser-microdissection microscopy (LMD). As a pilot study, we applied this workflow to analysis of 6 squamous cell and adeno carcinoma cases in which matched normal lung tissue, stromal and tumor cell parts can be isolated. ChIP-seq analyses of trimethyl histone H3 Lys4 (H3K4me3), acetyl histone H3 Lys27 (H3K27ac) and tri-methyl histone H3 Lys27 (H3K27me3) mark indicated cancer-associated activation of non-canonical promoter regions in lineage-specific transcription factors, driving expression of alternative forms of RNAs. These observations were further supported by strand-specific RNA-seq analysis of larger number of cases. These cryptic promoter sites are also frequently found in polycomb complex 2 binding sites as previously found in gastric cancer study. Overall, these results demonstrated that ectopic induction of key developmental regulators by cryptic promoter activation was common features in multiple types of human cancers.

To further study potential mechanisms of cancer-associated promoter and enhancer regulation, we analyzed single-nucleotide polymorphisms (SNPs) and somatic mutations. Our results indicated that majority of cryptic promoter activation events were not directly driven by somatic mutations. However, we could find many cancer-associated regulatory elements showing allele-specific histone modifications. This result indicated influence of genetic background in enhancer and promoter regulation and presence of functional SNPs.

From the perspectives of human genome research, LMD-based isolation of clinical cell samples and integrative ChIP-seq and RNA-seq approach could be highly useful for broader range of clinical studies as it allows effective filtering of candidate functional SNPs based on the accurate positioning of functional regulatory regions and allele-specific chromatin regulation.

Ethics Approval

This study was approved by University of Tsukuba's Ethics Board, approval numbers 1593-1 and H29-052.

Reference

1. Muratani M et al. Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements. *Nat Commun.* 2014; 5: 4361.

A51

Increased switch dynamics of *in vitro* synthesized mRNA using modified rNTP nucleosides

Callum Parr¹, Becki Y Kuang², Hirohide Saito², Jay Shin¹

¹Cell Conversion Technology Team, Division of Genomic Technologies Center for Life Science Technologies, RIKEN, Yokohama, Japan; Center for iPS Cell Research & Application, Kyoto University, Kyoto, Japan

Correspondence: Callum Parr (callum.parr@riken.jp)
Human Genomics 2018, **12**(Suppl 1):A51

mRNA-based switches (e.g., RNP L7Ae switches, and miRNA switches) and the gene circuits that they build are gaining in popularity in various applications due to the advantages of using RNA-only delivery, and the dynamic responsiveness of post-transcriptional based-switches. However, the necessity of using modified rUTP and rCTP in the exogenous mRNAs to evade immune surveillance comes at the price of decreased translational output and dynamic range between OFF and ON states of the encoded switch. This is likely resulting from the altered interaction with RNA binding proteins, including those of the ribosomal complex, and altered hybridization kinetics between complementary RNA strands. The relative low dynamic range compared to pDNA or replicon encoded circuits, means alternative modifications are highly sought after. We found a single uridine substitute modification can restore protein translation to that of unmodified exogenous mRNA, further still the same modification outperformed pseudouridine in respects to immune surveillance evasion. Finally, we tested modified rNTPs combinations and found an optimal modification that could provide higher dynamic range equal to that of

unmodified mRNA. Using miRNA switches to detect cell-type specific miRNA activity; this modification provides greater resolution in separating different cell types.

In future projects, we will further investigate the influence of various types of rNTP modifications on mRNA biogenesis using various 'omic' technologies, and RNA-targeted nuclease-dead CRISPR systems to detect and modify RNA modifications at single-base resolution. Using such systems will provide causative rather than correlative evidence of a particular type of RNA modification to RNA biogenesis and various biological processes.

A52

FANTOM6: Functional screening of long non-coding RNA for pluripotency regulation

Chi-Wai Yip, Chung-Chau Hon, Divya Sivaraman, Kayoko Yasuzawa, Youtaro Shibayama, Joachim Luginbuehl, Michiel de Hoon, Harukazu Suzuki, Masayoshi Itoh, Naoto Kondo, Takeya Kasukawa, Piero Carninci, Jay W. Shin
Division of Genomic Technologies, Center for Life Science Technologies, RIKEN, Yokohama, Japan

Human Genomics 2018, **12(Suppl 1):A52**

The discovering of extensive transcription of long non-coding RNA (lncRNA) and the studies on their functional roles suggested their involvement in gene regulation in different physiological processes, including pluripotency safeguarding and differentiation induction. This study aims to identify and characterize the lncRNAs, which play regulatory roles in pluripotency and differentiation.

Different categories of lncRNA including intergenic non-coding RNA (lincRNA), promoter upstream transcripts (PROMPTs) and enhancer lncRNA (eRNA) were selected for knockdown (KD) in the induced pluripotent stem (iPS) cells. For each lncRNA target (n=390), 5 or 10 LNA gapmeR anti-sense oligo (ASO) were designed to suppress the target specifically. A high-throughput screening method adopting lipofection in 96-well culture plate and real-time imaging on each well was applied to 2285 ASO KD in duplication. RNA was isolated at 48 hr post-transfection, followed by QPCR using 3 pairs of primers for each target. Targets with at least two ASO showing sufficient KD level were considered as successful KD (43%, n=166). Targets with two ASO showing KD and growth rate change revealed by real time imaging were considered as growth phenotype (8%, n=31). The KD efficiency was shown to correlate with cellular localization enrichment and exosome sensitivity. Among the successful KD targets, 123 of them were selected for transcriptomic analysis by low-quantity CAGE to unveil the underlying molecular phenotypes, focusing on self-renewal efficiency and differentiation potential.

The increasing number of lncRNAs being reported to regulate pluripotency remains limited. This study will identify key factors in stem cell pluripotency and differentiation, providing clues to improve reprogramming and cell conversion.

Keywords:

Long non-coding RNA

Pluripotency

Induced pluripotent stem cells

High-throughput screening

A53

Genetic landscape of Slovenians: past admixture and natural selection pattern

Damjan Glavač¹, Pierpaolo Maisano Delser^{2,3}, Massimo Mezzavilla^{4,5}

¹Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, Ljubljana, Slovenia; ²Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland; ³Department of Zoology, University of Cambridge, Cambridge, UK; ⁴Division of Experimental Genetics, Sidra Medical and Research Center, Doha, Qatar; ⁵IRCCS Burlo Garofolo, Trieste, Italy

Correspondence: Damjan Glavač (damjan.glavac@mf.uni-lj.si)

Human Genomics 2018, **12(Suppl 1):A53**

Background

The Slovenian territory played a crucial role in the past serving as gateway for several human migrations. Previous studies used

Slovenians as a source population to interpret different demographic events happened in Europe but not much is known about the genetic background and the demographic history of this population. Therefore, our aims are i) a detailed characterisation of the genetic structure of Slovenians in a broader European context using both Y chromosome and autosomal data, ii) a description of the past and present admixture pattern and iii) a survey of variants putatively under selection and associated with different traits/diseases.

Materials and Methods

Overall, 96 samples range from Slovenian littoral to Lower Styria were genotyped for 713,599 markers using the OmniExpress 24-V1 BeadChips. The Slovenian dataset has been subsequently merged with the Human Origin dataset (REF) for a total of 2163 individuals. Only population with a minimum samples size of 10 individuals were retained and related individuals were discarded. We also generated a dataset containing ancient genomes form.

Results

Y chromosome diversity splits into two major haplogroups R1b and R1a with the latter suggesting a genetic contribution from the steppe. Slovenian individuals are more closely related to Northern and Eastern European populations than Southern European populations even though they are geographically closer. This pattern is confirmed by an admixture and clustering analysis. We also identified a single stream of admixture events between the Slovenians with Sardinians and Russians around ~2630 BCE (2149-3112). We found a significant admixture event between the Yamnaya and the early Neolithic Hungarians dated around ~1762 BCE (1099-2426) suggesting a strong contribution from the steppe to the foundation of the observed genetic diversity in modern Slovenians. Our selection study reveals significant hits on markers associated mainly on lipid traits and eye pigmentation when compared to South Europeans such as HRC2 for blue eye colour and FADS1-FADS2 alleles, responsible for blue eye colour and synthesis of long-chain unsaturated fatty acids respectively.

Conclusions

Using our approach we found several genes linked to diet that could explain possible selection forces on these alleles that were introduced from different sources in the Slovenian genetic pool. Populations closely related to Yamnaya and early neolithic Hungarians contributed during the Bronze age to the foundation of the observed genetic variability in modern Slovenians. This means that also disease or specific traits alleles were introduced in the Slovenian genetic pool during this period, such as pigmentation alleles, lactose tolerance and immune genes. When the comparison was done with Eastern Europeans, we discovered significant signals in PKD2L1 and IL6R which are genes associated with taste and coronary artery disease.

A54

Diagnostic utility of clinical exome sequencing private genetics referral center of Thailand

Pawares Chitayanan¹, Nareenart Iemwimangsa², Bhakbhoon Panthan², Insee Sensorn², Sommon Klumsathian², Angkana Charoenyingwattana², Wasun Chantratita², Objoon Trachoo^{2,3}

¹Panthupark Genetics Clinic, Bangkok 10110, Thailand; ²Center for Medical Genomics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand; ³Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand

Human Genomics 2018, **12(Suppl 1):A54**

Abstract

Clinical Exome Sequencing (CES) carries the potential to facilitate the identification of disease causing gene. The test is not popular in developing countries due to the costly investment and requires time and well-trained medical staffs for pre-test genetic counseling. Currently, majority of Thai CES are performed by granting research in medical schools belonging to government sector. Here we presented the situation of CES in private sector where self-paid patients came for receiving the service. Fifty-two cases were referred to our private laboratory during January 2014–December 2017 from private hospitals around Thailand. The reasons for referral were categorized relied on the phenotype evaluated by the physician-in-charge: 1)

recognizable phenotype, 2) unrecognized and heterogeneous phenotype, 3) unaffected patient with strong family history of certain phenotype. Diagnostic utility of pathogenic and likely pathogenic variants was detected 100% in patients with known genetic syndrome by clinical diagnosis such as neurofibromatosis, tuberous sclerosis and Marfan syndrome. The second most diagnostic yield was 60% in patients referred for neurology phenotype. For unrecognizable phenotype, which was the distinctive group that required CES for gene discovery, we were able to detect pathogenic/likely pathogenic variant in 42% amongst referral samples. Unaffected patient with strong family history of certain phenotype such as cancer or sudden cardiac death could be detected only 1 in 3 cases. Hence, diagnostic utility of CES is maximal in cases affected by known genetic syndrome and has the power to increase the genetics diagnosis in patients with either recognizable or unrecognizable phenotype.

Key words: clinical exome sequencing, diagnostic utility, private genetics center, Thailand

A55

Construction of large scale gene regulatory networks

Chen Chen, Min Zhang, Dabao Zhang

Department of Statistics, Purdue University, West Lafayette, IN, USA

Correspondence: Dabao Zhang (zhangdb@purdue.edu)

Human Genomics 2018, **12(Suppl 1):A55**

Background

Constructing gene regulatory networks is crucial to understanding the genetic architecture of complex traits. While most of the gene regulatory network constructions only take gene expression information, here we propose using additional genomic polymorphism data to facilitate the network construction.

Materials and methods

Taking advantage of both gene expression and genomic data, we propose a two-stage penalized least squares (2SPLS) method to build large systems of structural equations for network construction. The system can be constructed using large numbers of endogenous and exogenous variables at the first stage, followed by consistent selection of regulatory effects at the second stage.

We conducted simulation studies to compare our method with the adaptive lasso (AL) based algorithm [1] and the sparsity-aware maximum likelihood (SML) algorithm [2]. Both acyclic and cyclic networks were simulated with each including 300 endogenous variables. On average, each endogenous variable has one regulatory effect for sparse networks and three regulatory effects for dense networks. The effects were simulated from a uniform distribution $(-1, -0.5) \cup (0.5, 1)$.

Results

For acyclic networks, 2SPLS has the greatest power when the sample size is small ($n=100$) and all three methods have similar performance when the sample size is large ($n=1000$). In addition, 2SPLS controls FDR under 20% while SML has low FDR (5%) for sparse acyclic networks with large sample size and large FDR (40%) for small sample size. Overall both 2SPLS and SML have better performance than AL in terms of FDR.

For cyclic networks, 2SPLS has greater power than SML and AL for all sample sizes and lower FDR when the sample size is not large. SML has similar power as 2SPLS for sparse cyclic networks, though its power is much lower than that of 2SPLS for dense cyclic networks. SML has much higher FDR for inferring dense networks when sample size is small while the FDR is small when the sample size is large.

Conclusions

The method is computationally efficient and allows for parallel implementation. We demonstrate the superior performance of the method with computer-based simulation studies. In addition, the method was applied to yeast data with both genomic and gene expression information.

References

1. Logsdon B, and Mezey J. Gene expression network reconstruction by convex feature selection when incorporating genetic perturbations. *PLoS Comput Biol.* 2010; 6(12): e1001014.
2. Cai X, Bazerque A, and Giannakis G. Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations. *PLoS Comput Biol.* 2013; 9(5): e1003068.

A56

Mitochondrial DNA analysis in Mexican women with breast cancer

Silvia Jiménez-Morales¹, Carlos Jhovani Pérez-Amado², Hugo Tovar-Romero³, Antonio Maffuz-Azziz⁴, Verónica Bautista-Piña V¹, Sergio Rodríguez-Cuevas⁴, Fredy Beltrán-Anaya¹, Alberto Cedro-Tanda¹, Luis Alfaro-Ruiz¹, Alfredo Hidalgo-Miranda¹

¹Laboratorio de Genómica del Cáncer, Instituto Nacional de Medicina Genómica, Ciudad de Mexico, 14610, Mexico; ²Programa de Maestría en Ciencias Bioquímicas, Universidad Nacional Autónoma de México, Ciudad de Mexico, 04510, Mexico; ³Unidad de Servicios Bioinformáticos, Instituto Nacional de Medicina Genómica, Ciudad de Mexico, 04980, Mexico; ⁴Instituto de Enfermedades de la Mama FUCAM, Ciudad de Mexico, 14610, Mexico

Correspondence: Silvia Jiménez-Morales (sjimenez@inmegen.gob.mx)

Human Genomics 2018, **12(Suppl 1):A56**

Keywords. Breast cancer, mitochondrial DNA mutations, Mexican, haplogroups

Background

It has been demonstrated that mitochondrial DNA (mtDNA) polymorphisms and mutations could contribute to cancer development and alter the metastatic behavior of cancer cells. To know the landscape of the mtDNA mutations in Mexican women with breast cancer we included cases collected from the Instituto de Enfermedades de la Mama FUCAM (Mexico).

Material and methods

Diagnosis and molecular classification were established using the current international histopathological criteria and microarray expression profiling (PAM50). None of the patients received chemotherapy or radiotherapy treatment before the surgery. Two overlapping primer sets were used to sequence the entire mtDNA from paired peripheral blood and tumor samples using MiSeq System. The study was approved by Ethics Committees of the participant institutions.

Results

Invasive ductal carcinoma was the predominant tumor type (74.2%) and luminal A molecular subtype (43.9%). The number of mtDNA germline mutations showed considerable heterogeneity across patients, ranged from 0 to 45 ($\bar{x} = 31$). Haplogroups A (44.7%), B (21.3%), C (14.9%) and D (8.5%) were the commonest and the remainder percentage (10.6%) were constituted by H, J, I and N haplotypes. We detected a total of 139 somatic mutations being single nucleotide substitution (SNS) more frequent than indels (94% vs 6%, respectively). Overall somatic mutations, seventy six (55.3%) were located in the coding region of which, 30 (39.5%) were no-synonyms and 46 (60.5%) were synonymous. Eight reported mutations were highly frequent (>10%) in tumor samples, and to our knowledge no one of these have been associated with breast cancer disease. The C8794T (H90Y) ATP6 is predicted to have an effect on protein

function, was present in 15% of the tumors. As well, seven very rare mutations, A6544G (N214S, *CO1*), G9838A (G211D, *CO3*), T11571C (M271T, *ND4*) G13333A (A333T, *ND5*), G14972A (G76S, *CYB*), A15328T (T194T, *CYB*) and G15612A (G289E, *CYB*) have not been reported previously. *CYB* and *ATP6* were more susceptible to mutation than other coding genes in tumor samples. Otherwise, *Her2* positive tumors presented the highest number of mutations in comparison with the remainder molecular subtypes.

Conclusions

According to our findings, the distribution of mtDNA germline mutations and mtDNA haplogroups in these patients is similar to that reported in healthy Mexican population. Additionally, a high heterogeneity in the mtDNA mutations among patients with breast cancer was observed. This study constitutes the first report where the complete sequences of the mtDNA have been analyzed in Mexican women with breast cancer.

Acknowledgements

This study was funded by Consejo Nacional de Ciencia y Tecnología (Grant. S0008-2016-1-0272618), Mexico and National Institute of Genomic Medicine (Grant. C2_19/2009), Mexico.

A57

A global view of the genetic architecture in human complex traits

Kyoko Watanabe¹, Sven Stringer¹, Tinca Polderman¹, Danielle Posthuma^{1,2}
¹Department of Complex Trait Genetics, VU University Amsterdam, Amsterdam, 1081 HV, The Netherlands; ²Department of Clinical Genetics, VU Medical Center, Amsterdam, 1081 HV, The Netherlands

Correspondence: Kyoko Watanabe (k.watanabe@vu.nl)

Human Genomics 2018, **12(Suppl 1):A57**

Background

Despite more than a decade of genome-wide association studies (GWAS), the proportion of segregating variants that contribute to variation of complex traits, the degree of polygenicity and the extent of pleiotropy are still largely unknown. Here we systematically analysed 3,795 GWAS summary statistics for 2,824 traits in 28 trait domains to (1) investigate the nature of the genetic architecture across hundreds of traits, and (2) chart the extent of pleiotropy on the single variant, gene and gene -set level. Findings are discussed in the context of the recently proposed omnigenic model [1].

Materials and methods

We compiled a database of 3,795 GWAS results for 2,824 unique traits across 28 domains. All trait names were manually synchronized, and GWAS results were checked using a standardized quality control pipeline. For within-trait comparisons, we used all GWAS results available per trait. For cross-trait comparisons, we used the largest available GWAS per trait, with at least 50,000 individuals.

Results

Within-trait analyses showed that generally, with increasing sample size, the number of detected loci and SNP-based heritability increases. However, a wide variety exists across traits, with some traits showing a plateau and others displaying a decrease in SNP-based heritability, most likely due to including heterogeneous phenotypes. Cross-trait analyses showed that 22% of detected lead SNPs have a minor allele frequency < 1%, yet there is large variety across traits. The majority of lead SNPs across all traits are in non-coding regions, while more pleiotropic SNPs have a higher probability of being coding SNPs. SNP-based heritability across all traits is not evenly distributed across the genome, showing enrichment in specific chromosomes, independent from physical length. The proportion of risk loci across all traits covers more than half the genome. Pleiotropy is

ubiquitous, 58% of all coding genes are associated with at least one trait and 45% of genes are pleiotropic. SNPs and genes that are more pleiotropic are less tissue-specific, and the size of gene sets associated with a single trait is significantly smaller than the size of pleiotropic gene sets, supporting an omnigenic model [1]. Genetic correlations between traits show widespread overlap, yet generally cluster within trait domains.

Conclusions

We show widespread variation in genetic architecture and widespread pleiotropy across hundreds of complex traits. Our results provide novel insights into how genetic variation contributes to trait variation, backed by a vast amount of empirical results.

Reference

- Boyle A, Li Y, Pritchard J. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017; 169:1177-1186.

A58

Variants at the enhancer-promoter interactions of human primary adipocytes drive adipose gene expression

David Pan^{1,2}, Kristina Garske¹, Arthur Ko^{1,3}, James Boockvar¹, Chelsea K. Raulerson⁴, Karen L. Mohlke⁴, Markku Laakso⁵, Päivi Pajukanta^{1,2,3}

¹Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA; ²Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA 90095, USA; ³Molecular Biology Institute at UCLA, Los Angeles, CA 90095, USA; ⁴Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; ⁵Department of Medicine, University of Eastern Finland, Kuopio 70210, Finland

Human Genomics 2018, **12(Suppl 1):A58**

Obesity is a global serious health epidemic and main risk factor for type 2 diabetes and non-alcoholic fatty liver disease. Obesity promotes hypertrophic dysfunctional adipocytes that predispose to ectopic storage of fat in the liver and epicardium, causing cardiometabolic disorders. The regulatory molecular mechanisms favoring intact adipose function are poorly understood in humans, preventing biology-based treatments of obesity. Our goal was to elucidate the regulation of human adipose transcriptomes to further the understanding of adipose biology. To this end, we first identified enhancer-promoter interactions in primary human white adipocytes, using promoter Capture Hi-C, to identify chromosomal regions regulating adipocyte gene expression. We then performed a *cis*-expression quantitative trait locus (eQTL) analysis using subcutaneous adipose RNA-sequence data from the METSIM and GTEx cohorts (total n=612) to identify variants regulating adipose gene expression. Finally, by overlapping the interaction and *cis*-eQTL data, we localized those adipose *cis*-eQTL variants that reside within the enhancer-promoter interactions in primary human adipocytes. These interactions were enriched for enhancer (H3K4me1, H3K4me3, and H3K27ac) and repressor (H3K27me3, H3K9me3) histone marks. Using public DNase I hypersensitive site (DHS) data and the HOMER software, we also found that the DHSs within the adipocyte chromatin interactions are significantly enriched for 26 transcription factor (TF) motifs (FDR<5%), including multiple key TFs in adipose biology, such as CEBPB and PPARG. To investigate whether the adipose *cis*-eQTL variants in the enhancer-promoter interactions explain a significant amount of local gene expression, we partitioned the heritability of local gene expression from human adipose tissue to multiple chromosomal categories using a modified version of the partitioned LD Score regression method. Our data show that open chromatin regions (i.e. DHSs) within these adipocyte chromosomal interactions

are significantly enriched in the heritability of local gene expression (p -value<0.002) and overall, the adipose *cis*-eQTL variants in the enhancer-promoter interacting DHS regions explain a significant portion (4.6%) of the heritability of local adipose gene expression even though they only contain 0.23% of variants genome-wide. Taken together, our integrative genomics data identify functionally highly relevant regulatory variants for local gene regulation and adipose biology. Our future studies are targeted to build regulatory looping units for each adipose expressed gene and associate the pooled set of variants within these regulatory units with obesity and related clinical traits in large human cohorts.

A59

Promoter-level expression atlas of skeletal muscle atrophy and recovery

Guzel Gazizova¹, Islam Nigmatzyanov¹, Oksana Tyapkina², Ruslan Deviatiiarov¹, Oleg Gusev^{1,3}

¹Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Tatarstan, Russia; ²Kazan Institute of Biochemistry and Biophysics, Kazan Scientific Center of the Russian Academy of Sciences, Kazan, Tatarstan, Russia; ³KFU-RIKEN Translational Genomics Unit, RIKEN, Yokohama, Kanagawa, Japan

Correspondence: Guzel Gazizova (rgazizova@gmail.com)

Human Genomics 2018, **12**(Suppl 1):A59

Background

Human life is often accompanied by a variety of physical inactivity: a low-active lifestyle, bed rest, limb immobilization, or zero-gravity conditions for astronauts. Muscle disuse is associated with a muscle atrophy, which leads to negative consequences for the locomotor system in mammals. To identify mechanisms underlying pathological changes in muscular system, complex molecular studies are required, thus, we aimed to study transcriptional changes in the muscles of rats during atrophy and subsequent recovery.

Materials and methods

Two types of muscles, "slow" (m. Soleus) and "fast" (m. EDL), were examined in rats in normal conditions, after 1, 3 and 7 days of anti-orthostatic hanging and after subsequent 1, 3 and 7 days of recovery using CAGE (Cap Analysis of Gene Expression) method followed Illumina HiSeq 2500 sequencing. After quality check and filtration, CAGE reads were mapped to the current rat genome assembly rn6 (2014) by using bwa and then clustered by python script. Further annotation of CAGE peaks, analysis of differential expression, and functional terms enrichment were proceeded through R environment. CAGE signal visualization for our experiments was done in Zenbu browser. The study was approved by the Kazan Federal University Animal Care and Use Committee guidelines (Permit Number 2, dated 5 May, 2015).

Results

We succeeded sequencing of CAGE libraries with depth >10M reads and mapping ratio ~80%. 9971 expressed CAGE clusters and 5766 associated genes in two muscle types were determined. Differential expression of genes and their promoter activity were strongly varied in m. EDL and m. soleus within hanging-recovery time course: "slow" m. soleus has no significant changes of transcriptional activity up to 7 days of hanging, but drastically shifted during recovery, while "fast" m. EDL shows quick and stable response to the stress and fast recovery after placing in normal conditions. The highest number of differentially expressed promoters and genes in m. soleus was found in the first day of recovery (~3k and 1.5k respectively) with strongly enriched oxidation-reduction processes associated with mitochondrion. In m. EDL the strongest early response to hanging was 10 times lower compared to changes in m. soleus. Besides, in functional terms, activated genes were mostly related to the muscle structure organization.

Conclusions

This study provides the first systematic annotation of promoters landscape and genes activated in "fast" and "slow" muscle types under induced atrophy and following recovery in rats. Our first results are consistent with known physiological effects of muscle disuse and might be used in further atrophy studies.

Acknowledgements

The reported study was funded by RFBR according to the research project № 17-00-00243 and is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

A60

Identification of population-stratified coding variants for low and high serum triglycerides in Mexicans

Arthur Ko^{1,2}, Elina Nikkola¹, Kristina Garske¹, Hortensia Moreno^{3,4}, Alicia Huerta-Chagoya^{3,4}, Teresa Tusie-Luna^{3,4}, Carlos A. Aguilar-Salinas³, Päivi Pajukanta^{1,2,5}

¹Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; ²Molecular Biology Institute at UCLA, Los Angeles, CA, USA; ³Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; ⁴Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico; ⁵Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA

Correspondence: Päivi Pajukanta

Human Genomics 2018, **12**(Suppl 1):A60

Hypertriglyceridemia is an independent risk factor for coronary heart disease, and a major constituent of multiple metabolic disorders. Hypertriglyceridemia is especially prevalent in populations with Amerindian background such as Mexicans. However, Latinos have been grossly underrepresented in genomic studies and they are unlikely to benefit from the current push for precision medicine unless genetic studies become more inclusive. Furthermore, the different haplotype structure and allele frequencies (AF) in admixed populations can provide better mapping resolution than homogenous groups, and rare risk variants in minority groups can shed light on the biological mechanism of disease. To this end, we investigated loci that we previously associated with lipid traits in Mexicans and showed to harbor variants with different frequencies between Latinos and Europeans. To assay low and rare frequency (AF>0.5%) coding variants of 28 genes that reside within these loci in Hispanics, we performed targeted and exome sequencing in 4,550 Mexicans. We focused on loss of function (LoF) and missense single-nucleotide variants (SNVs) that are more likely to have large effects, and then performed single-variant and gene-based association tests on hypertriglyceridemia (serum triglycerides (TGs) >200mg/dL and <150mg/dL for cases and controls, respectively). Overall, 89 SNVs were tested, and at the single-variant level, 10 SNVs passed the Bonferroni corrected P <0.05 threshold. Notably, all 10 SNVs were stratified in Latinos when compared to Europeans (AF difference >2X); 6 of the 10 were highly significant in the recent European exome-wide analysis (n >300,000), emphasizing the power of studying population-stratified regions in smaller diverse populations ascertained for the phenotype of interest. We also identified a new Mexican-specific TG risk variant, rs144966144, which is a missense variant (AF=5.3% in cases and AF=3.4% in controls) within *ZNF259*. Its frequency in Latinos is 3.7% (based on the gnomAD database), but occurs at much lower frequency in other populations (<0.5%). A protective LoF, rs328, in *LPL*, a key TG gene, is depleted in Mexicans versus Europeans (AF=4.7% versus 10.0%). This may explain some of the burden of hypertriglyceridemia in Mexicans and be clinically relevant given the new therapeutic interest in genetically lowered TGs by *LPL* variants. Of the 9 genes with ≥ 2 low frequency variants (0.5%<AF<5.0%) in our gene-based analysis, 7 genes, including known lipid genes

such as *LPL*, *PCSK7*, and *APOA4*, are highly significantly associated with TGs (Bonferroni $P < 0.05$). Overall, our study highlights the value of investigating diverse populations to identify risk variants in human disease.

