ESHG

EUROPEAN SOCIETY OF HUMAN GENETICS

European Human Genetics Conference 2013

June 8 - 11, 2013, Paris, France

Около 4000 участников

Основные направления:

1. NGS секвенирование (во всех областях генетики)
2. Неинвазивная диагностика

Темы:

1.Неинвазивная диагностика. Фирмы. Новая технология
2.NGS секвенирование. Фирмы. Направления исследований. Трио. Технологии. Обработка данных. Экзом. Стоимость.
3.Этногенетика
4.Другое. Геном Голландца. Варием.

Наиболее частые заболевания:
РМЖ
Шизофрения, аутизм, слабоумие, синдромы
Ожирение, СД2

ION Pronon

1. 8 образцов на 316 чипе (у нас такого нет)
2. CNV – Alu не идентифицируют
3. 23 часа (из них 1 час ручной работы)
4. 3*10 ng DNA
5. False positiv – 1:34 образцов (скрининг BRCA)


Ion AmpliSeq Community Panels
BRCA 1&2 Panel
2 genes
167 amplicons
Colon and Lung Cancer Panel
22 genes
90 amplicons
AML Panel
Coding regions of known mutations
21 genes
CFTR Panel
All exons, intron-exon boundaries, and UTRs
1 gene
Cardio Panel
All exons and UTRs
62 genes
TP53 Panel
All exons and UTRs
1 gene
ion torrent
life

Peter Ray
40 тыс анализов в год
Муковисцидоз
Для дозы гена-MLPA, Проблема 10 экзон


Exome Sequencing for Genetic Analysis of Myopathies
Exome data generation => Data Filtering on Regions of Interest (Gene Lists)
Overall harmonization of the analysis workflow
In our lab: adapted to diverse heterogeneous diseases
Company-validated Exome Kits
No custom-design validation
Easy updates of Gene Lists
Readiness for future Personal Genome Sequencing (Genetic disorders/predisposition)
Myopathies 70 genes
Peripheral Neuropathies 30 genes
Dystonia 29 genes
Intellectual Disability 18 genes
Epilepsy 36 genes
Premature aging 31 genes
Deafness 14 genes

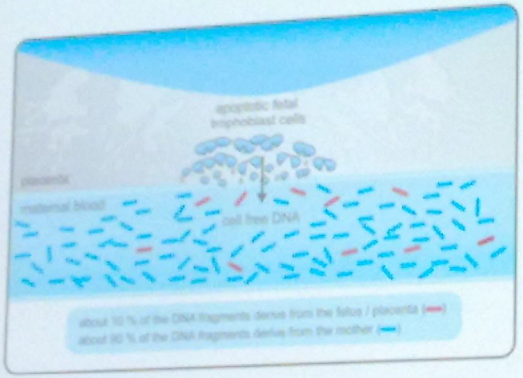
Неинвазивная диагностика.

1. 30000 SNP
2. 11-32 неделя
3. Нужно 30 млн 36 нуклеотид ридов
4. 466/468
5. 1 ложнопол и 1 ложноотриц
6. 1% - нет результатов

7. Фирмы:
8. www.harmonytest.com
9. www.panoramatest.com
10. www.lifecodexx.com
11. www.prenatalarray.org

Scientific Basis of NIPT – Cell Free Fetal DNA (cffDNA)

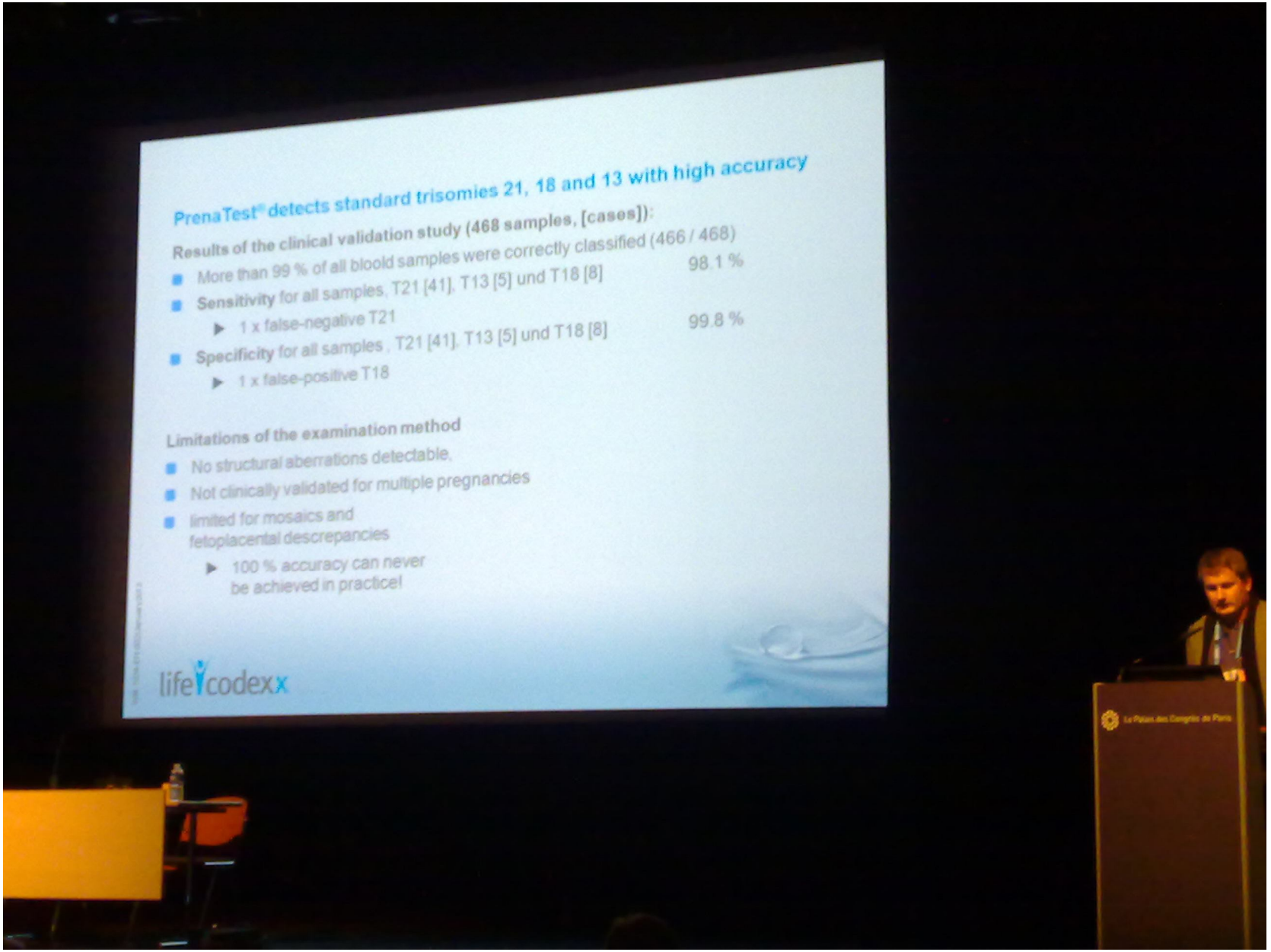
1997: detection of cffDNA fragments in maternal plasma



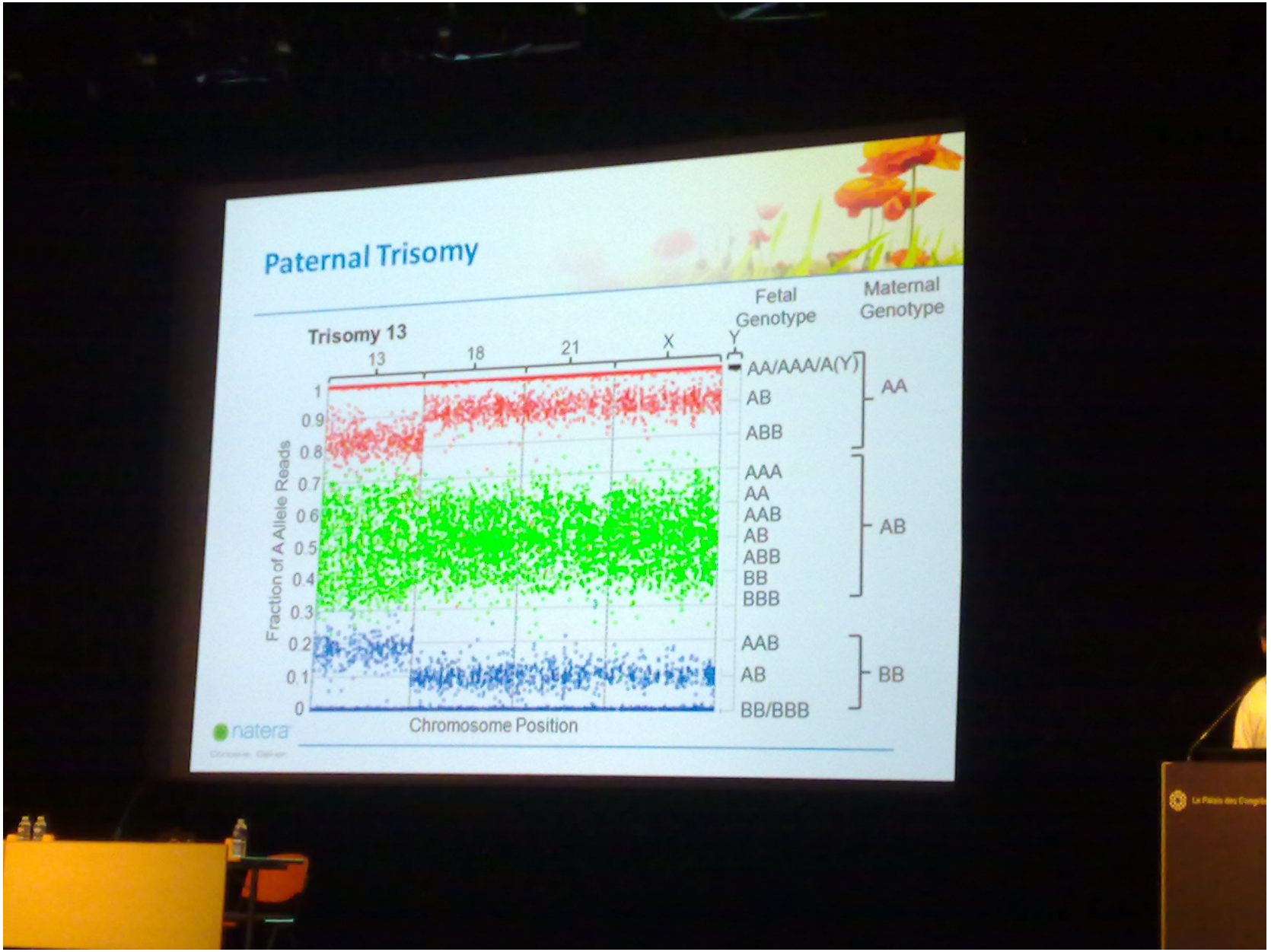
- **Detection:** starting week 4
- **Fetal fraction:** 2 – 40 %
- **Stability:** < 2 hours
- **Characteristics:**
  - short fragments, 80 % < 200 bp
  - released from trophoblast cells (placenta) by apoptosis / necrosis

