EUROPEAN SOCIETY OF HUMAN GENETICS

European School of Genetic Medicine

2nd Course in

Next Generation Sequencing

Bertinoro, Italy, May 17-20, 2013

Bertinoro University Residential Centre
Via Frangipane, 6 – Bertinoro

Course Director:
J. Veltman (Nijmegen, The Netherlands)

EUROPEAN GENETICS FOUNDATION

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

CENTRO RESIDENZIALE
UNIVERSITARIO
Bertinoro

FONDAZIONE
INTERNAZIONALE
MENARINI

2nd Course in

Next Generation Sequencing

Bertinoro, Italy, May 17-20, 2013

CONTENTS

2nd COURSE IN NEXT GENERATION SEQUENCING

Bertinoro University Residential Centre
Bertinoro (Italy), May 17-20, 2013

Arrival day: Thursday May 16

Friday, May 17

Morning Session: Introduction to Next Generation Sequencing

9.00 – 9.15	Introduction to the course	**G. Romeo**

9.00 – 9.15   Introduction to the course
G. Romeo

9.15 – 10.15   Next Generation Sequencing basics
J. Veltman

10.15 – 10.45   **Coffee Break**

10.45 – 11.45   Next generation Sequencing and its application to an isolated population
C. Sidore

11.45 – 12.45   Bioinformatic challenges in next generation sequencing
P. Robinson

12.45– 13.30   **Lunch Break**

Afternoon Session:

13.30 –14.00   Poster Viewing Session

14.00 – 16.00   Concurrent Workshops:
Computer practical: Variant identification (**C. Gilissen** & **T. Pippucci**)
Workshops by speakers

16.00-16.30   **Coffee Break**

16.30 – 18.00   Concurrent Workshops
Computer practical: Variant identification (**C. Gilissen** & **T. Pippucci**)
Workshops by speakers

Saturday, May 18

Morning Session: Next generation disease gene identification

9.00 – 10.00   Next generation disease gene identification
                      J. Veltman

10.00 – 11.00   Next generation consanguinity studies
                      G. Romeo

11.00 – 11.30   Coffee Break

11.30- 12.30   Exome sequencing to study rare and common variation in diabetes
                      Amelie Bonnefond

12.30 – 13.30   Mitochondrial genetics versus traditional Mendelian genetics : a challenge in the
                      analysis of NGS data.
                      G. Romeo

13.30 – 14.30   **Lunch Break**

Afternoon Session:

14.30 – 16.00   Concurrent Workshops
                      Computer practical: Disease gene identification (**C. Gilissen** & **T.Pippucci**)
                      Workshops by speakers

16.00-16.30   **Coffee Break**
16.30 – 18.00   Concurrent Workshops
                      Computer practical: Disease gene identification (**C. Gilissen** & **T.Pippucci**)
                      Workshops by speakers

Sunday, May 19

Morning Session: Diagnostic next generation sequencing

9.00 –10.00   Handheld diagnostics on nanowires
                      J. O'Halloran

10.00 – 11.00   Exome diagnostics in intellectual disability
                      A. Rauch

11.00 – 11.30   **Coffee Break**

11.30 – 12.30   Improving the prediction of disease-related non-synonymous
                      single nucleotides variants.
                      E. Capriotti

12.30 – 13:30   Sample Preservation and Trans-omic Studies Accelerate Scientific Research
                      L. Cheng

13:30 – 14.30	**Lunch Break**

<u>Afternoon Session:</u>

14.30 –15.00	Poster Viewing Session
15.00 – 16.30	Concurrent Workshops Computer practical: Diagnostic NGS (**C. Gilissen** & **T. Pippucci**) Workshops by speakers
16.30-17.00	Coffee Break
17.00 – 18.30	Concurrent Workshops Computer practical: Diagnostic NGS (**C. Gilissen** & **T. Pippucci**) Workshops by speakers

<u>Monday, May 20</u>

<u>Morning Session:</u>   <u>New frontiers in next generation sequencing</u>

9.00 – 10.00	Investigating single cells **J. Lundeberg**
10.00 – 11.00	Understanding the non-coding genome **N. Shomron**
11.00 – 11.30	**Coffee Break**
11.30 – 12.00	Best Posters Presentations by students
12.00 – 12.30	Wrapping up of the course (**J. Veltman**)
12.30	**Lunch**

Departure

ABSTRACTS OF LECTURES

Friday, May 17

Next Generation Sequencing basics

Joris A.Veltman
*Department of Human Genetics, Radboud University Nijmegen Medical Centre,
Nijmegen, The Netherlands*

There is considerable variation between the genetic code of two individuals, both at the single nucleotide and at the structural level. Identifying and studying the consequences of these variations, a core activity in human genetics research, is driven by technological innovations. Currently we are in the midst of one of the greatest technological revolutions in genomics. Novel DNA sequencing methods are dramatically increasing sequencing throughput to a level where it is soon possible to rapidly sequence an individual genome for an affordable price. If properly established, whole genome sequencing will have a major impact on the entire field of medicine; All genomic variation that can be linked to disease is detectable in a single experiment!

In this presentation I will introduce next generation sequencing technology and discuss its development and advantages over traditional sequencing technologies.

Recommended reading:

1. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 26: 1135-45 (2008).

2. Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. The next-generation sequencing technology and application. Protein Cell 1: 520-36 (2010). Review.

3. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 85:142-54 (2009). Review.

4. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. Clinical assessment incorporating a personal genome. Lancet 375: 1525-35 (2010).

Next Generation Sequencing
and its application to an isolated population

Sidore Carlo

CNR, Institute of Genetics Research, Sassari, Italy

In the past several years, genome-wide association studies have furthered our understanding of the molecular basis of many complex traits by finding, trough genotyping or imputation, the causal variants or their proxies. However such strategy, based on variants present in available chip platforms and imputation panels, may miss part of the variability due to rare and population specific variant. We will describe a large scale population approach based pn low pass sequencing. We will show its benefits and limitations and how we applied it to 2 main studies: 1) a cohort of >6,000 individuals recruited from an isolated population in the Lanusei Valley in Sardinia and mainly used for quantitative trait analysis; and 2) a cohort of >7000 individuals from a case control study on autoimmunity in Sardinia. Using phased haplotypes from whole genome low pass sequencing (coverage >3x) of 2120 Sardinians we identified and imputed >17M single nucleotide polymorphisms in the entire cohort. We then generate a Sardinian reference panel which allowed us to more accurately impute missing genotypes than an equal size reference panel including individuals from elsewhere in Europe, and in particular at low frequency variants (average r2 0.92 and 0.75, respectively, for frequency 1-3%). Furthermore, causal variants at some specific loci were missing in 1000G, and thus were detectable only by using our strategy. We will use some examples of Sardinian founder mutations to show their impact on the inherited component of complex traits.

Bioinformatic challenges in next generation sequencing

Peter Robinson
Institut für Medizinische Genetik Charité, Universitätsmedizin Berlin, Germany

In whole-exome sequencing (WES), target capture methods are used to enrich the sequences of the coding regions of genes from fragmented total genomic DNA, followed by massively parallel, 'next-generation' sequencing of the captured fragments. Since its introduction in 2009, WES has been successfully used in several disease-gene discovery projects, but the analysis of whole-exome sequence data can be challenging. In this overview, we present a summary of the main computational strategies that have been applied to identify novel disease genes in whole-exome data, including intersect filters, the search for de novo mutations, and the application of linkage mapping or inference of identity-by-descent (IBD) in family studies.

We will additional present new work on the use of model-organism phenotype and protein-protein interaction networks for identification of candidate genes in whole-exome sequencing projects.

Saturday, May 18

Next generation disease gene identification

Joris A. Veltman
Department of Human Genetics, Radboud University Nijmegen Medical Centre,
Nijmegen, The Netherlands

Next generation sequencing can be used to search for Mendelian disease genes in an unbiased manner by sequencing the entire protein-coding sequence, known as the exome, or even the entire human genome. Identifying the pathogenic mutation amongst thousands to millions of genomic variants is a major challenge, and novel variant prioritization strategies are required. The choice of these strategies depends on the availability of well-phenotyped patients and family members, the mode of inheritance, the severity of the disease and its population frequency. In this presentation, I will discuss the current strategies for Mendelian disease gene identification by exome and genome resequencing. We conclude that exome strategies are successful and identify new Mendelian disease genes in approximately 60% of the projects. Improvements in bioinformatics as well as in sequencing technology will likely increase the success rate even further. Exome sequencing is likely to become the most commonly used tool for Mendelian disease gene identification for the coming years.