A61

Functional analysis of sequence and structural features of SINEUP effector domain, a new class of positive post-transcriptional gene regulator non-coding RNAs

Harshita Sharma¹, Hazuki Takahashi¹, Kazuhiro Nitta¹, Silvia Zucchelli^{2,3}, Stefano Gustincich^{2,4}, Piero Carninci^{1,5}

¹RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, Japan; ²Area of Neuroscience, Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy;

³Department of Health Sciences, Università del Piemonte Orientale, Novara, Italy; ⁴Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy; ⁵TransSINE Technologies, Yokohama, Kanagawa, Japan

Correspondence: Piero Carninci (carninci@riken.jp)

Human Genomics 2018, **12(Suppl 1):A61**

Abstract

SINEUPs had introduced a new class of regulatory long non-coding RNAs (lncRNAs) which can bind to a target mRNA in a sequence-specific manner and increase its protein expression without altering the target transcript level [1,2]. This phenomenon was initially reported in a mouse lncRNA antisense (AS) to *Uchl1* (Ubiquitin C-Terminal Hydrolase L1) gene which upon binding to sense mRNA facilitated its association with polysomes and enhanced UCHL1 protein production [1]. This translation up-regulatory function was found to be due to an inverted SINEB2 repeat of B3 sub-family embedded at the 3' end of AS *Uchl1*. Thus, SINEUPs constitute two important regions- i) Binding domain (BD), the region overlapping to sequence within 5' UTR and few bases of CDS of target protein-coding mRNA; ii) Effector domain (ED), an inverted repeat of SINEB2 essential for SINEUP function [2]. In recent years, SINEUPs had proven their efficacy against a variety of exogenous and endogenous targets in a wide range of human and mouse cell lines and also in *in-vivo* system, still the functional features of SINEUPs and the mechanism of action are rather elusive [1,2,3]. In current study, we examined key sequence and structural features of SINEUP ED to get a grasp of underlying SINEUP biology.

Previously, we short-listed more than 30 natural antisense transcripts (NATs) in mouse FANTOM3 cDNA dataset with similar genomic features to that of AS *Uchl1* [1]. Here, we report the SINEUP effect of a number of diverse SINEB2 sequences isolated from some of these NATs which were tested in synthetic SINEUP-GFP in HEK293T cells. We found that not only the B3 sub-family but other sub-families of mouse SINEB2 also display the SINEUP effect suggesting vastness of SINEUP class. We also investigated combinatory and synergistic/additive effect of SINEB2 repeats over SINEUP function. As it is known that functions of various RNA classes are highly dependent on their structures, we further analyzed structure of SINEUPs in human cells using *in-vivo* click selective 2'-hydroxyl acylation and profiling experiment (icSHAPE). Thereby, we report structural domains crucial for SINEUP function and present consensus sequence and secondary structure model for SINEUP RNA class. Moreover, we show phylogenetic clustering of different SINEB2 repeats based on their sequences and RNA secondary structures to evaluate connection of SINE evolution and SINEUP function.

References

1. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012; 491(7424):454-457.
2. Zucchelli S, Fasolo F, Russo R, Cimatti L, Patrucco L, Takahashi H, et al. SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Front Cell Neurosci*. 2015; 9:174.
3. Indrieri A, Grimaldi C, Zucchelli S, Tammaro R, Gustincich S, Franco B. Synthetic long non-coding RNAs [SINEUPs] rescue defective gene expression *in vivo*. *Sci Rep*. 2016, 6:27315.

Acknowledgements

This study is supported by the Research Grant from MEXT to the RIKEN Center for Life Science Technologies and the AMED grant.

A62

Gene editing of the multi-copy H2A.B gene family by a single pair of TALENs in mice

Nur Diana Anuar¹, Matt Field^{1, 2}, Sebastian Kurscheid¹, Lei Zhang³, Edward Rebar³, Philip Gregory^{3,4}, Josephine Bowles⁵, Peter Koopman⁵, David J. Tremethick¹ and Tatiana Soboleva¹

¹The John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; ²Current address: James Cook University, PO Box 6811, Cairns, QLD 4870, Australia; ³Sangamo Therapeutics, 501 Canal Blvd, Richmond, CA 94804, USA; ⁴Current address: bluebird bio, 60 Binney St, Cambridge, MA 02142, USA; ⁵Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia

Correspondence: David J. Tremethick (David.Tremethick@anu.edu.au); Tatiana Soboleva (Tatiana.Soboleva@anu.edu.au)

Human Genomics 2018, **12(Suppl 1):A62**

Background

In view of the controversy related to the generation of off-target mutations by gene editing approaches, we tested the specificity of TALENs by disrupting a multi-copy gene family using only one pair of TALENs. We show here that TALENs do display a high level of specificity by knocking out simultaneously the function of the three genes (H2Afb3, Gm14920, H2Afb2, which are over 92% identical) that encode for H2A.B.3. This represents the first described knockout of this histone variant.

Results

We designed and validated 19 TALEN pairs; group 1 consisted of 12 specific to the H2Afb3 gene and group 2 (7 TALENs) targeting all three H2A.B.3 genes. Activity and specificity of TALENs were tested employing a Dual-Luciferase Single Strand Annealing Assay (DLSSA) and by Cel1 cleavage assay, confirming specificity of group 1 TALENs to only cleaving H2Afb3, while group 2 showed nuclease activity against all 3 H2A.B.3 encoding genes. One pair of group 2 was chosen to knock out the function of all three H2A.B.3 genes in mice. In total, we obtained 9 out of 19 pups which showed the desired KO. Female pups were used as founders for continued breeding with WT males, resulting in a pure non-mosaic mouse colony by G3. We then used a combination of exome sequencing and molecular techniques to investigate the presence of off-target effects. Single males from three consecutive generations (G1-G3) had their exomes sequenced and combined SNV and InDel analysis was performed. Off-target effects in the founder generation would lead to a "dilution" of InDels in each consecutive generation. This dilution effect was not evident in either exome sequencing data. Despite the lack of H2A.B.3, KO

mice are able to reproduce, but do so with much smaller litter sizes, an indicator that H2A.B.3 is required for fertility.

Conclusions

We successfully applied TALEN technology using one pair of TALENs to specifically and simultaneously disrupt three gene copies of a gene family. Many human diseases are caused by gene copy number variations and therefore the use of a limited number of genome editing tools might help in the future in the treatment of such diseases. The use of only one pair of TALENs, rather than multiple pairs, reduces the complexity of the TALEN approach (and hence the cost), would reduce the likelihood of mosaicism and most importantly, reduce the possibility of off-target mutations.

A63

Promoter-level heart transcriptome for ventricular remodeling studies

Ruslan Deviatiiarov¹, Ksenia Zaripova¹, Aaron Koppel², Christian Miccile², Grace Bakalag², Oleg Gusev^{1,3}, Igor R Efimov²

¹Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia; ²Department of Biomedical Engineering, George Washington University, Washington, DC, USA; ³Division of Genomic Technologies, Center for Life Science Technologies (CLST), RIKEN, Yokohama, Japan

Correspondence: Ruslan Deviatiiarov (ruselusabus@gmail.com)

Human Genomics 2018, **12(Suppl 1):A63**

Background

Pathological left ventricular (LV) remodeling followed by size, shape and functional changes and usually affected by aortic stenosis, myocardial infarction, hypertension, or valvular heart disease. Generally, it occurs in up to 15% of population and associated with older age, male sex, diabetes. This remodeling includes myocyte hypertrophy, dilatation, fibrosis, and inflammation, which increase the risks for further ischemia, cardiomyopathy and heart failure. High LV mass is a major complication in different heart diseases and has strong impact on longevity, but also an important factor affecting survival rate after surgery or transplantation. For example, patients with aortic stenosis and modified LV have 5 times lower survival rate, mortality rate after liver transplantation has 4 fold increase in patients with left ventricular hypertrophy. However, molecular mechanisms at DNA regulation level underlying maladaptive transition remain unclear mostly due to hearts availability.

Results

Our study is aimed to create a detailed map for human heart promoterome. It will include samples from separated left atrium (LA) and ventricle from male and female, sick and healthy patients of different age. At the time, by using CAGE (cap analysis of gene expression) sequencing of 15 samples we were able to identify 13100 heart promoters, where 12051 overlap promoters from FANTOM5 project and therefore could be annotated by cell type (8309) and related diseases (3158). A set of ion exchangers, connexins and other heart related genes (n=40) with known expression profiles (for example, KCNA5, KCNK3, GJD2, Irx4) were used for data validation. We defined LV and LA specific genes and showed that functional annotation of such groups is consistent with literature available data: LA contains "synapse", "neuron", "signaling", while LV - "sarcomere", "fiber", "heart process" terms. Enrichment analysis revealed that differentially expressed genes and promoters are responsible for diseases like ischemia, vascular disease, myopathy, aortic disease. *De novo* predicted heart enhancers (4388) have enriched cardiac differentiation, maturation and regulation motifs related to transcription factors like SRF, MEF2A, Foxk1.

Conclusions

These results indicate CAGE approach as a suitable tool for further heart studies on the promoter level and provides comprehensive and comparable data about molecular regulatory mechanisms in human heart. Next, we're planning to extend our project with aging time course and LV hypertrophy samples. Finally, it will be a part of the promoterome atlas project for mammalian muscles.

Ethics Approval

This study was approved by the ethical committee of UK National Research Ethics Service (07/H0715/101) and Washington University School of Medicine Ethics Committee.

A64

iGEM: A powerful statistical method integrating genotype and gene-expression data for dissecting genetic architecture of a disease

Sarmistha Das¹, Raghunath Chatterjee¹, Aditya Chatterjee², Partha P. Majumder¹ and Indranil Mukhopadhyay¹

¹Human Genetics Unit, Indian Statistical Institute, Kolkata, India;

²Department of Statistics, University of Calcutta, Kolkata, India

Correspondence: Indranil Mukhopadhyay (indranil@isical.ac.in, indranilm100@gmail.com)

Human Genomics 2018, **12(Suppl 1):A64**

Background

To decipher the genetic architecture of a human disease, various types of omics data are generated. Two common omics data types are (a) genotypes at a large number of marker loci, and (b) expression levels for a large number of genes, often generated at the genome scale using microarray. Although traditionally, data generated by various omics platform were analyzed separately, nowadays data integration and joint statistical analysis are emphasized to obtain robust inferences. But the problem in joint statistical analysis of multi-type omics data is that sample sizes of different data types vary. One reason is that RNA are less stable than DNA. Therefore, commonly in a disease genetic study genotype data are available from a larger number of individuals compared to data on gene-expression. Again, gene-expression assays are expensive, leading researchers to examine the genotype data first, arrive at a set of tentative inferences based on them, and then select a subset of individuals to be recruited for gene-expression assay, to analyze the specific research question in hand. Intuitively, the subset selected should carry the widest range of information subject to the research question but otherwise homogenous. Clearly, the gene-expression profile of the individuals plays no role in the subset selection criteria since, the selection is conducted before gene-expression assay. Thus, sample size variation poses a major challenge to the biologist community to do integrated analysis of omics data. This motivated us to develop a statistical method integrating the available gene-expression information with genotype data, generated in case-control study. Our method is entirely based on omics data collected from same individuals, and depends on no reference transcriptome data for imputation, that induces population-stratification-bias [1, 2].

Materials and Methods

We propose a novel two-step multi-locus association method using a latent variable, conjointly with logistic regression to integrate available genome-wide-marker information, with gene-expression data on a subset of individuals.

Results

We have derived asymptotic distribution of our test-statistic to calculate of p-value faster for real dataset. Our method detects novel single nucleotide polymorphisms (SNPs) associated with psoriasis along with already reported Genome wide association study (GWAS) SNPs. We

could identify these associated variants at a much smaller sample size compared to the published corresponding GWAS, using an additional gene-expression information of a subset of genotyped individuals.

Conclusion

Extensive simulations reveal that our method is robust and powerful. Based on a real dataset [dbGAP; phs000019.v1.p1], we could identify a few novel psoriasis-associated SNPs, that remain undetected by GWAS only.

References

1. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx B, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genet.* 2016; 48:245-252.
2. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV and Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nature Genet.* 2015; 47:1091-1098.

A65

Finding statistically significant marker combinations from genomic data for survival analysis

Raissa T. Relator¹, Aika Terada^{2,3,1}, Jun Sese^{1,4}

¹Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; ²PRESTO, Japan Science and Technology Agency, Saitama, Japan; ³Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan; ⁴AIST-Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory, Tokyo, Japan

Correspondence: Raissa T. Relator (rai.relator@aist.go.jp)

Human Genomics 2018, **12(Suppl 1)**:A65

Background

Analysis of genomic data can provide valuable information such as genetic markers for both personalized and preventive medicine. However, with high-dimensional data, large correction factors for multiple testing are usually required for exhaustive assessment of possible markers. Additionally, comprehensive evaluation of marker combinations, which may provide useful insight to complex diseases such as cancer, has also become infeasible. To address these challenges that discourage true novel discoveries from data, we developed an algorithm for finding statistically significant single and multiple-marker interactions that may affect survival of individuals.

Methods

We adapted the Limitless Arity Multiple-testing Procedure (LAMP), a previously developed p-value correction technique for statistical evaluation of combination of markers which employs significant pattern mining for finding the optimal correction factor. We extended LAMP by incorporating the log-rank test using a newly introduced theoretical lower bound of the log-rank p-value, making LAMP more suitable to handle time and censored information common in survival data.

Results

The proposed method was applied to The Cancer Genome Atlas (TCGA) mRNA expression data for urothelial bladder carcinoma (BLCA) and acute myeloid leukemia (LAML) with 129 samples and 19637 genes and 173 samples 18754 genes, respectively, with focus on the highly expressed genes. We limit our results to detected single and combination genes that are highly expressed in at least 10 individuals. The algorithm detected a total of 36 single and 2-gene combinations with statistically significant log-rank p-values for LAML, and 49 single and multiple-gene combinations of up to order 4 for BLCA. Identified LAML markers have a minimum unadjusted log-rank p-value of 4×10^{-08} , and 6×10^{-07} for BLCA. Among the selected genes are known oncogenes from the COSMIC Cancer Gene Consensus such as DCTN1 and CCDC6 in the LAML results, and EGFR and GMP5 in the BLCA results. Using the PPISURV tool from BioProfiling.de and the GSE32894 urothelial carcinoma data, we can verify the statistical significance of identified single and 2-gene combination markers from the BLCA data analysis. Out of 34 identified combinatorial markers also present in the GSE32894 data, 24 were statistically significant ($p < 0.05$) when considering genes expressed at both high and low levels.

Conclusion

The algorithm we present is especially advantageous in detecting multiple-gene combination markers and sets no theoretical limit on

the order of interaction. We expect that it can aid in the selection of potential prognostic markers for further clinical validation among large orders of candidate tests.

Reference

Terada, A., Okada-Hatakeyama, M., Tsuda, K., Sese, J.: Statistical significance of combinatorial regulations. *Proc. Natl. Acad. Sci. U.S.A.* 110(32), 12996–13001 (2013)

A66

A history of Middle East (ME) populations from a viewpoint of the integrated whole genome sequence data and its application to identification of ME – specific variant alleles for diabetes

Katsuhiko Mineta, Kosuke Goto, Takashi Gojobori

¹KAUST (King Abdullah University of Science and Technology), CBRC (Computational Bioscience Research Center), Thuwal, Saudi Arabia

Correspondence: Takashi Gojobori

Human Genomics 2018, **12(Suppl 1)**:A66

The draft sequence data of the whole human genome was released for the first time in 2000. Since then, genomic sequence analyses of human populations have been intensively conducted. In particular, the following three international projects of human genome sequencing are good examples for large-scale studies of human genome variations:

- 1) HapMap (1,417 individuals)
- 2) Human Genome Diversity Project (940 individuals)
- 3) 1000 genomes (2,504 individuals)

However, the human genome sequence data are not readily compared with each other because of the essential differences in data format and annotation. If we can integrate all the three data sets into a single volume of data, we should be able to conduct a more detailed analysis of human genome sequence variation. The simple addition of individual samples leads to a total number of 4,861 individuals (= 1,417+940+2,504 individuals). Moreover, though genomic data of the Middle East (ME) populations is limited, ME could be a key to understand human history because of its unique location as a crossroad of Asia, Europe and Africa. With the aim of elucidating an evolutionary history of ME populations, we have developed the computational tools of integrating these three human genome sequence data sets including ME populations into a single volume of data. Using those tools, we successfully integrated these data. Then, we constructed a phylogenetic tree of about 5,000 human individuals at the genome sequence level. As a result, we identified evolutionary clusters of the ME populations in relation to other major ethnic groups, with very interesting features. Here, we report the outcome of this kind of big data analyses with successful data integration, discussing evolutionary significance of human genomic variations. We also present how to identify the ME – specific variant alleles for particular diseases such as diabetes, using the Clinvar database.

A67

Assessment of a *GLO1* frameshift variant, c.365delC, in Japanese cases of schizophrenia and controls

Kanako Ishizuka, Branko Aleksic, Norio Ozaki

Department of Psychiatry, Nagoya University School of Medicine, Aichi, Japan

Correspondence: Kanako Ishizuka (ishizuka-k@med.nagoya-u.ac.jp)

Human Genomics 2018, **12(Suppl 1)**:A67

Enhanced carbonyl stress is closely related to neuropsychiatric disorders. The enzyme encoded by *GLO1* is responsible for the catalysis and formation of S-lactoyl-glutathione from methylglyoxal condensation and reduced glutathione. *GLO1* is localized to 6p21.2, where it is located emerged as a plausible autism-risk loci candidate. A common variant of *GLO1* (rs2736654) has been associated with autism. The reduced expression of *GLO1* has been detected in mood disorder patients. These findings strongly suggest that *GLO1* is a promising candidate risk gene for neuropsychiatric disorders because

neuropsychiatric disorders including autism, mood disorders and schizophrenia share genetic risk factors.

In this study, we focused on the association between rare genetic variants of *GLO1* and schizophrenia. First, we conducted exon-targeted resequencing of *GLO1* with next-generation sequencing technology in 370 Japanese schizophrenia cases. Three heterozygous variants with allele frequencies of $\leq 1\%$, including one variant in the 3'-UTR, one missense variant, and one frameshift variant, were identified. We regarded c.365delC, the frameshift variant, as functionally relevant. We then performed association analyses on 1,799 cases and 1,764 controls with this variant. The variant was found in four cases and eight controls. There was no statistically significant association between c.365delC and schizophrenia ($P=0.18$). This frameshift variant in *GLO1* might occur at near-polymorphic frequencies in Japanese populations, although further investigation by sample size expansion and biological analysis will be needed to exclude low-penetrance associations.

Ethics Approval

All procedures performed in this study involving human participants were approved by the Ethics Committee of the Nagoya University Graduate School of Medicine and conducted in accordance with the Helsinki declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from the participants and from the parents of the patients under 20 years old.

A68

Delivering personalized medicine through point-of-care exome sequencing and population-aware interpretation in the Middle East

Khalid A. Fakhro^{1,2}, Amal Robay², Juan Rodríguez-Flores³, Jason Mezey³, Ronald G. Crystal^{2,3}

¹Department of Human Genetics, Sidra Medicine, Doha, Qatar;

²Department of Genetic Medicine, Weill-Cornell Medicine in Qatar, Doha, Qatar; ³Department of Genetic Medicine, Weill-Cornell Medicine, New York, NY, USA

Correspondence: Khalid A. Fakhro

Human Genomics 2018, **12(Suppl 1)**:A68

The proliferation of next generation sequencing (NGS) into clinical research settings is critical for personalized medicine. As numbers and types of diseases studied increase, the demand for accurate and reproducible variant-interpretation pipelines has also surged. In the Middle East, variant interpretation is complicated by the significantly elevated fraction of novel or low-frequency alleles not seen before, due to the lack of large shared datasets of population-matched controls. These considerations suggest population-aware pipelines would be required to deliver personalized medicine at the point-of-care. To demonstrate this, we use whole exome sequencing (WES) to assess 48 patients with rare diseases from 30 families recruited as part of a Mendelian Disease pilot program in Qatar – a nation where multiple stakeholders are building expertise and capacity in genomic medicine. We enroll entire families (parents plus additional siblings), and develop population-aware pipelines to discover both *de novo* and recessive putative disease variants. Leveraging 1,370 Qatari sequenced controls – which we share in a public database – we demonstrate the impact of large population-specific allele resources in discriminating disease-causing from likely-benign variation. We show that for a given subject, even in the absence of sequenced family members, variant prioritization is improved up to 8-fold when our population-matched controls are considered, versus when other large public databases (e.g. dbSNP, ExAC, 1000 Genomes, gnomAD) are used, stressing the poor representation of Arabs in these and the importance of data sharing for personalized medicine. Importantly, we identify

candidate disease variants in 21 of 30 families. In 81% of these, we discover novel pathogenic variants in known disease genes, pointing to significant allelic heterogeneity and founder mutations in Arabs. For another 6 of 30 families, we discover variants in genes reported to cause only part of the observed clinical presentation, suggesting possible phenotypic-expansion of reported syndromes. Our results demonstrate the high diagnostic utility of WES in Arab populations and the dramatic improvement in variant prioritization conferred by using a large set of population-matched controls – which we share to improve personalized medicine in the Middle East.

Ethics statement

The study was approved by the Institutional Review Board of Weill Cornell Medicine in Qatar.

A69

Evaluation of tumor microenvironment by expression-based surrogate markers

Yi Huang, An Hsu, Pei-Yi Lin, Wan-Ting Huang, Hua-Chien Chen, Shu-Jen Chen

ACT Genomics Co, Ltd., Taipei, Taiwan

Correspondence: Shu-Jen Chen (sjchen@actgenomics.com)

Human Genomics 2018, **12(Suppl 1)**:A69

In recent years, several studies have revealed the impact of immune contexture on clinical outcome in human tumors. In addition, using the immunoscore for colorectal cancer (CRC) could predict the patient survival more accurate than using microsatellite instability (MSI). However, a quantitative approach that could evaluate the tumor microenvironment (TME) of different types of solid tumors is still lacking. We develop a QPCR-based detection panel consisted of a set of immune-related genes and specific cancer biomarkers as surrogate markers to evaluate of TME. The CAGE profiles retrieved from FANTOM 5 were used to assess the capability of distinguishing tissue types by this panel of genes. Without any feature engineering method, the accuracy is more than 60% to separate the samples into the individual type of primary cultured cells and 97% for primary cultured immune and non-immune cells by a random forest model. Currently, we are optimizing the gene content based on feature importance and developing a prediction model based on the expression-signatures to estimate immune contextures by the mixture of pure immune and cancer cells. Before using the mixture of pure immune and cancer cells, we are going to establish a theoretical model by the CAGE profiles of primary cultured immune cells mixed with cancer cell lines. Our ultimate goal is applied our TME detection panel for the clinical usages in future.

A70

Making supportive guidance for gathering and sharing clinical and -omics data to develop next-generation medicine under the recent legislative efforts to protect personal information

Natsuko Yamamoto, Noriko Ohashi and Kazuto Kato

Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Osaka, 565-0871, Japan

Correspondence: Natsuko Yamamoto (nyamamot@eth.med.osaka-u.ac.jp)

Human Genomics 2018, **12(Suppl 1)**:A70

Background

Recently the Japan Agency for Medical Research and Development (AMED) started new database programs to collect genomic variation and disease phenotypes at the national level. It aims to share the data among researchers and develop high-level medical research in the area of cancer and rare diseases. Simultaneously, legislative

efforts towards protecting personal information have been underway in Japan and other countries. Our ELSI team, with funding from AMED, has been working on investigating these legal movements and preparing guidance to support researchers who want to share their data to advance research in the field.

Materials and methods

We carefully examined the Personal Information Protection Acts and related government guidelines. We contacted the Personal Information Protection Commission and discussed a variety of relevant legal issues. We also held an international meeting to survey comparable regulations in other countries, especially those in the United States and the United Kingdom. While preparing guidance, we discussed with researchers about ways and means for efficient data sharing which match the needs of their profession.

Results

We classified the laws and guidelines for different kinds of data and usage, and compiled them in a chart. For example, the data were categorized by the existence of consent of the subject and by data attributes such as identifiability of individuals. In general, data which was obtained with informed consent or data which did not contain individually-identifying information was allowed to be used without additional procedures when sharing the data. Omics data such as genomic data or transcriptomic data were difficult to classify, because they contain many types and there seems to be a variety of rules in different countries.

Conclusions

Usability of high-quality data is essential to innovation in medicine. Integration of clinical data and biological data such as individual genomic data is one way to create such high-quality data. However, as an individual's data becomes more deeply integrated, the potential risks in terms of privacy increase, which would in turn increase the difficulty of obtaining consent. New mechanisms which complement the present scheme of privacy protection—which mainly consist of informed consent and the de-identification of the data—will be needed for the safe gathering and sharing of high-quality data, which is necessary for the development of next-generation medicine.

A71

Promoter-level atlas of diversity of gene expression in the mammalian skeletal muscles: scope and milestones of the project

Guzel Gazizova¹, Ruslan Deviatiiarov¹, Islam Nigmatzyanov¹, Alistair Forrest², Roman Deev³, Yoshihide Hayashizaki⁴, Elena Shagimardanova¹, Oleg Gusev^{1,4,5,6}

¹Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Tatarstan, Russia; ²Systems Biology and Genomics, Harry Perkins Institute of Medical Research, Nedlands, WA, Australia; ³General Medicine Faculty, Ryazan State Medical University, Ryazan, Russia; ⁴RIKEN Preventive Medicine & Diagnosis Innovation Program, Cluster for Science and Technology Hub, Yokohama, Japan; ⁵KFU-RIKEN Translational Genomics Unit, RIKEN Innovation Center, RIKEN, Yokohama, Kanagawa, Japan; ⁶Division of Genomic Technologies, Center for Life Science Technologies, RIKEN, Yokohama, Japan

Correspondence: Oleg Gusev (oleg.gusev@riken.jp)

Human Genomics 2018, 12(Suppl 1):A71

Background

Skeletal muscle plays an important role in many vital processes and is maintained by numerous pathways regulating protein synthesis and degradation. Loss of skeletal muscle mass and strength induces by the various conditions, including muscle disuse resulted in bed rest or unloading, age-related muscle weakness (sarcopenia), cachexia, multiple muscular dystrophies, myopathies, denervation and etc. Noticeably, muscle atrophy can affect specific fiber types and different muscles have the differential susceptibility to muscle wasting. However, the factors underlying these differences remain to be elucidated.