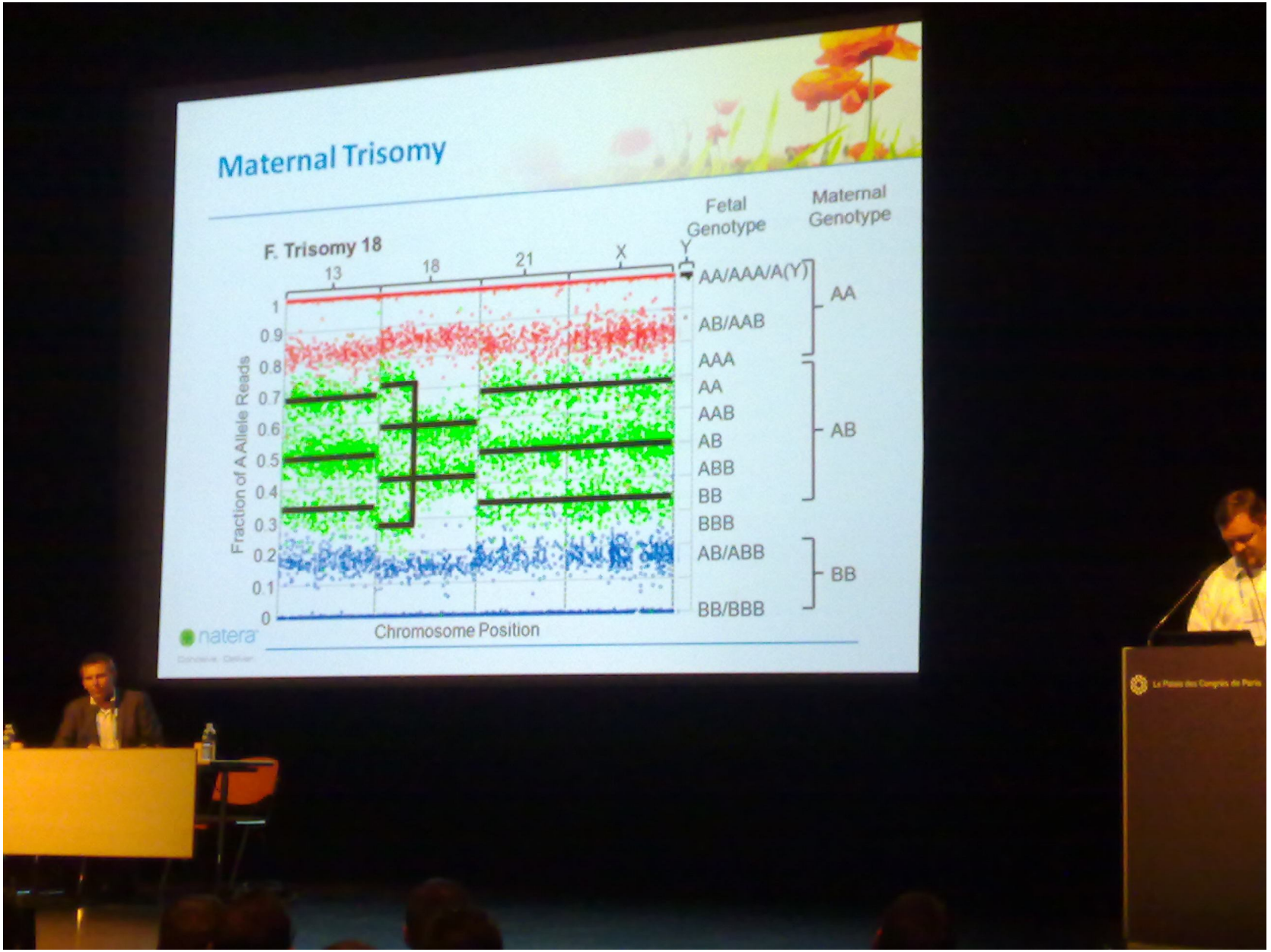
lifecodexx

* Lo et al., Am. J. Hum. Genet. 1997; 62: 768-775

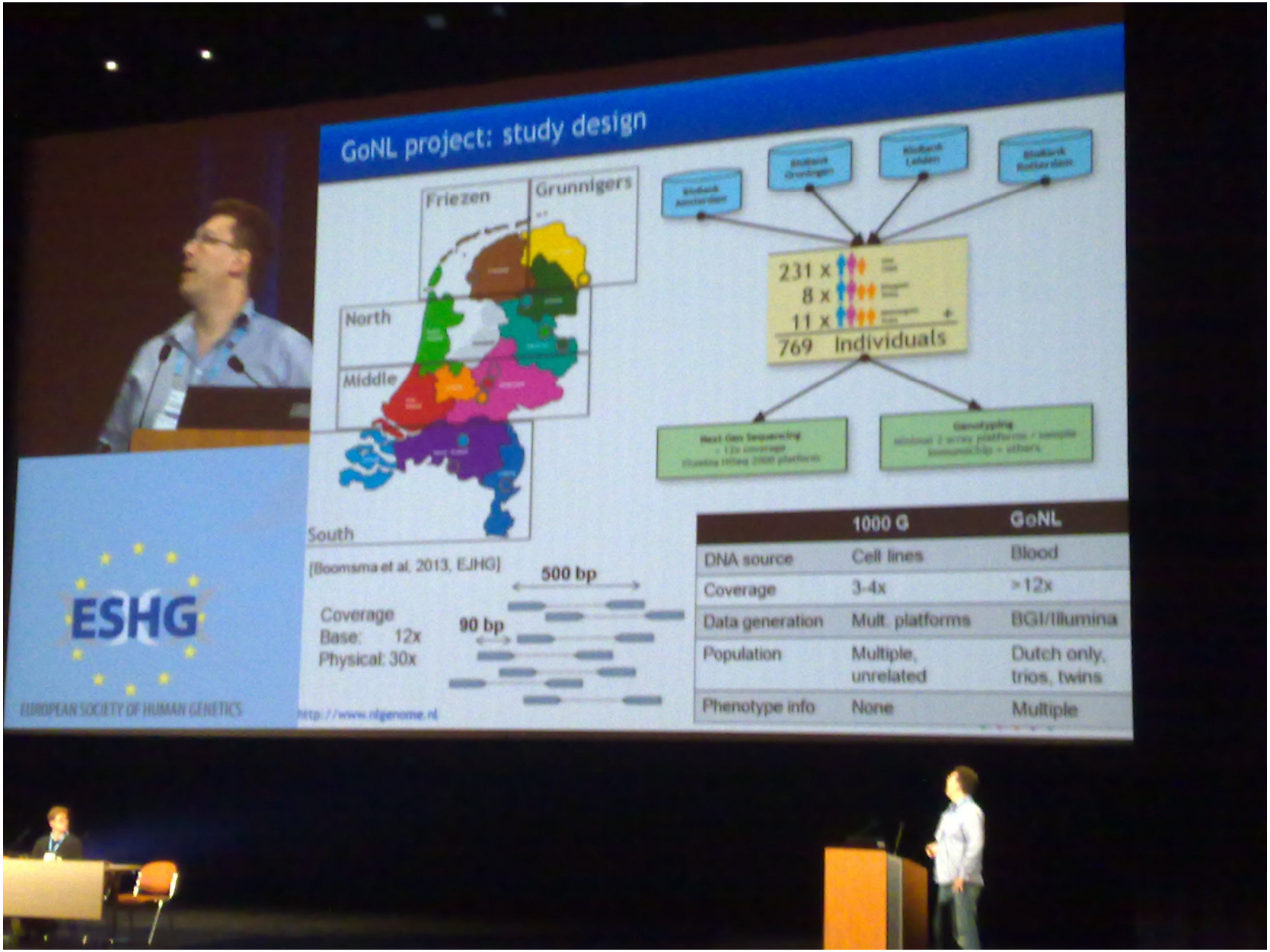

PrenaTest® detects standard trisomies 21, 18 and 13 with high accuracy
Results of the clinical validation study (468 samples, [cases]):
More than 99 % of all bloold samples were correctly classified (466 / 468) 98.1 %
Sensitivity for all samples, T21 [41], T13 [5] und T18 [8]
1 x false-negative T21
Specificity for all samples , T21 [41], T13 [5] und T18 [8] 99.8 %
1 x false-positive T18
Limitations of the examination method
No structural aberrations detectable,
Not clinically validated for multiple pregnancies
limited for mosaics and fetoplacental descrepancies
100 % accuracy can never be achieved in practice!
lifecodexx

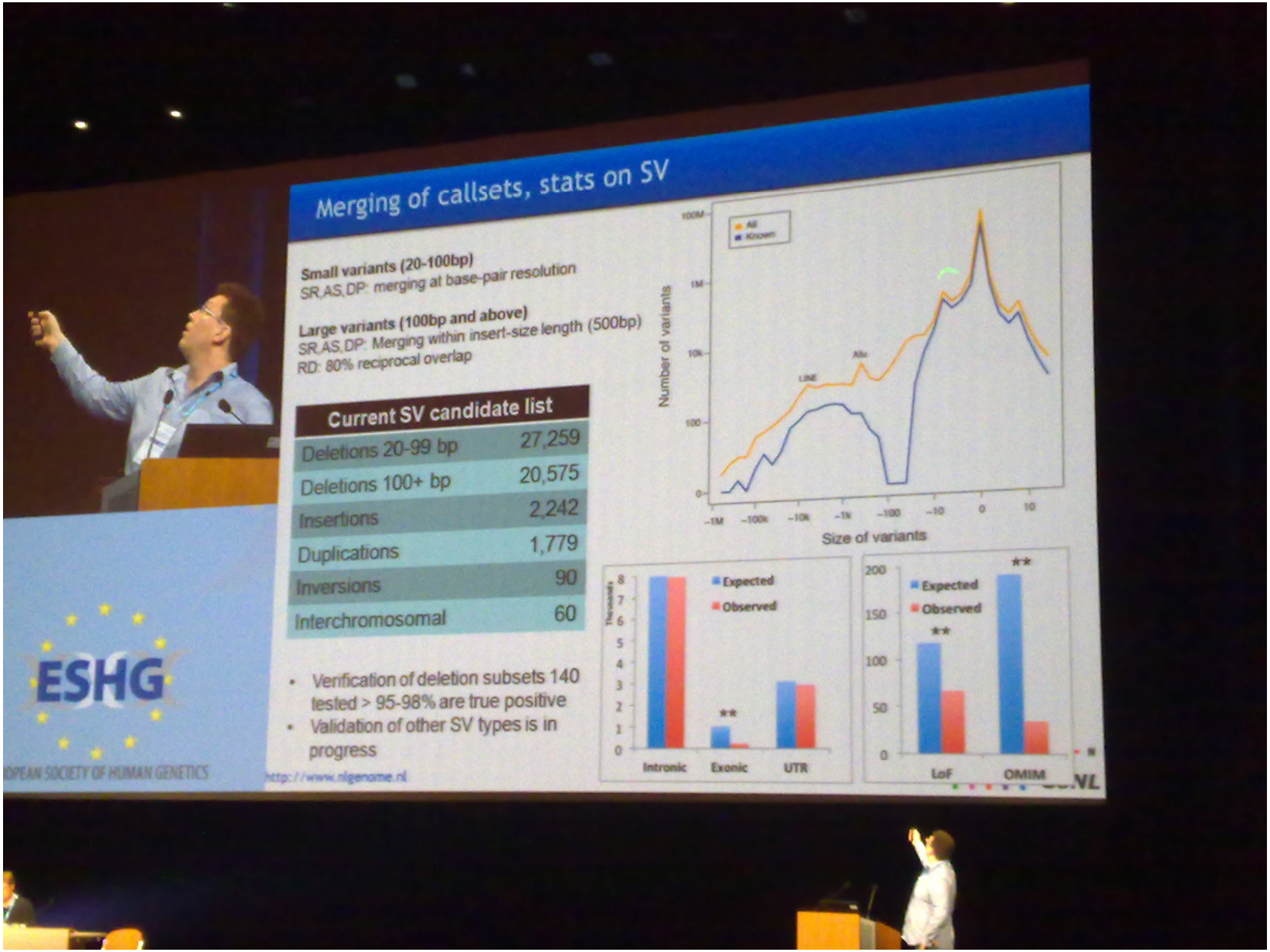

Experiences Since Market Launch in August 2012
More than 3500 blood samples analysed
in about 98 % the result was negative for T21
in about 1.3 % the result was positive for T21
in < 1 % „no result" could be reported due to a repeatedly low amount of fetal DNA
< 1%
98 %
1,3%
lifecodexx
Le Palais des Congrès de Paris


Paternal Trisomy
Trisomy 13
13
18
21
X
Y
Fetal Genotype
Maternal Genotype
AA/AAA/A(Y)
AB
ABB
AAA
AA
AAB
AB
ABB
BB
BBB
AAB
AB
BB/BBB
AA
AB
BB
Fraction of A Allele Reads
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Chromosome Position
natera


Maternal Trisomy
F. Trisomy 18
13
18
21
X
Y
Fetal Genotype
Maternal Genotype
AA/AAA/A(Y)
AB/AAB
AAA
AA
AAB
AB
ABB
BB
BBB
AB/ABB
BB/BBB
AA
AB
BB
Fraction of A Allele Reads
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Chromosome Position
natera

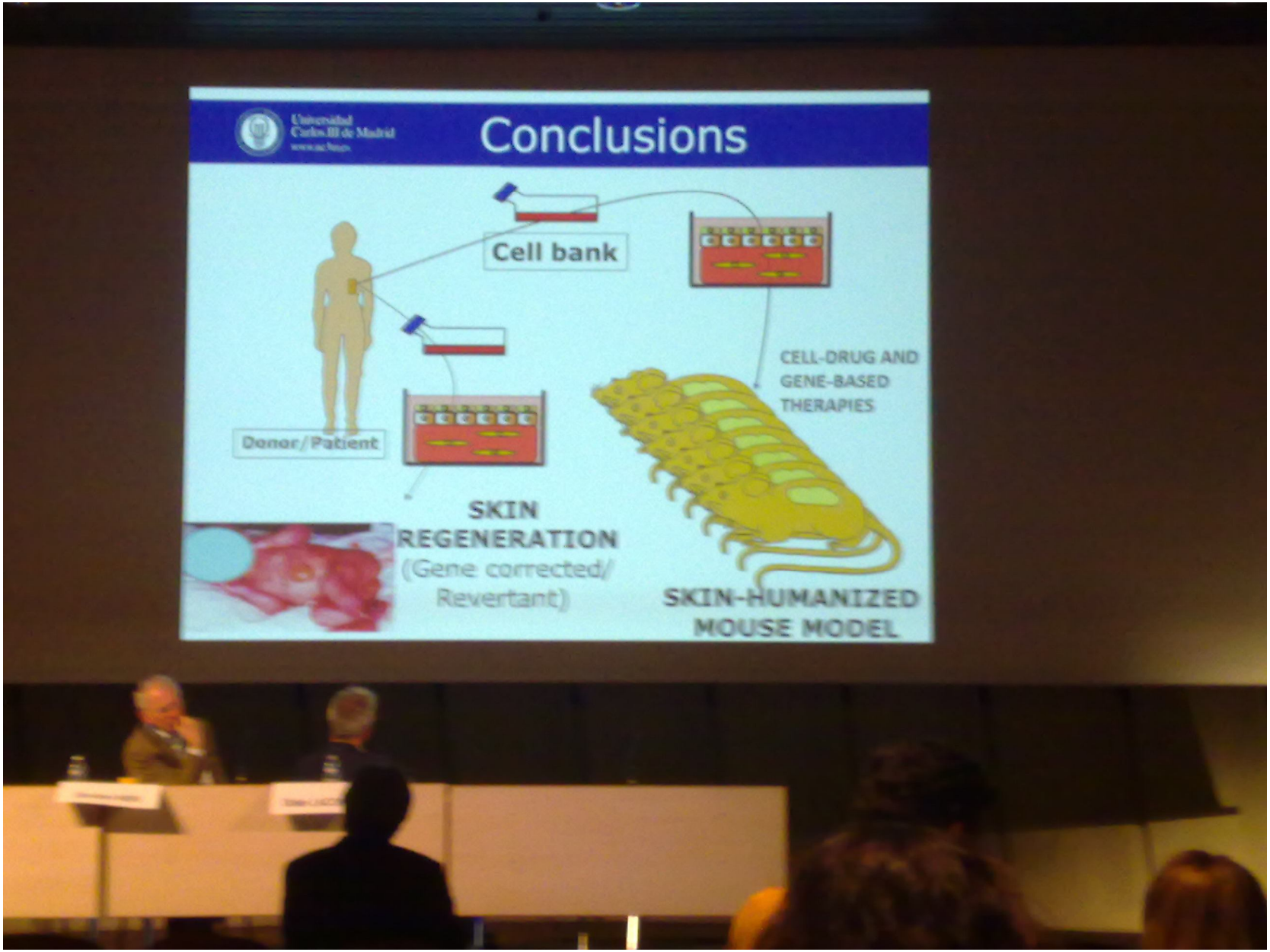

Summary
Developed molecular fetal karyotyping by targeted sequencing from pregnancy plasma DNA.
Targeting single-nucleotide polymorphisms obviates the need for a reference chromosome and allows incorporation of a large amount of information when making copy number calls.
By targeting SNPs, this approach identifies fetal ploidy state and parental origin. Detection of T21, T18, T13, sex chromosome aneuploidies, triploidy have been validated.
Correct ploidy can be called down to below 4.0 % fetal DNA.
Panorama was launched December 2012.
Approach also works on genomic DNA down to single cells: AF, POC, PGS.
Up to 30'000 SNPs have been successfully amplified.
natera
23


GoNL project: study design
Friezen
Grunnigers
North
Middle
South
231 x
8 x
11 x
769 Individuals
[Boomsma et al, 2013, EJHG]
500 bp
90 bp
Coverage
Base: 12x
Physical: 30x
1000 G
GoNL
DNA source
Cell lines
Blood
Coverage
3-4x
>12x
Data generation
Mult. platforms
BGI/Illumina
Population
Multiple, unrelated
Dutch only, trios, twins
Phenotype info
None
Multiple
ESHG
EUROPEAN SOCIETY OF HUMAN GENETICS
http://www.nlgenome.nl


Merging of callsets, stats on SV
Small variants (20-100bp)
SR,AS,DP: merging at base-pair resolution
Large variants (100bp and above)
SR,AS,DP: Merging within insert-size length (500bp)
RD: 80% reciprocal overlap
Current SV candidate list
Deletions 20-99 bp 27,259
Deletions 100+ bp 20,575
Insertions 2,242
Duplications 1,779
Inversions 90
Interchromosomal 60
Verification of deletion subsets 140 tested > 95-98% are true positive
Validation of other SV types is in progress
Number of variants
Size of variants
All
Known
LINE
Alu
Expected
Observed
Intronic
Exonic
UTR
LoF
OMIM
ESHG
EUROPEAN SOCIETY OF HUMAN GENETICS
http://www.nlgenome.nl


EuroGentest
EuroGentest wants to build on these available documents and the work of others, and then comment or amend where necessary.
* US: Nex-StoCT (CDC)
Nat Biotechnology 20 (2012): 1033-36
* UK: Targeted NGS Guidelines (CMGS)
* AUSTRALIA: Concept NGS Guidelines (MPS-WG)
* THE NETHERLANDS: Concept NGS Guidelines (VKGL)
* ESHG PPPC: Recommendations for WGS
Discussed at expert meeting, February 2013


Research
≠
Diagnostics


Universidad Carlos III de Madrid
Conclusions
Cell bank
Donor/Patient
CELL-DRUG AND GENE-BASED THERAPIES
SKIN REGENERATION
(Gene corrected/ Revertant)
SKIN-HUMANIZED MOUSE MODEL