Recommended reading:

1. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42: 30-5 (2010).

2. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328: 636-9 (2010).

3. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 362: 1181-91 (2010).

4. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. Genome Biol 12: 228 (2011).

5. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20: 490-7 (2012).

Next generation consanguinity studies

Giovanni Romeo

University of Bologna Medical School and European School of Genetic Medicine

The application of NGS technology to consanguinity studies has shown that the causative mutation for any rare autosomal recessive disorder can be identified by whole exome sequencing in a single experiment if, as an example, only two affected sibs born of first cousins can be studied (Pippucci T. et al.: Hum Hered 2011; 72: 45-53). But consanguinity is also a powerful tool for establishing the prevalence of rare disorders in inbred population groups. Classical studies of consanguinity have taken advantage of the relationship between the gene frequency for a rare autosomal recessive disorder (q) and the proportion of offspring of consanguineous couples who are affected with the same disorder. We recently developed a new approach for estimating q using mutation analysis of affected offspring of consanguineous couples based on the fact that the child born of consanguineous parents can carry the same mutation in double copy (homozygosity) or alternatively two different mutations in the same gene (compound heterozygosity) inherited through two different ancestors. The proportion of compound heterozygotes among children affected with a given autosomal recessive disorder born of consanguineous parents can be taken as an indicator of the frequency of the same disorder in the general population. Data collected by different molecular diagnostic laboratories in Mediterranean countries where the frequencies of consanguineous marriages is high show the validity of this approach which we called the HI (Homozygosity Index) method (Gialluisi A. et al. Ann Hum Genet 2012;76: 159-67). Initially we tested the HI method on different samples of patients affected with two autosomal recessive disorders, namely Familial Mediterranean Fever (FMF) and Phenylketonuria (PKU), born either to first cousins or to unrelated parents. More recently the HI method has been used to verify the high prevalence of Wilson disease (WD) in Sardinia where essentially four different mutations in the ATP7B gene (13q14.3, MIM 606882) are associated with alterations of copper metabolism resulting in pathological progressive copper accumulation in liver and other tissues. The worldwide prevalence (P) of WD is about 30/million, while in Sardinia it has been estimated to be in the order of 1/10000, one of the highest worldwide (Loudianos G. et al. Hum Mutat 1999; 14:294–303). All these estimates have been inferred through classical clinical approaches, and as such they are likely to suffer from an underdiagnosis bias. Indeed, a recent molecular neonatal screening in Sardinia reported a WD gene frequency of 1.92% with a resulting prevalence of 1/2707 live births (Zappu A et al. J Pediatr Gastroenterol Nutr 2008; 47:334–338). The HI method applied to a sample of 178 carefully characterized patients collected by the pediatricians of the Ospedale Regionale per le Microcitemie and University of Cagliari, confirms the results reported by Zappu et al. raising the interesting question of whether the high prevalence of WD in Sardinia is due to genetic drift or selection.

In conclusion, once the molecular defect of a given autosomal recessive disorders is identified thanks to the use of NGS technology, the molecular data generated in clinical diagnostic laboratories make it possible to use the consanguinity approach based on the HI method to establish priorities for screening and intervention policies, as in the case of WD in Sardinia. This is not a trivial result for communities where autosomal recessive "rare" disorders can be not so rare and have a strong social impact.

Exome sequencing to study rare and common variation in diabetes

Amelie Bonnefond

CNRS UMR8199, Lille Pasteur Institute, Lille II University, Lille, France

Type 2 diabetes (T2D) is a complex disorder that has developed into a major health problem responsible for early morbidities and mortality with the worldwide prevalence doubling every 15 years. Although T2D is considered preventable, it is usually not reversible and current medications are generally unable to stop disease progression, resulting in poor glucose control and degenerative complications. T2D is characterized by insulin secretory dysfunction of pancreatic β-cells combined with insulin resistance, but the primary mechanisms are still largely debated. A genetic contribution is well recognized in the diverse forms of both early-onset (mostly known as monogenic) and adult-onset diabetes. Through several subsequent genetic approaches, we know that these diverse forms of diabetes are characterized by vast genetic heterogeneity, highlighting the complexity of the signalling pathways that are pivotal for the biology of the pancreatic β-cell. Although recent genetic discoveries have provided new insights into β-cell physiopathology (via genome-wide association studies for polygenic forms of T2D and candidate gene studies for monogenic forms of diabetes), major gaps in our understanding of the genetic basis of diabetes remain; a proportion of monogenetic diabetes cases remain to be explained and the heritability of common T2D is much higher than the overall impact of the known genetic variants. Thus, a substantial genetic 'dark matter' remains to be elucidated.

Next-generation sequencing, including whole-exome sequencing (WES), has brought new fantastic expectation to elucidate most of this dark matter. Today, even if WES was quite successful in elucidating novel monogenic forms of diabetes (including neonatal diabetes associated with pancreatic agenesis or maturity-onset diabetes of the young [MODY]), the strategy did not lead to strong breakthrough in polygenic forms of T2D. I will discuss about the strength and the limit of WES in both monogenic and polygenic forms of diabetes.

Mitochondrial genetics versus traditional Mendelian genetics :

a challenge in the analysis of NGS data

Giovanni Romeo

University of Bologna Medical School and European School of Genetic Medicine

The mitocondrial genome is a circular double stranded DNA molecule, which encodes, among others, the subunits of the respiratory chain complexes, necessary for energy production in the cell. These complexes are formed of many subunits, mostly encoded by the nuclear and partly by the mitochondrial genome. Apart from beeing maternally inherited, the mitochondrial genome is characterized by the phenomenon of homoplasmy and heteroplasmy, as each cell/mitocohndrion possessess a large number of mtDNA copies. The wt and mutant species can coexist in the same cell and this leads to the peculiarity of mitochondrial genetics namely the threshold effect which makes mitochondrial genetic analysis totally different from that of mendelian genetics. In addition the analysis of mitochondrial DNA sequences faces the problem

originated by the PCR co-amplification of NumtS (Nuclear mitochondrial Sequences). NumtS are fragments of mitochondrial genomes within eukaryotic genomes, that maintain high nucleotide similarity with the mtDNA sequence of origin. The discovery of NumtS started with hybridization experiments between mtDNA and nuclear DNA (nuDNA) on mouse liver. Discerning between true and *false heteroplasmy* is very important in mitochondrial sequence analysis. The evaluation of the NumtS presence is therefore of outmost importance in the definition of the mtDNA mutation threshold and can be coupled with the NGS high-throughput techniques (Petruzzella V. et al., Hum. Mol. Genet. doi: 10.1093/hmg/dds182 HMG 2012)

The most well characterized clinical disorders due to mtDNA mutations are listed below.

Table 2 | **Clinical disorders that are caused by mutations in mitochondrial DNA**

Mitochondrial DNA disorder	Clinical phenotype	mtDNA genotype	Gene	Status	Inheritance	Reference
Kearns–Sayre syndrome	Progressive myopathy, ophthalmoplegia, cardiomyopathy	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	61,158
CPEO	Ophthalmoplegia	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	61, 64
Pearson syndrome	Pancytopoenia, lactic acidosis	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	65
MELAS	Myopathy, encephalopathy lactic acidosis, stroke-like episodes	3243A>G; 3271T>C Individual mutations	TRNL1 ND1 and ND5	Heteroplasmic Heteroplasmic	Maternal Maternal	159 160, 161
MERRF	Myoclonic epilepsy, myopathy	8344A>G; 8356T>C	TRNK	Heteroplasmic	Maternal	162
NARP	Neuropathy, ataxia, retinitis pigmentosa	8993T>G	ATP6	Heteroplasmic	Maternal	163
MILS	Progressive brain-stem disorder	8993T>C	ATP6	Heteroplasmic	Maternal	67
MIDD	Diabetes, deafness	3243A>G	TRNL1	Heteroplasmic	Maternal	164
LHON	Optic neuropathy	3460G>A 11778G>A 14484T>C	ND1 ND4 ND6	Hetero- or homoplasmic Hetero- or homoplasmic Hetero- or homoplasmic	Maternal Maternal Maternal	165 62 166
Myopathy and diabetes	Myopathy, weakness, diabetes	14709T>C	TRNE	Hetero- or homoplasmic	Maternal	167,168
Sensorineural hearing loss	Deafness	1555A>G Individual mutations	RNR1 TRNS1	Homoplasmic Hetero- or homoplasmic	Maternal Maternal	55 169,170
Exercise intolerance	Fatigue, muscle weakness	Individual mutations	CYB	Heteroplasmic	Sporadic	68
Fatal, infantile encephalopathy; Leigh/Leigh-like syndrome	Encephalopathy, lactic acidosis	10158T>C; 10191T>C	ND3	Heteroplasmic	Sporadic	66