Materials and Methods

In the frame of international consortium, we are aiming to reconstruct transcriptional network in variety of muscles in normal and atrophic conditions using Cap Analysis of Gene Expression (CAGE) and Small RNA expression analysis. For the analysis, we are selecting 20-30 target muscles according to several criteria. Firstly, we are

considering muscles which is frequently affected or not affected at all by disuse, sarcopenia, cachexia or muscle diseases. The second criterion is an availability of skeletal muscle to a biopsy. Since we are going to elicit age-dependent muscle atrophic changes, two or more age groups will be examined. Besides, male and female comparison will be carried out. It is assumed that as the research objects macaque, subset of human muscles and primary cell lines from patients with dystrophy will be used. As a result, the atlas of promoters and enhancers of RNA transcription in human and animal muscles in normal and pathological conditions will be created.

Results and Conclusion

The determination of the transcriptional factors and other genomic regulatory elements involved to differential response of muscle types may lead to a better understanding of the pathogenesis of the various muscle wasting conditions. Thus, it may help to find out new therapeutic targets for the specific disorders.

A72

Immuno-genomic subtype of liver cancer correlates with mechanisms of immune suppression and prognosis

Masashi Fujita¹, Seiya Imoto², Rui Yamaguchi³, Takanori Hasegawa², Shuto Hayashi³, Satoru Miyano^{2,3}, Hiroki Yamaue⁴, Kazuaki Chayama⁵, Hidewaki Nakagawa¹

¹Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan; ²Health Intelligence Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ³Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ⁴Second Department of Surgery, Wakayama Medical University, Wakayama, Japan; ⁵Department of Medicine & Molecular Science, Hiroshima University School of Medicine, Hiroshima, Japan
Human Genomics 2018, 12(Suppl 1):A72

Objectives

The recent clinical success of immune checkpoint inhibitors demonstrated an importance of understanding anti-tumor immunity, and it is required to explore the immunological and clinical significance of cancer genomics. We analyzed the transcriptomes and whole genomes of liver cancers in order to find immuno-genomic subtypes that provides mechanistic insights and prognostic power.

Methods

We have performed RNA-Seq of 234 liver cancers and 196 adjacent liver tissues, as well as whole-genome sequencing [1]. Cytolytic activity (CYT) within tissues was computed as the mean expression level of *GZMA* and *PRF1*. The immune cell fractions were computed by CIBERSORT on RNA-seq data. Neoantigens were predicted by NetMHCpan3.0 using somatic mutations profiles and HLA genotypes from whole genome sequencing.

Results

CYT in tumors was associated with better overall survival (hazard ratio (HR) = 0.46, 95% CI 0.27–0.78) (Fig. 1). In contrast, CYT in adjacent liver tissues was associated with poor disease-free survival (HR = 2.187, 95% CI 1.07–4.48), probably due to the development of meta-chronous tumors originated from chronic hepatitis. Tumor CYT was significantly lower than liver CYT ($p < 1 \times 10^{-15}$), indicating immune suppression by tumors. Liver fibrosis and the hepatitis C virus infection was associated with liver CYT but not with tumor CYT. A correlation between tumor CYT and liver CYT was positive but marginal (Pearson's correlation coefficient $r = 0.18$). Neither mutation burden nor the number of neoantigen was associated with tumor CYT ($p > 0.05$). To define immune subtype, we examined mRNA levels of immune suppressive genes. Some tumors had excessive expression of *IDO1*, *CSF2*, and *PTGS1* relative to tumor CYT and had poor prognosis, consistently with the involvement of myeloid-derived suppressor cells (MDSCs). *IL2RA* and *TNFRSF4* also showed similar trends, possibly reflecting regulatory T cells (Tregs). Independent estimation of Treg fraction using CIBERSORT was also associated with poor prognosis ($p < 0.01$). The absolute expression level of these genes did not have prognostic power, suggesting that the imbalance between positive and negative immune regulators is important for the immunity against liver cancers.

Conclusion

The local imbalance between cytotoxic lymphocytes, MDSC and Tregs could be used to define immune subtype of liver cancer. Association between the immune subtype and somatic alterations will be also discussed.

Reference

1. Fujimoto A, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet.* 2016; 48(5):500-509.

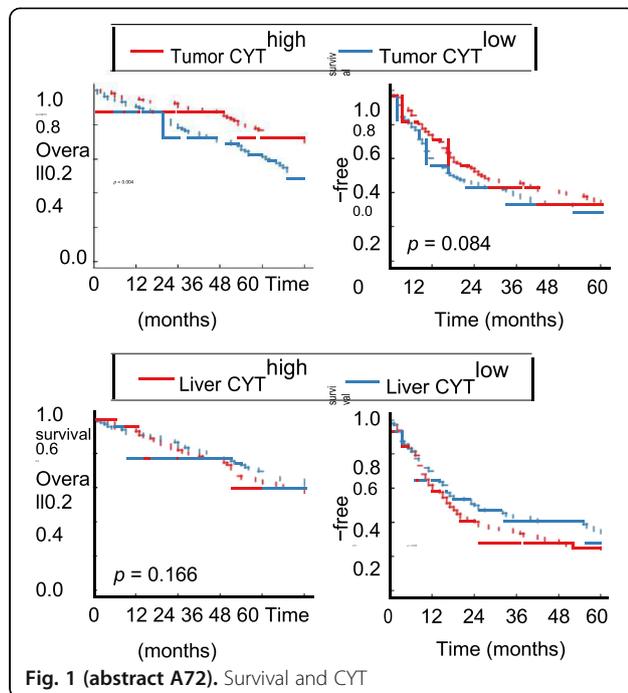


Fig. 1 (abstract A72). Survival and CYT

A73

Differences and similarities of international ELSI practices in genomic medicine

Hiroshi Toya, Kazuto kato
Department of Biomedical Ethics and Public Policy, Osaka University,
Osaka, 565-0871, Japan

Correspondence: Hiroshi Toya (hiroshi.toya@eth.med.osaka-u.ac.jp)

Human Genomics 2018, **12(Suppl 1)**:A73

Background

The concept of ELSI was developed originally with the human genome project in NIH and later spread around the world. Unlike traditional bioethics, ELSI is interdisciplinary field that studies the development of genomic medicine and considers its social implementation. However, during the last twenty years, the circumstances surrounding genomic medicine have changed drastically. International collaborative research is increasingly important, including data sharing and international biobanks. In order to respond to this situation, ELSI research must become international in its scope. One vital part of this is grasp how ELSI research is practiced in other countries and regions. Understanding the diversity of approaches in ELSI research will lay the groundwork for more fruitful collaboration between researchers around the globe.

Materials and methods

We surveyed major organizations from around the world, including National Human Genome Research Institute in US, Genome Canada in Canada, Economic and Social Research Council in UK, Netherlands Genomics Initiative in Nederland, ELSA Norway network in Norway, Australian Genomics Health Alliance in Australia and Japan Agency

for Medical Research and Development in Japan. Our primary method research was to survey the literature published by and about these institutions. We referred to papers and online information and compared mission statements, research subjects, and the achievements of each institution.

Results

As a result of the above examination, we were able to identify many of the differences and similarities of these institutions. One of our findings concerned the vast difference in the ELSI research agendas of different institutions, owing to the state of genetic research and the social conditions in the country. We were also able to confirm the importance of collaboration among foreign countries, as well as the benefits of cross-disciplinary programs between doctors and patients, experts and citizens, and the humanities and natural sciences

Conclusions

Based on our research, we conclude that, along with the development of genomic medicine, ELSI research must be globalized, taking into account common goals and values on the one hand, and respecting diversity on the other.

A74

Fine-mapping reveals complex genetic architecture underlying DNA methylation

V. Kartik Chandru¹, Riccardo E. Marioni^{2,3}, Ian J. Deary^{2,4}, Naomi R. Wray^{1,5}, Peter M. Visscher^{1,5}, Allan F. McRae¹

¹Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia; ²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK; ³Medical Genetics Section, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK; ⁴Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK; ⁵Queensland Brain Institute, University of Queensland, St Lucia, QLD 4072, Australia

Correspondence: Allan F. McRae (a.mcrae@uq.edu.au)

Human Genomics 2018, **12(Suppl 1)**:A74

Many fine-mapping methods have been developed that aim to select a small sets of variants underlying a genome-wide association peak with a high probability of containing the true causal variant. Due to a lack of real world traits with a known causal variant, these methods are primarily tested through simulations that may not accurately reflect the underlying biology of genetic associations. Recently, many genetic associations with DNA methylation levels have been identified, including the case where a SNP in the methylated CpG site disrupts DNA methylation. We used this rare case of a known causal variant to test three differing fine-mapping approaches – the J-test, BSLMM and BIMBAM – by constructing a 95% credible set for the causal genetic variant and testing for presence of the CpG site SNP. Using DNA methylation data from the Lothian Birth Cohorts of 1921 and 1936 (n=1366), we identified genetic associations with DNA methylation at 2266 CpG sites disrupted by a SNP with minor allele frequency greater than 10%. All three methods generated 95% credible sets containing the CpG site SNP in only ~50% of cases, compared to the expected 95% obtained from simulated data. The low coverage of the known causal variant in credible sets was replicated in an independent cohort (n=1886). A conditional analysis of the region surrounding the CpGs site determined 34% of methylation sites were associated with multiple independent genetic effects, invalidating assumptions made when attempting to fine-map a single causal variant. For those CpG sites with evidence for a single underlying genetic effect, the regional linkage disequilibrium structure showed a masking of the effect of the SNP at the CpG site. This occurs through the most associated SNP both affecting the methylation at the CpG site and having its allele associated with less methylation occurring at a higher frequency and in complete LD with the allele disrupting methylation the CpG site SNP. This masking was more evident in CpG islands, where the fine-mapping methods captured the CpG-SNP less often than in non-island regions (OR=1.6, $p=2 \times 10^{-3}$). The identified underlying genetic complexity when fine-mapping a

“trivial” case of SNPs within CpG sites altering DNA methylation has consequences for the fine-mapping of variants underlying higher order traits, particularly as many associations to complex traits are likely to act through effects on genomic regulation such as altered DNA methylation.

A75

Identification of chromosomal structural rearrangements by low-pass whole genome sequencing in recurrent miscarriage couples

Kwong Wai CHOY¹, Zirui DONG¹, Huilin WANG², Ye CAO³, Hui JIANG⁴, Sau Wai CHUENG⁵, Cynthia C. MORTON⁶

¹Department of Obstetrics & Gynaecology, The Chinese University of Hong Kong, Hong Kong; ²Bao'an Maternal-fetal Medicine Research Institute, Bao'an Women's and Children's Health, Hospital, Jinan University, Shenzhen, China; ³Baylor college of Medicine, Houston, TX, USA; ⁴BGI-Shenzhen, Shenzhen, China; ⁵Department of Molecular and Human Genetics, Baylor College of Medicine Houston, TX, USA; ⁶Brigham and Women's Hospital, Boston, MA, USA

Correspondence: Kwong Wai CHOY (richardchoy@cuhk.edu.hk)

Human Genomics 2018, 12(Suppl 1):A75

Background: Recent studies demonstrate that whole-genome sequencing (WGS) enables detection of cryptic rearrangements in apparently balanced chromosomal rearrangements (also known as balanced chromosomal abnormalities, BCAs) previously identified by conventional cytogenetic methods. In our pilot study, we demonstrated the feasibility of detecting BCAs using WGS data from the 1000 Genomes Project (<http://www.internationalgenome.org>) with our reported analytical tool without knowing which bands were affected. Here, we aimed to further evaluate the sensitivity and specificity of BCAs detection by low-pass WGS compared to conventional chromosome analysis.

Methods: Peripheral blood samples were collected from 100 couples with consecutive recurrent miscarriage (RM) in Hong Kong and we applied an optimized low-pass WGS approach to each DNA with sufficient physical coverage (8.25-fold) for BCA detection blinded to the results of the corresponding diagnostic cytogenetic study.

Results: With this approach, all the 100 RM couples were assessed by low-pass WGS, and seven balanced translocations were detected including one case not identified by chromosome analysis, but failed to detect two Robertsonian translocations. Cryptic deletions were found to involve the BCA breakpoints in two out of the seven reciprocal translocations. In addition, three inversions with size > 1 Mb and two likely pathogenic microdeletions/duplications were also detected among this group.

Conclusion: Our study demonstrates the extension of utilizing low-pass whole-genome sequencing for unbiased detection of BCAs including translocations and inversions in recurrent miscarriage couples.

Acknowledgements

This project is supported by the National Natural Science Foundation of China (No. 81370715, 81300075) and the Health and Medical Research Fund (HMRF, Project 04152666). C.C.M. acknowledges funding through NIH/NIGMS GM061354.

A76

Identifying chimeric and nonlinear RNAs from single-molecule CAGEScan data in cervical cancer cell lines

Sofia Khan¹, Anton Kratz², Kazunori Nagasaka^{3,4}, Charles Plessy⁵, Martin Frith^{1,6,7}

¹Computational Bio Big-Data Open Innovation Laboratory (CBBDO-IL), AIST, Tokyo, Japan; ²School of Medicine, University of California, San Diego, CA, USA; ³Department of Obstetrics and Gynecology, Graduate School of Medicine, Teikyo University, Tokyo, Japan; ⁴Department of Obstetrics and Gynecology, Graduate School of Medicine, The University of Tokyo, Japan; ⁵Division of Genomic Technologies, Center for Life Science Technologies, RIKEN, Yokohama, Japan; ⁶Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan; ⁷Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan

Correspondence: Sofia Khan

Human Genomics 2018, 12(Suppl 1):A76

Background

Nonlinear RNAs, defined here as transcripts with atypical exon order (e.g. trans-splicing or exon scrambling), have been identified in both tumor and normal cells. They have exciting potential as biomarkers or functional molecules, but they are not well characterized by typical transcriptome measurements. Similarly, the chimeric RNAs may serve as biomarkers and in the cervical cancer chimeric RNAs may have a viral component, because of the integration of the human papillomavirus (HPV) to the host genome that is characteristic for cervical cancer.

Materials and methods

Here, we used single-molecule CAGEScan data to characterize chimeric and nonlinear RNAs in cervical cancer cell lines (CaSki, HeLa, c33a and SiHa), normal immortalized keratinocyte cell line (HaCaT) and cervical epithelial cell line (W12E). This technique attaches unique molecular identifiers (UMIs), random sequences of bases, to the 5' end of each RNA, followed by PCR amplification, and paired-end sequencing of the 5' end together with a random downstream segment. The UMIs enable us to group read-pairs deriving from the same original molecule. We mapped the assembled reads (contigs) to both human and HPV genomes in order to identify putative chimeric transcript molecules originating from the viral integration to the host genome. The mapping allowing for *cis*- and *trans*-splicing enabled us to study the chimeric and nonlinear transcripts.

Results

To minimize false discoveries due to library construction and sequencing errors, we focused on the instances that were present in at least two replicate samples out of three replicate samples per each cell line (with the exception of W12E cell line for which we had two replicates). We identified multiple copies of one known chimeric transcript consisting of viral and human RNA: HPV16 + 13q22.1 in SiHa cell line, present in all the three replicate samples. We found several nonlinear transcripts both within one chromosome and across chromosomes.

Conclusions

Based on preliminary results it seems that it is possible to identify nonlinear and chimeric RNAs from single-molecule CAGEScan data.

A77

Genetic networks modulating cell fate specification and contributing to cardiac disease risk in hiPSC-derived cardiomyocytes at single cell resolution

Quan H. Nguyen¹, Samuel W. Lukowski¹, Han S. Chiu¹, Clayton E. Friedman¹, Anne Senabouth¹, Liam Crowhurst^{1,2}, Timothy J. C. Bruxner¹, Angelika N. Christ¹, James Hudson³, Jun Ding⁴, Ziv Bar-Joseph⁴, Patrick P L Tam⁵, Nathan J. Palpant¹, Joseph E. Powell^{1,6}

¹Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; ²School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia; ³School of Biomedical Sciences, The University of Queensland, Brisbane, QLD Australia 4072; ⁴Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA; ⁵School of Medical Sciences, Sydney Medical School, University of Sydney, NSW 2006 Australia; ⁶Queensland Brain Institute, University of Queensland, Brisbane, Australia

Correspondence: Quan H. Nguyen (quan.nguyen@uq.edu.au); Joseph E. Powell (joseph.powell@uq.edu.au)

Human Genomics 2018, 12(Suppl 1):A77

The majority of genetic loci underlying common disease risk act through changing gene expression, measured using bulk populations of mature cells. A crucial step that is missing is evidence of variation in the expression of genes as cells progress from a pluripotent to mature state. This is especially important for cardiovascular disease, as the majority of cardiac cells have limited properties for renewal post-neonatal. To investigate the dynamic changes in gene expression across the cardiac lineage, we generated RNA-sequencing data captured from 43,168 single cells progressing through in vitro cardiac-directed differentiation from pluripotency. We developed two single-cell transcriptomics analysis software, including a novel and generalized unsupervised cell clustering approach (Clustering at Optimal Resolution – CORE) and a machine learning method for

prediction of cell transition between clusters (single cell Global fate Potential of Subpopulations - scGPS). Using these methods, we were able to reconstruct the cell fate choices as cells transition from a pluripotent state to mature cardiomyocytes, uncovering intermediate cell populations that do not progress to maturity, and distinct cell trajectories that terminate in cardiomyocytes that differ in their contractile forces. These data reveal transcriptional networks underlying lineage derivation of mesoderm, definitive endoderm, vascular endothelium, cardiac precursors, and definitive cell types that comprise cardiomyocytes and a previously unrecognized cardiac outflow tract population. We identified the non-DNA binding homeodomain protein, HOPX, as functionally necessary for endothelial specification and dysregulation of HOPX during *in vitro* cardiac-directed differentiation underlies the molecular and physiological immaturity of stem cell-derived cardiomyocytes. Second, we identify new gene markers that denote lineage specification and demonstrate a substantial increase in their utility for cell identification over current pluripotent and cardiogenic markers. By integrating results from analysis of the single-cell lineage RNA-sequence data with population-based GWAS of cardiovascular disease and cardiac tissue eQTLs, we show that the pathogenicity of disease-associated genes is highly dynamic as cells transition across their developmental lineage, and exhibit variation between cell fate trajectories. Third, through the integration of single-cell RNA-sequence data with population-scale genetic data we have identified genes significantly altered at cell specification events providing insights into a context-dependent role in cardiovascular disease risk. This study provides insights into genetic networks governing lineage progression of hPSC cardiac directed differentiation by small molecule Wnt modulation, establishes a valuable data resource on cardiomyocyte differentiation, and presents novel analysis methods and software with broad applications.

A78

RISQ: A novel robust statistical approach for isoform quantification from RNA-seq data

Pronoy Kanti Mondal, Raghunath Chatterjee and Indranil Mukhopadhyay
Human Genetics Unit, Indian Statistical Institute, Kolkata, India

Correspondence: Indranil Mukhopadhyay (indranil@isical.ac.in, indranilm100@gmail.com)

Human Genomics 2018, **12(Suppl 1)**:A78

Background

Technological advances trigger the generation of massively parallel genome-wide transcriptome data, known as RNA-sequencing data. Correct quantification of transcripts is the first and foremost problem to analyze such data. However, it faces considerable challenges mainly due to problems in comprehending the distribution of reads and ambiguity in mapping reads to proper isoforms, thus leading to problems in modeling and estimation of transcript abundance. Moreover, estimation of isoform abundance using millions of reads needs to be done under very realistic conditions of the experiment. The reads that hold key information to this estimation problem may be of different types for single-end or pair-end reads.

Materials and Methods

In order to estimate this abundance, we have developed a statistical method for isoform level quantification based on maximum likelihood approach under very general conditions of the nature and distribution of reads. Our likelihood function is multinomial type with indicators as latent variables. We adopt EM algorithm to obtain exact estimates and we avoid approximations or plug-in estimates in maximizing the likelihood function unlike existing methods. Among other features, our proposed method is also robust to the assumption of the probability distribution of reads, thus requiring no prior knowledge about read distribution.

Results

We have studied our method extensively using simulated as well as real data. We did simulations under various models with different distribution of reads. Our method shows very promising result in comparison to other available methods [1,2,3,4,5,6]. It outperforms other methods significantly especially when the number of alternatively spliced isoforms is relatively large. It also shows superior

performance when some isoforms are extremely low abundant. Our method shows good correlation with other methods on real data especially with eXpress^[3] and Salmon^[4]. It shows high correlation with qRT-PCR estimates. The confidence intervals calculated using our method is narrower than Cufflinks [1] estimates.

Conclusion

We propose a novel statistical method for isoform level quantification from RNA-seq data. It surpasses other existing algorithms in terms of accuracy when number of alternately spliced isoforms is very large and also when some isoforms are extremely low abundant. It can work with mixture of paired and single end reads and it also produces confidence intervals. It is robust with respect to distribution of reads. Moreover, this method is scalable with respect to memory allocation and is computationally very fast, thus making it extremely useful and feasible approach in practical implementation with real data.

References

1. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 2010; 28(5):511-515.
2. Li B & Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics.* 2011; 12:323.
3. Roberts A & Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods.* 2013; 10(1)
4. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C. Salmon provides fast and bias aware quantification of transcript expression. *Nat. Methods.* 2017; 14(4): 417-419.
5. Patro R, Mount SM & Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* 2014; 32:462-464.
6. Bray NL, Pimentel H, Melsted P & Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 2016; 34:525-527.

A79

Pre-operative progesterone exposure benefits operable breast cancer patients by modulating surgical stress

Rohan Chaubal², Arindam Maitra¹, Nilesh Gardi², Amit Dutt², Sudeep Gupta², Rajendra A Badwe², Partha P Majumder¹, Priyanka Pandey¹

¹National Institute of Biomedical Genomics, Kalyani, West Bengal, India;

²Tata Memorial Centre/Hospital, Parel, Mumbai-400012, India

Correspondence: Sudeep Gupta; Rajendra A Badwe; Partha P Majumder; Priyanka Pandey

Human Genomics 2018, **12(Suppl 1)**:A79

Background

Operable breast cancer (OBC) patients undergoing surgery in the luteal, as opposed to the follicular, phase of the menstrual cycle were reportedly conferred with survival benefit. Clinically, a single dose of depot hydroxyprogesterone administered prior to surgery was shown to improve disease-free survival in OBC patients. We aimed to understand the molecular basis of the action of pre-operative hydroxyprogesterone in OBC patients.

Methods

Whole transcriptome sequencing (RNA-Seq) was performed on paired breast tumour samples that were obtained from 18 OBC patients who were exposed to hydroxyprogesterone before surgery and 13 patients who were unexposed to hydroxyprogesterone but only underwent surgery. RNA-Seq data were analysed using the Tuxedo protocol. Unpaired Student's t-test followed by Benjamini-Hochberg correction was performed for the differentially expressed genes between the two groups such that they qualified stringent filters in the exposed group. Significantly altered genes were evaluated for their functional annotation and some of these were validated using qPCR.

Results

Our study revealed 207 genes as significantly modulated due to hydroxyprogesterone. Of these, 142 genes were found up-regulated post-surgery among exposed patients, and down-regulated post-surgery among unexposed patients. Some of the major significantly enriched processes (with their genes forming an interaction network) included

(a) response to: progesterone, steroid hormone, drug, regulation of hormone levels, (b) negative regulation of: tumour necrosis factor production, nitric oxide biosynthetic process, inflammatory response, (c) cellular response to: osmotic stress; oxidative stress; external stimulus, radiation; zinc ion; reactive oxygen species, hypoxia, (d) protein targeting to: endoplasmic reticulum; mitochondria, protein oligomerization and (e) gene expression, nonsense-mediated decay, nucleosome assembly, protein-DNA complex assembly. These results suggest that cellular stress is modulated due to hydroxyprogesterone exposure. Network analysis predicted *UBC* as the central node suggesting a stress responsive behaviour due to hydroxyprogesterone in patients irrespective of their progesterone receptor status. A high correlation and a similar trend of expression level between RNA-Seq and qPCR assays were obtained for 10 out of 14 chosen genes. Validated genes included progesterone-modulated genes (*TSPO*, *PLN* and *CLDN4*).

Conclusions

Significantly modulated genes identified in our study provide an initial set of biological correlates to the finding of the mentioned clinical trial. Hydroxyprogesterone possibly primes tumours towards growth retardation after having sensitized them towards surgery-induced stress.

A80

Possible implications of REP522 repeat element family in cancer biology

Bogumil Kaczkowski, Kazuhide Watanabe, Piero Carninci
Division of Genomic Technologies (DGT), RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Correspondence: Bogumil Kaczkowski (bogumil.kaczkowski@riken.jp)
Human Genomics 2018, **12(Suppl 1):A80**

Background

REP522 is a small and elusive family of repeat elements with 368 copies in the genome. REP522 is an unclassified interspersed repeat of 1.8 Kb in size, with large palindrome of ~600nt in the center. Previously, we reported that many REP522 repeats harbor bidirectional promoters that are activated in multiple cancer types [1]. Such REP522 elements are silent in normal cells, but become active promoters in cancer cells and drive the expression of multiple long non-coding RNAs and pseudogenes. Later, we showed by integrating DNA methylation and histone modification data that REP522 is epigenetically activated in lung cancer by DNA hypomethylation and active promoter histone modification [2]. However, the activation mechanism and the role of REP522 in cancer remains unknown.

Materials and methods

We performed in silico analysis of REP522 sequences and explored their potential to act as promoter and/or exon sequences using latest transcriptome annotation data. We also analyzed genomic localization and clustering of REP522 elements on the genome. Additionally, we used RNA-Seq data from 21 tumor types profiled by The Cancer Genome Atlas (TCGA) and performed comprehensive gene activation and correlation analyses in 7916 primary tumors and 725 normal tissue controls.

Results

We show that the large palindrome of ~600nt in the center of REP522 repeats acts as a bidirectional promoter. Based on secondary structure prediction, the palindrome sequence can theoretically fold into a hairpin structure with ~300nt arm length, the implication of the predicted hairpin structure is unknown as the sequence is occupied by nucleosomes when given REP522 element is transcriptionally active in cancer.

Based on the RNA-seq data from TCGA, we show the frequency of REP522 activation across 21 cancer types and the correlations of expression of transcripts initiated from REP522 promoters.

Conclusions

The results from computational predictions and RNA-seq data analysis are encouraging and offer the first view into the role of REP522 in cancer. However, further studies, including wet-lab experiments and follow-up are required to elucidate the REP522 function and mechanism of activation in cancer.

References

1. Kaczkowski B, Tanaka Y, Kawaji H, Sandelin A, Andersson R, Itoh M, et al. Transcriptome Analysis of Recurrently Deregulated Genes across Multiple Cancers Identifies New Pan-Cancer Biomarkers. *Cancer Res. AACR*; 2016;76:216–26.
2. Horie M, Kaczkowski B, Ohshima M, Matsuzaki H, Noguchi S, Mikami Y, et al. Integrative CAGE and DNA Methylation Profiling Identify Epigenetically Regulated Genes in NSCLC. *Mol. Cancer Res. AACR*; 2017;15:molcanres.0191.2017–1365.