Abbott Molecular FMR1 TP-PCR

Journal of Molecular Diagnostics, Vol. 12, No. 4, July 2010

A Simple, High-Throughput Assay for Fragile X Expanded Alleles Using Triple Repeat Primed PCR and Capillary Electrophoresis

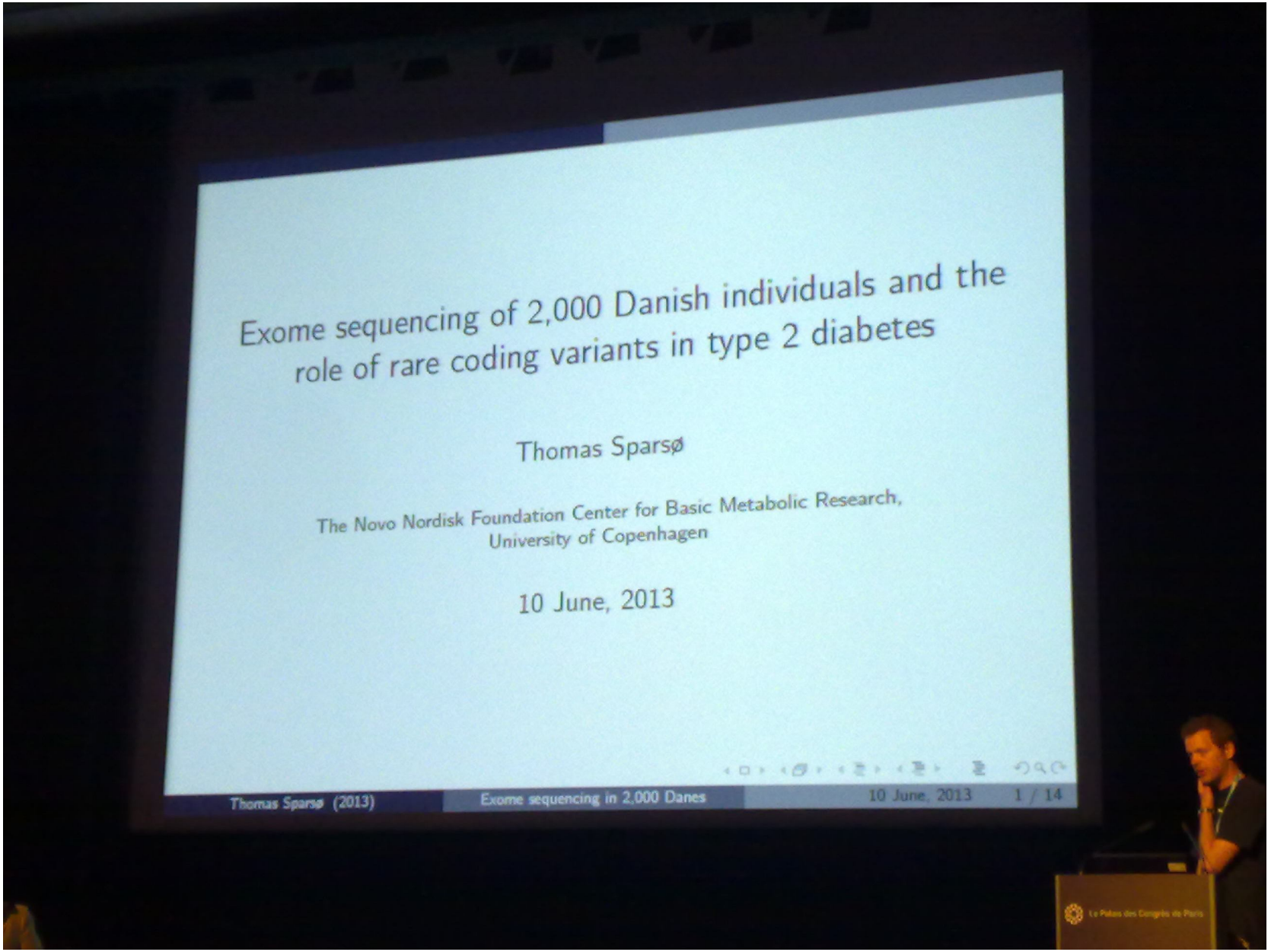
Elaine Lyon,* Thomas Laver,† Ping Yu,‡ Mohamed Jama,† Keith Young,† Michael Zoccoli,† and Natalia Markova‡

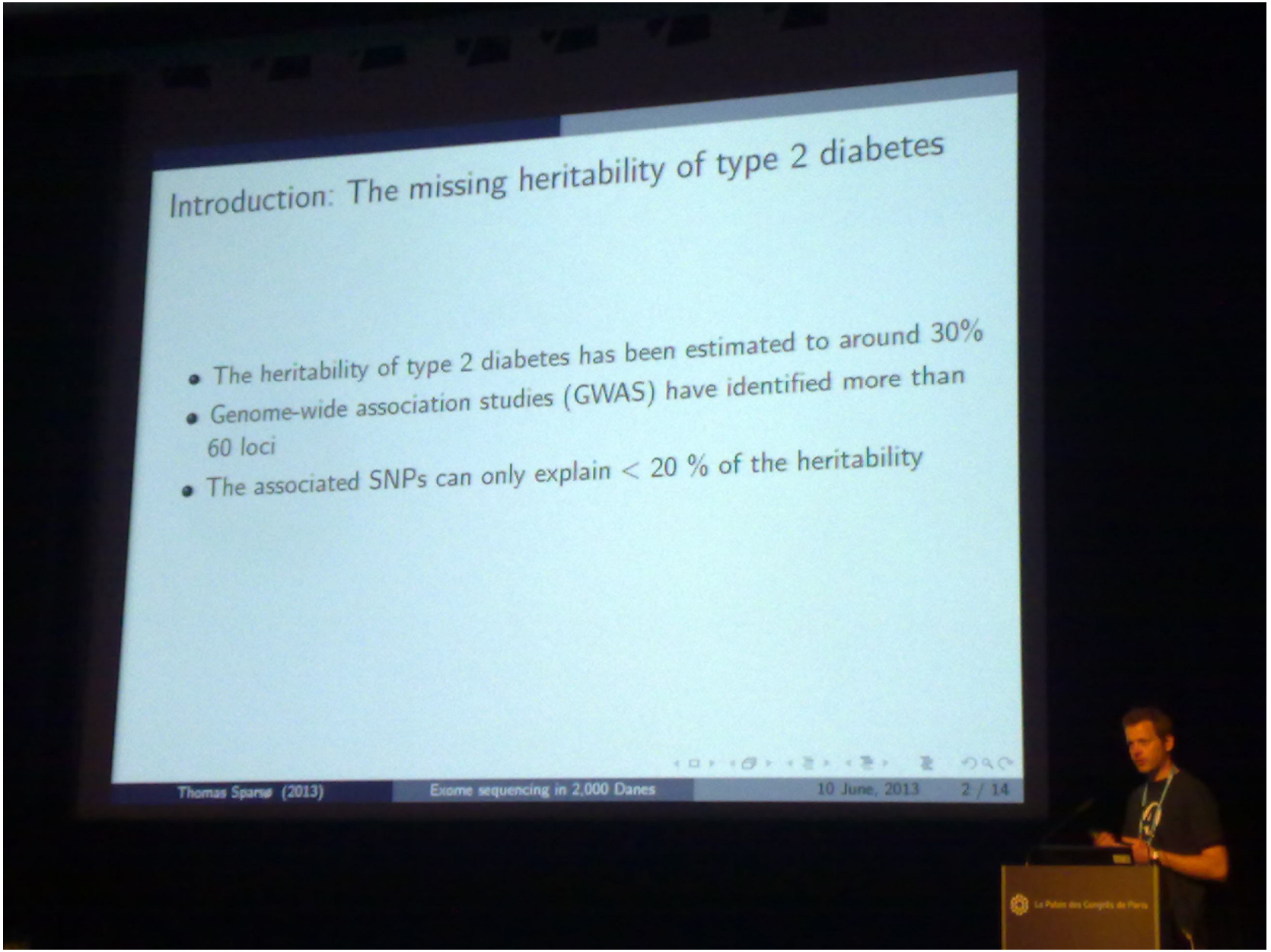
205 samples tested

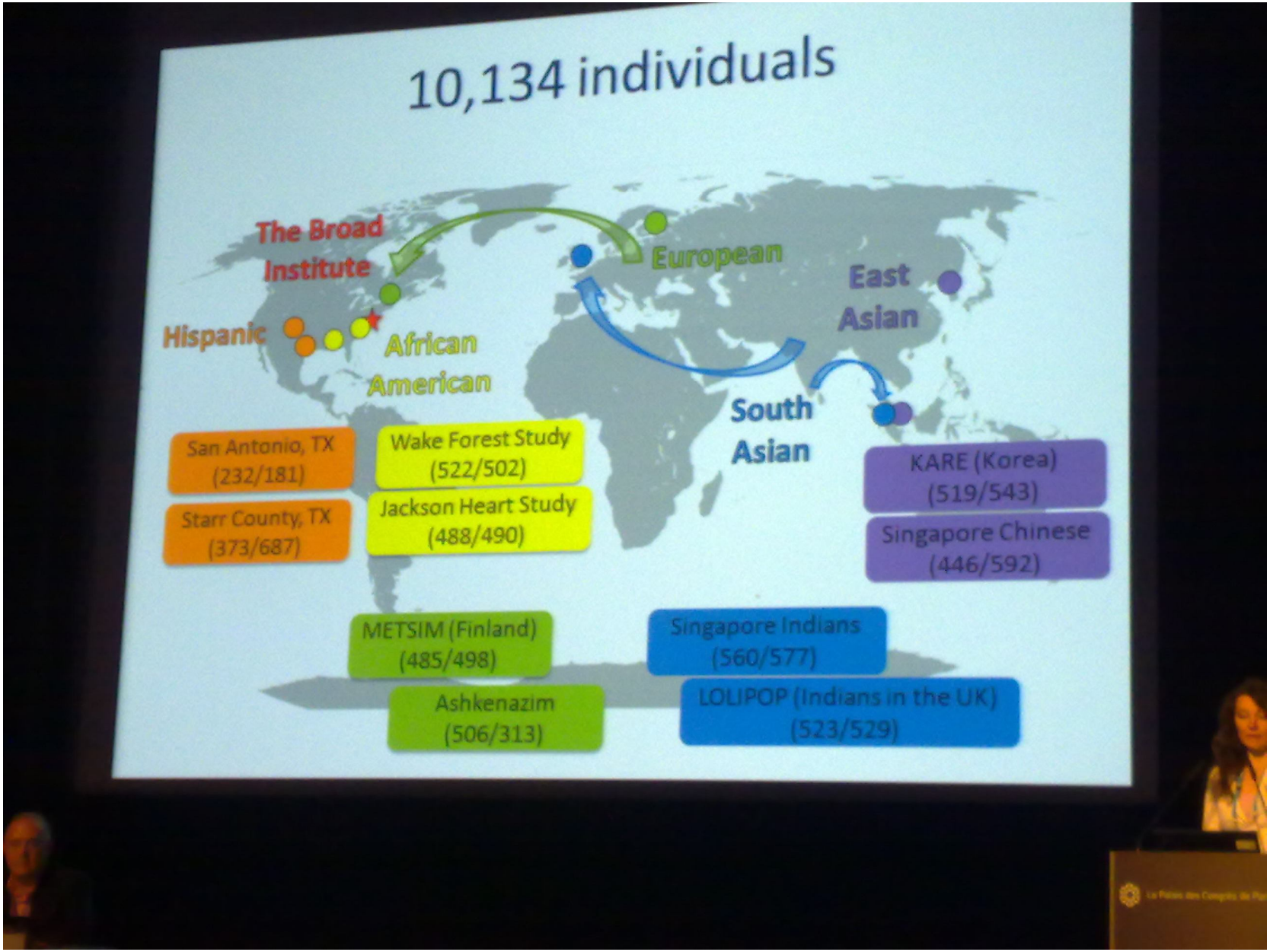
- ✓ 73 normal/intermediate, 59 PM, 71 FM, 2 mosaic
- ✓ 100% sensitivity/specificity versus PCR/Southern blot
- ✓ Works with blood spots
- ✓ Works with artificial mosaics (to 12.5% in the sample)

High sensitivity and specificity for detecting expanded alleles.
If no expansion is detected, no further testing is necessary.

Abbott
A Promise for Life


Exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes
Thomas Sparsø
The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen
10 June, 2013
Thomas Sparsø (2013)
Exome sequencing in 2,000 Danes
10 June, 2013
1 / 14


Introduction: The missing heritability of type 2 diabetes
• The heritability of type 2 diabetes has been estimated to around 30%
• Genome-wide association studies (GWAS) have identified more than 60 loci
• The associated SNPs can only explain < 20 % of the heritability
Thomas Sparsø (2013)
Exome sequencing in 2,000 Danes
10 June, 2013
2 / 14

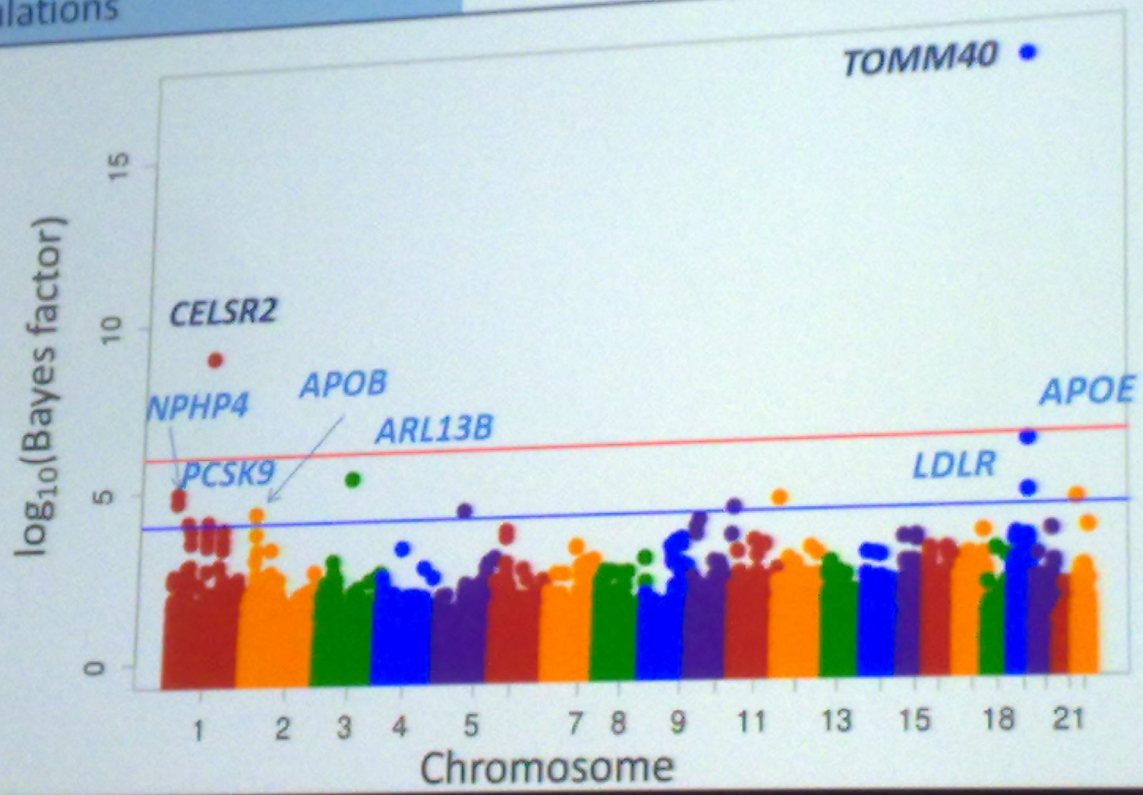

10,134 individuals
The Broad Institute
European
East Asian
Hispanic
African American
South Asian
San Antonio, TX (232/181)
Wake Forest Study (522/502)
Starr County, TX (373/687)
Jackson Heart Study (488/490)
KARE (Korea) (519/543)
Singapore Chinese (446/592)
METSIM (Finland) (485/498)
Singapore Indians (560/577)
Ashkenazim (506/313)
LOLIPOP (Indians in the UK) (523/529)