ATP6, ATPase 6; CPEO, chronic progressive external ophthalmoplegia; CYB, cytochrome b; LHON, Leber hereditary optic neuropathy; MELAS, mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes; MERRF, myoclonic epilepsy and ragged-red fibres; MIDD, maternally-inherited diabetes and deafness; MILS, maternally-inherited Leigh syndrome; ND1,3–6, NADH dehydrogenase subunits 1,3–6; NARP, neurogenic weakness, ataxia and retinitis pigmentosa; RNR1, 12S ribosomal RNA; TRNE, TRNK, TRNL1, TRNS1, mitochondrial tRNAs.

Moreover in the last few decades a high frequency of mtDNA mutations has been observed in all types of tumors. This led to revisit the phenomenon originally described by Otto Warburg (a Nobel prize winner in 1931) who observed that cancerous cells preferably utilize glycolysis for ATP production even when oxygen is available and respiration may function (a condition called pseudohypoxia). The glycolitic shift at the basis of the Warburg effect can have different causes like a mitochondrial impairment, the overexpression of glycolytic enzymes and the triggering of hypoxia or pseudohypoxia (that may occur before neovascularization of the tumour mass takes place). In presence of a mitochondrial mutation and through the

action of HIF (or Hypoxia Inducible factor), all these three causes may be called into play to explain what could be the selective advantage for transformed cells in the initial stage of tumor growth, when the tumoral mass generates hypoxic areas in cancer cells.

Certain tumor subtypes are characterized by abnormal accumulation of nonfunctional mitochondria in their cytoplasm and tumors which develop such a phenotype are called oncocytomas. They usually arise in tissues of epithelial origin such as thyroid, kidney or pituitary gland, but have also been reported in breast, endometrial, lung and colon cancers.

In order to progress towards malignancy, any solid tumor must go through hypoxic adaptation, a process tightly connected with metabolic reactions and directed by a transcription factor called Hypoxia Inducible Factor 1 α (HIF1α). Oncocytomas because of their inability to perform mitochondrial respiration, are not able to adapt to hypoxia and therefore generally maintain a low-proliferative, non-invasive state. Genetic changes underlying the benign nature of oncocytomas are those disruptive mtDNA mutations which when present in homoplasmy, i.e. in all or almost all mtDNA copies of a tumor cell, abolish mitochondrial respiration and prevent HIF1α assembly and therefore the subsequent progression to malignancy. However the anti-tumorigenic effect of disruptive mtDNA mutations is not observed when they are present in heteroplasmy in the cancer cell, thus giving rise to the concept of the double-faceted effect (oncojanus) of such disruptive mutations. This implies that a normally functioning respiratory complex I is essential to induce a Warburg profile in mitochondria-defective tumor cells (Calabrese et al. Cancer & Metabolism, in press, 2013)

Sunday, May 19

Handheld diagnostics on nanowires

Jonathan O'Halloran
QUANTUMDx GROUP LIMITED, Newcastle, UK

The DNA sequencing field has seen dramatic advances since the first draft of the human genome was published, with companies reporting ever faster and cheaper methods. However, despite the race to attain the $1000 genome producing a plethora of exciting technologies, CE is still being routinely used for targeted clinical sequencing for genetic testing and drug resistance testing for infectious disease. NGS, third generation and even "LastGen" offer much promise for the future of genomics, but their translation into the clinic is slow and initially appears to be a case of trying to fit a round peg through a square hole. QuantuMDx is developing two sequencing technologies to address this; a portable, handheld DNA sequencing device for PGx and infectious disease applications at the point of care, to provide an alternative to slow and relatively expensive CE DNA sequencing and a "LastGen 2.0" technology to compete Nanopore sequencing head on.

Exome diagnostics in Intellectual disability

Anita Rauch
University of Zurich, Institute of Medical Genetics, Germany

Intellectual disability is characterized by significant cognitive and adaptive deficits with an IQ below 70 and childhood onset. It affects about 2-3% of the population and its causes are very heterogeneous. Trisomy 21 is the most common single cause accounting for about 10% of patients, and further microscopically visible chromosomal aberrations are found in about 4%. About 14% of patients show submicroscopic aberrations, so called copy number variations, detectable by chromosomal microarray analysis. In about 10% of patients with ID the accompanying pattern of morphological or functional anomalies allows a clinical diagnosis of a well characterized monogenic disorder, which can be specifically tested by targeted Sanger sequencing of currently more than 400 different known ID genes. Thus the majority of patients currently remain undiagnosed unless exome sequencing is applied, which recently were shown to reveal likely deleterious de novo mutations causing ID in probably more than 50% of patients without chromosomal or clinical diagnosis. These studies also showed that the current perceptions of the phenotypes associated with certain genes are heavily biased towards the clinically recognizable spectrum. However, in populations with a high rate of consanguineous mating a plethora of different autosomal recessively inherited gene defects may be more common.

Identification of the underlying genetic defect in patients with ID is not only important for disease management but also for determination of the actual recurrence risk for further affected children within the family. Especially in severe ID, parents commonly experience a great personal burden and dare only to give birth to further children if recurrence risk is low or if reliable prenatal diagnostics can be offered. Therefore calling a gene mutation causative for a patient's ID is usually associated with a special social and legal responsibility. Since neither de novo occurrence nor segregation with the phenotype in small nuclear families, nor common in silico predication of pathogenicity or overlapping phenotypes in animal models are definite proves of the causative nature of individual gene variants, special considerations need to be taken. Reconciliation of genetic findings with a patient's phenotype may be very reassuring in well characterized disorders. Vice versa, observation of at least a few patients with a similar phenotype caused by similar mutations in the same gene may be the most reliable prove for genotype-phenotype correlation. Thus truncating mutations in known disease genes provide the highest causative evidence, followed by probably deleterious missense mutations in known disease genes with a fitting phenotype. In contrary, missense variants in known disease genes with atypical phenotype or any type of variants in novel genes in single cases remain speculative disease causes. A further level of complexity is present in offspring of highly consanguineous parents, which may suffer from more than one disease causing autosomal recessive disorder. Multiple hits may also be responsible for some patients with sporadic or familial mild ID, but given the large heterogeneity of ID, proving of such associations need a large number of cases and controls.

Suggested Reading:

Rauch et al. (2006) Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am J Med Genet A140:2063-74.

Ropers HH (2010) Genetics of early onset cognitive impairment. Annu. Rev. Genomics Hum. Genet. 11:161–87.

Topper et al. (2011) Exome sequencing and the genetics of intellectual disability. Clin. Genet. 80:117-26

Najmabadi H et al. (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478:57-63.

Rauch et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380:1674-82.

de Ligt J et al. (2012) Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. N. Engl. J. Med. 367:1921-9.

Improving the prediction of disease-related non-synonymous single nucleotides variants

Emidio Capriotti

Division of Informatics, Department of Pathology, University of Alabama at Birmingham, Birmingham (AL), United States of America

The most common form of human genetic variation is single nucleotide polymorphisms (SNVs) [1]. Although higher percentage of SNVs are expected to occur in the non-coding region, a smaller fraction of non-coding variants have so far been characterized as disease-causing than coding, non-synonymous SNVs (nsSNVs) [2].