A81

Identification of disease-associated SNPs in a population based on data integration of Human Genome Diversity Project and ClinVar

Kosuke Goto, Katsuhiko Mineta, Takashi Gojobori
KAUST (King Abdullah University of Science and Technology), CBRC (Computational Bioscience Research Center), Thuwal, Saudi Arabia

Correspondence: Takashi Gojobori
Human Genomics 2018, **12(Suppl 1):A81**

Due to the environmental and genetic background, it is known that some diseases show the population-specific tendency. In this study we focused on the genetic background for Middle Eastern population. Cardiovascular diseases are one of major death factors and actually causing a high mortal rate in Middle East such as in Saudi Arabia. We developed a method to detect the SNPs responsible to the diseases in a particular population and apply it to the Middle Eastern population. First, to reveal the genetic background of the diseases in the Middle Eastern population, we integrated a genomic variation database and human health information. Although there are some projects about human genomic variation, we use Human Genome Diversity Project (HGDP) database because they contain genetic variations of the Middle Eastern population. Also, ClinVar provides information about genomic variation and human health. We integrated these two databases. Then, we performed statistical analysis such as Fisher's exact test to reveal SNPs significantly associated with a population of interest. As an example, we focused on cardiovascular diseases in the Middle Eastern population. We examined allele frequencies of variations and found promising SNPs that are related to genes; *CHRNA3*, *CACNA1C*, *CACNB2*, and *MYH6*. Interestingly, these genes code for proteins involved in the nicotinic acetylcholine receptor-signaling pathway. They are related to nicotine addiction, which contributes to the development of cardiovascular diseases. Considering higher smoking ratio in Middle East, our results are convincing and our methods worked well. We can apply this method to other populations and diseases for further studies and clinical applications.

A82

Elucidating broad functions of long noncoding RNAs in human cells

Jordan Ramilowski¹, Saumya Agrawal¹, Chi-Wai Yip¹, Chung Chau Hon¹, Masayoshi Ito^{1,2}, Takeya Kasukawa¹, Naoto Kondo¹, Ken Yagi¹, Harukazu Suzuki¹, Jay W. Shin¹, Michiel de Hoon¹, Piero Carninci¹

¹Division of Genomic Technologies (DGT), RIKEN Center for Life Science Technologies, Yokohama, Japan; ²RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan

Correspondence: Piero Carninci (carninci@riken.jp)
Human Genomics 2018, **12(Suppl 1):A82**

Background

Majority of the human genome encodes for RNAs (lncRNAs) yet most efforts to-date have focused on functionally characterizing protein coding genes. Growing experimental evidence suggests, however, that lncRNAs are key regulators in many biological processes and are widely implicated in human disease. Further, our recent computational work on human atlas of lncRNAs shows their diverse genomic and epigenomic origins and provides functional evidence for nearly 20,000 lncRNA genes. It is thus clear that understanding of the function of the non-coding part of the genome is required for general understanding of the function and regulation of the whole human genome.

Results

The sixth edition of the FANTOM project (FANTOM6) aims to systematically elucidate the function of lncRNAs in human by subjecting them to high-throughput perturbations in multiple cell types followed by profiling of the resulted transcriptional changes with Cap Analysis of Gene Expression (CAGE) to uncover their “molecular phenotypes”. In the pilot phase, we used unbiased selection to choose hundreds of lncRNA genes (both published as well as novel entities discovered in our FANTOM5 project) and perform a loss of function experiments in human skin fibroblast. We found that more than half of the suppressed genes resulted in an appreciable and non-random transcriptional change. Further, integrating the perturbation data with transcripts features showed that the transcriptome response was independent of the lncRNAs expression levels, their cellular localization and the conservation also did not seem to play a role. Still, we found that some of the transcript properties such as polyadenylation or specificity to the cell type of study produced a higher response. We also noticed that perturbing antisense transcripts resulted in a larger number of perturbed genes. Pathway and gene ontologies enrichments of affected genes revealed that different lncRNAs are involved in distinct biological processes, yet some commonalities across them exist.

Conclusions

We present experimental evidence that long-ncRNA molecules are more biologically relevant than previously known by showing that their suppression in a cell type of interest non-randomly affected expression profiles of genes that function collectively in specific molecular/biological processes.

A83

Estimating frequency of pathogenic variants in a Japanese population by using the whole-genome reference panel of ToMMo

Yumi Yamaguchi-Kabata^{1,2}, Jun Yasuda^{1,2}, Osamu Tanabe^{1,2}, Yoichi Suzuki^{1,2,6}, Hiroshi Kawame^{1,2}, Nobuo Fuse^{1,2}, Masao Nagasaki^{1,2,3}, Yosuke Kawai^{1,2}, Kaname Kojima^{1,2}, Fumiki Katsuoka^{1,2}, Sakae Saito^{1,2}, Inaho Danjoh^{1,2}, Ikuko N Motoike^{1,2,3}, Riu Yamashita^{1,2,3}, Seizo Koshiba^{1,2}, Daisuke Saigusa^{1,2}, Gen Tamiya^{1,2,7}, Shigeo Kure^{1,2}, Nobuo Yaegashi^{1,2}, Yoshio Kawaguchi¹, Fuji Nagami¹, Shinichi Kuriyama^{1,2,5}, Junichi Sugawara^{1,2}, Naoko Minegishi^{1,2}, Atsushi Hozawa^{1,2}, Soichi Ogishima^{1,2}, Hideyasu Kiyomoto^{1,2,8}, Takako Takai-Igarashi^{1,2}, ToMMo Study Group¹, Kengo Kinoshita^{1,3,4}, Masayuki Yamamoto^{1,2}

¹Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryomachi, Aoba-ku, Sendai 980-8573, Japan; ²Graduate School of Medicine, Tohoku University, 2-1 Seiryomachi, Aoba-ku, Sendai 980-8575, Japan; ³Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki Aza-Aoba, Aoba-ku, Sendai 980-8579, Japan; ⁴Institute of Development, Aging and Cancer, Tohoku University, 4-1 Seiryomachi, Aoba-ku, Sendai 980-8575, Japan; ⁵International Research Institute of Disaster Science, Tohoku University, Aramaki Aza-Aoba 468-1, Aoba-ku, Sendai 980-0845, Japan; ⁶Present address: Ageo Central General Hospital, 1-10-10 Kashiwa-za, Ageo, 362-8588, Japan; ⁷Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan; ⁸Japan Agency for Medical Research and Development (AMED), 1-7-1 Otemachi, Chiyodaku, Tokyo 100-0004, Japan

Correspondence: Yumi Yamaguchi-Kabata (yamaguchi@megabank.tohoku.ac.jp)

Human Genomics 2018, **12**(Suppl 1):A83

Clarifying population frequencies of pathological variants is essential to construct infrastructure for genomic medicine for different populations; however, rare variants that may have biological or medical

effects have not been clarified for the Japanese population. Here, by analyzing a whole-genome reference panel of a Japanese population based on 2,049 individuals (2KJPN, available at iJGVD; <http://ijgvd.megabank.tohoku.ac.jp/>), we characterized the genomes of Japanese individuals in terms of allele frequency of functional and pathological variants. Among the 28 million autosomal single nucleotide variants (SNVs) in 2KJPN, 6,862 SNVs were identified as pathologically annotated variants that overlap with known pathogenic variants in the Human Gene Mutation Database and ClinVar. In particular, we focus on 1) actionable genes recommended by the American College of Medical Genetics and Genomics (ACMG) for returning results of secondary findings, 2) congenital metabolic disorders for newborn screening (NBS), and 3) identifying genome-wide pathogenic variants showing higher frequencies in Japanese compared with other ethnic groups. Our results by an automatic analysis on the 57 autosomal ACMG genes showed that 21% of the individuals were found to have at least one reported pathogenic allele. However, after reviewing variants through literature survey in several genes, we found that reevaluation is indispensable for constructing the information infrastructure of genomic medicine for the Japanese population. By focusing on 32 genes for 17 congenital metabolic diseases for newborn screening (NBS) in Japan, we identified reported pathogenic variants and estimated their carrier frequencies by variant filtering based on variant annotations and allele frequencies. Our results showed that variant frequencies were relatively higher in *PCCA* and *SLC25A13* among the genes examined. To understand genetic basis of different incidence rates among ethnic groups, further evaluation by appropriate variant interpretation is needed. In addition, we identified genome-wide pathogenic SNVs showing higher allele frequencies in 2KJPN than in other ethnic groups, and these SNVs included clinically important variants (e.g. *CDH1* for hereditary diffuse gastric cancer, *APRT* deficiency and *CETP* for cholesterol ester transfer protein deficiency) for personalized medicine and prevention for the Japanese. Our results and ongoing activities on variant curation would lay the foundation for i) evaluating the relationships of genomic variants and disease prevalence, and ii) improving diagnostic strategies and genetic testing in Japan and East Asia.

A84

Discovery of biased orientation of DNA motif sequences affecting human gene expression for prediction of enhancer-promoter interactions

Naoki Osato

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

Correspondence: Naoki Osato (naokiosato11@gmail.com)

Human Genomics 2018, **12**(Suppl 1):A84

Chromatin interactions have important roles for enhancer-promoter interactions and regulating the transcription of genes. CTCF and cohesin proteins are located at the anchors of chromatin interactions, forming their loop structures. CTCF has insulator function limiting the activity of enhancers into the loops. However, it is still unclear what proteins are involved in chromatin interactions.

I previously analyzed the functional enrichment and gene expression level of transcriptional target genes predicted in various cell types using the experimental data of open chromatin regions and gene expressions and DNA binding sequences of transcription factors [1]. I found that putative transcriptional target genes tend to include similar functions of genes. The normalized numbers of functional enrichments of putative transcriptional target genes changed according to the criteria of enhancer-promoter associations, and showed the same tendency in various cell types. DNA binding motif sequences of CTCF indicate their orientation bias at chromatin interaction anchors – the forward-reverse orientation is frequently observed. The normalized numbers of functional enrichments were significantly larger in enhancer-promoter associations shortened at the genomic locations of forward-reverse orientation of CTCF binding sites than enhancer-promoter associations shortened at CTCF binding sites without considering their orientation in various cell types. The numbers of chromatin interactions of Hi-C experiments overlapped with enhancer-promoter associations showed

the same tendency. The expression level of transcriptional target genes predicted by each transcription factor also changed, according to the criteria of enhancer-promoter associations.

To predict DNA binding motif sequences of transcription factors and repeat DNA sequences involved in chromatin interactions, here I examined the expression level of transcriptional target genes predicted by enhancer-promoter associations shortened at the genomic locations of DNA binding motif sequences of a transcription factor, instead of CTCF. I found that forward-reverse orientation of cohesin (SMC3 and RAD21) as well as CTCF significantly affected the expression level of putative transcriptional target genes in various cell types. DNA binding motif sequences of other transcription factors and repeat DNA sequences significantly affected the expression level of putative transcriptional target genes in forward-reverse or reverse-forward orientation.

Reference

- Osato N. Characteristics of functional enrichment and gene expression level of human putative transcriptional target genes. *BMC Genomics*. 2018;18(Suppl 1). in press.bioRxiv. preprint 2016; <https://doi.org/10.1101/085654>

A85

Open chromatin profiling of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis

Ye Liu^{1,2, +}, Jen-Chien Chang^{1, +}, Chung-Chau Hon³, Naoshi Fukui⁴, Yanfei Zhang⁵, Zhenya Zhang², Ming Ta Michael Lee^{5,6,7}, Aki Minoda¹
¹Epigenome Technology Exploration Unit, Division of Genomic Technologies, RIKEN Center for Life Science Technologies (CLST), Yokohama, Japan; ²Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan; ³Genome Information Analysis Team, Division of Genomic Technologies, RIKEN Center for Life Science Technologies (CLST), Yokohama, Japan; ⁴Clinical Research Center, National Hospital Organization Sagami Hospital, Kanagawa, Japan; ⁵Genomic Medicine Institute, Geisinger Health System, Danville, PA, USA; ⁶Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, ROC; ⁷Laboratory for International Alliance on Genomic Research, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Correspondence: Ming Ta Michael Lee; Aki Minoda
Human Genomics 2018, **12(Suppl 1)**:A85

+ Both authors contributed equally to this work

Background: Osteoarthritis (OA) is a common joint disorder with increasing impact in an aging society. To understand its pathogenesis, molecular profiling such as RNA expression and DNA methylation for diseased tissues have been reported. On the other hand, chromatin variations in OA are so far unexplored, mainly due to the technical difficulties in applying traditional epigenomic tools on clinical samples.

Methods: In this study, we employed an epigenomic method, Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq), to characterize the genome-wide chromatin alterations associated with knee OA by comparing damaged to intact primary cartilage tissues in patients.

Results: In total 109,215 open chromatin regions were detected, and 4,450 of which were differentially accessible between normal and diseased cartilage tissues. We further defined 2,318 coding and non-coding genes with overlapping differential accessible chromatin region at their promoters and enhancers as OA related, which significantly overlapped with those identified by RNA-seq in previous study using different donors. Further analyses indicated these genes are enriched for pathways regulating chondrogenesis, endochondral ossification, and mesenchymal stem cell differentiation. Consistently, analyses for differential accessible regions distal to the TSS showed bone-related enhancers and super-enhancers are dysregulated in OA diseased tissue. Moreover, integrating ATAC-seq data with public available database allowed us to identify altered cis-regulatory elements as well as transcription factors binding, and link them to important genes involved in these pathways. Furthermore, ATAC-seq data allows us to prioritize functional SNP relevant to OA discovered in previous GWAS studies. Taken together, abnormal self-healing in OA knee cartilage was observed in this study, suggesting induced

endochondral ossification-like cartilage-to-bone conversion is a characteristic of OA progression.

Conclusions: This study demonstrates a genome-wide investigation of accessible chromatin regions can be a powerful method by probing changes of regulatory genomic elements in clinical samples.

A86

Rewiring MIR retrotransposons in long non-coding RNAs to regulate translation of paired antisense genes: the case of MAPT-AS1

Simone R^{1,4}, Javad F^{1,4}, Emmett W^{3,5}, Ehteramyan M^{1,4}, Zuccotti P², Modelska A², Siva K², Hondhamuni G^{1,4}, Trabzuni D⁴, Ryten M⁴, Wray S^{1,4}, Preza E^{1,4}, Kia D⁴, Pittman A^{1,4}, Lees A^{1,4}, Hardy J^{1,4}, Denti M.A², Quattrone A², Warner T.T^{1,4}, Plagnol V³, Ule J^{4,5} and de Silva R^{1,4}
¹Reta Lila Weston Institute, UCL Institute of Neurology, Wakefield Street, London WC1N 1PJ, UK; ²CIBIO Centre for Integrative Biology, via Sommarive 9, Povo, 38123 Trento, Italy; ³UCL Genetics Institute, Darwin Building, Gower Street, London, WC1E 6BT, UK; ⁴Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK; ⁵The Francis Crick Institute, Midland Road, London, NW1 1AT, UK
Human Genomics 2018, **12(Suppl 1)**:A86

Background

Retrotransposons constitute ~42% of the human genome. They are non-randomly distributed within the transcriptome, enriched in introns and long non-coding RNAs (lncRNAs) compared to protein-coding genes. Little is known about the function of specific classes of retrotransposon-derived lncRNAs.

Materials and methods

We took an unbiased transcriptome-wide approach to investigate lncRNAs with exapted mammalian-wide interspersed repeats (MIR) and identified 1,197 lncRNA genes containing MIRs (MIR-lncRNAs). They form sense-antisense pairs with 1,045 protein-coding genes, most forming a network of interacting proteins. Gene ontology analysis revealed that MIR - lncRNAs overlap with genes commonly expressed in the nervous and immune systems. We focused on one neuronally restricted MIR-lncRNA, known as *MAPT-AS1*.

Using RNA sequencing and qRT-PCR, we assessed expression of *MAPT-AS1* and *MAPT* in brain and differentiating human induced pluripotent stem cells. In neuroblastoma cell lines, we either silenced or stably expressed *MAPT-AS1* splice variants and targeted deletion mutants to identify the essential functional domains of *MAPT-AS1* and characterized them by western blot, qRT-PCR, polysome profiling and luciferase reporter assays.

Results

As an example relevant to human disease, we focused on *MAPT-AS1*, an alternatively spliced, cytoplasmic-enriched antisense lncRNA that overlaps with the 5'-untranslated region of the tau gene, *MAPT*. Using loss-of-function experiments and cell lines stably expressing variants of *MAPT-AS1*, we showed that it specifically represses *MAPT* translation. This repression requires both the overlapping 5' region of *MAPT-AS1*, complementary to the *MAPT* internal ribosome entry site (IRES), and a 3' inverted MIR element. A truncated 90nt *MAPT-AS1* construct consisting of only the 5' overlapping region and the MIR retained the capacity to repress *MAPT* translation. Furthermore, two 7-mer motifs within the MIR, complementary or identical to sequences within 18S ribosomal RNA, are essential for *MAPT* translational repression.

Conclusions

Our data suggest that *MAPT-AS1* lncRNA functions by interfering with ribosome recruitment to the *MAPT* mRNA. MIR-lncRNAs frequently overlap with genes implicated in neurodegenerative diseases and may thus establish an additional layer of translational regulation, with implications for proteostasis in neurodegeneration.

A87

Eukaryotic genome annotation using RNA-seq and homology information Keywords: bioinformatics, eukaryotic genome, genome annotation

Kentarō Fukuta^{1,2}, Kondo Shinji^{1,2}, Ejima Fumiwo^{1,2}, Hideki Noguchi^{1,2}
¹Center for Genome Informatics, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan; ²Advanced Genomics Center, National Institute of Genetics, Japan

Correspondence: Hideki Noguchi
Human Genomics 2018, **12(Suppl 1)**:A87

Recent advances in the large-scale sequencing technology have expanded opportunities to perform the *de novo* sequencing of various organisms. The genome annotation is fundamental information for all genome-based analyses. RNA-seq data obtained from NGS and known protein sequences stored in the public databases are very useful for constructing annotations. However, the numbers of raw RNA-reads and all known proteins are large, additionally their similarities to the target genes are varied, so it is difficult to combine these data to construct a reliable genome annotation. Here we present the automatic genome annotation pipeline that integrates the RNA-seq and protein sequences derived from various species. In the current version, RNA-seq reads are assembled with Trinity [Grabherr, M.G., *et al.*, 2011] to determine the transcript sequences. The known protein sequences are obtained from the UniProt Knowledgebase [Chen, C. *et al.*, 2017]. All of these sequences are mapped to the genome to determine the loci of target genes using Ghostx [Suzuki, S. *et al.*, 2014], and the precise gene structures are determined by Spaln [Gotoh, O. *et al.*, 2008]. We evaluated performances of our pipeline using the benchmarking software of universal single-copy orthologs (BUSCO [Simao, F.A. *et al.*, 2015]) on three different species' genome (mammalian, insect, and fungi). The accuracies were 97%, 96%, and 88% of BUSCO-scores in the mammalian, insect, and fungi, respectively. Our pipeline archived high accuracies without complicated parameter settings, and will help to construct reliable genome annotations and to facilitate whole-genome analyses.

A88

GWAS meta-analysis (N=279,930) identifies new genes and functional links to general cognitive ability

Jeanne E Savage^{1#}, Philip R Jansen^{1,2#}, Sven Stringer¹, Kyoko Watanabe¹, Julien Bryois³, Christiaan A de Leeuw¹, Mats Nagel¹, Swapnil Awasthi⁴, Peter B Barr⁵, Jonathan R I Coleman^{6,7}, Katrina L Grasby⁸, Anke R Hammerschlag¹, Jakob Kaminski^{4,9}, Robert Karlsson⁵, Eva Krapohl⁶, Max Lam¹⁰, Marianne Nygaard^{11,12}, Chandra A Reynolds¹³, Joey W Trampush¹⁴, Hannah Young¹⁵, Delilah Zabaneh⁶, Sara Hägg³, Narelle K Hansell¹⁶, Ida K Karlsson³, Sten Linnarsson¹⁷, Grant W Montgomery^{8,18}, Ana B Muñoz-Manchado¹⁷, Erin B Quinlan⁹, Gunter Schumann⁹, Nathan Skene¹⁷, Bradley T Webb^{19,20,21}, Tonya White², Dan E Arking²², Deborah K Attix^{23,24}, Dimitrios Avramopoulos^{22,25}, Robert M Bilder²⁶, Panos Bitsios²⁷, Katherine E Burdick^{28,29,30}, Tyrone D Cannon³¹, Ornit Chiba-Falek³², Andrea Christoforou²³, Elizabeth T Cirulli³³, Eliza Congdon²⁶, Aiden Corvin³⁴, Gail Davies^{35, 36}, Ian J Deary^{35,36}, Pamela DeRosse³⁷, Dwight Dickinson³⁸, Srdjan Djurovic^{39,40}, Gary Donohoe⁴¹, Emily Drabant Conley⁴², Johan G Eriksson^{43,44}, Thomas Espeseth^{45,46}, Nelson A Freimer²⁶, Stella Giakoumaki⁴⁷, Ina Giegling⁴⁸, Michael Gill³⁴, David C Glahn⁴⁹, Ahmad R Hariri⁵⁰, Alex Hatzimanolis^{51,52,53}, Matthew C Keller⁵⁴, Emma Knowles⁴⁹, Bettina Konte⁴⁸, Jari Lahti^{55,56}, Stephanie Le Hellard^{23,40}, Todd Lencz^{37,57,58}, David C Liewald³⁶, Edythe London²⁶, Astri J Lundervold⁵⁹, Anil K Malhotra^{37,57,58}, Ingrid Melle^{40,46}, Derek Morris⁴¹, Anna C Need⁶⁰, William Ollier⁶¹, Aarno Palotie^{62,63,64}, Antony Payton⁶⁵, Neil Pendleton⁶⁶, Russell A Poldrack⁶⁷, Katri Räikkönen⁶⁸, Ivar Reinvang⁴⁵, Panos Roussos^{28,29,69}, Dan Rujescu⁴⁸, Fred W Sabb⁷⁰, Matthew A Scult⁵⁰, Olav B Smeland^{71,72}, Nikolaos Smyrnis^{51,52}, John M Starr^{35,73}, Vidar M Steen^{23,40}, Nikos C Stefanis^{51,52,53}, Richard E Straub⁷⁴, Kjetil Sundet^{45,46}, Aristotle N Voineskos⁷⁵, Daniel R Weinberger⁷⁴, Elisabeth Widen⁶², Jin Yu³⁷, Goncalo Abecasis^{76,77}, Ole A Andreassen^{46,71,78}, Gerome Breen^{6,7}, Lene Christiansen^{11,12}, Birgit Debrabant¹², Danielle M Dick^{5,21,79}, Andreas Heinz⁴, Jens Hjerling-Leffler¹⁷, M Arfan Ikram⁸⁰, Kenneth S Kendler^{19,20,21}, Nicholas G Martin⁸, Sarah E Medland⁸, Nancy L Pedersen³, Robert Plomin⁶, Tinca JC Polderman¹, Stephan Ripke^{4,81,82}, Sophie van der Sluis¹, Patrick F Sullivan^{3,83}, Henning Tiemeier^{2,80}, Scott I Vrieze¹⁵, Margaret J Wright^{16,84}, Danielle Posthuma^{1,85}

¹Department of Complex Trait Genetics, Center for Neurogenetics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands; ²Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands; ³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ⁴Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Campus Mitte, Berlin, Germany; ⁵Department of Psychology, Virginia Commonwealth

University, Richmond, VA, USA; ⁶Medical Research Council Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; ⁷NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London, UK; ⁸QIMR Berghofer Medical Research Institute, Herston, Brisbane, Australia; ⁹Berlin Institute of Health (BIH), 10178 Berlin, Germany; ¹⁰Institute of Mental Health, Singapore, Singapore; ¹¹The Danish Twin Registry and the Danish Aging Research Center, Department of Public Health, University of Southern Denmark, Odense, Denmark; ¹²Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark; ¹³Department of Psychology, University of California Riverside, Riverside, CA, USA; ¹⁴BrainWorkup, LLC, Los Angeles, CA; ¹⁵Department of Psychology, University of Minnesota, Minneapolis and Saint Paul, Minnesota, USA; ¹⁶Queensland Brain Institute, University of Queensland, Brisbane, Australia; ¹⁷Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; ¹⁸Institute for Molecular Biosciences, University of Queensland, Brisbane, Australia; ¹⁹Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; ²⁰Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; ²¹Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA; ²²McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ²³Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; ²⁴Psychiatry and Behavioral Sciences, Division of Medical Psychology, and Department of Neurology, Duke University Medical Center, Durham, NC, USA; ²⁵Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ²⁶UCLA Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA; ²⁷Department of Psychiatry and Behavioral Sciences, Faculty of Medicine, University of Crete, Heraklion, Crete, Greece; ²⁸Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ²⁹Mental Illness Research, Education, and Clinical Center (MISN 2), James J. Peters VA Medical Center, New York, NY, USA; ³⁰Department of Psychiatry - Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; ³¹Department of Psychology, Yale University, New Haven, CT, USA; ³²Department of Neurology, Bryan Alzheimer's Disease Research Center, and Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC, USA; ³³Human Longevity Inc, Durham, NC, USA; ³⁴Neuropsychiatric Genetics Research Group, Department of Psychiatry and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland; ³⁵Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; ³⁶Department of Psychology, University of Edinburgh, Edinburgh, UK; ³⁷Division of Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, NY, USA; ³⁸Clinical and Translational Neuroscience Branch, Intramural Research Program, National Institute of Mental Health, National Institute of Health, Bethesda, MD, USA; ³⁹Department of Medical Genetics, Oslo University Hospital, University of Bergen, Oslo, Norway; ⁴⁰NORMENT, K.G. Jebsen Centre for Psychosis Research, University of Bergen, Bergen, Norway; ⁴¹Neuroimaging, Cognition & Genomics (NICOG) Centre, School of Psychology and Discipline of Biochemistry, National University of Ireland, Galway, Ireland; ⁴²23andMe, Inc., Mountain View, CA, USA; ⁴³Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; ⁴⁴Helsinki University Hospital, Helsinki, Finland; ⁴⁵Department of Psychology, University of Oslo, Oslo, Norway; ⁴⁶Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; ⁴⁷Department of Psychology, University of Crete, Greece; ⁴⁸Department of Psychiatry, Martin Luther University of Halle-Wittenberg, Halle, Germany; ⁴⁹Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; ⁵⁰Laboratory of NeuroGenetics, Department of Psychology & Neuroscience, Duke University, Durham, NC, USA; ⁵¹Department of Psychiatry, National and Kapodistrian University of Athens Medical School, Eginition Hospital, Athens, Greece; ⁵²University Mental Health Research Institute, Athens, Greece; ⁵³Neurobiology Research Institute, Theodor-Theohari Cozzika Foundation, Athens, Greece; ⁵⁴Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA; ⁵⁵Department of

Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland; ⁵⁶Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland; ⁵⁷Department of Psychiatry, Hofstra Northwell School of Medicine, Hempstead, New York, USA; ⁵⁸Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, New York, USA; ⁵⁹Department of Biological and Medical Psychology, University of Bergen, Norway; ⁶⁰Division of Brain Sciences, Department of Medicine, Imperial College, London, UK; ⁶¹Centre for Integrated Genomic Medical Research, Institute of Population Health, University of Manchester, Manchester, UK; ⁶²Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; ⁶³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK; ⁶⁴Massachusetts General Hospital, Center for Human Genetic Research, Psychiatric and Neurodevelopmental Genetics Unit, Boston, Massachusetts, USA; ⁶⁵Centre for Epidemiology, Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Manchester, UK; ⁶⁶Division of Neuroscience and Experimental Psychology/ School of Biological Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Salford Royal NHS Foundation Trust, Manchester, UK; ⁶⁷Department of Psychology, Stanford University, Palo Alto, CA, USA; ⁶⁸Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland; ⁶⁹Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁷⁰Robert and Beverly Lewis Center for Neuroimaging, University of Oregon, Eugene, OR, USA; ⁷¹NORMENT, K.G. Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo and Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; ⁷²Department of Neuroscience, University of California San Diego, La Jolla, California, USA; ⁷³Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK; ⁷⁴Lieber Institute for Brain Development, Johns Hopkins University Medical Campus, Baltimore, MD, USA; ⁷⁵Campbell Family Mental Health Institute, Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada; ⁷⁶Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; ⁷⁷Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; ⁷⁸Institute of Clinical Medicine, University of Oslo, Oslo, Norway; ⁷⁹College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, VA, USA; ⁸⁰Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; ⁸¹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; ⁸²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; ⁸³Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; ⁸⁴Centre for Advanced Imaging, University of Queensland, Brisbane, Australia; ⁸⁵Department of Clinical Genetics, section Complex Trait Genetics, Neuroscience Campus Amsterdam, VU Medical Center, Amsterdam, the Netherlands

Correspondence: Danielle Posthuma (d.posthuma@vu.nl)
Human Genomics 2018, 12(Suppl 1):A88

Background

Cognitive ability is highly heritable [1] and a major determinant of human health and well-being [2]. Recent genome-wide meta-analyses have identified 24 genomic loci linked to variation in general cognitive ability [3-7], but much about its genetic underpinnings remains to be discovered. Here, we present the largest genetic association study to date (N=279,930) of cognitive ability.