Meta-analysis LDL-C

Variants, have more similar effects in more similar populations

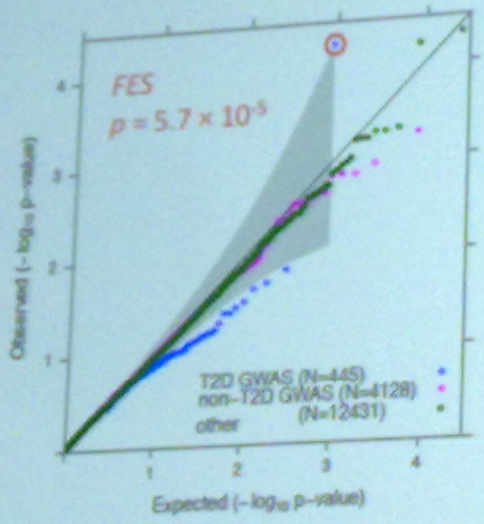
MANTRA
(Morris. Genet Epidemiol. 2011)

Similar genetic effect for closely related populations, and heterogeneity between groups

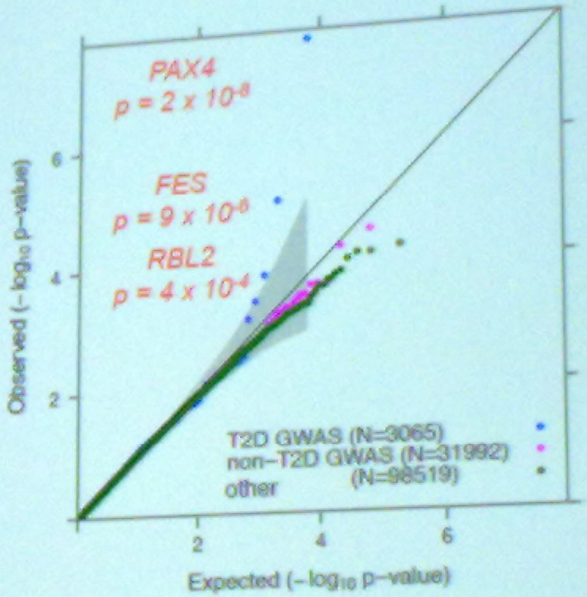


Are T2D GWAS regions enriched for rare, deleterious variants associated with T2D?

Gene-level mega-analysis
Deleterious variants, MAF < 1%.

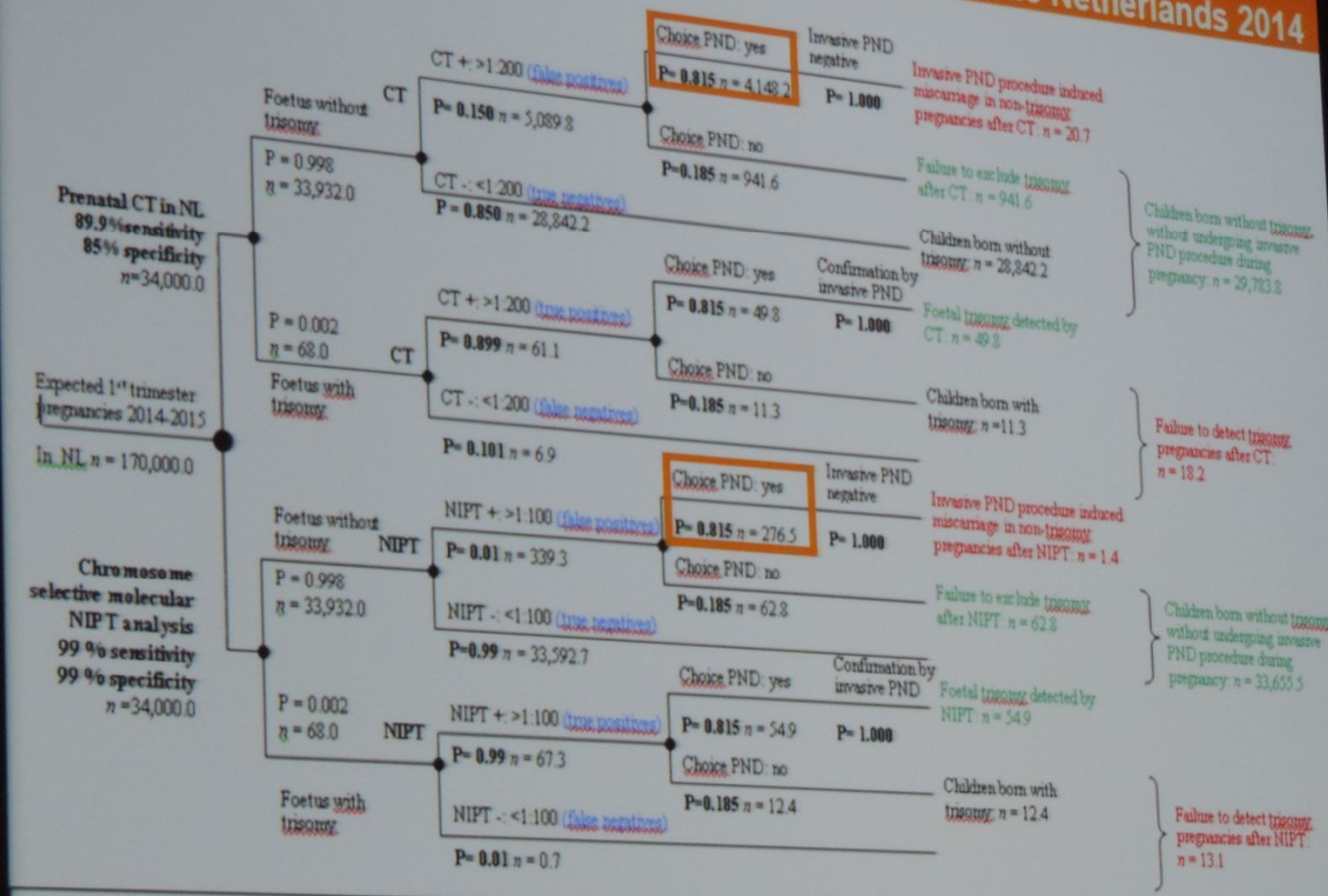


Single variant mega-analysis
Deleterious variants, MAF < 1%.



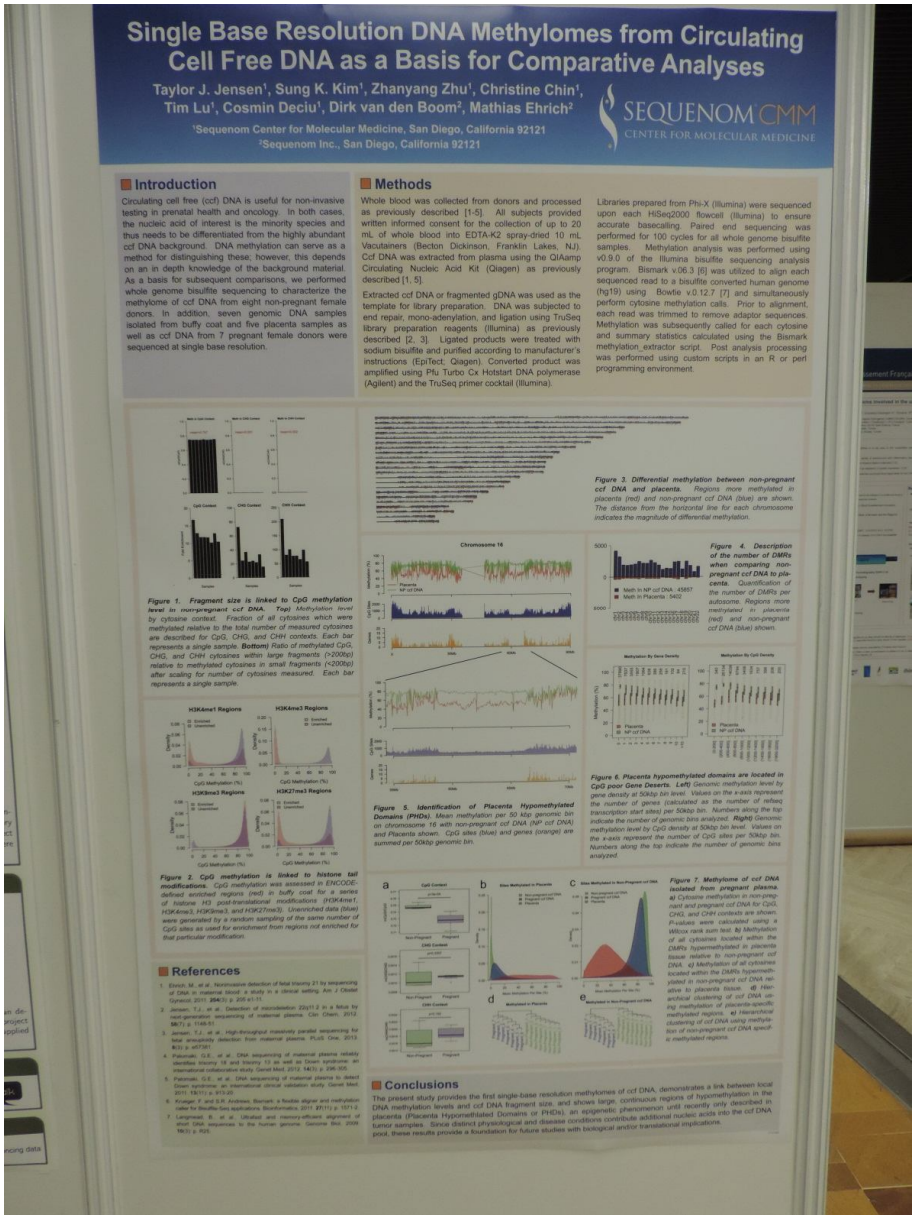
Focus on regions with prior knowledge of relationship with T2D (GWAS, MODY).

...cted outcome CT compared to NIPT in The Netherlands 2014



Legend to decision tree of expected first trimester pregnancies in 2014 [based on average year figures 2012 in routine prenatal care] in The Netherlands. CT: Combined Test (ultrasound nuchal translucency measurement combined with serum screening analysis which leads to an individual risk calculation result for trisomy pregnancy). NIPS: non-invasive prenatal screening. P: probability. n = absolute number of pregnancies. PND: prenatal diagnosis (CVS or AC; both invasive procedures).

Maastricht UMC+



Single Base Resolution DNA Methylomes from Circulating Cell Free DNA as a Basis for Comparative Analyses

Taylor J. Jensen[1], Sung K. Kim[1], Zhanyang Zhu[1], Christine Chin[1], Tim Lu[1], Cosmin Deciu[1], Dirk van den Boom[2], Mathias Ehrich[2]

[1]Sequenom Center for Molecular Medicine, San Diego, California 92121
[2]Sequenom Inc., San Diego, California 92121

SEQUENOM CMM CENTER FOR MOLECULAR MEDICINE

Introduction

Circulating cell free (ccf) DNA is useful for non-invasive testing in prenatal health and oncology. In both cases, the nucleic acid of interest is the minority species and thus needs to be differentiated from the highly abundant ccf DNA background. DNA methylation can serve as a method for distinguishing these; however, this depends on an in depth knowledge of the background material. As a basis for subsequent comparisons, we performed whole genome bisulfite sequencing to characterize the methylome of ccf DNA from eight non-pregnant female donors. In addition, seven genomic DNA samples isolated from buffy coat and five placenta samples as well as ccf DNA from 7 pregnant female donors were sequenced at single base resolution.

Methods

Whole blood was collected from donors and processed as previously described [1-5]. All subjects provided written informed consent for the collection of up to 20 mL of whole blood into EDTA-K2 spray-dried 10 mL Vacutainers (Becton Dickinson, Franklin Lakes, NJ). Ccf DNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) as previously described [1, 5].

Extracted ccf DNA or fragmented gDNA was used as the template for library preparation. DNA was subjected to end repair, mono-adenylation, and ligation using TruSeq library preparation reagents (Illumina) as previously described [2, 3]. Ligated products were treated with sodium bisulfite and purified according to manufacturer's instructions (EpiTect, Qiagen). Converted product was amplified using PfuTurbo Cx Hotstart DNA polymerase (Agilent) and the TruSeq primer cocktail (Illumina).

Libraries prepared from Phi-X (Illumina) were sequenced upon each HiSeq2000 flowcell (Illumina) to ensure accurate basecalling. Paired end sequencing was performed for 100 cycles for all whole genome bisulfite samples. Methylation analysis was performed using v0.9.0 of the Illumina bisulfite sequencing analysis program. Bismark v.06.3 [6] was utilized to align each sequenced read to a bisulfite converted human genome (hg19) using Bowtie v.0.12.7 [7] and simultaneously perform cytosine methylation calls. Prior to alignment, each read was trimmed to remove adaptor sequences. Methylation was subsequently called for each cytosine and summary statistics calculated using the Bismark methylation_extractor script. Post analysis processing was performed using custom scripts in an R or perl programming environment.

Figure 1. Fragment size is linked to CpG methylation level in nonpregnant ccf DNA. Top) Methylation level by cytosine context. Fraction of all cytosines which were methylated relative to the total number of measured cytosines are described for CpG, CHG, and CHH contexts. Each bar represents a single sample. Bottom) Ratio of methylated CpG, CHG, and CHH cytosines within large fragments (>200bp) relative to methylated cytosines in small fragments (<200bp) after scaling for number of cytosines measured. Each bar represents a single sample.

Figure 2. CpG methylation is linked to histone tail modifications. CpG methylation was assessed in ENCODE-defined enriched regions (red) in buffy coat for a series of histone H3 post-translational modifications (H3K4me1, H3K4me3, H3K9me3, and H3K27me3). Unenriched data (blue) were generated by a random sampling of the same number of CpG sites as used for enrichment from regions not enriched for that particular modification.

Figure 3. Differential methylation between non-pregnant ccf DNA and placenta. Regions more methylated in placenta (red) and non-pregnant ccf DNA (blue) are shown. The distance from the horizontal line for each chromosome indicates the magnitude of differential methylation.

Figure 4. Description of the number of DMRs when comparing non-pregnant ccf DNA to placenta. Quantification of the number of DMRs per autosome. Regions more methylated in placenta (red) and non-pregnant ccf DNA (blue) shown.

Figure 5. Identification of Placenta Hypomethylated Domains (PHDs). Mean methylation per 50 kbp genomic bin on chromosome 16 with non-pregnant ccf DNA (NP ccf DNA) and Placenta shown. CpG sites (blue) and genes (orange) are summed per 50kbp genomic bin.

Figure 6. Placenta hypomethylated domains are located in CpG poor Gene Deserts. Left) Genomic methylation level by gene density at 50kbp bin level. Values on the x-axis represent the number of genes (calculated as the number of refseq transcription start sites) per 50kbp bin. Numbers along the top indicate the number of genomic bins analyzed. Right) Genomic methylation level by CpG density at 50kbp bin level. Values on the x-axis represent the number of CpG sites per 50kbp bin. Numbers along the top indicate the number of genomic bins analyzed.