In the last decade, several algorithms have been developed to predict disease-related and functionally deleterious variants [3]. The majority of the current methods rely on the manually curated collections of disease-associated variants from OMIM [4], SwissVar [5], and, more recently, the dbSNP [6] clinical SNV collections. Once the data sets are collected, all methods use some combination of the affected protein sequence/structure features and functional annotations to look for patterns indicative of disease involvement. In general, all the predictors rely on evolutionary information that is extracted using different procedures. For example, the SIFT algorithm [7] exploits the information contained in sequence alignments to calculate the probability that a mutation of a residue in a given sequence position is tolerated. PhD-SNP [8] is a machine learning approach that takes as input the frequencies of wild-type and mutant residues from a sequence profile calculated with the BLAST algorithm [9]. PolyPhen [10] evaluates a substitution score calculating the Position Specific Independent Count (PSIC) matrix. More sophisticated methods such as MutPred [11] and SNPs&GO [12] also include outputs of other predictors and/or a functional score calculated using the Gene Ontology.

In this talk, I will summarize the recent advances in the prediction of deleterious SNVs, discussing the protein structural and functional features that allow to improve the accuracy of the available algorithms.

REFERENCES

1. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature, 467(7319), 1061-1073 (2010).

2. Stenson PD, Ball EV, Mort M et al. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat, 21(6), 577-581 (2003).

3. Capriotti E, Nehrt NL, Kann MG, Bromberg Y. Bioinformatics for personal genome interpretation. Brief Bioinform, 13(4), 495-512 (2012).

4. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res, 37(Database issue), D793-796 (2009).

5. Mottaz A, David FP, Veuthey AL, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. Bioinformatics, 26(6), 851-852 (2010).

6.   Sherry ST, Ward MH, Kholodov M et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res, 29(1), 308-311 (2001).

7.   Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res, 31(13), 3812-3814 (2003).

8.   Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 22(22), 2729-2734 (2006).

9.   Altschul SF, Madden TL, Schaffer AA et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25(17), 3389-3402 (1997).

10.   Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res, 30(17), 3894-3900 (2002).

11.   Li B, Krishnan VG, Mort ME et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics, 25(21), 2744-2750 (2009).

12.   Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat, 30(8), 1237-1244 (2009).

Sample Preservation and Trans-omic Studies Accelerate Scientific Research

Le Cheng, Meiru Zhao, Qian Wan, Xin Zhou, Ying Sun, Puyi Qian, Zhixiang Yan, Yong Zhang

China National Genebank, BGI-Shenzhen, Shenzhen, China

Fast development of sequencing technology with decreasing cost has substantially promoted large cohort study, population health screening, and epidemiological prevention study on genomic level. China National Genebank (Shenzhen) was approved by Chinese government and operated by BGI, which aims at combining large biospecimen collection and trans-omic (genomic, transcriptomic, proteomic, epigenomic etc.) data production to accelerate scientific discovery and translation to clinical use. The Genebank has accumulated genetic variation information of more than 2,000 monogenetic diseases by collecting suspected pedigree samples countrywide and worldwide. Based on this database over ten papers have been published on discoveries of new variation spots. In addition, China National Genebank has collected samples for 1000 Genomes Project, Chinese Cancer Genome Project and Human Metagenome Project, which made substantial contribution in establishing databases of global human genetic variation, cancer genetic variation and human inhabited microbial composition. These databases are of great significance in our understanding of human disease and development of personalized medicine in the future.

Monday, May 20

Investigating single cells

Joakim Lundeberg

*Science for Life Laboratory, Division of Gene Technology, Royal Institute of Technology (KTH),
Box 1031, SE-17121, Solna, Sweden*

Introduction

Cells in a particular tissue are not identical. Instead, cells that are identical from a genomic point of view may have considerable variation in gene expression profile and protein levels[1-3], giving rise to a heterogeneous collection of cells with different behavior and appearance. Even the genomic DNA sequence may vary slightly between neighboring cells within a tissue[4,5]. Tissues are commonly made up from several cell types with varying functions. When studying multiple cells, or a mixture of genomic DNA from several cells, the average characteristics of the bulk are usually obtained and information on rare cells may be lost[6]. Traditionally, and in many cases currently, studies have been carried out on a collection of cells; in studies aimed at global analysis and/or investigating large/average differences between samples this approach is still valid. However to be able to address particular biological questions, investigating single cells may be the only option.

Cell-cell variability is an interesting aspect to study in many biological contexts. In some diseases, a combination of traits is known to give a cell its particular disease phenotype and obtaining information on cell heterogeneity within a tissue is sometimes crucial. For instance, it has been argued that knowing more about the co-occurrence of several different mutations within the same cell is essential to fully understand the nature and evolution of cancer[7,8]. In other cases, a cell population may be made up of two subpopulations with distinctly varying traits, making the information obtained in bulk analysis, dependent upon their proportions in the mixed sample. Circulating tumor cells in cancer patients and fetal cells in the blood of pregnant women are two examples of rare cells that are difficult to study in their natural habitat and therefore need to be isolated for proper characterization of cancer genotype for prognostic evaluation or fetal aneuploidy, respectively. In microbial analysis, many species cannot be cultured in the laboratory making single cell analysis the only option[9]. By monitoring single cell proliferation over time, the response of a heterogeneous cell population to various agents can be evaluated[10]. Also, multiplication and analysis of single cell clones is a common way to increase the genetic material available for analysis[11].

In order to get information from single cells one need to isolate, study, and sometimes culture them, separately. In the following chapter a selection of tools that can be used to single out and isolate particular cells of interest will be described.

Working with single cells

The field of cell biology has always been dependent on technology development. Development of high-resolution microscopy in combination with specific staining methods in 17th century made it possible to observe tissues and individual cells for the first time[12,13]. Nowadays, microscopy is used daily in most biological laboratories, to investigate, e.g., the intracellular localization of particular proteins and even the division of living cells, when using live-cell imaging in time-lapse microscopy[14,15]. Methods such as electrophysiology, where the voltage over a single cell or ion channel is measured by attaching a pipette containing a volt-meter to the surface of a cell, give information about one cell at a time[16,17]. The same is true for flow cytometry, by which the sizes of, and fluorescent signals from, individual cells in a stream can be monitored. These methods have been around for a long time and by their nature give information about the properties of single cells. However, when it comes to biochemical approaches for investigating DNA sequences, gene expression levels or the amount of small molecules or protein in a cell, there is a need for tools offering higher sensitivity[18].

One way to increase sensitivity is through miniaturization. Since human cells are about 10 µm in size and occupy a volume of less than a fl[19], the standard volumes of 50 µl and more, commonly used in laboratories, can be excessive when studying small numbers of cells. A reduction in reaction volume has been identified as a way to enhance the sensitivity and speed of reactions by allowing shorter diffusion distances and smaller dilution factors[20-22]. Another obvious advantage of smaller reaction volumes is a decrease in reagent costs. In recent times, the numbers of publications in the fields of genomics, transcriptomics and proteomics describing single cell analysis utilizing miniaturization techniques have dramatically increased. Regardless of the biomolecule of interest, single cell analysis has some general technical features, such as picking up single cells (possibly based on certain traits such as cell surface markers) and putting these cells in a container where the downstream analysis can be carried out. There are several readily available platforms for these types of tasks.

Genetic analysis in single cells

When researchers at Stanford successfully attached small amounts of DNA oligonucleotides to a glass surface in an ordered array and used it to monitor the transcription of 45 genes, the initial step towards a highly parallel analysis of transcriptomes was taken. Today, new and more comprehensive massively parallel DNA sequencing (MPS) technologies are starting to replace DNA microarrays for many applications, among other in-depth transcriptome analysis. The new advances in high-throughput genomics have reshaped the biological research landscape and have enabled the study of the full transcriptome in a digital and quantitative fashion. The bioinformatics tools to visualize and integrate these comprehensive sets of data have also been significantly improved during recent years.

The majority of gene expression studies based on microarrays, or massively parallel sequencing of RNA, use a representative sample containing many cells. Thus the results will represent the average expression levels of the investigated genes. The separation of cells that are phenotypically different has been used in some cases together with the global gene expression platforms (Tang F *et al*, RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010;5(3):516-35.) and resulted in very precise

information about cell-to-cell variations. However, high throughput methods to study transcriptional activity with high resolution <u>in intact tissues</u> have until now not been available.