Materials and methods

We performed a genome-wide meta-analysis of 16 independent cohorts of European ancestry. Cognitive ability was assessed using various measures of performance in fluid domains of cognitive functioning. Despite sample and methodological variations, genetic correlations (rg) between cohorts were considerable (mean=0.63), and there was no evidence of heterogeneity between cohorts in the single nucleotide polymorphism (SNP) associations. GWAS was conducted following a standardized quality control procedure, including e.g. checks on missingness, genotyping errors, relatedness and corrections for possible population stratification. Meta-analysis was carried out in Metal.

Results

We detected 12,701 genome-wide significant variants, represented by 246 independent lead SNPs in 206 genomic loci (191 novel). Roughly half of the significant SNPs are located within a gene, and 4.4% are highly likely to have a regulatory function. The results implicated 1041 genes via positional, eQTL and chromatin interaction mapping, and gene-based analyses. 89 genome-wide significant hits were exonic nonsynonymous variants, with clear implications on gene function. Stratified LD Score regression showed enrichment for heritability of SNPs located in conserved regions of the genome, coding regions, H3K9ac histone regions, and super-enhancers. While conserved regions have been implicated for cognitive ability and neuropsychiatric phenotypes, enrichment in coding regions has not previously been linked to variation in cognitive ability. Pathway analyses showed independent associations of three biological pathways (neurogenesis, neuron differentiation and synaptic structure), several brain tissues, and specific brain cell types (cortical and hippocampal pyramidal neurons, and striatal medium spiny neuron cells). We confirm previously reported genetic correlations with several neuropsychiatric traits, and used Mendelian randomization to test for credible causal associations. We found a strong protective effect of cognitive ability on ADHD and Alzheimer's dementia, bidirectional causation and strong pleiotropy between cognitive ability and schizophrenia, and a risk effect of cognitive ability on autism.

Conclusions

This study provides the most extensive meta-analysis of genetic variants associated with cognitive ability to date, provides novel insight into the underlying neurobiology and into causal relationships with other traits. The results are a rich resource of functionally relevant outcomes, which can provide starting points for further functional work.

References

1. Polderman, T. J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 47, 702-709, doi:10.1038/ng.3285 (2015).
2. Wraw, C., Deary, I. J., Gale, C. R. & Der, G. Intelligence in youth and health at age 50. *Intelligence* 53, 23-32, doi:10.1016/j.intell.2015.08.001 (2015).
3. Davies, G. *et al.* Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry* 20, 183-192, doi:10.1038/mp.2014.188 (2015).
4. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* 21, 758-767, doi:10.1038/mp.2016.45 (2016).
5. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* 49, 1107-1112, doi:10.1038/ng.3869 (2017).
6. Trampush, J. W. *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry* 22, 336-345, doi:10.1038/mp.2016.244 (2017).
7. Zabaneh, D. *et al.* A genome-wide association study for extremely high intelligence. *Mol Psychiatry*, doi:10.1038/mp.2017.121 (2017).

A89

Genetic predisposition to cancer due to germline mutations in DNA repair-associated genes among Russian woman

Leyla Shigapova¹, Marat Gordiev², Maxim Druzkov², Rafael Enikeev², Alexei Nikitin³, Yana Leksina¹, Oleg Gusev^{1,3}, Elena Shagimardanova¹

¹Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Tatarstan, Russia; ²Tatarstan cancer center, Kazan, Tatarstan, Russia; ³Federal Research and Clinical Center FMBA, Moscow, Russia; ³KFU-RIKEN Translational Genomics Unit, RIKEN, Yokohama, Kanagawa, Japan

Correspondence: Elena Shagimardanova (ryukula@gmail.com)
Human Genomics 2018, 12(Suppl 1):A89

Background

Hereditary breast and ovarian cancer (HBOC) is one of the most frequent monogenic disorders, what led to identification of number of high-penetrance (BRCA1 and BRCA2, PTEN, TP53, CDH1, and STK11), moderate-penetrance (PALB2, ATM, BRIP1, and CHEK2) genes. Rapid

development of high throughput sequencing and decrease in cost of the analysis led to the utilization of gene panel sequencing for clinical research and fast accumulation of pathogenic mutations identification. Present study is the first attempt to use NGS for hereditary cancer study in different ethnic groups in Russia.

Materials and methods

Individuals were chosen to be included in this study according to the following criteria: (1) young age of disease onset, (2) the presence of relatives with breast or ovarian cancer diagnosis. The NimbleGen SeqCap EZ Choice kit ("Roche") was used for target enrichment, and sequencing was performed using Illumina MiSeq ("Illumina") using paired-end 2 × 251 nucleotide single-index sequencing. HGMD Professional 2017.1 and BIC databases were used to identify pathogenic mutations.

Results

We have launched "Hereditary Oncogenomics Russia" project to study hereditary cancer syndromes in Russian Federation aiming to create biobank of blood samples and mutation database of pathogenic mutations in cancer related genes. In order to collect samples from patients suffering from hereditary cancer according to NCCN recommendations several cancer centers were recruited. All participants provided written informed consent. On the first step we analyzed 378 samples from patients with breast or ovarian cancer. Of the 378 patients, 26% carried mutation in BRCA1 gene, 17% in BRCA2 gene. Thus, less than 50% of cases might be explained by pathogenic BRCA1 or BRCA2 mutation. However, in the majority of BRCA-negative cases mutation in other reparation related genes, including CDK12, FANCL, APC, ATM, CDH1, MUTYH, PALB2, FANCI, CHEK2, RAD1C, was found.

Conclusions

Among patients with fulfilling NCCN criteria of hereditary breast or ovarian cancer, in addition to well-known founder mutations in BRCA1 and BRCA2 genes, we found other frequent pathogenic alteration in these genes. Panel sequencing revealed that in BRCA negative cases, one of the reparation related genes might play leading role in cancer development. Our first results stress out the critical importance of country-wide national project for identification of spectrum of germline pathogenic mutations in Russia, as a multi-ethnic country, and following updating of clinical recommendations for cancer-associated mutation screening strategy.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University

A90

Forward genetic analysis revealed a causal gene responsible for susceptibility to rat retinal photic injury

Kentaro Ohishi¹, Masafumi Ohtsubo¹, Katsuhiko Hosono², Akira Obana^{3,4}, Yoshihiro Hotta², Tadahisa Hiramitsu¹, Shinsei Minoshima¹

¹Department of Photomedical Genomics, Institute for Medical Photonics Research, Preeminent Medical Photonics Education & Research Center, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan; ²Department of Ophthalmology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan; ³Department of Medical Spectroscopy, Institute for Medical Photonics Research, Preeminent Medical Photonics Education & Research Center, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu 431-3192, Japan; ⁴Department of Ophthalmology, Seirei Hamamatsu General Hospital, 2-12-12 Sumiyoshi, Naka-ku, Hamamatsu 430-8558, Japan

Correspondence: Shinsei Minoshima (mino@hama-med.ac.jp)
Human Genomics 2018, 12(Suppl 1):A90

Background

Long-time sun exposure is thought as a risk factor for age-related macular degeneration (AMD) which is a leading cause of the blindness in developed countries. In rats, intense light exposure induces retinal degeneration resembling AMD pathology (an experimental retinal photic injury model). The susceptibility to the degeneration differs in different rat strains. Therefore, we tried to identify the

gene(s) which causes the susceptibility difference and might be related to AMD onset.

Materials and methods

The retinal photic injury was induced by white light exposure (3000 lux, 3 hrs). The susceptibility was evaluated by Morris water maze test for eyesight and pathological analysis of the eyes for retinal degeneration. The genotyping of each rat individual was performed using polymorphic microsatellite and SNP markers which can distinguish the strain specificity. A targeted exome analysis was performed with a next generation sequencer.

Results

A rat strain (WKY) susceptible to and another strain (LEW) resistant to the retinal photic injury were crossed to produce F1 offspring. The F1 was susceptible, meaning that the susceptibility was a Mendelian autosomal dominant trait to the resistance. We named the trait "the susceptibility to retina photic injury" and the responsible gene "*Rpi1*". In order to identify the *Rpi1* gene, the first backcross (BC1) offspring was produced by mating F1 with the recessive LEW. Using the Morris water maze test before/after the light exposure, susceptible individuals in the BC1 generation were selected to produce BC2 offspring. Thus, as the backcross and selection of susceptible individuals were iterated, the fraction of WKY genome was gradually decreased. An obtained BC11 offspring had a 3.8-Mb WKY genome region within chromosome 5q36 as a sole WKY component, in which the *Rpi1* gene must exist. We performed the targeted exome analysis for this region to identify *Rpi1* and found 18 loci with non-synonymous base changes in coding regions of 15 genes. Using additional backcross individuals and another photic injury-susceptible strain F344, recently we could narrow down to only one candidate gene (tentatively *Gene R*) with reasonable relationship between genotype and phenotype.

Conclusions

We consider that the *Gene R* is *Rpi1* itself. The *Gene R* has never been reported as the susceptibility gene of AMD. We are now extensively making efforts to give more direct evidence for the *Gene R* as the causal gene of rat retina photic injury susceptibility difference and possible relationship with AMD.

Acknowledgements

S. Minoshima will present this research on behalf of our study group.

A91

The genetic basis of invasive meningococcal disease revealed thorough whole exome sequencing

Evangelos Bellos¹, Jethro Herberg¹, Victoria Wright¹, Daniela Klobassa², Meg Mashbat¹, Rahmeen Rahman¹, Luregn Schlapbach³, Richard Pouw⁴, Taco Kuijpers⁵, Michael Levin¹, Vanessa Sancho-Shimizu^{1,6*}

¹Department of Paediatrics, Division of Medicine, Imperial College London, London, UK; ²Department of General Paediatrics, University Clinic of Paediatrics and Adolescent Medicine, Medical University Graz, Graz, Austria; ³Infectious Diseases and Vaccinology, University Children's Hospital Basel, Basel, Switzerland; ⁴Department of Immunopathology, Sanquin Research and Landsteiner laboratory of the Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; ⁵Department of Experimental Immunology, Academic Medical Centre, PO Box 22660, 1100 DD Amsterdam, the Netherlands; ⁶Department of Virology, Division of Medicine, Imperial College London, London, UK

Correspondence: Vanessa Sancho-Shimizu (v.sancho-shimizu@imperial.ac.uk)

Human Genomics 2018, 12(Suppl 1):A91

Background

Invasive meningococcal disease (IMD) is a rare condition affecting children and young adults due to infection with *Neisseria meningitidis*, resulting in meningitis or sepsis. Although the majority of the general population is colonized by *N. meningitidis*, only a small minority go on to develop IMD, suggesting that those that succumb to invasive disease may possess an underlying genetic susceptibility. The notion of a genetic contribution to disease manifestation is supported by the finding that patients with congenital complement deficiencies are susceptible to recurrent IMD, yet these conditions are rare. With the aim of identifying other genetic factors underlying

IMD, we carried out whole exome sequencing (WES) of approximately 300 IMD patients from an extensive and well characterized cohort of >2,000 childhood IMD patients.

Materials and Methods

The WES analysis focused on rare variants (MAF<0.01) predicted to have a detrimental effect on protein function. We undertook analysis of seven multiplex families identifying IBD variants. The rest of the index cases were analysed as a cohort and put through gene and pathway burden tests to identify any genes/pathways that were enriched in the collection of patients.

Results

These analyses revealed a number of patients harbouring mutations in known primary immunodeficiency genes (approx. 10%) as well as a novel configuration of mutations underlying the complement genes. Furthermore, novel mutations in the coagulation pathway and mucosal immunity genes were identified and functionally confirmed.

Conclusions

The identification of genes involved in IMD through WES has demonstrated the complex genetic architecture of meningococcal immunity revealed some novel and unexpected genes/pathways that modulate disease susceptibility and severity. The results from this study provide us with a more comprehensive understanding of IMD pathogenesis.

A92

Single-cell assays in a simulated peritumoral microenvironment

Christian Maercker^{1,2}

¹Esslingen University of Applied Sciences, Esslingen, Germany; ²Genomic and Proteomic Core Facilities, German Cancer Research Center (DKFZ), Heidelberg, Germany

Human Genomics 2018, **12(Suppl 1)**:A92

Metastasis and tumor growth are very much dependent on the peritumoral microenvironment. Cells in the peripheral cortex send out growth factors and generate an extracellular matrix which are decisive for infiltration of neoplastic cells and, finally, the formation of a functioning tumor tissue. To figure out specific functions of different cell types and their various interactions in in vitro or in vivo assays, which are possibly relevant for the identification of new drug targets, is a challenge, because we deal with a very complex multifactorial network. Therefore, our goal is to develop test systems with a limited complexity, but which still are informative in terms of molecular tumor development.

To reach this, we first have to singularize cells from tumor tissue. Within surgical tissue, tumor core tissue has to be discriminated from surrounding tissue by MRI or immunohistochemistry. Afterwards, microdissected tissue samples have to be dissociated by protease digest. In a next step, the mixture of cells has to be separated into the different cell types by immunopanning and/or FACS sorting. RNAseq helps to identify gene expression profiles specific for individual cells within the tumor compared to the tumor microenvironment.

For us the most important parameters are the adhesion and migration activities of the various cell types of different origin on different matrices. To establish a test system as a starting point for a more sophisticated experimental setup to characterize tumorigenic tissues, we firstly isolated mesenchymal stem cells (MSC) from fat tissue or bone marrow, before transfer to microwells covered with electrodes to monitor cell adhesion by electrochemical impedance sensing. We found that pre-coating of the microwells with either collagen type I or IV induced different impedance profiles, corresponding to distinct differentiation pathways which was confirmed by histochemical staining as well as protein and gene expression analyses.

In another approach, we used impedance sensing to measure prostate or breast cancer cell migration towards the secretome of mesenchymal stem cells. MSC cell culture supernatant was fractionated by size exclusion and ion exchange chromatography. We identified type I and III collagen, fibronectin and Laminin 421 as extracellular matrix

proteins to be the main drivers of rapid cancer cell migration requiring as little as two hours for a full migration response. The challenge now is to adjust this experimental approach to individual cell types isolated from different tumor types and also to different matrices simulating the tumor microenvironment.

A93

omicsCNN: a general deep learning framework for omics data modeling and classification

G. Jurman¹, V. Maggio¹, I. Landi^{1,2}, M. Francescato¹, M. Chierici¹, M. De Domenico^{1,3}, C. Furlanello¹

¹Fondazione Bruno Kessler, Trento, Italy; ²University of Trento, Trento, Italy; ³Universitat Rovira e Virgili, Tarragona, Spain

Correspondence: M. Francescato

Human Genomics 2018, **12(Suppl 1)**:A93

Deep Learning, as a class of machine learning algorithms and neural network architectures, shows high promise in biomedicine. In particular, Convolutional Neural Networks (CNNs) have been successfully applied in a wide range of domains, including the prediction of transcription factor binding sites, the identification of genomic variants, and bioimaging. CNN architectures are highly efficient in modeling 2D spatial information, discovering features of interest through layers of convolutional filter units inspired by the structure of the brain visual cortex. The convolution filters require a distance or neighbourhood function in the input feature space, which is a limitation to their general application to omics data. Overcoming existing approaches that convert omics data into images for subsequent CNN processing, we propose a new framework that uses direct feature remapping into appropriate metric spaces.

We recently developed omicsCNN, a generalized CNN solution for omics data. The framework is implemented as a novel Keras layer, OmicsConv. On metagenomics 16S sequencing data, we applied the patristic distance to define the embedding and build a classifier of distinct subtypes of Inflammatory Bowel Disease, with superior accuracy to other machine learning methods.

Here we present a further development of the omicsCNN architecture customized for application to transcriptomics data. We applied the omicsCNN solution for transcriptomics to expression data (both microarrays and RNAseq) from the MicroArray/Sequencing Quality Control study (MAQC/SEQC) neuroblastoma dataset, including 498 patients (176 high risk) to predict event free and overall survival endpoints. The framework uses Gene Ontology (GO) mutual semantic similarities combined with a gene co-expression measure, obtaining a multilayer network, whose layers are computed from the three GO categories and co-expression. For each layer, the metric between two genes is given by the diffusion distance, which quantifies the likelihood that two random walkers originating in the two genes meet in another node of the net. The global distance in the network is defined as the L2 Euclidean product of the diffusion distances of each layer, thus enabling the OmicsConv layer. We evaluate the solution in terms of predictive accuracy, also comparing with a set of machine learning methods. We discuss the impact of OmicsCNN as a natural deep learning setup for transcriptomics data modelling and evaluate insights from features discovered by the model.

A94

Integrated transcriptome and methylome analysis of mucinous colorectal adenocarcinoma

Nurul-Syakima Ab Mutalib¹, Rasyidah Baharuddin¹, Noraidatulakma Abdullah¹, Luqman Mazlan², Rahman Jamal¹

¹UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia; ²Department of Surgery, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia

Correspondence: Nurul-Syakima Ab Mutalib

Human Genomics 2018, **12(Suppl 1)**:A94

Colorectal cancer (CRC) is the 3rd most common cancer diagnosed in males and second in females, with approximately around 1.4 million cases of CRC were diagnosed and over 690,000 deaths were reported in 2012. The majority of CRC are classified as adenocarcinoma (AC) and approximately 10-15% of CRC patients are presented with mucinous carcinoma (MC) subtype. MC is characterised by secretion of mucin in which the presence of mucin causes the cancer cells to spread faster and become more aggressive compared to AC subtype. Therefore, colorectal MC has been associated with poor response to treatment, low survival rates, worse prognosis among patients and present at more advanced stage of disease than non-mucinous adenocarcinoma. However, the knowledge on the epigenome-wide methylation profile associated with this type of malignancy is scarce. This study aims to determine differentially expressed and methylated genes in colorectal MC as compared to AC, to integrate the methylation status with gene expression profiles and to correlate with clinicopathologic features. We used the TCGA-generated methylation data based on Illumina Methylation 450 beadchip in β value format retrieved from and level 3 RNA sequencing dataset in RSEM (RNA-Seq by Expectation Maximization) format downloaded from Firebrowse. After filtering for only patients with paired methylation and gene expression datasets, a total of 289 CRC patients were identified. Two hundred fifty patients were diagnosed with AC and 39 were diagnosed with MC histology. Additionally, gene expression data from 173 normal colon and methylation data from 45 normal colon were included as comparisons. Unpaired t-test with a Benjamini Hochberg (BH) false discovery rate (FDR) multiple testing correction was performed using Bioconductor in R. We identified 1336 genes which were significantly demethylated in MC versus AC; 788 were hypermethylated and 548 were hypomethylated with 295 and 131 genes had $\Delta\beta \geq 0.1$ and ≤ -0.1 , respectively. From RNAseq analysis we identified 3645 significant differentially expressed genes with 2025 were upregulated and 1620 were downregulated. *EIF6* is the most significantly hypermethylated (22.8% hypermethylation; $p = 9.92E-07$) in MC versus AC and its expression was downregulated -1.631 fold ($p = 1.90E-04$). On the other hand, *BAG3* is the most significantly hypomethylated (20.8% hypomethylation, $p = 1.65E-03$) despite modest but significant upregulation of expression (0.309, $p = 8.20E-03$). Both genes were involved in carcinogenesis processes. These preliminary findings suggest that the aberrant DNA methylation of *EIF6* and *BAG3* might be involved in MC development.

A95

Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depressive disorder

Maciej Trzaskowski¹, Naomi Wray¹, The PGC MDD Consortium, Patrick Sullivan²
¹Institute for Molecular Bioscience, The University of Queensland, St Lucia, 4072, QLD, Australia; ²University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA

Correspondence: Maciej Trzaskowski (m.trzaskowski@uq.edu.au)
 Human Genomics 2018, 12(Suppl 1):A95

Background

Major Depressive Disorder (MDD) is a notably complex illness with a lifetime prevalence of 14%. It is often chronic or recurrent and is thus accompanied by considerable morbidity, excess mortality, substantial costs, and heightened risk of suicide. MDD is a major cause of disability worldwide. Twin studies attribute ~40% of the variation in liability to MDD to additive genetic effects, and may be greater for recurrent, early-onset, and postpartum MDD. GWA studies of MDD have had notable difficulties in identifying loci. Previous findings suggest that an appropriately designed study should identify susceptibility loci.

Materials and methods

We conducted a Genome-Wide Association (GWA) meta-analysis in 130,664 MDD cases and 330,470 controls. We used the standard GWA mega/meta-analysis methods of the core PGC quality control, imputation, and analysis pipeline (i.e., "ricopili").

Results

We identified 44 independent loci that met criteria for statistical significance. We can make a strong case for the identification of

RBFOX1 and NEGR1. The genetic findings were associated with clinical features of MDD, and implicated prefrontal and anterior cingulate cortex in the pathophysiology of MDD (regions exhibiting anatomical differences in MDD cases vs controls). Genes that are targets of antidepressant medications were strongly enriched for MDD association signals ($P=8.5e-10$), suggesting the relevance of these findings for improved pharmacotherapy of MDD. Sets of genes involved in gene splicing and in creating isoforms were also enriched for smaller MDD GWA P-values, and these gene sets have also been implicated in schizophrenia and autism. Genetic risk for MDD was correlated with that for many adult and childhood onset psychiatric disorders. Our analyses suggested important relations of genetic risk for MDD with educational attainment, body mass, and schizophrenia: the genetic basis of lower educational attainment and higher body mass were putatively causal for MDD whereas MDD and schizophrenia reflected a partly shared biological etiology.

Conclusions

We present extensive analyses of these results which provide new insights into the nature of MDD. All humans carry lesser or greater numbers of genetic risk factors for MDD, and a continuous measure of risk underlies the observed clinical phenotype. MDD is not a distinct entity that neatly demarcates normalcy from pathology but rather a useful clinical construct associated with a range of adverse outcomes and the end result of a complex process of intertwined genetic and environmental effects. These findings help refine and define the fundamental basis of MDD

A96

Single-cell RNA transcriptome analysis elucidates diverse heterogeneity of gene regulation in cardiomyocytes

Akira Watanabe, Masahiro Nakamura, Shunsuke Funakoshi
 Yoshinori Yoshida Center for iPSC Cell Research and Application, Kyoto University

Human Genomics 2018, 12(Suppl 1):A96

Human pluripotent stem cell-derived cardiomyocytes show promise for clinical application, especially disease modeling and drug screening. To reach this potential, however, improved maturity and reduced cellular heterogeneity of in vitro differentiated cardiomyocytes is needed. Here, we addressed transcriptional heterogeneity in induced pluripotent stem cell (iPSC)-derived cardiomyocytes (iPSC-cardiomyocytes) by single-cell RNA sequencing. We successfully defined differentiation states at the single-cell level and clarified heterogeneities in the gene expression patterns of iPSC-cardiomyocytes, especially those cultured over long periods. We developed a novel maturation index to perform in silico sorting and score the maturation level of the cells. Additionally, we identified a new cell surface marker of mature iPSC-cardiomyocytes by which we could enrich well-matured iPSC-cardiomyocytes from a heterogeneous population of iPSC-cardiomyocytes. Our single-cell RNA sequencing approach could help clarify variation of in vitro differentiated cells and provide a new strategy to purify cells of interest.

A97

Systematic analysis of copy number variation associated with congenital diaphragmatic hernia

Qihui Zhu^{1,5}, Frances A. High^{2,3,4,5}, Chengsheng Zhang^{1*}, Eliza Cerveira¹, Meaghan Russell², Mauro Longoni^{2,4}, Mallory Ryan¹, Adam Mil-homens¹, Lauren Bellfy¹, Caroline Coletti², Pooja Bhayani², Regis Hila², Patricia K. Donahoe^{2,4*}, Charles Lee^{1*}

¹The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032, USA; ²Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA 02114, USA; ³Department of Surgery, Boston Children's Hospital, Boston, MA 02115, USA; ⁴Harvard Medical School, Boston MA 02115, USA

Correspondence: Patricia K. Donahoe; Charles Lee
 Human Genomics 2018, 12(Suppl 1):A97

⁵ Co-first authors

*: Co-senior authors

Congenital diaphragmatic hernia (CDH), characterized by malformation of the diaphragm and hypoplasia of the lungs, is one of the most common and severe birth defects. This phenotype is associated with high mortality rates and long-term morbidity among survivors. There is a growing body of evidence demonstrating that genetic factors contribute to many forms of CDH in humans, although the pathogenesis remains largely elusive. Small variants such as single nucleotide polymorphisms (SNPs) have been studied in recent whole exome sequencing (WES) efforts but larger copy number variants (CNVs) have never been studied on a large scale at the genome level. In this study, to capture CNVs within CDH candidate regions in a cost-effective manner, we developed and tested a targeted array comparative genomic hybridization (aCGH) platform to identify CNVs within 140 candidate CNV regions across 196 patients with CDH and 987 normal control samples. As a result, we identified 6 significant CNVs from this patient-control study. Three of these 6 significant CNVs were found in multiple patients with CDH, but not in any controls, increasing the likelihood that they are disease-causing. These CDH-specific CNVs reveal high-priority candidate genes including *HLX*, *LHX1*, and *HNF1B*. The remaining 3 significant CNVs were found in multiple patients and a low frequency in controls, but were significantly enriched in patients with CDH. The candidate genes within these significant CNVs form functional networks with other known CDH genes and play putative roles in both DNA binding/transcription regulation and embryonic organ development. Investment in functional studies will further our understanding their contribution to the pathogenesis of CDH.

ACKNOWLEDGMENTS

We would like to thank the patients with CHD and their families for participating in this study, as well as the physicians at MassGeneral Hospital for Children and Boston Children's Hospital for their continued support. Funding was provided by the National Institute of Child Health and Human Development (NICHD/NIH, www.nichd.nih.gov) P01HD068250.

A98

Changes of RNA alternative splicing in striatum of MPTP mouse model of Parkinson's disease

Jian-Huan Chen¹, Zhou-Heng Xu¹, Mengfei Sun², Zhi-Yuan Wei¹, Ke-Wei Wang¹, Yanqin Shen²

¹Laboratory of Genomic and Precision Medicine, Wuxi School of Medicine, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China; ²Center for Neurosciences, Wuxi School of Medicine, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China

Correspondence: Jian-Huan Chen (cjh_bio@hotmail.com)

Human Genomics 2018, **12**(Suppl 1):A98

Background: The mouse model of Parkinson's disease (PD) using the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) provides an important and reliable tool to study the disease¹. Herein We used RNA sequencing (RNA-Seq) to evaluate the role played by alternative exon usage and splicing events in the MPTP mouse model of PD.²

Materials and methods: MPTP and saline were administered to adult mice by intraperitoneal injection for MPTP-treated and sham groups respectively. Whole striatum was collected from both MPTP-treated and sham groups at Day 4, and total RNA was extracted using TRIzol followed by phenol/chloroform. RNA-Seq libraries were prepared from total RNA using the ribosomal RNA depletion method. RNA-Seq was performed on the Illumina HiSeq X10 platform by sequencing constructed individual cDNA library with a configuration of 90 million 150 bp pair-end reads per sample. Reads were aligned to the mouse full genome (UCSC version mm10) and splice junctions were identified by using RNA STAR version 2.5. Differential exon usage of RNA-Seq data was detected using the R language package DEXseq. Significance level was set at 5% false discovery rate. Functional analysis including gene ontology (GO) enrichment and KEGG pathway was performed using DAVID bioinformatics resources 6.8. Selected

exons and transcripts were validated using reverse transcription polymerase chain reaction.