Figure 7. Methylome of ccf DNA isolated from pregnant plasma. a) Cytosine methylation in non-pregnant and pregnant ccf DNA for CpG, CHG, and CHH contexts are shown. P-values were calculated using a Wilcox rank sum test. b) Methylation of all cytosines located within the DMRs hypermethylated in placenta tissue relative to non-pregnant ccf DNA. c) Methylation of all cytosines located within the DMRs hypermethylated in non-pregnant ccf DNA relative to placenta tissue. d) Hierarchical clustering of ccf DNA using methylation of placenta-specific methylated regions. e) Hierarchical clustering of ccf DNA using methylation of non-pregnant ccf DNA specific methylated regions.

References

1. Ehrich, M., et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol, 2011. 204(3): p. 205 e1-11.
2. Jensen, T.J., et al., Detection of microdeletion 22q11.2 in a fetus by next-generation sequencing of maternal plasma. Clin Chem, 2013. 59(7): p. 1145-51.
3. Jensen, T.J., et al., High-throughput massively parallel sequencing for fetal aneuploidy detection from maternal plasma. PLoS One, 2013. 8(3): p. e57381.
4. Palomaki, G.E., et al., DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. Genet Med, 2012. 14(3): p. 296-305.
5. Palomaki, G.E., et al., DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med, 2011. 13(11): p. 913-20.
6. Krueger, F. and S.R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics, 2011. 27(11): p. 1571-2.
7. Langmead, B., et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 2009. 10(3): p. R25.

Conclusions

The present study provides the first single-base resolution methylomes of ccf DNA, demonstrates a link between local DNA methylation levels and ccf DNA fragment size, and shows large, continuous regions of hypomethylation in the placenta (Placenta Hypomethylated Domains or PHDs), an epigenetic phenomenon until recently only described in tumor samples. Since distinct physiological and disease conditions contribute additional nucleic acids into the ccf DNA pool, these results provide a foundation for future studies with biological and/or translational implications.

Phillipos Patsdis

Ссуществуют разные сайты метилирования у плода и матери
АТ к ДНК плода и сравнение мат и плодовой ДНК методом CGH
Есть около 30 000 DMRs
Выбрали 12
Др вариант – реалтайм метод (MeDip-qPCR)
7 маркеров Tsaliki, 2012
99,2% чувс-ть, 100%-спец-ть
300-400евро

Проект метилом map – 5 эмбрионов и 5 женщин

Экзом

Шизофрения – экзом 55 новых мутаций
RGS12 ген –вовлечен в функцию мозга
Слабоумие – 35 новых генов

Проект 6700 экзомов (популяция, опросник) цель- экзом чип

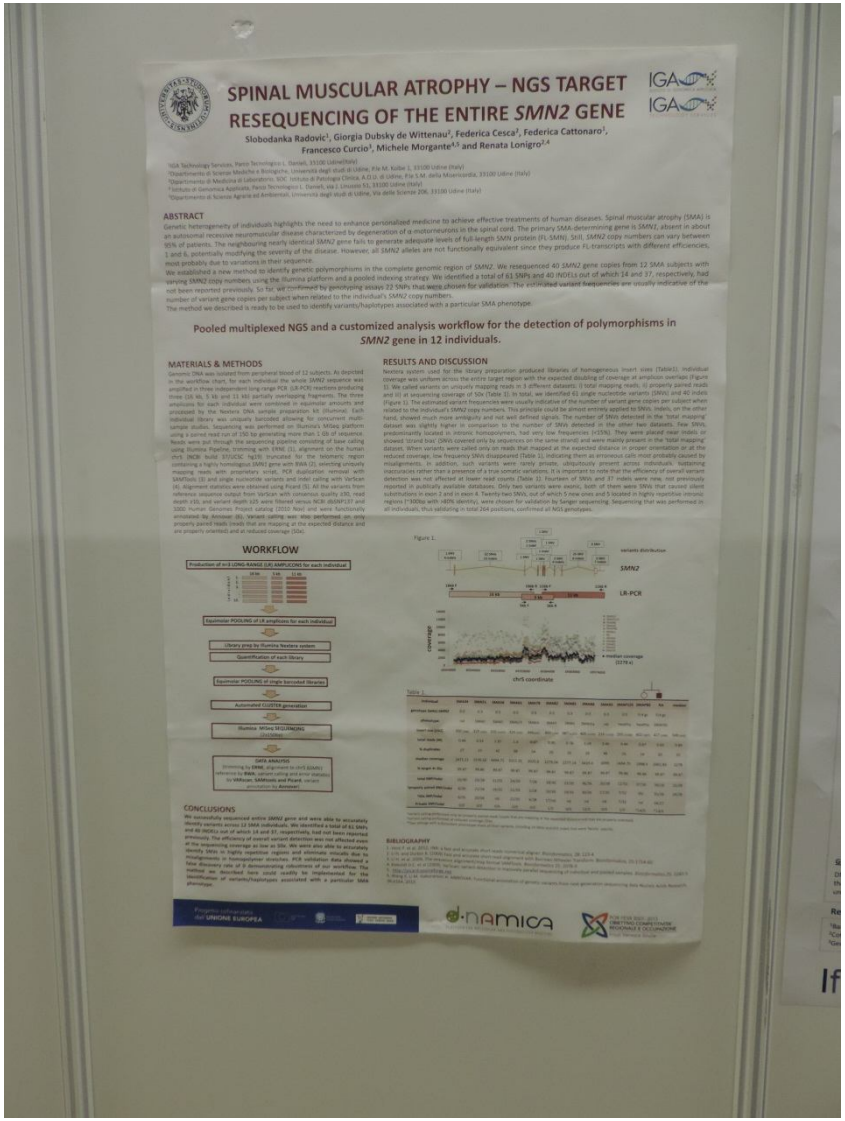
Цена -999евро без аннотации в Китае

1. Если частота мутации девиации 15% - игнорируется
2. Игнорируется ins/del концах, если осталось 2-3 буквы
3. Подход по фильтрации данных:

А) сначала все SNP
Б) потом минус негенные и интроны
В) исключит известные
Г) анализировать кандидаты de novo

Agilent – Haloplex – 98 генов нейромыш заболеваний (2447) регионов Все гены болезней, а не кандидаты
Ест и другие наборы

Ген и псевдоген различают по числу ридов



http://raindancetech.com/

RainDance technologies — Ян...

Raindance Technologies: T...

Search

NEWS | EVENTS | ABOUT US | SUPPORT | CONTACT | PURCHASE

RainDance Technologies

The Digital Biology™ Company

Home | Core Technology | Targeted DNA Sequencing | Digital PCR | Customers | Publications | Blog

RainDrop™

Digital PCR System

The Next Generation of Digital PCR is Here

LEARN MORE

Customer Interview

Dr. Olivier Harismendy from University of California, San Diego

Latest Webinar

Madhuri Hegde, Ph.D., FACMG, & Christin Collins, Ph.D., FACMG, from Emory Genetics Laboratory

Inside Our Technology

Darren Link, Co-Founder of RainDance Technologies

Customer Interview

Dr. Andrew G. Hadd, Ph.D., Asuragen, Inc.

Latest Tweets

Press Releases

RainDance Attracts Top-Tier

Latest Blog Posts

A New Home for

Media Coverage

Boston Business Journal:

http://raindancetech.com/digital-pcr-tech/raindrop-digital-pcr-system/

RU 18:49 25.06.2013

AmpliSeq™ RNA: Targeted sequencing of genes on the PGM ™

ion torrent by life technologies™

Kelli Bramlett, Angie Chang, Ron Abruzzese, Laura Chapman, Natalie Hernandez, Luming Qu, Jeoffrey Schageman, Dan Williams, Emily Zeringer, Brian Sanderson, Mike Kennemer, Richard Fekete, Susan Magdaleno and Robert Setterquist. Life Technologies, 2130 Woodward St., Austin, TX, USA, 78744

ABSTRACT

As next generation sequencing becomes readily available, more information becomes accessible for translational researchers. RNA profiling can be clinically useful in regards to diagnostics and treatment of patients. However, in many cases, whole transciptome sequencing is not required whereas targeting a subset of genes provides all the relevant information, while data analysis is significantly faster. This approach offers many advantages over qPCR and microarray technologies - such as single nucleotide information which can differentiate smaller changes in gene expression; lower input amounts can be used; clinically relevant samples such as formalin-fixed paraffin embedded (FFPE) tissues can be utilized.

Apoptosis is a fundamental process important in the development and homeostasis of living organisms. This process is crucial in cancer and other diseases. We have selected a comprehensive panel of genes involved in the cellular apoptosis pathway and designed specific RT-PCR amplicons to measure this pathway in a single multiplex Ion Torrent sequencing assay. We compared FFPE tumor and normal adjacent tissue (NAT) from a lung cancer patient. We also tested matched sections that were not fixed - for a comparative profile study. Using only 10ng of RNA, cDNA was prepared, followed by target amplification using primers designed to our select panel of genes. The resulting amplicons were used with the Ion AmpliSeq™ technology and sequenced on the Ion Torrent PGM™ with an Ion 318 chip.

We have demonstrated that targeted RNA sequencing methodology allows one to identify a unique apoptotic signature discriminating normal and tumor samples. This valuable information could provide insights into how tumors avoid cellular apoptotic death. This new multiplex RNA AmpliSeq™ technology has broad applications for accurate measurements of hundreds of specified transcripts, and expands the utility of Ion Torrent RNA applications.

MATERIALS AND METHODS

RNA was isolated from FFPE samples using Ambion's RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE and from unfixed samples using *mir*Vana ™ miRNA Isolation Kit. The RNA samples were then processed using the Ion AmpliSeq™ RNA Library Kit and Ion AmpliSeq™ RNA Apoptosis Panel. Libraries were quantified using Agilent's Bioanalyzer on a High Sensitivity DNA chip and prepared for sequencing on Ion Torrent's One Touch™ System with Ion One Touch™ 200 Template Kit v2. Samples were then sequenced on the Ion PGM™ using Ion PGM™ 200 Sequencing Kit. After sequencing, data was analyzed to give the number of counts per amplicon in the Apoptosis Panel.

Concurrently, RNA was made into cDNA using Invitrogen's SuperScript® VILO™ cDNA Synthesis Kit. TaqMan® Gene Expression Assays for the same genes in the Apoptosis Panel were used to interrogate the cDNA using Applied Biosystems' TaqMan® Gene Expression Master Mix and 7900 HT Fast Real Time PCR System.

Fig 2 – Seventy genes in the Apoptosis Panel were also tested using TaqMan assays on cDNA from the same RNA samples. TaqMan assays with CTs >35 and RNA AmpliSeq genes with <10 counts were not used. Fold change was calculated for both platforms and plotted to show correlation. Panel A shows the fold change correlation to unfixed lung samples while Panel B is FFPE. Of the 70 genes tested, 68 had detectable signal in both platforms for unfixed and 60 for fixed.

RESULTS

Figure 1. Workflow of the Ion AmpliSeq RNA Library Kit

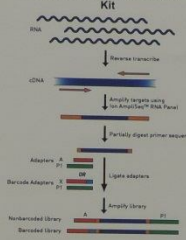


Fig. 1 – Ion AmpliSeq RNA Library Kit work flow. RNA is reverse transcribed into cDNA. The cDNA is then amplified with a pool of specific gene targets, in this case the Apoptosis Panel. Primers are partially digested and adapters ligated onto the amplicons. The library is amplified and purified and then ready for sequencing.

Table 1. Sequencing statistics for FFPE and unfixed RNA samples

	Total Reads	Percent On Target	Percent Mapped Reads	Amplicons Detected (of 267)
FFPE NAT Rep-1	5.18M	99.17%	97.35%	251 (94.0%)
FFPE NAT Rep-2	5.69M	99.53%	97.87%	257 (96.3%)
FFPE NAT Rep-3	4.10M	99.16%	96.25%	254 (95.1%)
FFPE Tumor Re-1	4.76M	99.64%	98.13%	254 (95.1%)
FFPE Tumor Rep-3	5.78M	99.34%	97.24%	252 (94.4%)
Unfixed NAT Rep-1	5.25M	92.72%	99.56%	261 (97.8%)
Unfixed NAT Rep-2	4.70M	99.88%	99.51%	262 (98.1%)
Unfixed NAT Rep-3	6.00M	99.55%	99.57%	260 (97.4%)
Unfixed Tumor Rep-1	5.68M	99.95%	99.53%	260 (97.4%)
Unfixed Tumor Rep-2	4.80M	99.57%	99.45%	262 (98.1%)
Unfixed Tumor Rep-3	5.39M	99.95%	99.63%	261 (97.8%)

Table 1 – The table above reports sequencing statistics for each sample from PGM runs. The total number of reads column is expressed in "millions". Percentage of on target and mapped reads was calculated based on total reads. Average read length is taken from the post filtered reads and Amplicons Detected is the number of amplicons detected of the 267 interrogated from the Apoptosis Panel.

Figure 2. TaqMan and RNA AmpliSeq Correlation

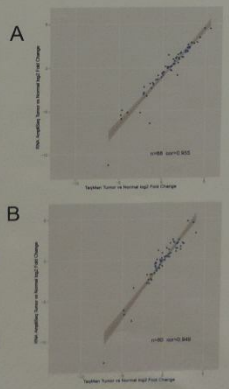


Figure 3. Genes highly differentiated between tumors and NAT.

Fig 2 – The plot above shows genes that are greater than 2 log(2) fold differentially expressed in both fixed and unfixed samples when comparing tumor and NAT tissues.

CONCLUSIONS

The Ion AmpliSeq RNA workflow allows users to generate gene expression data for hundreds of genes in a single tube

- Barcoding allows for multiplexing to get more information from a single sequencing reaction
- High quality sequencing, even for FFPE samples
- Ion AmpliSeq™ RNA results are highly correlated to results using individual TaqMan assays for both unfixed and FFPE RNA samples

Of the genes in the Apoptosis panel, 37 were highly differentiated between tumor and NAT samples

- Several of the induced genes are growth factors, such as FGFR1 & 2, HFG, NGFR, PDGFB and TGFBR3
- Another group of genes induced are involved in cell growth and proliferation, like AKT3, FGF2, NRG1
- A few genes, NTRK3, TP53AIP1 and PRKCE, have previously been associated with cancers
- The most highly induced genes are NTRK3, IL2, BIRC8 and BMP7. Interestingly, BICR8 protects cells against apoptosis.
- Of the down regulated genes, many of these are apoptosis inducing genes: BCL2L11, BIK, CASP14, PERP, PMAIP1 and SLC25A5
- Others are involved in cell cycle checkpoints, like CCNA2, CCNB1 & E1, CHEK 1&2 and TGFA
- The most highly down regulated are BIRC5, CASP14, the Cyclins (CCNA2, B1, E1 and CDKN2A) and TERT.

While many of these genes have been associated with cancer, they have not all been shown in lung cancer. This data shows a gene signature specific to lung cancer that is consistent for both unfixed and FFPE RNA.