Spatial Transcriptomics

The parallel development of increasingly advanced histological and gene expression techniques has resulted in the separation of imaging and transcriptome analysis, as there has <u>not been any feasible method available for global transcriptome analysis with spatial resolution.</u> Transcriptome analysis is typically performed on mRNA extracted from tissues, and methods for collecting smaller tissue areas or individual cells for transcriptome analysis are typically labor intensive, costly and have low precision.

We have recently devised a simple strategy that enables global gene expression analysis in histological tissue sections, yielding transcriptomic information with two-dimensional spatial resolution. This enables the identification of individual transcriptomes of cells while maintaining the positional information of those cells in the tissue.

This is possible by arraying reverse transcription (RT) primers with unique positional tags on object slides, onto which thin tissue sections can be placed. mRNA molecules diffuse from the tissue and hybridize locally to the array. A reverse transcription reaction is performed in the tissue section on the object slide to obtain cDNA, which, due to the unique positional tags in the RT primers, carries information about the position of the template mRNA in the tissue section. The cDNA is sequenced and a transcriptome with exact positional information is obtained.

The sequencing data is then visualized in the computer together with the tissue section, for instance to display the expression pattern of a gene of interest across the tissue. It is also easy to mark different areas of the tissue section on the computer screen and obtain information on differentially expressed genes between any selected areas of interest.

Current methods to study mRNA populations in a tissue context are based on in-situ hybridization and provide only relative information of the expression of one gene at a time. Obvious advantages with our approach over current in-situ technologies are thus the global gene expression information, the co-expression information and the quantitative estimates of transcript abundance. This is a generally applicable strategy available for the analysis of any tissue in any species but also cells.

The current Spatial Transcriptomics (ST) arrays range from 135 k to 10k features and each feature has the approximate dimension of a cell inferring that the equal number of cells can be analysed in a parallel. The binding capacity of each feature enables the capture of the entire mRNA population of a cell. Currently we have used the method to study coding of transmission of olfactory information to the brain (Ståhl et al, manuscript in preparation), dissecting the stroma/tumor interaction in prostate cancer (Salmen et al, manuscript in preparation) and we also have used the ST array to analyse tumor cells (liquid biopsy) to enable parallel analysis of 1000s of individual cells (Vickovic et al, manuscript in preparation).

References

1   Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836-840, (2002).

2   Raj, A., Peskin, C., Tranchina, D., Vargas, D. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS biology* **4**, e309, (2006).

3   Paszek, P. *et al.* Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11644-11649, (2010).

4   Salipante, S. J. & Horwitz, M. S. Phylogenetic fate mapping. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5448-5453, (2006).

5   Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, (2011).

6   Carlo, D. D. & Lee, L. P. Dynamic Single Cell Analysis for Quantitative Biology. *Analytical chemistry* **78**, 7918-7925, (2006).

7   d'Amore, F. *et al.* Clonal evolution in t(14;18)-positive follicular lymphoma, evidence for multiple common pathways, and frequent parallel clonal evolution. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 7180-7187, (2008).

8   Riethdorf, S., Wikman, H. & Pantel, K. Review: Biological relevance of disseminated tumor cells in cancer patients. *International Journal of Cancer* **123**, 1991-2006, (2008).

9   Marcy, Y. *et al.* Nanoliter Reactors Improve Multiple Displacement Amplification of Genomes from Single Cells. *PLoS Genetics* **3**, e155, (2007).

10   Frisk, T. W., Khorshidi, M. A., Guldevall, K., Vanherberghen, B. & Onfelt, B. A silicon-glass microwell platform for high-resolution imaging and high-content screening with single cell resolution. *Biomedical microdevices* **13**, 683-693, (2011).

11   Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. *PLoS computational biology* **1**, e50, (2005).

12   Hooke, R. (1665). Micrographia; Some physiological descriptions of minute bodies made by magnifying glasses. With observations and inquiries thereupon., London: J. Martyn and J. Allestry.

13   Bode, P. K. *et al.* MAGEC2 is a sensitive and novel marker for seminoma: a tissue microarray analysis of 325 testicular germ cell tumors. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc* **24**, 829-835, (2011).

14   Barbe, L. *et al.* Toward a confocal subcellular atlas of the human proteome. *Molecular & cellular proteomics* **7**, 499-508, (2008).

15   Koshkin, V. & Krylov, S. N. Single-cell-kinetics approach to compare multidrug resistance-associated membrane transport in subpopulations of cells. *Anal Chem* **83**, 6132-6134, (2011).

16   Neher, E., Sakmann, B. & Steinbach, J. H. The extracellular patch clamp: a method for resolving currents through individual open channels in biological membranes. *Pflügers Archiv: European journal of physiology* **375**, 219-228, (1978).

17   Mitsushima, D., Ishihara, K., Sano, A., Kessels, H. W. & Takahashi, T. Contextual learning requires synaptic AMPA receptor delivery in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12503-12508, (2011).

18   Newman, J. R. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846, (2006).

19   Rubakhin, S. S., Romanova, E. V., Nemes, P. & Sweedler, J. V. Profiling metabolites and peptides in single cells. *Nature methods* **8**, S20-29, (2011).

20   Manz, A., Graber, N. & Widmer, H. M. Miniaturized Total Chemical Analysis Systems: a Novel Concept for Chemical Sensing. *Sensors and Actuators* **B1**, 244-248, (1990).

21   Kricka, L. J. & Wilding, P. Microchip PCR. *Analytical and bioanalytical chemistry* **377**, 820-825, (2003).

22   Zare, R. & Kim, S. Microfluidic platforms for single-cell analysis. *Annual Review of Biomedical Engineering* **12**, 187-201, (2010).

Understanding the non-coding genome

Noam Shomron

Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre

Nijmegen, The Netherlands

Understanding viral infections in clinical samples is critical for effective patient health care. The study of a transcriptome is a powerful tool for perfecting gene expression and RNA-based regulation in any organism. Our laboratory uses massive parallel sequencing technologies (Deep sequencing or Next Generation Sequencing) that generate billions of reads per experiment, to study the RNA-mediated regulatory mechanisms during cellular processes. We are currently focusing on emerging principles controlling non-coding RNA-mediated regulatory mechanisms during pathogenesis caused by viral infection. Understanding the impact of this regulation on viral life cycles together with host interactions, might provide insight into future novel targets for intervention. Our genome-wide analyses will bring us closer towards having a full non-coding RNA 'signature map' during infection from both the viral and host perspectives. These findings will make significant contribution in the field of systems pathogenicity.

2nd Course in

Next Generation Sequencing

FACULTY Who is Who

Family Name	Name	Affiliation	Work e-mail
Bonnefond	Amelie	CNRS UMR8199, Lille Pasteur Institute, Lille II University, Lille, France	amelie@good.ibl.fr
Capriotti	Emidio	Division of Informatics, Department of Pathology University of Alabama at Birmingham (USA)	emidio@uab.edu
Cheng	Le	China National Genebank, BGI-Shenzhen, Shenzhen, China	chengle@genomics.cn
Gilissen	Christian	Department of Human Genetics - Nijmegen Centre for Molecular Life Sciences- Radboud University Nijmegen Medical Centre, the Netherlands	c.gilissen@antrg.umcn.nl
Lundeberg	Joakin	Karolinska Institute, SciLifeLab , Uppsala, Sweden	joakim.lundeberg@scilifelab.se
O'Halloran	Jonathan	QUANTUMDx GROUP LIMITED, Newcastle, U.K.	Jonathan.OHalloran@quantumdx.com
Pippucci	Tommaso	University of Bologna, Italy	tommaso.pippucci@unibo.it

Rauch	Anita	University of Zurich Institute of Medical Genetics, Sweden	anita.rauch@medgen.uzh.ch
Robinson	Peter	Institut für Medizinische Genetik Charité - Universitätsmedizin Berlin	peter.robinson@charite.de
Romeo	Giovanni	University of Bologna, Italy	romeo@eurogene.org
Shomron	Noam	Tel Aviv University, Tel Aviv, Israel	nshomron@post.tau.ac.il
Sidore	Carlo	CNR, Institute of Genetics Research, Sassari, Italy	scarlino81@gmail.com
Veltman	Joris	Department of Human Genetics- Nijmegen Centre for Molecular Life Sciences Radboud University Nijmegen Medical Centre	J.Veltman@antrg.umcn.nl