Results: In total 1048 exons in 873 genes showed significant differential usage that resulted in an absolute value of log₂ (fold change) greater than or equal to 0.5, with 409 exons up-regulated and 639 exons down-regulated in the MPTP-treated group compared to the sham group. Among the genes with differential usage, 774 were protein-coding genes, and 99 were non-coding. KEGG pathway analysis using DAVID showed that these genes were mostly enriched in neurotransmission and axon guidance pathways, especially in the dopaminergic synapse pathway (Table 1 and Fig. 1), which had a 3.5-fold enrichment of genes. GO analysis in Table 2 revealed that the top three enriched biological processes were transport, nervous system development, and positive regulation of synapse assembly, which were in line with the KEGG pathway findings.

Conclusions: In the current study, our results revealed remarkable exon usage change of the striatum of the MPTP mouse model of PD, suggesting that alternative splicing could play a substantial role in synapse assembly, especially for dopaminergic synapses. The current study hence provided new insight into the importance of alternative splicing in PD.

Acknowledgements

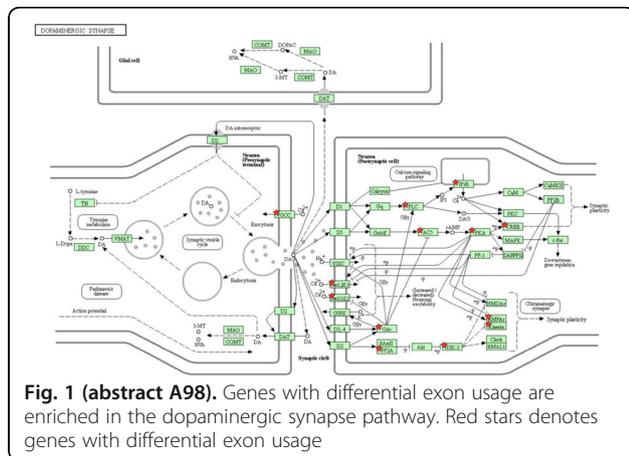
The current study is supported in part by grants from The Natural Science Foundation of China (No. 31671311, 81371033 and 81000397), the "Six Talent Peak" Project of Jiangsu Province (SWYY-127), the Guangdong High-level Personnel of Special Support Program, the Yangfan Plan of Talents, and the Fundamental Research Funds for the Central Universities (JUSRP51712B).

Table 1 (abstract A98). KEGG pathways enriched in genes with differential exon usage

KEGG pathway	Gene Count	FDR
Dopaminergic synapse	18	0.003
Glutamatergic synapse	16	0.003
Thyroid hormone signaling pathway	14	0.026
Synaptic vesicle cycle	10	0.029
Type II diabetes mellitus	9	0.024
Axon guidance	14	0.045
cAMP signaling pathway	18	0.043
Calcium signaling pathway	17	0.038
Adherens junction	10	0.038

Table 2 (abstract A98). Biological Processes enriched in genes with differential exon usage in GO analysis.

Biological process	Gene Count	FDR
transport	121	8.2×10^{-7}
nervous system development	40	1.3×10^{-5}
positive regulation of synapse assembly	16	3.3×10^{-5}
chemical synaptic transmission	24	8.0×10^{-5}
neuron migration	17	0.009
ion transport	44	0.008
negative regulation of neuron projection development	12	0.009
learning	12	0.011
potassium ion transport	16	0.025
social behavior	10	0.044



References

- Jackson-Lewis, V. & Przedborski, S. Protocol for the MPTP mouse model of Parkinson's disease. *Nat Protoc* 2, 141–151 (2007).
- Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63 (2009).

A99

Identification of gene expression signatures in human leukemia cell lines using Nanostring gene expression profiling platform

Imilia Ismail^{1,2}, Syed Ahmad Tajudin Tuan Johari³, Wan Rohani WanTaib⁴, Mohd Nizam Zahary¹, Kah Keng Wong⁵, Rosline Hassan²

¹School of Biomedicine, Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300, Kuala Terengganu, Terengganu, Malaysia; ²Department of Hematology, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150, Kubang Kerian, Kelantan, Malaysia; ³Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Besut Campus, 22200, Besut, Terengganu, Malaysia; ⁴Institute for Community (Health) Development, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300, Kuala Terengganu, Terengganu, Malaysia; ⁵Department of Immunology, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150, Kubang Kerian, Kelantan, Malaysia

Correspondence: Imilia Ismail (imilia@unisza.edu.my)

Human Genomics 2018, **12(Suppl 1)**:A99

Background

Gene expression profiles have been examined extensively in diseases including hematological malignancies. Although the diagnostic tests that sub classified leukemia have been improved, leukemia patients occasionally exhibit different responses to treatment. In order to find more precise molecular markers, we performed differential gene expression profiling in human acute myeloid leukemia (AML) and chronic myeloid leukemia (CML) cell lines (HL60 and K562, respectively) using Nanostring nCounter[®] MAX Analysis System (Nanostring Technologies, Seattle, WA).

Materials and methods

Total RNA was extracted from HL60 and K562 cell lines using innuPREP RNA mini kit (AJ Innuscreen GmbH, Germany) according to the manufacturer's instructions. The RNA quality was assessed on the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and the concentration was determined with a Nanodrop spectrophotometer (ND-1000, Thermo Scientific, Wilmington, MA, USA). We performed

gene expression profiling of 230 human cancer-related genes with six internal reference genes using nCounter GX Human Cancer Reference Kit (NanoString Technologies). A total of 100 ng of RNA for each sample was prepared as per the manufacturer's instructions under the high sensitivity protocol. Normalization and subsequent data processing were performed by using the nSolver[™] Analysis Software v2.6 (Nanostring Technologies). Differentially expressed mRNAs were identified through fold change (≥ 2.0) and p values < 0.05 filtering.

Results

We identified distinctive gene expression patterns in K562 and HL60 cell lines. The most significantly up regulated genes in K562 cells included *FGFR3*, *WT1*, *CCNA2*, *FGF2* and *HSP90AB1* while *CSF3R*, *BCL2A1*, *TNFSF10*, *AKT1* and *GNAS* were most significantly down regulated. In HL 60 cells, *WT1*, *CCNA2*, *PRKAR1A*, *MYB* and *CHEK1* were the most significantly up regulated while *FOS*, *AKT1*, *GNAS*, *TP53* and *IL1B* were the most significantly down regulated genes. Several genes that were up regulated in K562 were found to be down regulated in HL60 such as *FGF2*, *GATA1*, *IL6* and *PIM1*. *FGFR1* and *SPI1* were significantly up regulated in HL60 but were found to be significantly down regulated in K562 cell line.

Conclusions

In conclusion, our results suggest that gene expression profiling identified *FGF2*, *GATA1*, *IL6*, *PIM1*, *FGFR1* and *SPI1* to be differentially expressed between AML and CML cells. These findings may also help to assess future markers in developing therapies targeting mRNA.

A100

SFRP1 protein promotes growth of prostate cancer xenoinjert by modulation of TMPRSS2-ERG fusion

C. D. Cruz-Hernandez, S. Cortes-Ramirez, M. Cruz-Burgos, A. Alor, A. Losada-García, M. Rodríguez-Dorantes

Oncogenomics, Imegen, Mexico City, México

Correspondence: M. Rodríguez-Dorantes

Human Genomics 2018, **12(Suppl 1)**:A100

Objectives: Was to determine the effect of SFRP1 protein on progression of prostate cancer xenoinjert established with cells expressing TMPRSS2-ERG fusion

Methods: Cell culture. LNCaP and VCaP cell lines from ATCC, were grown in RPMI 1640 and DMEM medium. Confluent cells were trypsinized and 3.5×10^5 cells were inoculated to balb/c nu/nu mice with 4-6 week of age. Growth of xenoinjerto was registered by measuring volume with next formule: $V = \pi/6 \times (\text{large diameter} \times [\text{short diameter}]^2)$. Blood of mice was taken from their tail for determine prostate specific antigen (PSA) levels during experiment. PSA was determine by ELISA assay with abnova kit designed for measuring human PSA. When volume of xenoinjert was around 300 mm³ a castration surgery was made. One week after surgery of castration mice were randomized into two groups: 1) control group (vehicle) and 2) problem group (SFRP1 administration). SFRP1 protein and vehicle was administrated subcutaneously around tumor tissue during experiment one time per week. At the end of experiment mice were sacrificed and tumor tissue was collected into two tubes for RNA and protein isolation for expression assays and one container with formalin for immunohistochemistry assays. Results: the SFRP1 protein administration in xenoinjerto with cells expressing TMPRSS2-ERG fusion promoted growth of tumor tissue compared to vehicle. Furthermore, levels of PSA in blood increased with SFRP1 administration. Interestingly we observe that SFRP1 administration promoted expression of TMPRSS2-ERG, KLK3 and LEF-1 genes.

Conclusion: SFRP1 protein play an important role in stromal-epithelial communication promoting prostate cancer progression. However, prostate cancer tumors are not classified for TMPRSS2-ERG fusion and SFRP1 expression. We show evidence in animal model

A102**Reference transcription start site (refTSS): an annotated reference dataset for TSS**

Shuhei Noguchi¹, Imad Abugessaisa¹, Akira Hasegawa¹, Hideya Kawaji², Takeya Kasukawa¹

¹RIKEN Center for Life Science Technologies, Yokohama, Kanagawa, Japan; ²RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Saitama, Japan

Human Genomics 2018, 12(Suppl 1):A102

Background

In genomic research, many researchers are using reference datasets, such as genome references (e.g. hg38, hg19) for genome mapping of sequence reads and gene references (e.g. GENCODE, RefSeq) for quantification of gene expression levels. Standardized reference datasets are much important for comparing and integrating various data resources generated by many different research groups. However there is no reference dataset for transcription start sites (TSS). In the FANTOM5 (The functional annotation of the mammalian genome 5) project, they show that many TSS did not completely match with 5'-ends of reference genes, and this implies that we need an independent dataset about TSS. In this study, we attempted to create reference TSSs (refTSS) for human and mouse by collecting, re-processing and integrating published TSS datasets. Further, we also aim to integrate various datasets based on TSS, including TSS activity in various samples, gene annotation, transcriptional regulation, epigenetic markers, and genomic segmentation information.

Materials and methods

In our strategies for the construction of refTSS, we first integrate TSS regions derived from published 5'-end sequencing data, such as CAGE (Cap Analysis of Gene Expression) and TSS-Seq. Secondly, we annotate gene information and transcriptional regulatory information to the integrated TSS sets. Thirdly, we manually curate consistency of TSSs and annotated information.

Results

The current release of the refTSS consists of TSSs from FANTOM5 promoter atlas [<http://fantom.gsc.riken.jp/5/>], DBTSS [<https://dbtss.hgc.jp/>], EPD [http://epd.vital-it.ch/EPDnew_database.php], ENCODE CAGE, and RAMPAGE [<https://www.encodeproject.org/>]. To increase the use of the refTSS, we annotated information in published genomics databases. We associate each TSS with genes, transcripts, and proteins. We found total 225,052 of TSS in the latest human genome (hg38), and about 50% of TSS were associated with HGNC gene symbols. Furthermore, integrated TSS datasets were evaluated by various methods, and released to the public refTSS [<http://refTSS.cst.riken.jp/datafiles/>].

Conclusions

We created and published integrated TSS datasets, refTSS. refTSS will be useful tools for researchers using TSS related data.

A103**Whole exome sequencing identifies BPIFA1 mutation underlying invasive meningococcal disease**

Bayarchimeg Mashbat¹, Evangelos Bellos¹, Fadil Bidmos¹, Robert Tarran³, Yaxuan Lu¹, Victoria Wright¹, Jethro Herberg¹, Paul Langford¹, Luregn Schlapbach⁴, EUCLIDS Consortium, Ming-Shi Li¹, Peter Di⁵, Michael Levin¹ and Vanessa Sancho-Shimizu^{1,2}

¹Departments of Paediatric infectious diseases, Imperial College London, London, Norfolk Place, UK; ²Virology, Division of Medicine, Imperial College London, London, Norfolk Place, UK; ³Marsico Lung Institute, Cystic Fibrosis/Pulmonary Research and Treatment Centre, University of North Carolina, 7102 Marsico Hall, Chapel Hill, North Carolina 27599-7248, United States; ⁴Mater Research Institute-UQ, Faculty of Medicine, South Brisbane, Qld 4101, Australia; ⁵Department of Environmental and Occupational Health, University of Pittsburgh, Pennsylvania, USA

Human Genomics 2018, 12(Suppl 1):A103

Background

Neisseria meningitidis (Nm) is a commensal gram-negative bacterium that colonizes the upper respiratory nasopharynx and only in rare cases causes life-threatening invasive meningococcal disease (IMD).

Genetic analysis of whole exome sequencing undertaken in familial IMD cases revealed a novel missense mutation in *BPIFA1* gene. The encoding protein is specifically expressed in the human nasopharyngeal mucosa as part of the host innate immune defence against infectious agents in the airway. Previous studies have shown that *BPIFA1* has a role in host defence against other Gram-negative bacteria and viruses, possibly through inhibition of biofilm formation or import of viral particles into epithelial cells. The objective of this study was to characterise the role of *BPIFA1* in the context of meningococcal infection and investigate whether the novel missense mutation leads to increased susceptibility to IMD in the carriers.

Method

We tested the recombinant *BPIFA1* activity on Nm biofilm formation as well as bacterial adhesion and invasion of airway epithelial cells. We used transfection in cultured cells, with recombinant expression plasmids encoding WT or mutant *BPIFA1* to assess the functional effect at protein and mRNA levels.

Results

We showed that recombinant *BPIFA1* significantly inhibits Nm biofilm formation. Using *in-vitro* human colonisation models, we demonstrated that purified *BPIFA1* is protective against Nm infection by deterring bacterial adherence and invasion of the epithelial cells. In addition, the mutant protein had impaired ability in anti-biofilm and adhesion and invasion defences of the epithelial cells.

Conclusion

Our results demonstrated that *BPIFA1* plays an important host immune defence role against Nm infection. The identified missense mutation may have the potential to predispose carriers to invasive pathogenic infections such as IMD.

A104**DNA replication-coupled acetylation of cohesin complex by Esco2 acetyltransferase in heterochromatic region**

Mai Ishibashi¹, Atsunori Yoshimura², Masashige Bando², Toyonori Sakata², Takashi Sutani², Katsuhiko Shirahige²

¹Grad. Sch. of Agri. and Lifesci., The University of Tokyo, Tokyo, Japan;

²Inst. of Mol. and Cell. Biosci., The University of Tokyo, Tokyo, Japan

Correspondence: Mai Ishibashi

Human Genomics 2018, 12(Suppl 1):A104

The cohesin complex plays a key role in modulating higher-order chromosome structure, thereby mediating sister chromatid cohesion and controlling developmental gene expression. Stable binding of cohesin to chromosomes requires acetylation on a cohesin subunit Smc3, and in vertebrates, two paralogues Esco1 and Esco2 are responsible for this reaction. It is widely believed that the cohesin complex engaged in sister chromatid cohesion is acetylated in a DNA replication -coupled manner. Previous studies suggest that Esco2 is expressed specifically during S phase and shows physical interaction with replication machinery, implying its primary role in formation of cohesion. Here we provide the first evidence that Esco2 acetylates Smc3 in heterochromatic regions in S phase cells, by quantitative ChIP-seq analysis utilizing spike-in control. Conventional ChIP-seq analysis of acetylated Smc3 identified very few binding sites in HeLa cell genome, which marks a sharp contrast with a binding profile of Esco1. To understand the genome regions where Esco2 functions in the process of DNA replication progression, we compared quantitative ChIP-seq profiles of acetylated Smc3 between control and Esco2 knock-down cells. Optimization of binning size enabled us to detect large-scale genome (a few Mb) regions where Esco2-dependent Smc3 acetylation occurred. These regions coincided with lately-replicated regions, or heterochromatic regions, and Esco2 was bound evenly to the same regions in early-to-mid, but not in late, S phase cells. Interestingly, MCM replicative helicase also showed very similar binding dynamics in S phase cells, implying that Esco2 binding to heterochromatin depends on pre-replication complex. Taken together, our results suggest that Esco2 fulfills replication-coupled cohesin acetylation primarily in heterochromatic regions, whereas Esco1 is rather enriched in transcriptionally active euchromatic region (Minamino et al., 2015).

A105**Genetic analysis of the cause of semaphorin 5A downregulation in human glioblastomas and exploration of its tumor suppressor function for brain cancer treatment**

Amrutha Menon, Fathima R Shah Jahan, Janice WS Law and Alan YW Lee
School of Pharmacy, Monash University Malaysia, Bandar Sunway,
Selangor, Malaysia

Correspondence: Alan YW Lee
Human Genomics 2018, **12(Suppl 1)**:A105

BACKGROUND

Semaphorins and their receptors plexins are families of axon guidance molecules. Accumulating evidence suggests their involvement in cancer development. We have recently demonstrated the effects of semaphorin 5A (Sema5A) in inhibiting migration, invasion, proliferation and tumorigenesis in human glioblastoma cells. Notably, analysis of clinical specimens of astrocytoma revealed a marked decline in Sema5A protein level from low- to high-grade. This led to our hypothesis that reduced expression of Sema5A causes a loss of its tumor suppressor function in glioblastomas, hence contributing to brain cancer development.

OBJECTIVES

To examine the mechanism underlying downregulation of Sema5A protein expression in high-grade glioblastomas, and to identify the anti-tumorigenic domains of Sema5A.

METHODS

Sema5A mRNA transcripts from human astrocytoma of different grades were subjected to quantitative real-time PCR and sequencing analysis to determine its transcript level and identify genetic aberrations, respectively. Candidate microRNAs (miRNA) that target Sema5A transcripts were predicted in-silico and their effect in downregulating Sema5A were validated by forced expression in glioblastoma cells. Anti-tumorigenic domains of Sema5A were mapped by systematic truncation and functional assays.

RESULTS

While neither mutation nor significant change in mRNA levels of Sema5A was observed across different grades of astrocytoma specimens, a number of alternatively spliced variants of Sema5A transcript were found to be expressed at comparable level as the wild-type counterpart. Ectopic expression of these frameshift splice variants in HEK293 cells produced premature truncation isoforms of Sema5A protein, which are unstable and readily targeted for proteasomal degradation. In-silico analysis of Sema5A gene predicts it as target of several miRNAs that are strongly up-regulated in glioblastomas. Forced expression of these candidate miRNAs in glioblastoma cells significantly reduces Sema5A protein level, which can be restored by treating the cells with corresponding anti-miR inhibitors. These findings suggest that both alternative splicing and translational repression by miRNA lead to a decline in Sema5A protein expression in glioblastomas. Importantly, both reconstitution of expression and exogenous supply of Sema5A is sufficient to suppress invasiveness and anchorage-independent growth of glioblastoma cells. To explore the therapeutic potential of Sema5A against glioblastomas, domains that exhibit high potency of anti-tumorigenic function have been identified.

CONCLUSIONS

Sema5A is downregulated in glioblastomas by aberrant splicing and miRNA-mediated gene silencing, which causes a loss of its tumor suppressor function. Identification of the anti-tumorigenic domains of Sema5A in this study allows further exploration of its therapeutic potential for brain cancer treatment.

KEYWORDS

Glioblastomas, semaphorins, alternative splicing, micro RNA, tumor suppressor

Disclosure of interest

This work is supported by the Fundamental Research Grant Scheme, Ministry of Education and the eScienceFund, Ministry of Science, Technology and Innovation, Malaysia. Presentation of findings in this project in conferences must be in the form of oral presentation, as stated by the funding agencies.

A106**Database for reprocessed public single-cell transcriptomics data (SCPortalen)**

Imad Abugessaisa¹, Shuhei Noguchi¹, Michael Böttcher¹, Akira Hasegawa¹, Tsukasa Kouno¹, Sachi Kato¹, Yuhki Tada², Hiroki Ura², Kuniya Abe², Jay W. Shin¹, Charles Plessy¹, Piero Carninci¹, Takeya Kasukawa¹
¹RIKEN Center for Life Science Technologies (CLST), Yokohama, Kanagawa 230-0045, Japan; ²RIKEN BioResource Center (BRC), Tsukuba, Ibaraki 305-0074, Japan

Correspondence: Takeya Kasukawa (takeya.kasukawa@riken.jp)
Human Genomics 2018, **12(Suppl 1)**:A106

Single-cell RNA-seq is a method to quantify expression profiles in the single-cell resolution, and it is widely used in various biomedical research projects. Single-cell transcriptome profiles obtained in these projects have been published in public repositories by the original data producers. These profiles are also useful for investigators who want to solve their own questions not targeted in the original project. However, there are several hurdles to reuse the published single-cell resources. One issue is that these data tend to miss a part of essential metadata, for example, sample treatment methods and sequencing specifications. Another issue is that these datasets are processed by different computational methods, and it makes hard to compare and integrate among datasets obtained by different groups.

To solve these issues, we reprocessed public single-cell RNA-seq data and constructed a new public database named "SCPortalen" (<http://single-cell.clst.riken.jp/>) to provide the reprocessed single-cell resources¹. In our reprocessing, we manually curated the original metadata and fixed missing items by reading original papers and sometimes directly contacting to authors. In this curation process, we also annotated sample ontology terms (cell ontology, Uberon, and so on) to single-cell samples to improve searchability and comparability. We then performed a unique quality assessment procedure to all single-cell datasets in order to provide mutually comparable quality metrics. After that, we reprocessed every single-cell transcriptome data using a unique pre-processing methods including mapping to the latest reference genome, quantification of expression levels, and standard analyses like PCA and t-SNE clustering. The all curated metadata and reprocessing results are available in the SCPortalen database. Users can browse them in their web browser and can download these files.

In the big-data era, how the large-scale data including published ones by other research groups could become more important strategies in research projects. We hope that our approach can be a hint for reuse the published data for the new science and the database can contribute to many research projects.

1. Abugessaisa I, Noguchi S, Böttcher M, Hasegawa A, Kouno T, Kato S, Tada Y, Ura H, Abe K, Shin JW, Plessy C, Carninci P, Kasukawa T. SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.* 2017; gkx949.

A107**The molecular basis of the cellular taxonomy of the human body**

Alessandra Breschi^{1,2}, Carrie Davis³, Sarah Djebali⁴, Jesse Gillis³, Dmitri D. Pervouchine⁵, Anna Vlasova¹, Alex Dobin³, Chris Zaleski³, Jorg Drenkow³, Cassidy Danyko³, Alexandra Scavelli³, Manuel Munoz¹, Diego Garrido^{1,2}, Ferran Reverter¹, Thomas R. Gingeras³, Roderic Guigo^{1,2}

¹Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain; ²Universitat Pompeu Fabra (UPF), Barcelona, Spain; ³Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11742, USA; ⁴GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, Castanet Tolosan, France
Human Genomics 2018, **12(Suppl 1)**:A107

The collection of RNA sequencing experiments produced by the ENCODE project from many cell lines and bulk tissues constitutes a comprehensive catalogue of the expression programs utilized by human cells. However, the relationship between the transcriptomes of tissues and the transcriptomes of the constituent primary cells, and how these impact tissue phenotypes has not been well established. Here we have produced RNA sequencing data for a number of primary cells from multiple human body locations. The analysis of this data, together with additional epigenetic data also produced by the ENCODE project for a total of 146 primary cells, indicate that most cells in the human body belong to five major cell types: epithelial, endothelial, mesenchymal, neural and blood cells. These redefine, based on gene expression, the basic histological types in which tissues are usually classified. We identified genes specific to these cell types, including a core set of transcription factors (TFs). Cell type specific genes, particularly when lying in open chromatin domains, are enriched for motifs for these cell type specific TFs, suggesting that they are potential candidates to drive cell type specificity. We estimated the relative proportion in tissues of the different cell types based on the transcriptional profiles obtained from bulk tissue sections. This inferred cellular composition is a characteristic signature of tissues and reflects.

A108**Know thy cells: the classification of tumor models by tissue, disease, sex, and species**

Heather Selby¹, Mark Kon^{1,2}, John Quackenbush^{3,4,5}, and Benjamin Haihe-Kains^{5,6,7,8,5}

¹Boston University, Bioinformatics Program, Boston, MA, USA; ²Boston University, Mathematics and Statistics, Boston, MA, USA; ³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA USA; ⁴Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA USA; ⁵Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada; ⁶Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; ⁷Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; ⁸Ontario Institute of Cancer Research, Toronto, Ontario, Canada

Correspondence: Heather Selby

Human Genomics 2018, **12(Suppl 1)**:A108

Cell lines and xenografts are models of patient tumors. Since the first patient-derived tumor cell line, HeLa, was created, new cell lines have been repeatedly established from patients. Large cell line collections now broadly capture the genomic diversity of patient tumors, and provide valuable insight into drug response. Patient-derived xenografts (PDXs) have begun to replace cell lines as more predictive experimental tumor models for drug development. PDXs are directly established from patient tumors in immune-deficient mice, and better resemble the actual patient tumor than cell lines. Recently, the U.S. National Cancer Institute (NCI) retired their NCI-60 panel of

patient-derived tumor cell lines to refocus their anti-cancer drug screening on PDXs.

Misidentified cell lines and xenografts are invalid models of patient tumors. The growth of cells in culture or xenografts makes tumor models susceptible to cross-contamination and/or mislabeling in spite of best laboratory practice. Although cell line misidentification is now widely-recognized, and authentication by short tandem repeat (STR) recommended, misidentified cell lines continue to be pervasive problems in cancer research. The International Cell Line Authentication Committee (ICLAC) has documented as many as 451 misidentified cell lines. Misidentified tumor models are often not from their original donor, and may even come from another tissue, disease, sex, or species.

To address the problem of misidentified tumor models, we developed a new classifier, Top Scoring-Binary Gene Pair (TS-BGP). The TS-BGP classifier, trained on tumor gene expression profiles, was used to predict the tissue, disease, sex, and species of both cell lines and PDXs. Binary gene pair signatures were extracted from the 31 tissues in the Genotype-Tissue Expression (GTEx) project, 32 diseases in The Cancer Genome Atlas (TCGA), 2 sexes in the GTEx project, and 2 species in the Encyclopedia of DNA Elements (ENCODE). The predictive accuracies of the leave-one-out-cross validations were 97.51% (GTEx; n=9115) for tissue, 90.89% (TCGA; n=10088) for disease, and 99.10% for sex (GTEx; n=9115), and 100% for species (ENCODE; n=622).

The TS-BGP classifier identifies tumor models that accurately resemble patient tumors. With tumor models that resemble patient tumors, data generated in pre-clinical research can successfully identify translatable drugs and therapies that demonstrate clinical efficacy. The TS-BGP classification of tumor models by tissue, disease, sex, and species helps ensure the credibility, reproducibility, and translation of pre-clinical research.