REFERENCES

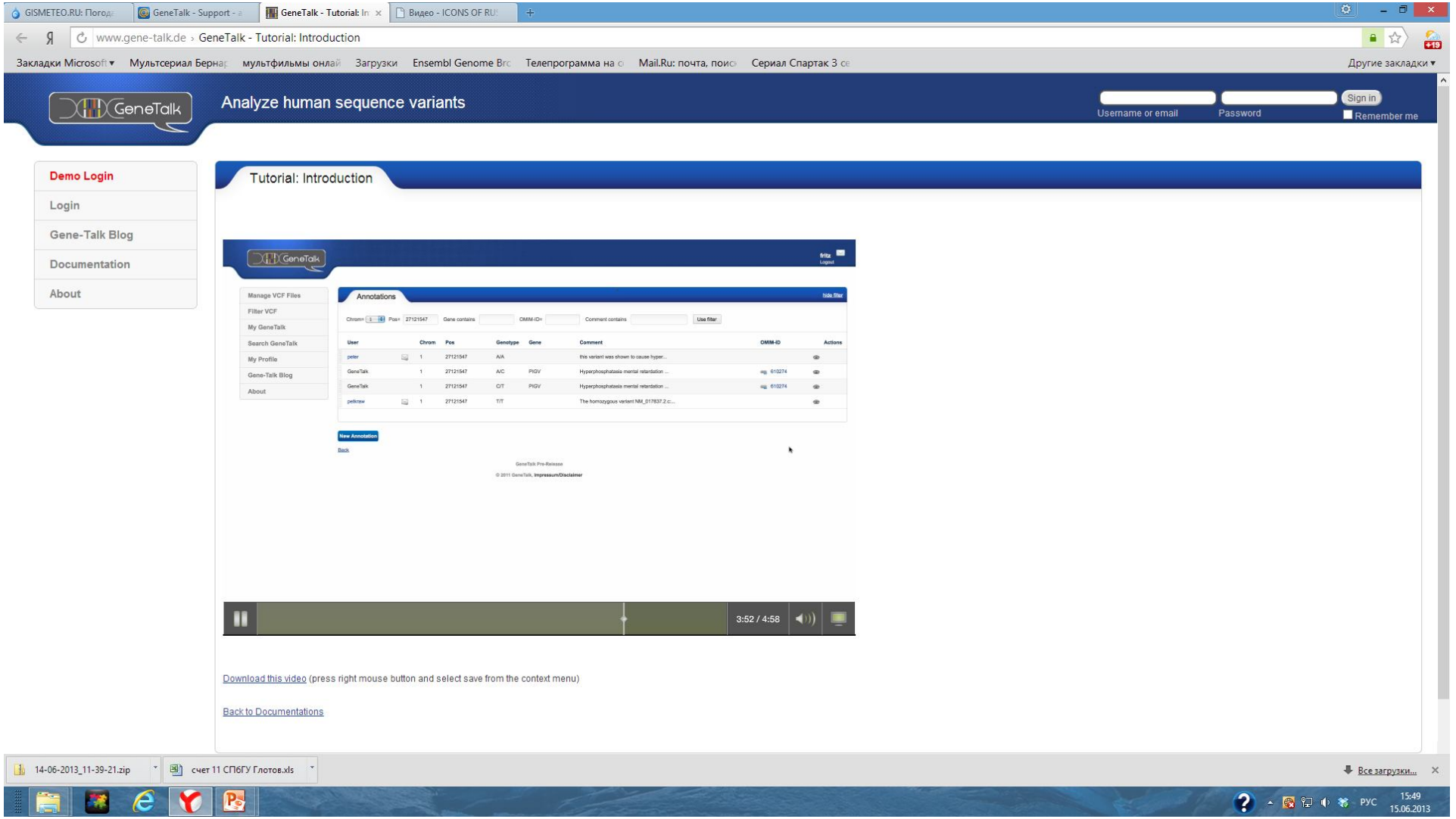
1. HUGO Gene Nomenclature Committee, Gene Symbol Report, www.genenames.org
2. NCBI, Genes & Expression Gene Database, www.ncbi.nlm.nih.gov
3. NTRK3 as a novel tumor suppressor gene in colorectal cancer. Luo, Yanxin; Kaz, Andrew; Kanngurn, Samonmas; Grady, William M. Fred Hutchinson Cancer Research Center, Seattle, WA. J Clin Oncol 30, 2012 (suppl 4; abstr 469)

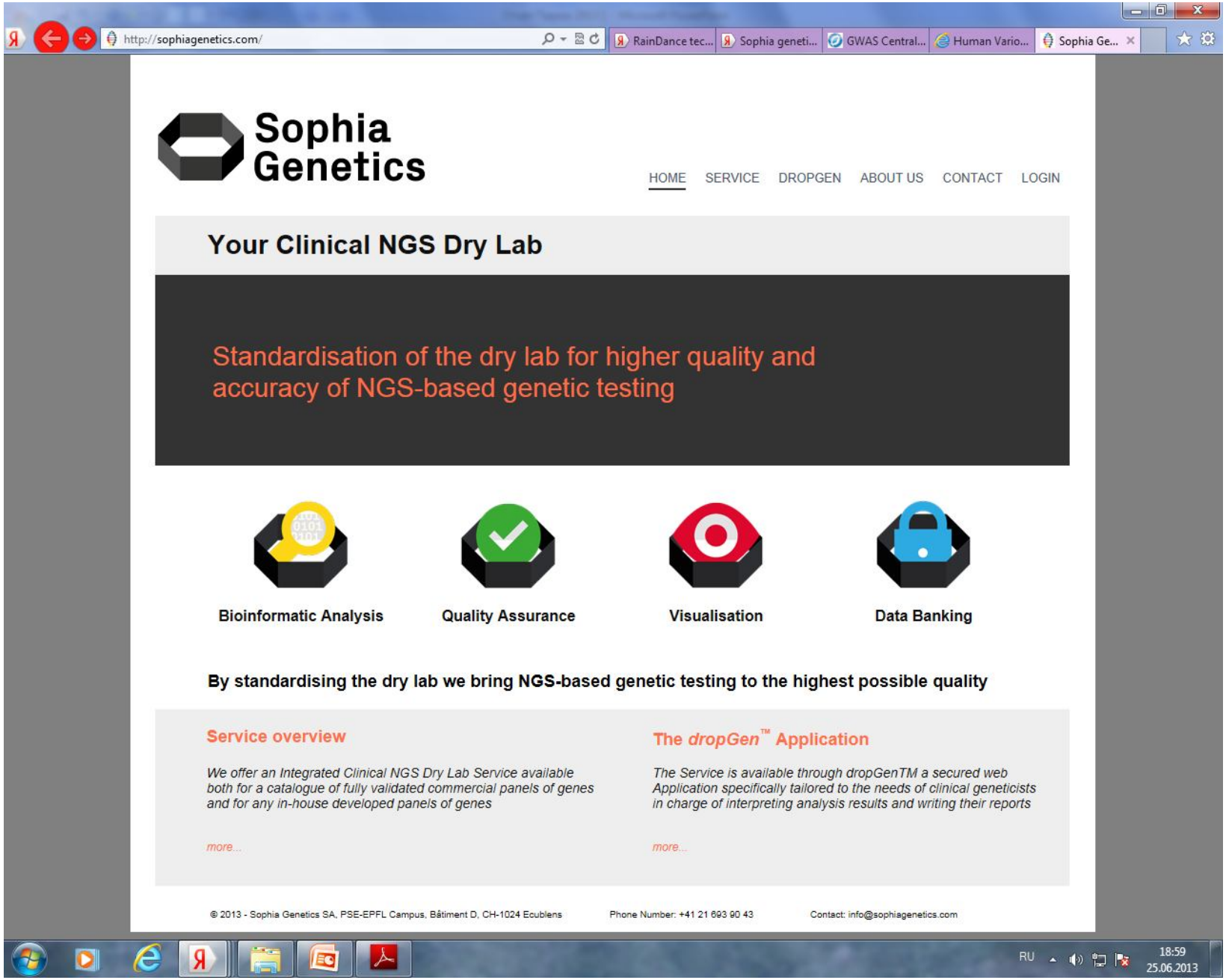
For Research Use Only. Not For Use in Diagnostic Procedures.

Программы;
PolyPhen 2
Condel
SIFT
BioSkope
etasy

https://www.gene-talk.de/

Analyze human sequence variants




http://sophiagenetics.com/
RainDance tec...
Sophia geneti...
GWAS Central...
Human Vario...
Sophia Ge...
Sophia Genetics
HOME
SERVICE
DROPGEN
ABOUT US
CONTACT
LOGIN
Your Clinical NGS Dry Lab
Standardisation of the dry lab for higher quality and accuracy of NGS-based genetic testing
Bioinformatic Analysis
Quality Assurance
Visualisation
Data Banking
By standardising the dry lab we bring NGS-based genetic testing to the highest possible quality
Service overview
We offer an Integrated Clinical NGS Dry Lab Service available both for a catalogue of fully validated commercial panels of genes and for any in-house developed panels of genes
more...
The dropGen™ Application
The Service is available through dropGenTM a secured web Application specifically tailored to the needs of clinical geneticists in charge of interpreting analysis results and writing their reports
more...
© 2013 - Sophia Genetics SA, PSE-EPFL Campus, Bâtiment D, CH-1024 Ecublens
Phone Number: +41 21 693 90 43
Contact: info@sophiagenetics.com
RU
18:59
25.06.2013

P15.32

HelmholtzZentrum münchen
German Research Center for Environmental Health

Exome Database

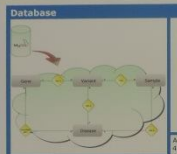
Thomas Schwarzmayr[1], Thomas Wieland[1], Anett Schmittfull[1], Elisabeth Graf[1], Thomas Meitinger[1,2], Tim M. Strom[1,2]

[1] Helmholtz Zentrum München, Institute of Human Genetics, Neuherberg, Germany
[2] Technische Universität München, Institute of Human Genetics, Munich, Germany

SNVs and indels derived from exome sequencing are usually filtered by minor allele frequency and mode of inheritance in order to identify candidate variants for Mendelian disease. We developed a database that stores both, pedigree information and variant data generated by an analysis pipeline. A corresponding web interface allows to filter the data in an interactive way. Standard queries include searches for rare heterozygous, compound heterozygous, homozygous, de novo and somatic variants. Variants are connected with a number of annotations such as position in transcripts, amino acid change, functional predictions, and presence in dbSNP, HGMD and the 1000 Genomes data. Furthermore, quality scores, mapping scores, read depth and links to external databases are provided. In addition to publically available frequency data, variants identified in in-house exomes of individuals with unrelated phenotypes are used as controls. In addition to frequency bases searches, one can also perform gene and disease-based searches, thus displaying the variation content of genes or focusing the search on known disease genes or HGMD mutations. The database also contains data about the average read depth per exon and the proportion by which each exon is covered at least 20-times. These data can be used in a diagnostic context to define the regions of a gene which have been appropriately analyzed and to identify genes which possibly carry deletions of entire exons. Lastly, the database is connected with a LIMS and provides basic quality statistics, such as amount of sequence, read depth, sample contamination and discordance between specified and identified sex.

Feature Overview

Database



The figure shows a snippet of the database schema. All variants called by the analysis pipeline are stored in an in-house MySQL database.

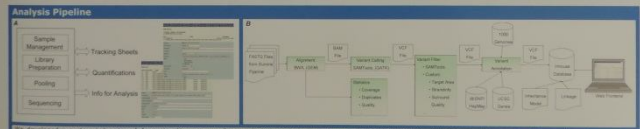
5,014 alleles (2,507 individuals)

	Singletons	MAF < 1%	MAF < 5%	All variants
Non-synonymous variants	228,287	379,616	391,588	411,199
Synonymous variants	113,279	205,318	215,228	235,805
All variants	339,279	580,756	602,458	642,303

~1 variant per 60 bp coding sequence

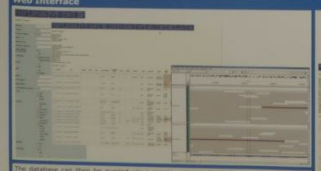
Altogether we called over 642.000 variants within coding regions of which 411.000 were non-synonymous and 236.000 were synonymous. The vast majority of these variants have a minor allele frequency of below 1%. More than half of the variants even occur in only a single sample. On average, we discovered about 1 variant per 60 bp of coding sequence. For comparison: The Human Gene Mutation Database (HGMD) includes roughly 137.000 variants, which gives about 1 variant per 277 bp.

Analysis Pipeline

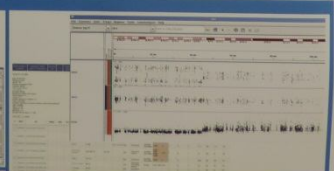


We developed a custom Laboratory Information Management System (LIMS) that provides modules to document sample management, library preparation, pooling and sequencing (Figure A). It is capable of importing data from several different sources, for instance sample information from external collaborators, quality/quantity metrics from different lab devices, pool compositions from 2D barcoded tubes, etc.
The LIMS database also provides meta information for our analysis pipeline (Figure B). The pipeline is a collection of Perl scripts which facilitate analysis, starting with FASTQ files generated by the Illumina software. It performs read alignment to the reference sequence using BWA, variant calling using SAMtools and functional annotation and filtering of variants using custom developed Perl scripts. Additionally, detailed quality metrics to assess performance of the sequencing and capture experiments are calculated.

Web Interface



The database can then be queried via a simple web interface which provides forms for pedigree or gene based searches. The Integrated Genomics Viewer (IGV) is used to summary statistics and quality metrics.



Furthermore, it is possible to detect putative copy-number variations (CNVs) in tumors. Thanks to links on the corresponding mask one can simply open the desired information in the IGV and investigate the results for specific samples.

Institute of Human Genetics

HELMHOLTZ ASSOCIATION

P15.58

VariantDB : A flexible annotation and filtering portal for NGS data

Geert Vandeweyer[1,2,3], Lut Van Laer[1,2], Tim Van Den Bulcke[3], Frank Kooy[1,2],
[1] University of Antwerp and [2] University Hospital of Antwerp, Belgium
[3] Biomedical Informatics Research Center Antwerp, University and University Hospital of Antwerp, Belgium

Introduction

Causes of many hereditary disorders can be identified through detailed inspection of the patients genome using Next Generation Sequencing technologies. However, interpretation of the vast amount of data, typically over 20.000 variants per patient, became a bottleneck for routine implementation. To help streamlining NGS data analysis through various trimming, mapping, annotation and filtering tools, centralised web-based interfaces, such as the Galaxy platform, have been developed. Although Galaxy provides many bioinformatic tools, it lacks flexibility in downstream annotation and filtering of the resulting variants.

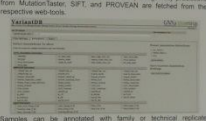
Methods

The platform is implemented in MySQL, with a PHP/CGI/ajax based frontend, compatible with all major browsers. Data can be uploaded using a direct link from Galaxy, using a custom tool, or by FTP upload. On-site storage of BAM and VCF files is supported for direct visualisation in IGV. All submitted data is private at user level, and can be shared between collaborators. Annotations are added using ANNOVAR, snpEff, GATK and custom annotation scripts using Perl wrappers. Additional annotation sources and filtering options can be specified in global XML-based configuration files. Local installation and automatic updating is supported through Mercurial.

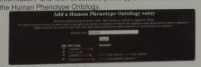
Platform Structure

Annotation

Variants are automatically annotated using ANNOVAR and snpEff. Furthermore, QC-values from the GATK Unified genotyper are also stored upon VCF submission. Available annotations include occurrence information (dbSNP, ESP, 1KG, private samples), effect on transcripts (UCSC, Ensembl and RefSeq based) and predicted impact on gene function (snpEff, PolyPhen, Grantham, GERP ...). In addition, Gene Ontology associations and pathogenicity predictions from MutationTaster, SIFT, and PROVEAN are fetched from the respective web-tools.



Samples can be annotated with family or technical replicate information, and with structured phenotypic information taken from the Human Phenotype Ontology.