Who is Who: Participants

Name	Family Name	Affiliation	Nationality	City	E-mail
Vellingiri	Balachandar	Bharathiar University	India	Coimbatore	geneticlaba@yahoo.co.in
István	Balogh	University of Debrecen, Department of Laboratory Medicine	Hungary	Debrecen	baloghòmed.unideb.hu
Sarka	Bendova	Department of Biology and Medical Genetics	Czech Republic	Prague	sarka.bendova@lfmotol.cuni.cz
Nikola	Bila	University hospital Motol, Dept. of Biology and Medical Genetics	Czech Republic	Prague	nikola.bila@fnmotol.cz
Amina	Chaouch	Newcastle University	England	Newcastle	amina.chaouch@ncl.ac.uk
Roberto	Ciccone	Univ. of Pavia	Italy	Pavia	ciccone@unipv.lt
Renato	Colognato	TOMA Advanced Biomedical Assays S.p.A.	Italy	Busto Arsizio	dmosti@tomalab1.com
Milena	Crippa	Istituto Auxologico italiano	Italy	Cesano Maderno	m.crippa@auxologico.it
Johanna Christina	Czeschik	Institut für Humangenetik, Unievrsität Duisburg-Essen	Germany	Essen	christina.czeschik@uni-due.de
Jakob	Ek	Copenhagen University Hospital, Rigshospitalet	Denmark	Copenaghen	jakob.ek@regionh.dk
Mingyan	Fang	BGI	China	Shenzen	fangmingyan@genomics.cn
Regina	Fillerova	Faculty of Medicine and Dentistry, Palacky Unievrsity	Czech Republic	Olomouc	regak@seznam.cz
Dominique	Germain	Univ. of Versailles	France	Garches	Dominique.germain@rpc.aphp.fr
Francesca Romana	Grati	TOMA Advanced Biomedical Assays S.p.A.	Italy	Busto Arsizio	dmosti@tomalab1.com
Simonetta	Guarrera	Human Genetics Foundation	Italy	Turin	simonetta.guarrera@hegf-torino.org
Rebeqa	Gunnarsson	Lund University	Sweden	Lund	rebeqa.gunnarsson@med.lu.se
Dicky	Halley	Erasmus mc/clinical genetics	The Netherlands	Rotterdam	d.halley@erasmusmc.nl

Wagner	Holger	University Hospital Bonn	Germany	Bonn	hwagner1@uni-bomm.de
Ramunas	Janavicius	Vilnius University Hospital Santariskiu clinics	Lithuania	Vilnius	ramunas.janavicius@santa.lt
Anne Marie	Jelsig	University Hospital of Odense	Denmark	Odense	anne.marie.jelsig@ouh.regionsyddanmark.dk
Christeen Ramane	Jesuthasan Pedurupillay	Dept. Of Medical Genetics, Oslo University	Norway	Oslo	ramane_p@yahoo.com
Ingrid	Kalev	Tartu Unviersity	Estonia	Tartu	ingrid.kalev@ut.ee
Anju	Karingadampallil Philips	Dept. Of Medical Genetics, University of Helsinki	Finland	Helsinki	anju.philips@helsinki.fi
Heba	Kassem	Magdi Yacoub Hearth Foundation	Egypt	Cairo	Heba.kassem@gmail.com
Irene	Konstanta	N.C.S.R. Demokritos	Greece	Athens	eikonstanta@ipta.demokritos.gr
Eva	Kriegova	Faculty of Medicine and Dentistry, Palacky University	Czech Republic	Olomouc	eva.kriegova@email.cz
Anne Bruun	Kroeigaard	Dept. of Clinical Genetics, University Hospital	Denmark	Odense	anne.kroeigaard@ouh.regionsyddanmark.dk
Marzena	Kucharczyk	The Children's Memorial Health Institute, Dept. of Medical Genetics	Poland	Warsaw	m.kucharczyk@czd.pl
Alexander	Lavrov	Research Centre for Medical Genetics of RAMS	Russia	Moscow	avlavrov@yandex.ru
Stefania	Lenarduzzi	IRCCS Burlo Garofolo Trieste	Italy	Trieste	ragioneria@burlo.trieste.it
Kristina	Lundin	Dept. of Clinical Genetics	Sweden	Lund	kristina.lundin@med.lu.se
Azam	Maleeha	Radboud University Nijmegen Medical Centre	The Netherlands	Nijmegen	h.brunner@gen.umcn.nl
Maria	Maleeha	Radboud University Nijmegen Medical Centre	The Netherlands	Nijmegen	m.maria@gen.umcn.nl
Filippo	Martinelli Boneschi	Ospedale San Raffaele Srl	Italy	Milan	martinelli.filippo@hsr.it
Lilian	Martinez	Institute of Human Genetics	Germany	Cologne	landreamartinez@yahoo.com
Karmen	Meško Meglič	University Clinic of Respiratory and Allergic Diseases	Slovenia	Golnik	karmen.mesko-meglic@klinika-golnik.si
Elena	Mocan	The State University of Medicine and Pharmacy "Nicolae Testemitanu"	Moldova	Chisinau	helen_mokan@yahoo.com

Doris	Mueller	Dept. Of Human Genetics, General Women's and Children's Hospital	Austria	Linz	doris.mueller@gespag.at
Zuzana	Musova	Dept. of Biology and Medical Genetics	Czech Republic	Prague	zuzana.musova@fnmotol.cz
Deepti Akkihebbal	Narasimhaiah	Université Catholique de Louvain	Belgium	Brussel	deepti.narasimhaiah@uclouvain.be
Linda	Olsson	Lund University Clinical Genetics	Sweden	Malmo	linda.olsson@med.lu.se
Elsebet	Ostergaard	Copenhagen University Hospital Rigshospitalet	Denmark	Copenaghen	elsebet.ostergaard@dadlnet.dk
Jana	Paderova	FN Motol	Czech Republic	Prague	jana.inprague@gmail.com
Paraskevi	Apostolou	NCSR Demokritos	Greece	Athens	apostolouv@hotmail.com
Ilze	Radovica	Latvian Biomedical Research and Study Centre	Lettonia	Riga	ilze@biomed.lu.lv
Mario	Reiman	Human Molecular Genetics Group, Institue of Molecular and Cell Biology, University of Tartu	Estonia	Tudulinna Vald	kodu14@gmail.com
Fabio	Rosa	Human Genetics Foundation	Italy	Turin	fabio.rosa@hugef.torino.org
Laura	Santangelo	Sapienza University	Italy	Rome	sant.lori@yahoo.it
Marco	Savarese	TIGEM, Seconda Università di Napoli	Italy	Pagani	msavarese@tigem.it
Petra	Schneiderova	Faculty of Medicine and Dentistry, Palacky Unviersity	Czech Republic	Olomouc	petra.schneiderova@fnol.cz
Ying	Sheng	Oslo Univeristy Hospital, Dept. of Medical Genetics	Norway	Olso	ying.sheng@medisin.uio.no
Amihood	Singer	Barzilai Medical Centre Ashkelon	Israel	Tel Aviv	mihood@netvision.net.il
Tove	Skodje	Dept. of Medical Genetics, Oslo University Hospital	Norway	Oslo	tskodje@ous-hf.no
Anna	Sokolenko	N.N. Petrov Institute of Oncology	Russia	Saint-Petersburg	annasokolenko@mail.ru
Alexandra	Stambergova	University Hospital Motol	Czech Republic	Prague	alexandra.stambergova@fnmotol.cz
Markus	Storbek	Institute of Human Genetics	Germany	Cologne	markus.storbeck@uk-koeln.de
Evgeny	Suspitsin	St. Petesburg State Pediatric Medical Unvieristy	Russia	Saint-Petersburg	evgeny.suspitsin@gmail.com
Jack	Tarleton	Mission Health System – Fullerton Genetics Lab	U.S.A.	Asheville	jack.tarleton@msj.org

Fotis	Tsetsos	Dmocritus University of Thrace	Greece	Alexandroupolis	ftsetsos@mbg.duth.gr
Chana	Vinkler	Inst. Of Medicl Genetics Wolfson Medical Centre Holon Israel	Israel	Tel Aviv	vinkler2@gmail.com
Maj	Westman	Copenhagen University Hospital	Denmark	Copenaghen	maj.westman@regionh.dk
Martina	Wilke	Erasmus MC/Clinical Genetics	The Netherlands	Rotterdam	m.wilke@erasmusmc.nl
Thilini Hansamali Gamage	Yaddehi Gamage		Sri-lanka	Oslo	thilinigamage87@yahoo.com
Andrea	Zauli	Ospedale San Raffaele	Italy		zauli.andrea@hsr.it