A109**RIKEN Ageing Resource Data Project: Single cell transcriptome and epigenetic changes in mice**

Tommy Terooatea^{1,4}, Tsukasa Kouno¹, Matteo Guerrini², Yasutaka Motomura², Naoko Sato², Takeshi Matsui², Sidonia Fagarasan², Hironobu Fujiwara³, Kazuyo Moro², Hiroshi Ohno², Ichiro Taniuchi², Kosuke Hashimoto¹, Piero Carninci¹, Aki Minoda^{1,4}

¹RIKEN CLST-DGT, Yokohama, Japan; ²RIKEN IMS, Yokohama, Japan; ³RIKEN CDB, Kobe, Japan; ⁴Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan

Correspondence: Tommy Terooatea; Aki Minoda

Human Genomics 2018, **12(Suppl 1)**:A109

Bulk analyses of high-throughput genomic and proteomic technologies have produced invaluable publicly available data on ageing. However, these data are limited in terms of accurately defining cellular states, development and disease state. Single-cell omics on the other hand help overcome these obstacles and promises greater understanding of cellular processes. Consequently, we are currently generating a **resource dataset for ageing studies** with a focus on single cell RNA-seq. We present here our initial results produced from various tissues taken at different ages of the mouse. One aspect of our analysis will be to determine whether we observe dynamic evolution of cells in heterogeneity throughout ageing, which may reveal key regulatory pathways that are gradually deregulated during ageing.

We will also present our plan for obtaining single cell ATAC-seq and multi-omic datasets (DNA methylation, CAGE, translome (ribosome profiling), proteomics and metabolomics) for the selected cell types of interest and integration analysis. Additionally, a unique aspect of our dataset is the utilization of both germ-free and SPF mouse models, which will enable us to introduce and access the effects of

various environmental factors in the ageing process. Taken together, these data will help us better understand the overall genomic and epigenomic landscape changes taking place in the cell during ageing and we hope that these will serve as invaluable resources for the scientific community.

A110

The single cell PBMC transcriptome of supercentenarians

Kosuke Hashimoto¹, Norihito Hayatsu¹, Tsukasa Kouno¹, Tomokatsu Ikawa², Haruka Yabukami¹, Tommy Terooatea¹, Akiko Minoda¹, Takashi Sasaki³, Yasumichi Arai³, Nobuyoshi Hirose³, Piero Carninci¹

¹RIKEN Center for Life Science Technologies, Yokohama, Japan; ²RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; ³Keio University School of Medicine, Tokyo, Japan

Correspondence: Nobuyoshi Hirose; Piero Carninci
Human Genomics 2018, 12(Suppl 1):A110

The supercentenarian, over age 110, is a great model of healthy ageing. We aim to understand successful ageing process using technologies for transcriptome analysis. Here, we report our preliminary results of the single cell transcriptome analysis on peripheral blood mononuclear cells (PBMC), derived from Japanese supercentenarians, their offspring, and unrelated controls. In total, 50,000 single cells from 11 donors were profiled using 10X Genomics single cell libraries. We successfully classified major immune cell types including T cells, B cells, monocytes, and dendritic cells based on the single-cell expression profiles. The cell population analysis confirmed that B cells and naive T cells are decreased in supercentenarians whereas memory T cells tend to be increased in supercentenarians. The population changes were validated using FACS with canonical surface markers, such as CD3, CD19, and CD56.

A111

Whole genome – and targeted capture – sequencing analysis of dual lung metastases and circulating tumour DNA (ctDNA) in a rare Primary Adrenocortical Carcinoma (ACC) case

Mark J McCabe^{1,2,3}, Mark Pinese^{2,4}, Nisa Sheriff^{1,5}, Monika Fazekas⁵, Ann I McCormack^{1,5}, Marcel E Dinger^{2,3,6}, Mark J Cowley^{2,3}

¹Hormones and Cancer Group, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; ²Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; ³St Vincent's Clinical School, UNSW Sydney, Darlinghurst, NSW, Australia; ⁴Cancer Research Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; ⁵St Vincent's Hospital, Darlinghurst, NSW, Australia; ⁶GenomeOne, Darlinghurst, NSW, Australia

Correspondence: Mark J McCabe
Human Genomics 2018, 12(Suppl 1):A111

Background ACC is an aggressive, rare malignancy with poor prognoses and limited treatment options. In 2015, a patient presented at our clinic with ACC and dual lung metastases. We performed whole genome sequencing (WGS) plus targeted-sequencing of both metastases. Furthermore, we developed a liquid biopsy assay via ctDNA, which we have applied quarterly and tracked with droplet digital PCR. We hypothesised that a genomics-based strategy may identify optimal treatment options for this patient, and that serial liquid biopsy may detect relapse earlier than imaging based approaches.

Materials and Methods WGS was performed by HiSeq – X of germline (~30X) and 2 metastases (~90X), followed by a custom Roche/Nimblegen panel designed to target ~400 cancer-associated genes (~0.9Mb) and applied to germline (~900X), metastases (~570X) and ctDNA (~300X) which had been subjected to library preparation using KAPA or Rubicon protocols, prior to sequencing on a HiSeq-2500. Droplet digital PCR analysis against variants of interest was performed using Bio-Rad's QX200 system. Bioinformatics followed GATK best practices for the identification of germline variants, with Strelka, Manta, Sequenza, and MANTIS analysis in the tumour (WGS), or Connor, Strelka and CNVkit for targeted sequencing.

Results No germline cancer predisposition variants were detected. Targeted-capture revealed 1306 variants within the metastases, of

which 362 were shared. WGS of both metastases revealed hyperdiploid cancers with extensive loss-of-heterozygosity but few structural rearrangements. Loss-of-function mutations were identified in cancer-associated genes including *TP53* and *PTEN*, and a splice-site mutation observed in *MSH2*, resulting in a tumour mismatch repair and consequent microsatellite instability. The quarterly ctDNA mutation profiles are quietas confirmed by droplet digital PCR, which is consistent with the patient being in remission.

Discussion We present an extensive WGS/capture-sequencing analysis of an ACC patient with lung metastases. Encouragingly, a critical splice-site mutation detected in *MSH2* appears to be a promising therapeutic target through immunotherapy, and we anticipate that ctDNA will sensitively reveal if the patient relapses.

A112

High-resolution characterization of drug-induced cellular response

Andrew T. Kwon, Masaaki Furuno, Takahiro Arakawa, Satoshi Takizawa, Efthymios Motakis, Yi Huang, Harukazu Suzuki, Erik Arner
¹RIKEN Center for Life Science Technologies, Yokohama, Japan

Correspondence: Erik Arner
Human Genomics 2018, 12(Suppl 1):A112

Drug response profiling is a powerful technique for drug mode of action analysis and repositioning, which allows for the characterization of cellular drug responses at a molecular level. However, the existing methods have been hampered by their low resolution, at both genomic and cellular scales, and their inability to account for the inherent cellular heterogeneity. We now have an opportunity to overcome these limitations with the recent advances in single cell high throughput transcriptomics approaches, as we can now resolve the minute differences in cellular make up and internal states by measuring the expression levels of both existing and novel promoters, enhancers, and lncRNA in an unbiased manner. Taken together, we can investigate the effects of population heterogeneity on their response to drugs and perform a more detailed mode of action analysis. In this study, we measure the drug response in 3 different cell types (fibroblasts, MCF7, and HepG2) to 2 types of histone deacetylase inhibitors (TSA and VPA) and examine how differential responses are produced. We employ the newly developed C1 CAGE method, which can robustly detect the expression of individual promoters and enhancers at a genome-wide scale, in a single experiment.

A113

Frequency and nature of CARD15/NOD2 mutations in patients with Crohn's disease in Kuwait

Fawziah Mohammed¹, Fatma Alyatama¹, Alaa Elshafey²
¹Faculty of Allied Health, Department of Medical Laboratory Sciences, Kuwait University, Kuwait City, Kuwait; ²Kuwait Medical Genetic Center, Ministry of Health, Kuwait City, Kuwait

Correspondence: Fawziah Mohammed
Human Genomics 2018, 12(Suppl 1):A113

Crohn's disease (CD) is a multi-factorial inflammatory bowel disease resulting from defects in mucosal immunity and intestinal epithelial barrier function. Genetic predisposition is important in susceptibility to CD disease. *CARD15/NOD2* was the first CD susceptibility gene identified. It has a protective effect in healthy individual by recognizing lipopolysaccharide components of the cell wall of the pathogenic bacteria responsible for CD. In the present study we investigated 9 polymorphic SNPs in the *CARD15/NOD2* gene in 64 CD Kuwaiti patients and 36 apparently normal controls of matched age and sex. The distribution of individual alleles between the tested groups were compared using the chi-square test. Only one SNP (IVS8⁺¹⁵⁸) showed a statistically significant lower (P value < 0.0009) representation in patient group assigning this SNP as a protective one in our population. To the best of our knowledge this is the first study investigated the frequency and nature of *CARD15/NOD2* mutations in Kuwaiti patients with CD. Further studies adding more SNPs are recommended to clarify the genetic nature of this devastating disorder in Kuwait.

Key Words: Crohn's disease, *CARD15/NOD2* gene, single nucleotide polymorphism, Kuwait

A114**Epigenetic dysregulation of AFP-producing gastric cancer**

Shihang Chen, Aya Nonaka, Takanori Fujita, Genta Nagae, Akihiro Suzuki, Amane Tagashira, Hiroyuki Aburatani

¹Genome Science Laboratory Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

Human Genomics 2018, **12(Suppl 1)**:A114

All over the world gastric cancer is the fourth common cancer and the second most common cause of cancer death. Alpha fetoprotein (AFP)-producing gastric cancer is a highly malignant subclass of the gastric cancer and was initially reported by Bourreille et al. in 1970. Most of studies demonstrated that AFP-positive gastric cancer had more aggressive clinical behavior than AFP-negative gastric cancer due to high incidence of lymphatic invasion, venous invasion and liver metastasis. Furthermore, high incidence of distant metastasis led to relatively poor overall prognosis. The underlying molecular mechanisms, which could explain why AFP-positive cancer shows hepatoid differentiation and more aggressive behavior, have not yet been elucidated. The aim of this study is to clarify the mechanism of transcriptional dysregulation which would be a key of its tumorigenesis and aggressive behavior in AFP-producing gastric cancer. AFP-positive gastric cancer cell lines (Fu97, Takigawa, GCIY) and an AFP-negative gastric cancer cell line (MKN7) were used to identify significant transcription factors by integrating RNA-seq, Chromatin immunoprecipitation followed by sequencing (ChIP-seq), and Assay for transposase-Accessible Chromatin using sequencing (ATAC-seq) data. Furthermore, by comparing them to RNA-seq and ChIP-seq data from human live cancer cell line (HepG2), we observed that a similar set of transcription factors (TFs) were overexpressed in AFP-producing gastric cancer, indicating the cause of AFP-producing gastric cancer. We are currently exploring the candidate genes targeted by these transcription factors and the mechanism for aberrant activation of those TFs in gastric epithelial cells.

A115**Alterations in the Rho pathway contribute to EBV-induced lymphomagenesis in immunosuppressed environments**

Sung-Yup Cho¹, Chang Ohk Sung^{2,3}, Jeessoo Chae⁴, Jieun Lee⁵, Deukchae Na¹, Wonyoung Kang^{2,5}, Jinjoo Kang⁵, Seoyeon Min⁵, Ahra Lee⁵, Eunhye Kwak⁵, Jooyoung Kim⁵, Boram Choi⁵, Hyunsoo Kim², Jeffrey H. Chuang², Hyo-Kyung Park^{3,6}, Chan-Sik Park³, Sanghui Park⁷, Young Hyeh Ko⁸, Dakeun Lee⁹, Jin Roh³, Min-Sun Cho⁷, Seongyeol Park¹⁰, Young Seok Ju¹⁰, Yun-Suhk Suh¹¹, Seong-Ho Kong¹¹, Hyuk-Joon Lee^{11,12}, James Keck¹³, Jacques Bancheureau², Edison T. Liu², Woo-Ho Kim^{12,14}, Hansoo Park¹⁵, Han-Kwang Yang^{11,12}, Jong-Il Kim^{4,12,16}, and Charles Lee^{1,2,5}

¹Ewha Institute of Convergence Medicine, Ewha Womans University Mokdong Hospital, Seoul, Korea; ²The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; ³Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea;

⁴Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; ⁵Department of Life Science, Ewha Womans University, Seoul, Korea; ⁶Cell Dysfunction Research Center, University of Ulsan College of Medicine, Seoul, Korea; ⁷Department of Pathology, Ewha Womans University College of Medicine, Seoul, Korea;

⁸Department of Pathology and Translational Genomics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; ⁹Department of Pathology, Ajou University School of Medicine, Suwon, Korea; ¹⁰Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea;

¹¹Department of Surgery, Seoul National University College of Medicine, Seoul, Korea; ¹²Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea; ¹³The Jackson Laboratory, Sacramento, CA, USA; ¹⁴Department of Pathology, Seoul National University College of Medicine, Seoul, Korea; ¹⁵Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, Korea; ¹⁶Medical Research Center, Genomic Medicine Institute (GMI), Seoul National University, Seoul, Korea

Correspondence: Sung-Yup Cho; Chang Ohk Sung

Human Genomics 2018, **12(Suppl 1)**:A115

Epstein-Barr virus (EBV)-positive diffuse large B cell lymphomas (EBV⁺-DLBLs) tend to occur in immunocompromised patients, such as the elderly or those undergoing solid organ transplantation. The pathogenesis and genomic characteristics of EBV⁺-DLBL are largely unknown because of limited availability of human samples and lack of experimental animal models. We observed the development of 25 human EBV⁺-DLBLs during engraftment of gastric cancer adenocarcinomas into immunodeficient mice. An integrated genomic analysis of the human-derived EBV⁺-DLBLs revealed enrichment of mutations in Rho pathway genes, including RHPN2, and Rho pathway transcriptional activation. Targeting the Rho pathway using the ROCK inhibitor, fasudil, decreased tumor growth markedly in EBV⁺-DLBL patient-derived xenograft (PDX) models. Thus, alterations in the Rho pathway appear to contribute to EBV-induced lymphomagenesis in immunosuppressed environments.

A116**Interethnic methylation difference on SLC7A5 explains ethnic disparity of response to tamoxifen therapy in breast cancer patients**

Hsin-Chou Yang, Shih-Kai Chu

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Correspondence: Hsin-Chou Yang (hsinchou@stat.sinica.edu.tw)

Human Genomics 2018, **12(Suppl 1)**:A116

Background

African-descendant breast cancer patients have exhibited a lower sensitivity to tamoxifen than do European-ancestry patients. Recently, we found the influence of interethnic DNA methylation difference on drug absorption, distribution, metabolism and excretion (ADME), suggesting that ethnicity should be carefully accounted for in pharmacogenomics [1]. This study aims to establish a mechanism for the differential sensitivity to tamoxifen in breast cancer patients from different ancestry populations through an interethnic difference of DNA methylation in ADME genes.

Materials and methods

We analyzed the whole DNA methylome and RNA transcriptome data of primary tumor tissues of 84 African-descendant and 508 European-descendant breast cancer patients from The Cancer Genome Atlas. A differential methylation analysis was performed to identify ethnicity-associated CpGs (E-CpGs) and a differential expression analysis was performed to identify differentially expressed genes. We constructed Bayesian networks by using PC algorithm and Tabu-search-based algorithm to examine a causal effect of the identified E-CpGs on differentially expressed genes related to drug ADME.

Results

We identified 35 E-CpGs located in or close to 24 ADME genes, where E-CpG cg05834639 on *SLC7A5* exhibited an interesting pattern: cg05834639 had a significantly lower methylation level in the African-descendant patients than that in the European-descendant patients ($p < 0.021$). The methylation level of cg05834639 was negatively associated with the gene expression of *SLC7A5* (correlation < -0.501); thus, the expression level of *SLC7A5* was significantly higher in the African-descendant patients than that in the European-descendant patients ($p < 5.25 \times 10^{-5}$). In the Bayesian network analysis, the results showed a stable directed link from E-CpG cg05834639 to the *SLC7A5* expression – a reproducibility rate of 66.8% by using the PC algorithm and a reproducibility rate of 99.8% by using the Tabu-search-based algorithm.

Conclusions

Our results showed that cg05834639 on ADME gene *SLC7A5* was an E-CpG and directly regulated the *SLC7A5* expression. It was reported that higher *SLC7A5* expression caused a poorer response to tamoxifen therapy in breast cancer patients [2]. Based on the evidences, this study has established a hypothetical mechanism about ethnic disparity of response to tamoxifen therapy in breast cancer patients through an interethnic difference of DNA methylation in ADME gene *SLC7A5*.

References

- [1] Chu, S.-K. & Yang, H.-C. Interethnic DNA methylation difference and its implications in pharmacoepigenetics. *Epigenomics*. 2017; 9, 1437-54.
- [2] Mihaly, Z., Kormos, M., Lanczky, A., et al. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer.

A117

Genetics of Childhood-Onset Psychosis

M.S Sato and A.C Need

Centre for Psychiatry, Division of Brain Sciences, Department of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London, W12 0NN, UK

Human Genomics 2018, **12(Suppl 1)**:A117

Childhood-Onset Psychosis (COP) is a very rare and debilitating disorder characterised by the onset of psychotic symptoms before age 14. Here we report early findings from a study exploring the potential of exome sequencing as a diagnostic tool in child psychiatry clinics. We hypothesise that some occurrences of COP are the result of known Mendelian disorders manifesting primarily with psychiatric symptoms.

In this study, we will extract saliva DNA from 100 trios, run pedigree analysis, examine rare schizophrenia-associated copy number variants (CNVs) and perform exome sequencing to identify rare (MAF<1%) inherited or *de novo* SNVs or INDELS.

To date we have recruited 27 families (12 male and 15 female probands) and documented 3-generation pedigrees per case. All probands have a primary diagnosis of a psychotic illness with 12/27 having a comorbid neurodevelopmental disorder. According to family report 9/27 probands have a family history of a psychiatric diagnosis. Cognitive assessment (WASI-II) on the first 17 probands shows a variation in their full-scale intelligence quotient (FSIQ) whereby 6/17 scored extremely low (FSIQ<70), 7/17 scored borderline to low average (FSIQ 70-90) and 4/17 scored average (FSIQ 90-110). In addition, 8/27 probands are on Clozapine, highlighting their disease severity.

Genetic analysis of the first 7 probands indicates that none have common schizophrenia-associated CNVs, however exome sequence analysis revealed one proband to be harbouring a very rare pathogenic variant in gene RET, giving rise to a Pheochromocytoma – a phenotype associated with the Mendelian disease, multiple endocrine neoplasia type 2a that can manifest with psychosis. Determination of the pathogenicity of the RET variant in causing psychosis in the probands will be investigated with further clinical assessments.

The identification of rare highly penetrant variants that predispose patients to psychosis suggests that exome sequencing may be useful in the child psychiatric clinics as a diagnostic tool, as well as providing a platform for understanding the neurobiology underlying COP.

A118

Microarray molecular profiling to define pathogenesis and biomarkers of atherosclerosis

Maria S. Nazarenko^{1,2,3}, Aleksei A. Sleptcov¹, Anton V. Markov¹, Igor N. Lebedev^{1,2}, Nikolay A. Skryabin^{1,2}, Anton N. Kazancev³, Olga L. Barbarash³, Ludmila P. Nazarenko^{1,2}, Valery P. Puzyrev^{1,2}

¹Research Institute of Medical Genetics of Tomsk NRMС, Tomsk, 634050, Russia; ²Tomsk State University, Tomsk, 634050, Russia; ³Research Institute for Complex Problems of Cardiovascular Diseases, Kemerovo, 650002, Russia

Correspondence: Maria S. Nazarenko (maria.nazarenko@medgenetics.ru)

Human Genomics 2018, **12(Suppl 1)**:A118

Background

We hypothesize that microarray molecular profiling of vascular and blood cells of patients with severe atherosclerosis will be important to clarify the disease mechanisms and biomarkers.

Materials and methods

The Agilent SurePrint G3 Human CGH+SNP 2x400K and Illumina HumanMethylation27 BeadChip microarrays were used for DNA testing from advanced atherosclerotic plaques of carotid and coronary arteries, and healthy internal thoracic arteries. The results for certain loci were verified by quantitative real time PCR and bisulfite pyrosequencing with blood and vascular tissues. The study was approved by Research Institute of Medical Genetics' Ethics Board, approval number 0145.

Results

In total, we identified 90 high confidence copy number variations (CNVs) in arteries. We found that 16% CNVs regions were enriched with genes previously associated with mainly CAD risk factors but not the disease as a whole. The identified CNV genes were associated with immune/inflammation responses, olfactory transduction and metabolic pathways. The gain in chromosomal region 10q24.31 (*ERLIN1*) was not listed in the Database of Genomic Variants. Furthermore, two patients contained the gain in 10q24.31 (*ERLIN1*) that affected only the blood DNA. Overall, the disease-related DNA methylation effect size was relatively modest. Only a minority of differentially methylated genes have previously been linked to atherosclerosis-related diseases in gene based association studies. The advanced atherosclerotic plaques in comparison with the healthy arteries were characterized by the predominant DNA hypermethylation changes. These genes were annotated with muscle system process and positive regulation of cytosolic calcium ion concentration in Gene Ontology terms. In contrast, hypomethylated genes encode molecules belonging to different biological processes such as development, immune/inflammation responses, lipid storage, and programmed cell death. In advanced atherosclerotic plaques the most pronounced hypomethylation was registered in 2q31.1 (*HOXD4/HOXD3/MIR10B*). Moreover, methylation changes at this locus in blood cells were consistently associated with smoke and ischemic stroke.

Conclusions

Analysis of DNA extracted from blood of patients with severe atherosclerosis indicated a possible somatic origin for CNV. The vast majority of CpG-sites were hypermethylated in the atherosclerotic plaques and connected with genes involved in smooth muscle cell function. Functional enrichment analysis of CNVs and differentially methylated genes demonstrated that common term was related to immune/inflammation responses. Our data support the hypothesis that CNVs and DNA methylation could be new biomarkers to uncover susceptibility loci for future atherosclerosis association studies with larger cohorts, paired tissues and more sensitive single cells methods.

Acknowledgements

The study was supported by the Russian Science Foundation (16-15-10150).

A119

Array-CGH analysis of cells dissected from atherosclerotic plaque

Aleksei A Sleptcov¹, Maria S Nazarenko^{1,2,3}, Nikolay A Skryabin^{1,2}, Evgeny V Denisov⁴, Liubov A Tashireva⁴, Igor N Lebedev^{1,2}, Valery P Puzyrev^{1,2}

¹Research Institute of Medical Genetics of Tomsk NRMС, Tomsk, 634050, Russia; ²Tomsk State University, Tomsk, 634050, Russia; ³Research Institute for Complex Problems of Cardiovascular Diseases, Kemerovo, 650002, Russia; ⁴Cancer Research Institute of Tomsk NRMС, Tomsk, 634050, Russia

Correspondence: Aleksei A Sleptcov (alexei.sleptcov@medgenetics.ru)

Human Genomics 2018, **12(Suppl 1)**:A119

Background

Macrophages and smooth muscle cells (SMCs) play a significant role in atherosclerosis. Genomic alteration in cells arising during an atherosclerotic process can be important for plaque instability. Our purpose was an assessment of genomic alterations in cells obtained from atherosclerotic plaque using single-cell microarray technique.

Materials and methods

Laser capture microdissection of eight fresh frozen autopsies of coronary arteries with atherosclerosis was performed to obtain macrophages (n = 30-40 cells per sample; Anti-Human CD68 Antibody) and smooth muscle cells (Alpha-SMA Antibody, 1A4). Whole genome amplification (GenomePlex® WGA Kit) of these cells along with white blood cells from the same persons was performed with subsequent array-CGH (SurePrint G3 Human CGH 8x60K Microarray). The study was approved by Research Institute of Medical Genetics' Ethics Board, approval number 0159.

Results

We found several aneuploidies in macrophages (9 trisomies and 3 monosomies) and SMCs (1 trisomy and 8 monosomies). In two patients, SMCs had a normal karyotype. The ratio of duplications and deletions in macrophages and SMCs was 1:0.8 and 1:7, respectively. Comparative analysis of genomic rearrangements in macrophages demonstrated that duplication in chromosome region 9q34.13-q34.2, about 2Mb in size, was detected in macrophages in 6 patients.

Conclusions

Our study indicates that genomic alterations are diverse in their structure and are widely represented in cells involved in the atherosclerotic process. In macrophages obtained from atherosclerotic plaques, aneuploidy with a predominance of trisomy was found, while monosomies prevailed in SMCs. A significant prevalence of deletions in SMCs was found. Duplication in macrophages in chromosomal region 9q34.13-q34.2 contains genes that can potentially be associated with atherosclerosis.

A120

Identification of novel RNA biomarkers through transcriptomic analysis of Parkinson's disease patients

Matthew N.Z. Valentine¹, Kosuke Hashimoto¹, Shinji Saiki², Takeshi Fukuhara², Keiichi Ishikawa², Nobutaka Hattori² & Piero Carninci¹

¹Division of Genomic Technologies, RIKEN CLST, Yokohama, 230-0045, Japan;

²Department of Neurology, Juntendo University, Tokyo, 113-8431, Japan

Correspondence: Matthew N.Z. Valentine (matthew.valentine@riken.jp)

Human Genomics 2018, **12**(Suppl 1):A120

Background

Parkinson's disease (PD) is an age-related, chronic and progressive neurodegenerative disorder characterized by a loss of dopaminergic neurons, resulting in both non-motor and motor symptoms. While several genetic and environmental contributory risk factors have been identified, more exact methods for diagnosing and assessing prognosis of PD have yet to be discovered. Identifying altered expression of genes (both coding and non-coding) and pathways by analyzing transcriptomes of early-stage PD patients reveals potential novel biomarkers for PD diagnosis.

Methods

Changes in expression at the transcriptome level were assessed using no-amplification non-tagging cap analysis of gene expression (nAnt-iCAGE) over two years, starting with 22 PD patients and 10 healthy controls in the first year and with 17 PD patients and 10 healthy controls in the second year. Sequencing data was mapped to the hg38 human annotation, CAGE tags were clustered and differential expression analyses were performed with edgeR and DESeq2. Neural stem cell (NSC) and dopaminergic (DA) neuron pairs derived from iPS were studied in a similar

manner. Gene ontology, gene set enrichment and network analyses were also performed.

Results

Blood transcriptome changes between PD patients and healthy controls are minimal, though a number of differentially expressed transcripts can be detected in early stage PD, including a number of non-coding RNAs. Differential expression was more readily apparent in iPS-derived NSC and DA neurons, though key differences appear between the different cell types. Some differentially regulated transcripts have known links to PD or other neurodegenerative disorders, though many are novel in nature, and may constitute putative biomarkers.

Conclusions

Our data suggests that though the blood transcriptome is largely similar between disease and healthy states, there is still benefit to thorough analyses for the study of neurodegenerative disease. We identified a number of potential candidate biomarkers that are now being validated.

Ethics Approval

The study was approved by the ethics evaluation committee of Juntendo University (Approval Number: 15-104) and the ethics review committee of RIKEN (H26-27).