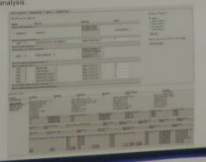


Galaxy Integration

A tool is available for direct communication between a Galaxy server and VariantDB, allowing automatic transmission of VCF files to VariantDB upon workflow completion. It can be installed on local deployments without further dependencies and supports multiple VariantDB instances.

Filtering & Presentation

Filtering is performed using drop-down menus on all avaiable annotations in realtime. Resulting variants and selected annotations are presented in tabular overview with links for visualisation in IGV. Results can also be exported as CSV files for further off-line analysis



Results

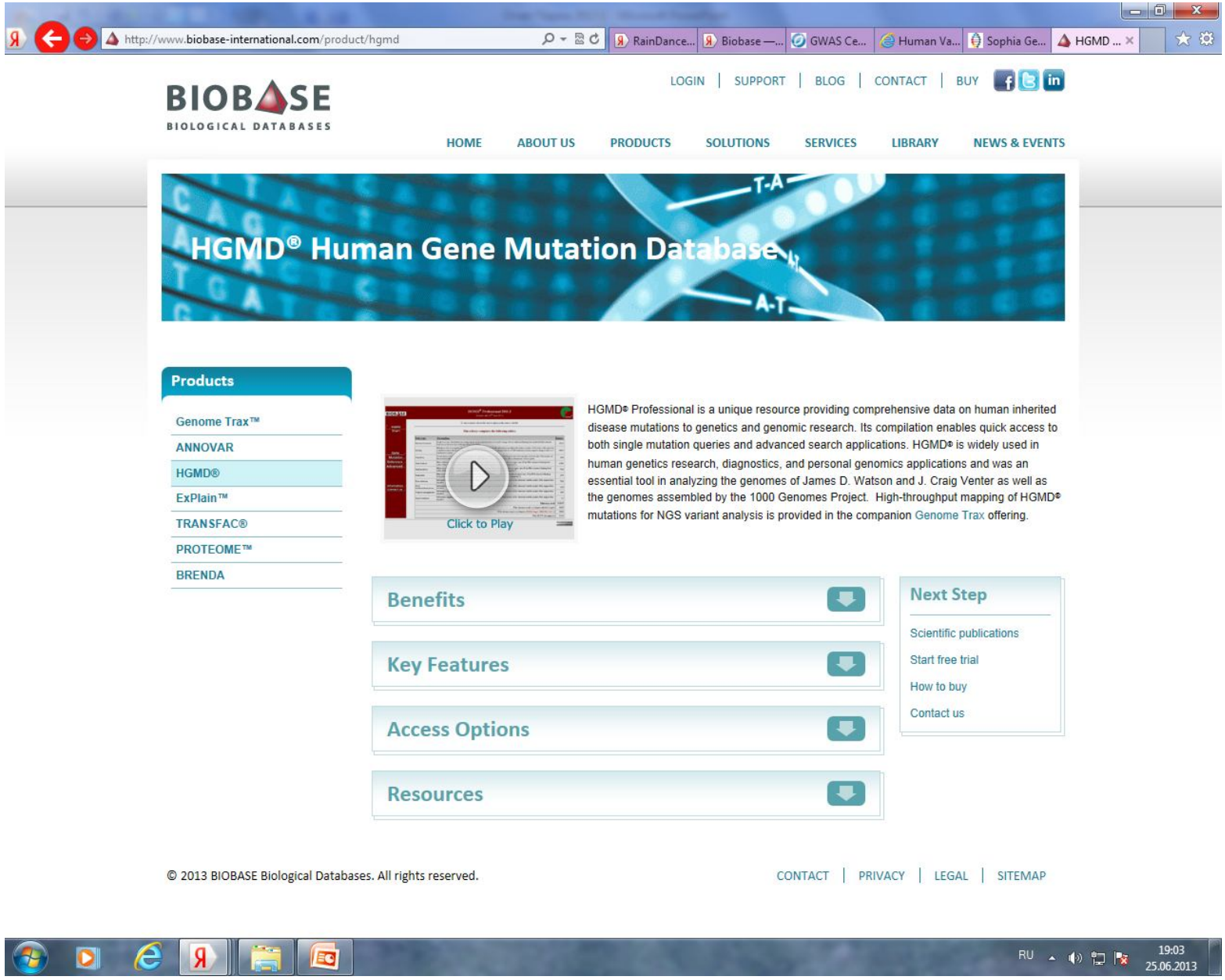
To complement the limited filtering options available in Galaxy, we built a web-based platform, that is able to directly interact with available primary analysis pipelines, and provides dynamic downstream annotation and filtering options for the end-user.

Filtering options comprise both nucleotide and amino acid level annotations, and the option to filter for inherited or *de novo* variants. Variants can also be filtered on genotype conformation, including compound heterozygosity or multiallelic variants.

Availability

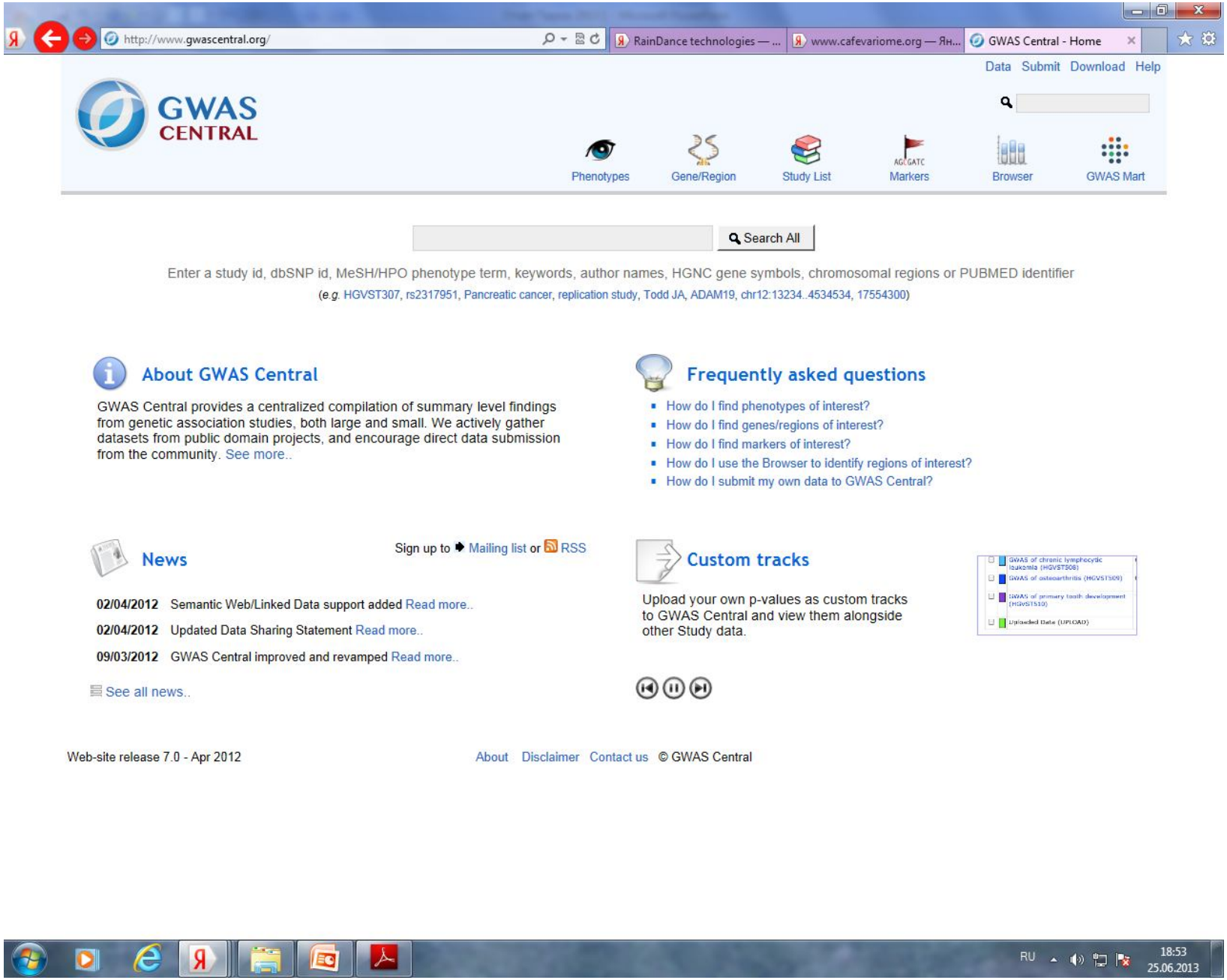
http://www.biomina.be/app/variantdb

UZA Cognitive Genetics

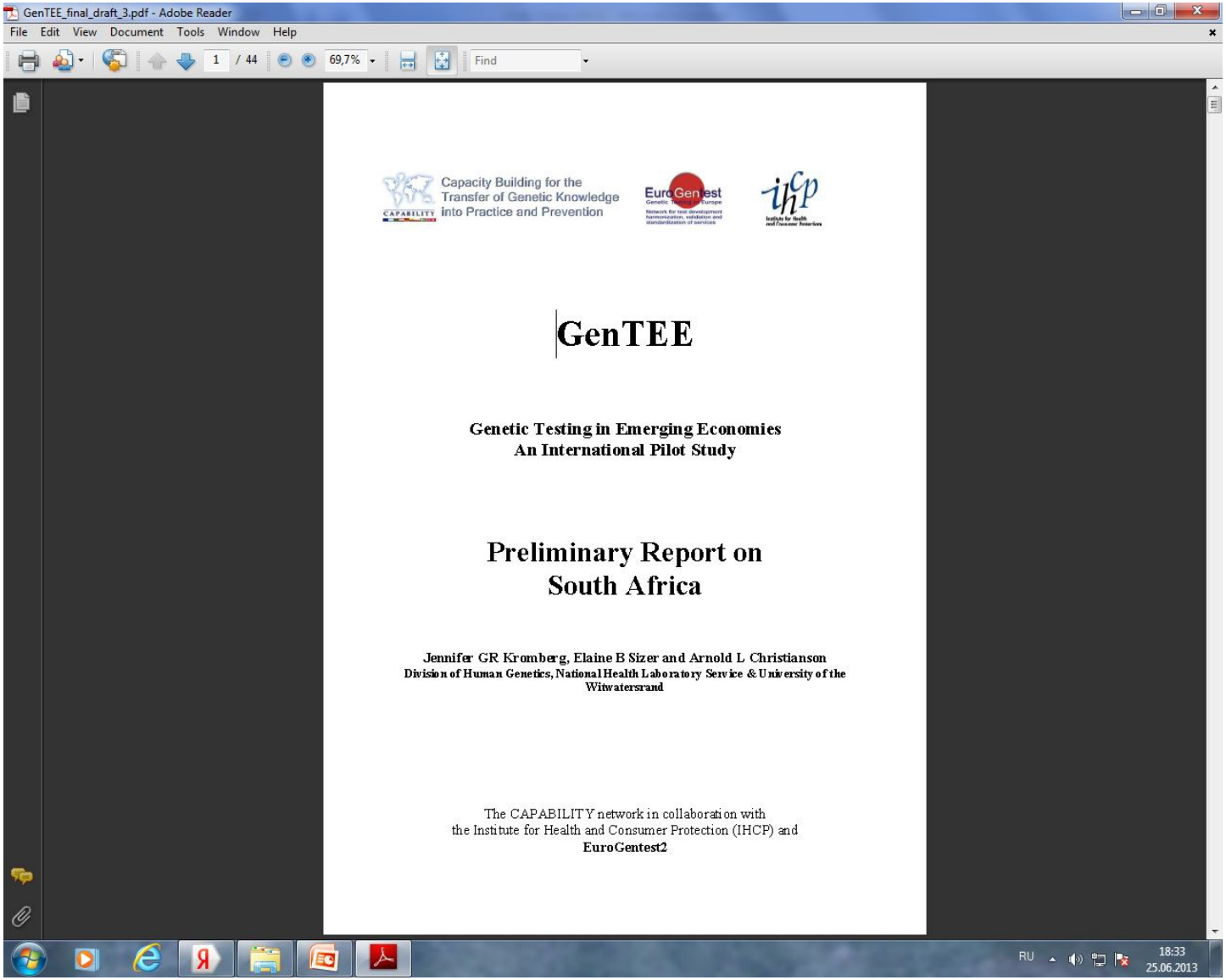

http://www.biobase-international.com/product/hgmd
RainDance...
Biobase —...
GWAS Ce...
Human Va...
Sophia Ge...
HGMD ...
BIOBASE
BIOLOGICAL DATABASES
LOGIN | SUPPORT | BLOG | CONTACT | BUY
HOME
ABOUT US
PRODUCTS
SOLUTIONS
SERVICES
LIBRARY
NEWS & EVENTS
HGMD® Human Gene Mutation Database
T-A
A-T
Products
Genome Trax™
ANNOVAR
HGMD®
ExPlain™
TRANSFAC®
PROTEOME™
BRENDA
Click to Play
HGMD® Professional is a unique resource providing comprehensive data on human inherited disease mutations to genetics and genomic research. Its compilation enables quick access to both single mutation queries and advanced search applications. HGMD® is widely used in human genetics research, diagnostics, and personal genomics applications and was an essential tool in analyzing the genomes of James D. Watson and J. Craig Venter as well as the genomes assembled by the 1000 Genomes Project. High-throughput mapping of HGMD® mutations for NGS variant analysis is provided in the companion Genome Trax offering.
Benefits
Key Features
Access Options
Resources
Next Step
Scientific publications
Start free trial
How to buy
Contact us
© 2013 BIOBASE Biological Databases. All rights reserved.
CONTACT | PRIVACY | LEGAL | SITEMAP
RU
19:03
25.06.2013

Этногенетика

1. Существуют гены, способствующие повышенной репродукции
2. Самое большое разнообразие внутри популяции (80%) – между континентами (85)-внутри континентов (4%) по SNP
3. Irish Plex – 6 генов на цвет глаз
4. Цвет волос -13 генов за 46 цветов (Польская популяция)
5. Все -180 генов
6. Морфология волос – TCHH (прямые Европа), EDAR (тонкие Азия)


http://www.gwascentral.org/
RainDance technologies — ...
www.cafevariome.org — Ян...
GWAS Central - Home
Data Submit Download Help
GWAS CENTRAL
Phenotypes
Gene/Region
Study List
Markers
Browser
GWAS Mart
Search All
Enter a study id, dbSNP id, MeSH/HPO phenotype term, keywords, author names, HGNC gene symbols, chromosomal regions or PUBMED identifier
(e.g. HGVST307, rs2317951, Pancreatic cancer, replication study, Todd JA, ADAM19, chr12:13234..4534534, 17554300)
About GWAS Central
GWAS Central provides a centralized compilation of summary level findings from genetic association studies, both large and small. We actively gather datasets from public domain projects, and encourage direct data submission from the community. See more..
Frequently asked questions
How do I find phenotypes of interest?
How do I find genes/regions of interest?
How do I find markers of interest?
How do I use the Browser to identify regions of interest?
How do I submit my own data to GWAS Central?
News
Sign up to Mailing list or RSS
02/04/2012 Semantic Web/Linked Data support added Read more..
02/04/2012 Updated Data Sharing Statement Read more..
09/03/2012 GWAS Central improved and revamped Read more..
See all news..
Custom tracks
Upload your own p-values as custom tracks to GWAS Central and view them alongside other Study data.
GWAS of chronic lymphocytic leukemia (HGVST308)
GWAS of osteoarthritis (HGVST309)
GWAS of primary tooth development (HGVST310)
Uploaded Data (UPLOAD)
Web-site release 7.0 - Apr 2012
About Disclaimer Contact us © GWAS Central
RU
18:53
25.06.2013


http://www.orphadata.org/cgi-bin/index.php/
RainDance...
Orpha...
GWAS Ce...
Human Va...
Sophia Ge...
HGMD - H...
homepage
downloads
about
legal issues
contact
my orphadata
Free access data from Orphanet
orphanet
orphadata
Tuesday, 25 June, 2013
Welcome To Orphadata
The mission of Orphadata is to provide the scientific community with a comprehensive, high-quality and freely-accessible dataset related to rare diseases and orphan drugs, in a reusable format.
See "About Orphadata" for more information
Home
About Orphadata
About Orphanet
Access Orphanet[→]
Contact
Freely accessible datasets
Diseases, cross referenced with other nomenclatures
Epidemiological data
Orphanet classifications
Diseases with their clinical signs
Thesaurus of clinical signs, cross referenced with other terminologies
Diseases with their associated genes
Documentation
User's guide
Legal issues
EN
19:05
25.06.2013


GenTEE_final_draft_3.pdf - Adobe Reader
File Edit View Document Tools Window Help
1 / 44
69,7%
Find
Capacity Building for the Transfer of Genetic Knowledge into Practice and Prevention
CAPABILITY
EuroGentest
ihcp
GenTEE
Genetic Testing in Emerging Economies
An International Pilot Study
Preliminary Report on South Africa
Jennifer GR Kromberg, Elaine B Sizer and Arnold L Christianson
Division of Human Genetics, National Health Laboratory Service & University of the Witwatersrand
The CAPABILITY network in collaboration with the Institute for Health and Consumer Protection (IHCP) and EuroGentest2
RU
18:33
25.06.2013

http://www.humanvariomeproject.org/

RainDance technol... | human variome pr... | GWAS Central - Ho... | Human Variom...