<div style="border:2px solid black; text-align:center">

Participants - ABSTRACTS

</div>

Phenet: Network Analysis for Gene Prioritization of Exome Sequencing Results in Syndrome Patients

Johanna Christina Czeschik

[1]J.C. Czeschik, [2]D. Beisser, [3]Y. Alanay, [4]P.O. Simsek-Kiper, [1]D. Wieczorek, [2]S. Rahmann

[1] Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany

[2] Genominformatik, Institut für Humangenetik, Universität Duisburg-Essen, Essen, Germany

[3] Pediatric Genetics, Department of Pediatrics, Acibadem University School of Medicine, Istanbul, Turkey

[4] Ihsan Dogramaci Children's Hospital, Clinical Genetics Unit, Ankara, Turkey

Exome sequencing has been successfully established as a useful tool for scientific discovery of new disease genes and diagnosis of patients with unknown syndromes. When the exome of a single human individual is sequenced and compared to a reference genome, 20.000 to 50.000 variants are usually identified. After filtering of synonymous and non-coding variants and presumably non-pathogenic variants present in public databases, several hundred variants remain. Therefore, methods are required to narrow down the number of variants that are considered candidates in the search for the disease-causing variants. Because manual review of all variants by a human expert is a very time-consuming task, a variety of computational methods to predict disease-causing variants from the set of all variants have been suggested, so-called gene prioritization methods. Among these, network analysis of protein-protein interaction networks has been proven successful in the prediction of disease genes. Here, we propose a new method for the prediction of disease genes of genetic syndromes based on 1.) the HPRD protein-protein interaction network and 2.) a set of disease genes derived from patients' phenotypes, using phenotype-gene associations available as

part of the Human Phenotype Ontology. The method is validated using 1.) 100 computationally generated patient data sets, and 2.) exome sequencing data of four patients with two genetic syndromes (Nager syndrome and Coffin-Siris syndrome) whose causative genes were unknown at the time of analysis, but have been identified in the meantime.

A balanced *de novo* inv(7)(p14.3q22.3) disrupting *PDEC1* and *ATXN7L1* in a 12-year old developmentally delayed boy
Thilini H. Gamage

Thilini H. GamageI, Doriana MisceoI, Madeleine FannemelI, Eirik FrengenI

I Department of Medical Genetics, Oslo University Hospital and University of Oslo, Norway

Abstract

We report a 12 year old male patient ascertained for developmental delay, carrying a *de novo* pericentric inversion on chr(7)(p14.3q22.3). Sequencing revealed that the breakpoints overlap a LTR sequence on 7q22.3 and a LINE on 7p14.3. A TTTAAA motif was found in proximity of the breakpoints on both arms. In addition the sequencing detected several small micro-rearrangements, deletion, duplication, insertion, at the breakpoints. No significant sequence identity exists between the 7p14.3 and 7q22.3 breakpoints. These features at the breakpoint junctions suggest that the inversion was triggered by TTTAAA motif, LTR and LINE and healed by a Non Homologous End Joining (NHEJ) mechanism. The genes *ATXN7L1* and *PDE1C* are disrupted by the inversion. *PDE1C* is responsible for the hydrolysis of the second messenger molecules cAMP and cGMP and is highly expressed in the human heart and certain brain regions. In mice, *Pde1c* is expressed in migrating neuronal cells within the central nervous system during early embryo development. Although neuronal migration disorder was not seen in our patient, this is the first patient described with haploinsufficiency of *PDE1C* possibly causing developmental delay.

Analysis of acute ischemic stroke DNA markers in moldavians
Elena Mocan

Mocan Elena, Barbacar Nicolae, Lisii Leonid

The State University of Medicine and Pharmacy "Niolae Testemitanu"

The aim of study was to evaluate the efficiency of gene candidate polymorphisms as markers of ischemic stroke (IS) in the population of the Republic of Moldova.

Tasks: to analyze the prevalence of polymorphisms of *F2*, *F5*, *GP1BA*, *PAI-1*, *ACE*, *ALOX5AP* and *PDE4D* candidate genes in studied groups; to examine the relationship of candidate genes polymorphisms with values of blood pressure, some biochemical parameters depending on age and sex; based on the patterns of genes to appreciate significance of key-genes of ischemic stroke; to describe the influence of smoking, behavioral risk factor of IS in the interaction between candidate genes.

Materials and methods: Using a case-control design, we studied 156 patients with ischemic stroke and 89 unrelated population-based controls, age- and sex-matched. The studied polymorphisms were genotyped by the PCR, technology TaqMan®.

Scientific results: For the first time the comprehensive analysis of the genetic susceptibility to IS in the population of the Republic of Moldova was held. It was detected the contribution of studied polymorphic gene variants *F2, F5, GP1BA*, *PAI-1*, *ACE*, *ALOX5AP* and *PDE4D* and some modifying risk factors (smoking, indices of lipid metabolism) in the susceptibility to ischemic stroke in the population of the Republic of Moldova. Statistically significant differences were found in biochemical parameters such as total cholesterol, triglycerides, HDL-cholesterol in group with ischemic stroke compared with the control group. The multivariate logistic regression model was used to exclude the influence of the conventional vascular risk factors on stroke. It's proved that AC haplotype, which includes alleles A rs10507391 locus and C allele rs9551963 locus of *ALOX5AP* gene is associated with increased levels of triglycerides and a decreased concentration of HDL-cholesterol with age and determine the increased risk to ischemic stroke. To specify the high and low-risk genotypes of genes-modifiers and to seek a genetic model (either single or multilocus) that predicts prognosis of ischemic stroke we also performed the multifactor dimensionality reduction analysis. There were created the 4-locus model of the following combinations of polymorphic loci rs4646994 *ACE* (Ins>Del), rs9551963 *ALOX5AP* (A>C), rs2910829 *PDE4D* (G>A), rs1799768 *PAI-1* (5G> 4G) with sufficient reproducibility (60%) and the high sensitivity and specificity, 0.75 and 0.69 (p<0.0001), respectively. Also we described the impact of smoking on the type and power of intergenic interactions between loci rs9551963 *ALOX5AP* (A>C) and rs2910829 *PDE4D* (G>A).

Conclusions: It's recommended to identify the carriers of D allele of *ACE* gene for men and A allele of locus rs10507391 of *ALOX5AP* gene for women for early detection of individuals with high risk of cerebrovascular diseases and prenosological prevention.

Immune infiltration and choroidal melanomas

Deepti Narasimhaiah

Narasimhaiah D[1], Remark R[2], Damotte D[2], Legrand C[3], De Potter P[4], Coulie PG[5], Vikkula M[6], Godfraind C[1]

[1]Laboratory of Pathology, [3]Institute of Statistics, Biostatistics and Actuarial Sciences, [4]Department of Ophthalmology, [5]de Duve Institute , [6]Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Belgium. [2] Centre de Recherche des Cordeliers, Laboratoire microenvironnement immunitaire et tumeurs, Inserm UMRS 872, Paris, France.

Introduction

Choroidal melanomas are the most common primary, intra-ocular malignant tumors in adults. According to the literature, about 50% of them metastasize to liver, with less than one year post-metastatic survival. In these tumors, the presence of CD8+ T lymphocytes is associated with bad prognosis. This is in contrast to most others solid tumors, including cutaneous melanomas, where it is related to good prognosis.

Materials and Methods

The study cohort comprised of 89 untreated choroidal melanomas, with and without ciliary body involvement. They were analysed (1) for gene expression using Affymetrix U133 plus 2.0 array (n=15) (2) by immunohistochemistry (n=89) and (3) using descriptive statistics and time-to-events data (SPSS v20).