A121

Three-dimensional structure of human genome at the population scale

Michał Sadowski^{1,¶}, Przemysław Szalaj^{1,¶}, Agnieszka Kraft^{1,¶}, Zhonghui

Tang², Paul Michalski², Yijun Ruan², Dariusz Plewczynski^{1,3}

¹Centre of New Technologies, Warsaw University, Warsaw, Poland; ²The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; ³Faculty of Mathematics and Information Science, Warsaw University of

Technology, Warsaw, Poland

Correspondence: Dariusz Plewczynski (d.plewczynski@cent.uw.edu.pl)

Human Genomics 2018, **12**(Suppl 1):A121

#contributed equally

The recent insights in the higher order chromatin organisation in human nucleus [1] allowed us to propose the 3D-GNOME (3-Dimensional GeNOme Modeling Engine), a complete computational pipeline for 3D simulation using Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) data [2]. 3D-GNOME consists of three integrated components: a graph-distance-based heatmap normalization tool, a 3D modeling platform, and an interactive 3D visualization tool. Using ChIA-PET and Hi-C data derived from human B-lymphocytes, we demonstrate the effectiveness of 3D-GNOME in building 3D genome models at multiple levels, including the entire genome, individual chromosomes, and specific segments at megabase (Mb) and kilobase (kb) resolutions of single average and ensemble structures. Further incorporation of CTCF-motif orientation and high-resolution looping patterns in 3D simulation provided additional reliability of potential biologically plausible topological structures.

Additionally, we developed 3D-GNOME web service [3], which generates 3D structures from 3C data and provides tools to visually inspect and annotate the resulting structures, in addition to a variety of statistical plots and heatmaps which characterize the selected genomic region. 3D-GNOME simulates the structure and provides a convenient user interface for further analysis. Alternatively, a user may generate structures using published ChIA-PET data for the GM12878 cell line by simply specifying a genomic region of interest. 3D-GNOME is freely available at <http://3dgenome.cent.uw.edu.pl/> providing unique insights in the topological mechanism of human variations and diseases.

Finally, we present yet not published the catalogue of topological variability for human domains as observed in 2,504 genomes

sequenced in the 1000 Genome Project [4]. We use a computational method for modelling three-dimensional structure of chromatin on multiple scales reaching the high-resolution scale of individual chromatin loops to model the high-resolution chromatin structure. We propose the first approach to modelling conformational changes induced by structural variants at the level of chromatin loops. We show that structural variants, which influence strongly gene transcription levels, localize preferably in genomic sites brought together by RNA polymerase II to create loops and show evidence that they have a significant influence on CTCF-mediated chromatin loops. The results are validated using Hi-C datasets for trios of 1000 Genomes Project from Bing Ren laboratory. Our unpublished results assess the impact of structural variants on the organization of chromatin topological domains and sheds light on the mechanisms regulating gene transcription, among which the local arrangement of chromatin loops seems to be the leading one.

References

- [1] "CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription" Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, Michalski P, Piecuch E, Wang P, Wang D, Tian SZ, Penrad-Mobayed M, Sachs LM, Ruan X, Wei CL, Liu ET, Wilczynski GM, Plewczynski D, Li G, Ruan Y. *Cell*. 2015 Dec 17;163(7):1611-27. doi: 10.1016/j.cell.2015.11.024.
- [2] "An integrated 3-Dimensional Genome Modeling Engine for data-driven simulation of spatial genome organization" Szalaj P, Tang Z, Michalski P, Pietal MJ, Luo OJ, Sadowski M, Li X, Radew K, Ruan Y, Plewczynski D. *Genome Res*. 2016 Dec;26(12):1697-1709.
- [3] "3D-GNOME: an integrated web service for structural modeling of the 3D genome" Szalaj P, Michalski PJ, Wróblewski P, Tang Z, Kadlof M, Mazzocco G, Ruan Y, Plewczynski D. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W288-93. doi: 10.1093/nar/gkw437.
- [4] "A global reference for human genetic variation" by 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. *Nature*. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393.
- [5] "Multi-platform discovery of haplotype-resolved structural variation in human genomes" by Mark JP. Chaisson, Ashley D. Sanders, Xuefang Zhao, Ankita Malhotra, ..., Ryan E. Mills, Tobias Marschall, Jan Korbel, Evan E. Eichler, Charles Lee. *BioRxiv* 2017 Sept 23 <https://www.biorxiv.org/content/early/2017/09/23/193144>. Submitted to Nature 2017. doi: <https://doi.org/10.1101/193144>

A122

The FANTOM5 atlas of microRNAs and their promoters in vertebrates

Michiel J.L. de Hoon^{1,2}, Akira Hasegawa^{1,2}, Takeya Kasukawa¹, Timo Lassmann^{1,2,3}, Shohei Noma^{1,2}, Hiroshi Tarui^{1,2}, FANTOM Consortium, Guojun Sheng^{5,6}, Yoshihide Hayashizaki^{2,4}, Piero Carninci^{1,2}, Hideya Kawaji^{1,2,4}, Alistair R.R. Forrest^{1,2,7}

¹Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan; ²RIKEN Omics Science Center (OSC), Yokohama, Japan; ³Telethon Kids Institute, The University of Western Australia, Subiaco, Western Australia, Australia; ⁴RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Japan; ⁵International Research Center for Medical Sciences (IRCMS), Kumamoto University, Kumamoto, Japan; ⁶RIKEN Center for Developmental Biology, Kobe, Japan; ⁷Harry Perkins Institute of Medical Research, and the Centre for Medical Research, University of Western Australia, QEII Medical Centre, Perth, Western Australia, Australia

Correspondence: Alistair R.R. Forrest
Human Genomics 2018, 12(Suppl 1):A122

MicroRNAs (miRNAs) are short non-coding RNAs that act as negative regulators by binding to complementary sequences, usually located in the 3' UTR of messenger RNAs. As part of FANTOM5 (the fifth edition of the Functional Annotation of Mammalian Genome project), we previously created a broad expression atlas of miRNAs in human and mouse, based on a diverse set of 500 short RNA sequencing libraries, including many from human primary cell types

[1] Each short RNA library in this analysis had a matching Cap Analysis Gene Expression (CAGE) library produced from the sample RNA sample, allowing us to identify the transcription start site and hence the promoter region of miRNAs.

We have now extended this atlas with additional short RNA sequencing libraries from rat, dog, and chicken samples, each again with a matching CAGE library. Using the miRDeep2 program [2], we analyzed the short RNA libraries to identify novel miRNAs in rat, dog, and chicken. In addition, we applied our miRNA promoter identification pipeline to the CAGE data to annotate the transcription start sites and thus the promoter regions of primary miRNAs in these three organisms. The combined data form the basis for detailed study and comparative analysis of the genomic sequence regions that orchestrate miRNA expression across vertebrate species.

Acknowledgements

FANTOM5 was made possible by the following grants: Research Grant for RIKEN Omics Science Center from MEXT to YH; Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan to YH; Research Grant from MEXT to the RIKEN Center for Life Science Technologies; Research Grant to RIKEN Preventive Medicine and Diagnosis Innovation Program from MEXT to YH. ARRF is supported by a Senior Cancer Research Fellowship from the Cancer Research Trust, funds raised by the MACA Ride to Conquer Cancer, and the Australian Research Council's Discovery Projects funding scheme (DP160101960). We would also like to thank GeNAS for data production.

References

1. De Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, Åström G, Babina M, Bertin N, Burroughs AM, Carlisle AJ, Daub CO, Detmar M, Deviatiiarov R, Fort A, Gebhard C, Goldowitz D, Guhl S, Ha TJ, Harshbarger J, Hasegawa A, Hashimoto K, Herlyn M, Heutink P, Hitchens KJ, Hon CC, Huang E, Ishizu Y, Kai C, Kasukawa T, Klincken P, Lassmann T, Lecellier CH, Lee W, Lizio M, Makeev V, Mathelier A, Medvedeva YA, Mejhert N, Mungall CJ, Noma S, Ohshima M, Okada-Hatakeyama M, Persson H, Rizzu P, Roudnický F, Sætrum P, Sato H, Severin J, Shin JW, Swoboda RK, Tarui H, Toyoda H, Vitting-Seerup K, Winteringham L, Yamaguchi Y, Yasuzawa K, Yoneda M, Yumoto N, Zabierowski S, Zhang PG, Wells CA, Summers KM, Kawaji H, Sandelin A, Rehli M, FANTOM Consortium, Hayashizaki Y, Carninci P, Forrest ARR, De Hoon MJL. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnol*. 2017; 5: 872-878.
2. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel miRNA genes in seven animal clades. *Nucleic Acids Res*. 2012, 40: 37-52.

A123

New insights into the human heart development using a combined spatial and single-cell transcriptomics approach

Stefania Giacomello^{1,2}, Michaela Asp², Eva Wårdell³, Johan Reimegård⁴, Fredrik Salmén², Karl - Henrik Grinnemo⁴, Agneta Månsson-Broberg³, Cecilia Österholm Corbascio³, Christer Sylvén³, Patrik Ståhl², Matthias Corbascio⁵, Joakim Lundeberg²

¹Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; ²Department of Gene Technology, Royal Institute of Technology, SciLifeLab, Stockholm, Sweden; ³Department of Medicine, Karolinska Institute, Stockholm, Sweden; ⁴Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; ⁵Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

Correspondence: Stefania Giacomello
Human Genomics 2018, 12(Suppl 1):A123

The heart is the first organ that becomes functional in the human embryo, beginning to beat and pump blood three weeks after fertilization. Although functional at this early stage, the developmental process that the nascent heart must follow is complex. Currently, how mature cardiomyocytes develop is not completely understood. Moreover, no defined stem cell, which can give rise to all components of the adult heart, exists. Here we studied the human fetal heart development in a spatiotemporal approach by applying both single-cell and spatial RNA-sequencing.

Specifically, we analyzed a 6.5-week-old human fetal heart studying ~4,000 dissociated single cells by using the 10x Chromium technology. Principal component analysis of the single cell gene expression clearly identified three clusters, which known marker genes assigned to outflow tract, atria and ventricles. Dimensionality reduction of the same information identified instead 17 clusters revealing a high heterogeneity of the outflow tract. Cell fate trajectory inference identified three clear branches of cell differentiation. Progenitor cells and cardiomyocytes were ordered early and late in pseudotime, respectively, as expected, while the third branch was localized early in pseudotime, and contained cells characterized by stemness.

We confirmed and extended these results by using Spatial Transcriptomics, a technology to visualize the gene expression information in tissue sections at 100 μm resolution. We studied the spatial gene expression of a second 6.5-week-old human heart, and of two additional human fetal hearts at 5 and 9 weeks of development. We identified a clear gene expression pattern determined by the heart structure, where the outflow tract had a distinctive gene expression signature compared to atria and ventricles in all three time points.

To conclude, the combination of Spatial Transcriptomics and single-cell sequencing demonstrates functional clusters related to the anatomical architecture of the heart. Moreover, our combined approach provides the spatial location of cells with stem cell signatures and complete gene profiles of progenitors committed to the cardiomyocyte lineage. Altogether, these findings offer new insights in the field of human embryology and cardiovascular medicine opening up questions for future studies in the field of myocardial regeneration.

A124

Accurate prediction of chromatin conformation status using deep neural network model

Hidetaka Uryu¹, Kenichiro Hata²

¹Department of Medical Genome Center, National Research Institute for Child Health and Development, Tokyo, Japan; ²Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan

Human Genomics 2018, 12(Suppl 1):A124

It has been strongly suggested that chromatin conformations in nuclei play an important role in regulating gene expression, and many reports have demonstrated that destructions of these three-dimensional (3D) structures lead to disease status. Recent advances of high-throughput sequencing have enabled it to display 3D proximities among each genomic locus on the chromatin contact matrix using Hi-C analysis method. However, large-scale computational resources are indispensable in current protocols of Hi-C analysis for high-resolution contact map, and the repeatability of those contact matrix are still uncertain in the current practical procedures. These defects prevent frequent use of Hi-C analysis.

To obtain repeatable matrix of chromatin conformations without heavy computational task, we developed a new algorithm to predict chromatin 3D structures using deep neural network model with the feature vectors based on genomic sequences, map positions on reference genome sequence, and epigenetic data from chromatin immunoprecipitation-sequencing (ChIP-Seq).

To evaluate our algorithm, we establish a model to predict the chromatin 3D structure of nucleus in a human lymphoblastoid cell line GM12878, and assessed its ability to predict the existence of chromatin looping conformations. The information of map position on human reference sequence (UCSC hg19), and DNA sequences were obtained from prior reports (Rozowsky et al., 2012), and ChIP-Seq data associated with epigenetic status were obtained from the ENCODE datasets. These alignment data were vectorized and normalized before inputting into our models. For supervised learning of neural network model, we prepared the combination of genomic loci tagged as chromatin-loop-positive ($N = 8526$) using HiCCUPs protocol as previously reported (Rao et al 2014), and also prepared randomly selected genomic loci combinations other than positively-tagged loci as chromatin-loop-negative ($N = 131365$). These data were randomly split into training, validation, and test data set. We configured the binary

classification model with the RMSprop optimizer, and the binary cross entropy loss function. Other hyper-parameters of the model were optimized using Bayes optimization method. Model trainings were continued until the validation loss no longer improves. We showed ROC curves and measured the areas under the ROC curve (AUC) to quantify the prediction performance. The model we selected as the best-performed one displayed its high accuracy (96.25%), sensitivity (99.35%), and specificity (95.22%). The AUC of the model was 0.997 against a test data set.

These results suggest that our algorithm to predict chromatin conformation using neural network is an effective method for analyzing chromatin 3D structures substituted for Hi-C analysis.

Performance of models are measured by the accuracy, sensitivity, specificity, and areas under the receiver-operating characteristic curve (AUC) on whole test data set. The training and prediction of each model was performed as follows: model A with only mapping positions on reference genome as input data, model B with only genomic sequences, model C with only epigenomic data, and model D with mapping positions, genomic sequences, and epigenomic data. A cut-off value of 0.5 was used to determine their classification results. We displayed the representative results of each model. The ROC curve of the best-performed model against whole test data set is displayed with a red-colored solid line. A blue-colored dotted line means a random guessing line.

Table 1 (abstract A124). See text for description

Model types	A	B	C	D
Model accuracy on whole test data	92.34%	91.53%	91.87%	98.25%
Sensitivity	97.20%	95.35%	96.86%	99.35%
Specificity	93.77%	94.79%	92.81%	99.62%
AUC	0.912	0.897	0.923	0.997

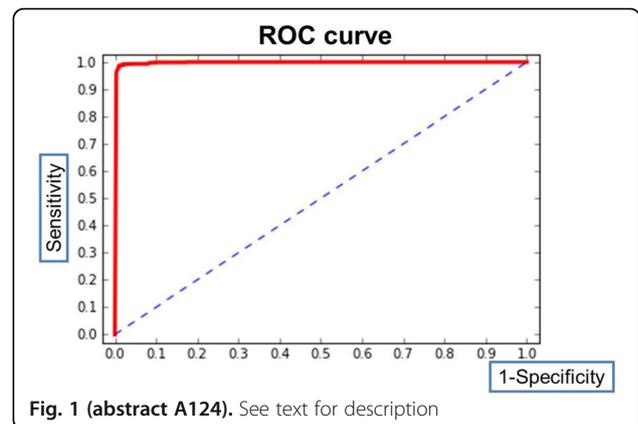


Fig. 1 (abstract A124). See text for description

A125

Distinctive gene expression signatures and *FLT3-ITD* mutation in acute promyelocytic leukemia patients

Imilia Ismail^{1,2}, Nor Azimah Abdul Azize³, Yusnita Yakob³, Sarina Sulong⁴, Muhammad Farid Johan², Rosline Hassan²

¹School of Biomedicine, Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300, Kuala Terengganu, Terengganu, Malaysia; ²Department of Hematology, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150, Kubang Kerian, Kelantan, Malaysia; ³Molecular Diagnostics and Protein Unit, Specialised Diagnostic Centre, Institute for Medical Research, 50588, Jalan Pahang, Kuala Lumpur, Malaysia; ⁴Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150, Kubang Kerian, Kelantan, Malaysia

Correspondence: Imilia Ismail (imilia@unisza.edu.my)

Human Genomics 2018, 12(Suppl 1):A125

Background

Acute promyelocytic leukemia (APL) is a subtype of acute myeloid leukemia (AML). It is characterized by a balanced reciprocal translocation t(15;17) (q22;q11-12) resulting in the fusion between the promyelocytic leukemia (*PML*) gene and the retinoic acid receptor α (*RAR α*) gene. *FLT3-ITD* mutation was reported to be frequently associated with APL as compared to other AML subtypes. Gene expression profiling has been extensively studied to enhance our understanding of the disease biology and to identify novel therapeutic targets. In this study, we analyzed the gene expression profiles and *FLT3-ITD* mutation in newly diagnosed APL patients that molecularly confirmed *PML-RAR α* fusion gene.

Materials and methods

We analyzed the gene expression profile of 9 newly diagnosed APL patients using Nanostring nCounter system (NanoString Technologies, USA). Mutation analysis of *FLT3-ITD* gene at exon 14 was performed by PCR and direct DNA sequencing. Data for gene expression analysis and DNA sequencing were analyzed using nSolver software v2.6 (NanoString Technologies, USA) and SeqScape software v3 (Applied Biosystem, USA), respectively.

Results

Among 770 genes analyzed, 252 genes were found to be differentially expressed in newly diagnosed APL patients ($p < 0.05$, > 2.0 -fold change). In total 252 differentially expressed genes, of which 98 were up regulated, while 154 were down regulated. The most significantly up regulated genes in newly diagnosed APL patients included *MPO*, *CCNA1*, *PPARG*, *FLT3* and *PLAU*, while *TNFRSF10C*, *PRKAR2B*, *HSPA6*, *LEF1* and *ITGB3* were the most significantly down regulated genes. In addition, up regulated of *FLT3* was detected in all newly diagnosed APL patients (at least 15-fold increase compared to normal controls) with *FLT3-ITD* mutation. Interestingly, up regulated of *FLT3* was also found in 5 patients without *FLT3-ITD* mutation. White blood cell count was significantly ($p < 0.05$) higher in patients with *FLT3-ITD* mutation.

Conclusions

In summary, this study showed the distinctive gene expression signatures in newly diagnosed APL. Further studies with a larger number of samples are required to investigate the clinical and biological significance of up regulated of *FLT3* in the absence of *FLT3-ITD* mutations.

A126

Characterization and functional investigation of alternative splicing of P4HB gene

Daniela Kajihara^{1,2}, Chung-Chau Hon², Kosuke Hashimoto², Piero Carninci², Francisco RM Laurindo¹

¹Vascular Biology Laboratory, Heart Institute (Incor), University of São Paulo School of Medicine, São Paulo, Brazil; ²RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, Japan

Correspondence: Daniela Kajihara (daniela.kajihara@riken.jp)

Human Genomics 2018, **12**(Suppl 1):A126

Alternative splicing (AS) is a process that produces different mRNAs from a primary transcript, resulting in the production of protein isoforms which can exhibit distinct functional properties and subcellular localization (Giampietro et al., 2015). More than 90% of human genes undergo AS (Pan et al., 2008; Wang et al., 2008). P4HB (also known as PDIA1) is the prototype of protein disulfide isomerase family with dithiol/disulfide oxidoreductase chaperone properties, primarily from the endoplasmic reticulum (ER), in which it supports redox protein folding. PDIA1 is largely expressed and one of the most active proteins to support ER function (Lovat et al., 2008; Galligan et al., 2012). Our group has characterized a number of PDIA1 functions in redox signaling, homeostasis and biomechanical responses, in some cases involving additional subcellular protein locations such as in the cell-surface or extracellular milieu (Laurindo et al., 2012; Tanaka et al.,

2016). The isomerase activity of PDI requires the presence of all its catalytic redox domains (a and a') and the non-redox domains (b and b'), while distinctly structured isoforms could as oxidoreductases or just as chaperones. However, there is very little information about PDI isoform transcript regulation, expression and the localization of different protein products. Recent advances in the NGS (Next Generation Sequencing) and proteomic data provided more accurately information from alternative splicing isoforms. Our objective is to investigate the occurrence and possible functional implications of P4HB isoforms derived from alternative splicing since PDI is a multi-protein domain. The FANTOM5 (Functional Annotation of Mammalian Genomes collaborative project) gathered biological information including RNA-seq data and the ENCODE/Caltech RNA-seq data allowed us to investigate the expression of P4HB gene isoforms in humans. The advantage of using data from FANTOM5 is the availability of information about the TSS (transcriptional start sites) of genes employed by Cap Analysis of Gene Expression (CAGE). We evaluated different expression of 14 P4HB isoforms in human genome from two different databases. All the isoforms are supported by ENSEMBL. The most representative samples considering all isoforms were whole blood and different cancer cell lines. The P4HB-029 and P4HB-026 isoforms depict parts of b and b' domain of PDI protein, allowing one to propose their possible roles as non-redox chaperones. Other isoforms as P4HB-027 and P4HB-02 are representative in neural cells. Investigation of alternative spliced isoforms of P4HB may yield relevant information regarding yet unknown issues related to PDI biology, including subcellular location, secretion and specific functions.

A127

Whole genome sequencing in a clinical setting is expanding our understanding of the phenotypic spectrum of rare diseases

Georgina Hollway^{1,2}, Madhura Bakshi^{1,2,3}, Alison Colley³, Lisa Ewans^{1,2}, Thomas Ohnesorg¹, Eric Lee¹, Edwin Kirk¹, Chiyan Lau¹, Kathy Wu⁴, Ebony Richardson¹, Aaron Statham¹, Leslie Burnett^{1,2}, Ben Lundie^{1,2} and Marcel Dinger^{1,2}

¹Genome.One, Darlinghurst, Sydney, NSW, Australia; ²Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia; ³Clinical Genetics Services SWSLHD, Liverpool Hospital, Liverpool, NSW, Australia; ⁴Clinical Genomics Unit, St Vincent's Hospital, Darlinghurst, Sydney, NSW, Australia

Correspondence: Georgina Hollway (georgina.hollway@genome.one)
Human Genomics 2018, **12**(Suppl 1):A127

Traditionally the majority of rare diseases have been clinically diagnosed through descriptions of recognisable constellations of observable characteristics. The increasing use of whole genome sequencing and analysis for diagnosis has expanded our understanding of the spectrum of phenotypes for a variety of rare disorders.

Genome.One is an ISO15189 accredited clinical laboratory service, applying whole genome sequencing to genetically diagnose a wide range of disorders. Since 2016, clinical reports have been issued on more than 240 families, comprising over 400 individuals.

Clinical diagnoses are largely driven by the presence of features that are considered essential for the diagnosis of a specific rare disorder. However, the paradigm for rare disease diagnosis is being challenged by the increasing use of genomics. Molecular diagnoses for rare genetic diseases are now made in cases that were not previously clinically diagnosable due to lack of classical characteristics. Genomic technologies are thus enabling both increased diagnosis rates and the expansion of clinical phenotypes associated with specific genes.

We demonstrate that the classic phenotypes of some rare diseases represent only the most severe end of what is in fact a wider spectrum of disease, through examples from cases

investigated at GenomeOne. We describe a number of cases where molecular diagnoses have been made through whole genome sequencing that have provided new and unexpected clinical diagnoses, ending some patient diagnostic odysseys of many years. These cases highlight the importance of assessing all variants in human disease genes identified through genomic testing in rare disease, even if the classical picture of that disease does not immediately appear to correspond to the patient phenotype. Adjusting our perception of what constitutes essential phenotypic features will likely have widespread practice implications and transform our understanding of the biology of many genetic diseases.

A128

ZENBU-reports: interactive analysis data visualization and website builder

Jessica Severin¹, Jordan Ramilowski¹, Masayoshi Ito², Takeya Kasukawa¹, Ken Yagi¹, Harukazu Suzuki¹, Jay W Shin¹, Michiel De Hoon¹, Piero Carninci¹

¹RIKEN Center for Life Science Technologies (CLST), Yokohama, JAPAN;

²RIKEN Preventive Medicine & Diagnosis Innovation Program (PMI), Yokohama JAPAN

Correspondence: Jessica Severin (jessica.severin@riken.jp)

Human Genomics 2018, **12(Suppl 1):A128**

ZENBU (Severin *et al.*, 2014) is an integrated data analysis system comprising a genome browser and analysis modules, which allows the interactive visualization and analysis of large sets of next-generation sequencing data. ZENBU is optimized for use in large international consortia by offering a collaboration system where users can create secured accounts allowing uploaded data to be either kept private or be shared with specific collaborators. ZENBU's on-demand data processing and analysis capabilities allow direct manipulation of 1000s of sequence alignments BAM files providing a range of visualization options such as raw sequence alignment visualization, signal histograms, clustering and peak calling, or collating counts onto gene models. As the main visual portal to the FANTOM5 data (Forrest *et al.*, 2014), the ZENBU interface includes several convenient transcriptome visualization capabilities.

Building on the ZENBU system, we here present ZENBU-Reports. Nowadays, scientific publications often times come with dedicated websites that manage and organize the data, showcase the analysis results, and offer tools for additional analysis. This is particularly important for consortia projects that produce atlases of genome-wide data by large-scale data production. Usually, such accompanying websites are constructed manually for each publication separately, which requires the services of a software engineer and increases the maintenance burden as well as data storage needs. To address this, we are developing ZENBU-Reports, an extension to ZENBU that will change the paradigm of data presentation and data integration. ZENBU-Reports will provide the graphical elements that are commonly used in supplementary websites, such as ordered lists, tables, scatter plots, bar graphs, network graphs, heatmaps and other scientific visualizations, as well as data download facilities. These elements can be user configured and arranged using a graphical interface, and can be linked to each other and to data already loaded in the

ZENBU database, allowing dynamic updates to the data presented. As ZENBU-Reports is built upon the ZENBU infrastructure and makes use of the same database, double storage of genome-wide data is prevented. In the next edition of FANTOM, (FANTOM6), we expect that many collaborators will use ZENBU-Reports as a convenient tool to create supplementary websites for satellite papers, while providing a consistent design across the FANTOM6 project.

Severin, J. *et al.* Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nature Biotechnology* **32**, 217–219 (2014).

Forrest, A.R.R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

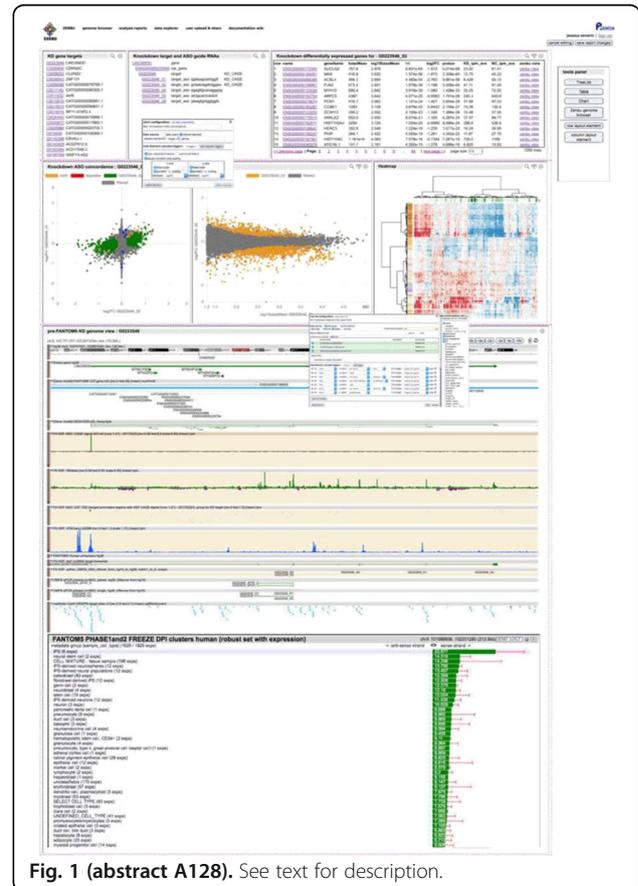


Fig. 1 (abstract A128). See text for description.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.