Results

Analysis of gene expression profiles led to the identification of a gene signature in a subset of choroidal melanomas. This signature comprised of 39 genes which were upregulated in these tumors. The pathways associated with these genes were: antigen processing and presentation, interferon-gamma signaling, cell adhesion molecules and chemokine signaling. On immunohistochemistry, based on semi-quantitative assessment, the signature positive tumors demonstrated a dense intra-tumoral infiltrate of HLA-DRA+CD163+ macrophages and CD3+CD8+ T-cells. CD3+CD4+ T lymphocytes were absent. The tumors lacking the signature showed mild to moderate infiltrate of macrophages and few to absent T-cells. Thus, the mRNA signature was translated into immunohistochemistry and applied to the rest of the series. This resulted in a total of

19 melanomas with high immune infiltrate and 70 tumors with low immune infiltrate. These immune cells were localized at lateral edges, apex, centre and base, with a heterogeneous distribution. In tumors with high immune infiltrate, the macrophages and T-cells were observed in all five regions. In contrast, in low immune infiltrate tumors, there were inter-regional differences, with the edges being localized first and the base last (P<0.001, Kruskal-Wallis test). Also, the assessment of predictive accuracy of each region showed the tumor centre to have the highest sensitivity (100%) and specificity (90%). The Kaplan-Meier plots demonstrated distinct differences in disease-free survival between immune high and immune low melanomas (log-rank P < 0.001). The tumors with high immune infiltrate had shorter disease-free survival.

Conclusions

We report the identification of a gene signature in choroidal melanomas associated with high infiltrate of lymphocytes and macrophages. This immune infiltrate is interferon-gamma induced and develops in the context of T_H1 orientation with CD8 effectors. There is heterogeneity in topographic localization of immune infiltrate and "tumor centre" is the most suitable region for assessment of immune infiltrate in a biopsy specimen. The presence of high immune infiltrate also correlates with shorter time to metastases.

X-exome sequencing revealed a Novel Mutation in the CUL4B gene underlying Cabezas syndrome

Anju Philips

A.K. Philips[1], K.Avela[2], M.Somer[2] V.Kalscheuer[3], S.Haas[3], I.Järvelä[1]
1. Department of Medical Genetics, Helsinki, Finland
2. Department of Medical Genetics, Family Federation of Finland, Helsinki, Finland
3. Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany

Background

CUL4B gene encodes a member of Cullin-RING ubiquitin ligase complex. *CUL4B* is an ubiquitin E3 ligase subunit implicated in the regulation of several biological processes. Mutations in *CUL4B* are known to underlie X-Linked Intellectual disability.

Methods

We performed Agilent SureSelect enrichment of all X-chromosome specific exons from the index patient followed by massively parallel sequencing. The novel mutation was confirmed by Sanger sequencing and segregated among the family members except for one patient. All variants were filtered against publicly available database.

Results

We identified a novel splice site mutation c.2493+3A>G in the *CUL4B* gene that has previously shown to cause Cabezas syndrome. This novel splice site mutation leads to aberration in *CUL4B* splicing. The mutation c.2493+3A>G ruins the splice donor site of exon 20, which leads to the activation of two cryptic splice donor sites that were present within exon 20, that both resulted in a frameshift and a premature termination codon in exon 20.

Conclusions

Our findings elucidate the functional significance of *CUL4B* in human cognition and in other aspects of human development. This study illustrates the importance of X-exome sequencing in discovering disease causing mutations.

Ilze Radoviča

The heritability of high-density lipoprotein cholesterol (HDL-C) level is estimated at approximately 50%. Recent genome-wide association studies have identified genes involved in regulation of high-density lipoprotein cholesterol (HDL-C) levels. The precise genetic profile determining heritability of HDL-C however are far from complete and there is substantial room for further characterization of genetic profiles influencing blood lipid levels.

Here we report an association study comparing the distribution of 139 SNPs from more than 30 genes between groups that represent extreme ends of HDL-C distribution. We genotyped 704 individuals that were selected from Genome Database of Latvian Population.

10 SNPs from *CETP* gene showed convincing association with low HDL-C levels (rs1800775, rs3764261, rs173539, rs9939224, rs711752, rs708272, rs7203984, rs7205804, rs11076175 and rs9929488) while 34 SNPs from 10 genes were nominally associated ($p < 0.05$) with HDL-C levels. We have also identified haplotypes from *CETP* with distinct effects on determination of HDL-C levels.

Our conclusion: So far the SNPs in *CETP* gene are identified as the most common genetic factor influencing HDL-C levels in the representative sample from Latvian population.

The diagnosis of neuromuscular disorders: an exhaustive "Next Generation" approach.

Marco Savarese

In the last few years, Next Generation Sequencing (NGS) has been revolutionizing biological research. In particular, in the field of heterogeneous genetic conditions, like muscular dystrophies, NGS could be considered a powerful tool to detect mutations in well-known genes, to identify novel causative genes and to characterize the effects of the mutations on mRNA expression and/or splicing.

We have developed a NGS-based workflow for an exhaustive analysis of patients affected by neuromuscular disorders. Up to now, we have recruited over 400 cases of familial or sporadic nonspecific limb-girdle muscular dystrophies or congenital myopathies. For all the DNA samples, 2447 exons of 98 causative genes have been investigated after a custom enrichment by Haloplex technology (*Motor HaloPlex*). Using different pooling strategies, several hundred different mutations have been identified and confirmed by Sanger sequencing.

A subset of undiagnosed familial cases has been analyzed by exome sequencing. Finally, available muscular biopsies have been selected for RNA sequencing: in fact, we have designed a customized RNA Target Enrichment System (*Motor RNA Seq*), focused on the transcripts of the 98 causative genes described above.

Our preliminary results confirm the ability of NGS to detect clinically significant variations and its utility in routine diagnosis. Moreover, they evidence the great clinical and genetic heterogeneity present in neuromuscular disorders, allowing us to investigate their molecular basis.

Next Generation Sequencing for Diagnostics in Department of Medical Genetics, Oslo University Hospital

Ying Sheng

Ying Sheng, Tove Skodje, Hanne S Sorte, Eidi Nafstad, Timothy Hughes, Yvan Strahm, Gregor Gilfillan, Benedicte Paus, Tony Håndstad, Olaug K Rødningen, Lars Rettersdøl, Beate Skinningsrud, Robert Lyle, Dag Erik Undlien Department of Medical Genetics, Oslo University Hospital, Norway

The department of Medical Genetics (DMG) in Oslo University Hospital (OUS) is the largest medical genetic facility in Norway. The main functions of the department are clinical genetic assessment, genetic counseling of individuals and families with a suspected genetic disease, and

guidance in connection with prenatal diagnosis. The department also conducts research in most areas of human genetics and modern molecular biology to find the molecular basis of diseases.

The Norwegian High-Throughput Sequencing Center (NSC) (www.sequencing.uio.no) is a national technology core facility offering highthroughput sequencing services from different next-generation sequencing technologies (e.g. Illumina, PacBio). It is a consolidation of the sequencing platforms at the DMG, OUS and Center for Ecological and Evolutionary Synthesis, University of Oslo (UiO). The DMG node has a particular focus on targeted resequencing and functional genomics applications. The node currently has 2 HiSeq (HiSeq 2000 and HiSeq 2500) and 2 MiSeq machines. It has successfully sequenced 10,000 giga base pairs of DNA data since 2008.

By application of Illumina technology, the department started to develop and implement gene panel based whole exome sequencing (WES) for clinical diagnostics from 2011. Up until the end of 2012, we had a pilot project which only clinicians in the department could refer samples to. This test was performed after a board assessed that the patient's medical history and physical examinations strongly suggested that there was an underlying genetic etiology. Exome enrichment was performed using the Agilent SureSelect Human All Exome kit and bioinformatic analysis was performed by an in-house pipeline. The data were filtered and only specific genes related to the patients clinical manifestations were analyzed. A total of 56 cases (patients) have been analyzed with a success rate of 28% (reported with probably causal variants).

Another 34% were reported to have variants of uncertain clinical significant. In 2013, the department continues to use the same exome enrichment protocol and analysis pipeline, however, the submission strategy is changed to only offer fixed gene lists. The first offered gene list is used for diagnostics of Epileptic Encephalophathy diseases. Another four gene lists (for Ciliopathies, Hereditary spastic paraplegia and taxa (HSP and HA), Ichthyosis and Connective Tissues Disease) are under development. We are also investigating Halo capture technology for targeted sequencing. The two pilot projects are focused on cardiomyopathies and inherited breast cancer.