Sitemap Contact Us Search Login

THE HUMAN VARIOME PROJECT

sharing data · reducing disease

Home About Activities Publications Recommendations Meetings News Links Members

Latest News

Ensuring the Free and Open Sharing of Clinically Relevant Genome Variants - A Statement by the Human Variome Project Consortium

Genetic tests for diagnostic, predictive and screening purposes are a routine part of clinical care in most modern health care systems, and as we discover more about the genetic determinants of our health, we can expect genetic testing to become more prevalent. At the same time, new technological advancements, lower costs and increased training and education will see genetic testing spread rapidly into routine clinical practice in countries with, to date, less advanced health care systems.

Read more...

A Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data

Over 70 leading health care, research, and disease advocacy organizations, including the Human Variome Project announced today that they have taken the first steps to form an international alliance dedicated to enabling secure sharing of genomic and clinical data to improve research into the genetic basis of human disease.

Read more...

Join the Consortium

Individuals wishing to apply to join the Human Variome Project consortium can register here.

Sign up to the *HVP Transcript*

Sign up to the HVP Transcript to stay up to date with the progress and achievements of the Human Variome Project.

Upcoming Events

International Society for Gastrointestinal Hereditary Tumours - 5th Biennial Meeting

28-31 August, 2013
(HVP Workshop 28 August)
Cairns Convention Centre
Cairns
Australia

HVP5: The 5th Biennial Meeting of the Human Variome Project Consortium

19-23 May, 2014
UNESCO Headquarters
Paris, France

YouTube facebook twitter Google+ flickr slideshare

RU 18:55 25.06.2013

Development of a comprehensive gene screen for dilated cardiomyopathy using next-generation sequencing

S Reid[1], K Thomson[1], J Woodley[1], J Hayesmoore, M Shanks[1], J Taylor[1], E Blair[2], H Watkins[3], A Seller[1]

[1] Oxford Molecular Genetics Laboratory, [2] Department. of Clinical Genetics, Churchill Hospital, Oxford, [3] Oxford University Department of Cardiovascular Medicine, Oxford, United Kingdom.

Background:

- Dilated cardiomyopathy (DCM) is a heart muscle disease with an estimated prevalence of 1 in 2500. It is familial in 30-40% of cases.
- Familial DCM exhibits considerable clinical heterogeneity both within and between families: penetrence is incomplete and age dependant.
- The genetic basis of DCM is complex with over 30 genes reported to be involved, few are associated with any differentiating phenotype and the majority of genes account for less than 1% of cases [1] .
- This clinical and genetic heterogeneity means that the traditional analysis strategies, undertaking testing on a gene by gene basis, have had limited utility in a clinical setting.
- Recent advances in next generation sequencing technology means that it is now feasible to develop a comprehensive and sensitive genetic test incorporating analysis of the most common DCM genes in a single test.

Gene	Protein	Protein Location	Associated Disease	Size (KB)	Average Coverage x30 (%)
ACTC1	Actin	Sarcomere	HCM/DCM	1.25	96.1
ACTN2	Alpha-actin-2	Z disc	HCM/DCM	3.1	99.4
ANKRD1	Ankyrin repeat domian containing-1	Sarcoplasmic Reticulum	HCM/DCM	1.1	99.9
CRYAB	Alpha-crystallin B chain	Cytoplasm	DCM	0.6	99.8
CSRP3	Cysteine and glycine-rich protein 3	Z disc	HCM/DCM	0.7	100
DES	Desmin	Desmosome	ARVC/DCM	1.6	95.6
DSC2	Desmocollin-2	Desmosome	ARVC/DCM	3.1	98.9
DSG2	Desmoglein- 2	Desmosome	ARVC/DCM	3.6	98
DSP	Desmoplakin	Desmosome	ARVC/DCM	9.1	99
FHL1	Four and half LIM domains-1	Z disc	HCM	1.3	99.4
FHL2	Four and half LIM domains-2	Z disc	DCM	0.9	100
GLA	α-Galactosidase	Lysosome	HCM	1.4	98.6
JUP	Junction plakoglobin	Desmosome	ARVC	2.5	85.8
LAMP2	Lysosome associated membrane protein 2	Lysosome membrane	HCM/DCM	1.7	98.3
LMNA	Lamin A/C	Nuclear Membrane	DCM	2.4	90.6
MYBPC3	Myosin-binding protein C	Sarcomere	HCM/DCM	4.9	98.2
MYH7	Myosin-7	Sarcomere	HCM/DCM	6.5	94.9
MYL2	Essential myosin light chain	Sarcomere	HCM	0.6	88.8
MYL3	Regulatory myosin light chain	Sarcomere	HCM	0.7	99.9
PKP2	Plakophilin-2	Desmosome	ARVC/DCM	2.9	99.7
PLN	Phospolamban	Sarcoplasmic Reticulum	HCM/DCM	1.8	99.2
PRKAG2	5'-AMP-activated protein kinase subunit gamma-2	Cytoplasm	HCM	2	94
SCN5A	Sodium channel protein type 5 subunit α	Calcium channel	DCM	6.7	97.5
TMEM43	Transmembrane protein 43	Nuclear Membrane	ARVC	1.4	100
TNNI3	Troponin I	Sarcomere	HCM/DCM	0.8	98.8
TNNT2	Troponin T	Sarcomere	HCM/DCM	1.3	98.4
TPM1	Tropomyosin 1 alpha	Sarcomere	HCM/DCM	1.6	98.3
TTN	Titan	Sarcomere	DCM	116.4	99

Table 1 Genes sequenced in the Oxford Molecular Genetics DCM panel

Method:

- Haloplex™ enrichment technology (Agilent) was used to target 28 DCM genes (~200kb) in 17 individuals with familial DCM (Table 1).
- Analysis of 13 of these genes (highlighted in grey in Table 1 above) had been previously undertaken in our diagnostic laboratory; no pathogenic variants had been detected.
- Sequencing was carried out on the MiSeq bench top sequencer by the Oxford Genomic Centre (OGC) at the Welcome Trust Centre for Human Genetics.
- Data was demultiplexed by OGC, alignment and variant calling was performed by NextGene® Software (SoftGenetics).
- Variants were annotated and filtered against EVS frequencies and our in house variant database.
- Analysis was undertaken to determine the coverage at a depth of x30 per nucleotide within the regions of interest.
- Sanger sequence was carried out to confirm variants.

References

[1] Hershberger and Siegfried, J Am Coll Cardiol 2011; 57:1641-9; Update 2011: clinical and genetic issues in familial dilated cardiomyopathy.

[2] Mattocks et al,Eur J Hum Genet. 2010 Dec;18(12):1276-88; A standardized framework for the validation and verification of clinical molecular genetic tests.

[3] CMGS Best practise guidelines for targeted next generation sequencing analysis and interpretation 2012.

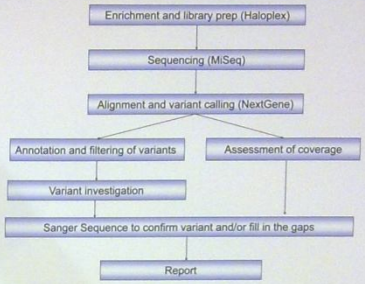


Figure 1 Next Generation Sequencing Workflow for DCM testing

Results:

- On average, 93.3% of quality filtered reads mapped to the Human Genome (GRCh37/hg19), of which 85.3% mapped to targeted regions.
- Coverage of genes was consistent between samples, the majority of genes were covered >95% (Table1).
- 97.4% of the targeted region was consistently covered to x30 depth.
- ~97 variants were identified per sample. After filtering this reduced to 2-5 variants per sample.

In the previously analysed genes (grey background Table 1),

- 151/151 known variants were detected, where read depth was greater than x10. Of these 62 were unique variants (Analytical sensitivity ≥95% (95% confidence).[2],[3]

In the previously unanalysed genes (white background, Table 1),

- 8/17 (47%) samples had a variant considered highly likely or likely to be pathogenic (see Figure 2).
- 5/17 (29%) had a variant of uncertain pathogenicity (UV) (see Figure 2). 2 of these 5 also had a variant considered likely to be pathogenic.

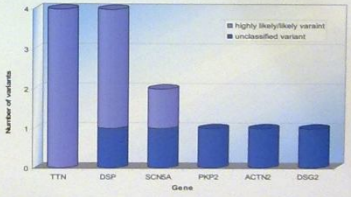


Figure 2 Number of highly likely/likely variants and variants of uncertain pathogenicity identified.

Discussion:

- Rapid advances in NGS technology have enabled the development of genetic tests for heterogeneous diseases such as DCM.
- Haloplex provides a rapid and robust method of enrichment producing good coverage across target regions.
- The results in our initial validation cohort indicate that this method has high analytical sensitivity and specificity.
- In this selected cohort of DCM patients, a putative pathogenic was detected in 47% of samples.
- Variants of uncertain pathogenicity were detected; further investigations will be required to clarify the clinical significance of these variants.

Oxford Biomedical Research Centre
Enabling translational research through partnership

UNIVERSITY OF OXFORD

Oxford University Hospitals NHS
NHS Trust

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS UPMC

Multiplicom

Molecular Diagnosis of Monogenic Diabetes and Congenital Hyperinsulinism applied to Next-Generation Sequencing

S. Clauin[1], L. Heyrman[2], S. Beaufils-Langlois[1], G. Leroy[1], A. Rotthier[2], C. Saint-Martin[1], J. Del-Favero[2], C. Bellanné-Chantelot[1]

[1] Department of Genetics, AP-HP Groupe Hospitalier Pitié-Salpêtrière, Université Pierre et Marie Curie, Paris, France

[2] Multiplicom N.V., Galileilaan. B-2845 Niel, Belgium

INTRODUCTION

Maturity-onset diabetes of the young (MODY) is the most frequent monogenic form of diabetes mellitus characterised by autosomal dominant inheritance, a young age of onset (< 40 years of age) and pancreatic β-cell dysfunction[1]. MODY is caused by mutations in at least 7 genes *GCK, HNF1A, HNF4A, HNF1B, ABCC8, KCNJ11* and *INS* (Fig. 1).

Mutations of these genes are also responsible for **Congenital hyperinsulinism (CHI)** characterized by an inappropriate oversecretion of insulin. CHI is most commonly associated with mutations of the β-cell ATP-sensitive potassium channel genes encoded by *ABCC8, KCNJ11* and more rarely with mutations of *GCK, HNF4A* and *HNF1A*[2]. The knowledge of the molecular aetiology for both disorders has important implications for clinical management, prognosis and genetic counseling. We co-developed with Multiplicom a multiplex PCR with amplicon specific primers (MASTR™ assay) followed by sequence analysis on the Roche 454 GS Junior or Illumina MiSeq NGS instruments.

[1] Murphy R. et al. *Nature Clinical Practise*, 2008. [2] Sinder KE. et al. *J. Clin Endocrin. Metab.* 2013

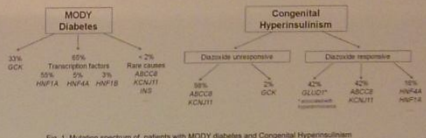


Fig. 1 Mutation spectrum of patients with MODY diabetes and Congenital Hyperinsulinism

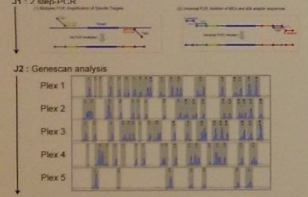
METHODS

• Patients : First, 22 DNA samples with know mutations were analyzed to validate the MASTR assay design and workflow. Second, 32 DNA samples (24 MODY, 8 CHI) were analysed in parallel with Sanger sequencing analysis.

• MASTR assay design

- ✓ **99 amplicons** covering all coding exons, flanking regions (at least 30 bp) of *GCK, HNF1A, HNF4A, HNF1B, ABCC8, KCNJ11, INS* and minimal promoter region of *HNF1A, HNF4A* and *INS* corresponding to 40316 bp with amplicons ranging from 300 bp to 400 bp.
- ✓**21 control amplicons** for copy nunber variants analysis.
- ✓ **5 multiplex reactions;** 50 ng of high quality DNA was used by multiplex reaction.

• Flowchart of MASTR assay and sequencing



• Data analysis : by the Amplicon Variant Analyzer (AVA2) software from Roche Diagnostics and by SeqNext software from JSI Medical Systems.

This flowchart allows the complete analysis (from 1st Multiplex PCR to mutation analysis) of 16 patients in two weeks on the GS Junior instrument.

RESULTS

• Coverage

Seven sequencing runs (54 samples) were performed on the GS junior. An average of 106 read counts per run (Min 73X - Max 147X) was obtained.

Patterns of amplification were stable within multiplex reactions. Six amplicons except one (ABCC8, exon 17, 54%GC) characterized by a high GC-content (62 to 70%) did not reach 40X coverage on the GS junior instrument and should be optimized.

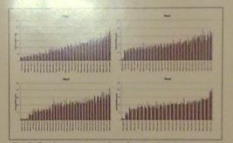


Fig. 2: Readcounts per amplicon for plex 1 to 4 determined with AVA2 script

• GS junior vs. MiSeq sequencing run

One sequencing run (2x 250 bp) including 16 MODY samples combined with 16 BRCA samples was performed on MiSeq. An average of 3960 reads per run per amplicon was obtained. Patterns of amplification after MiSeq sequencing analysis were similar to those obtained with the GS junior. On the MiSeq run, only 3 amplicons (HNF1A, ex1; ABCC8, ex1 and ex17 did not reach 40X coverage).

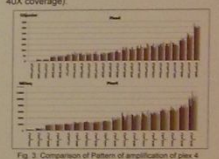
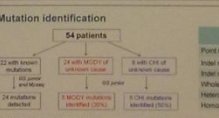


Fig. 3: Comparison of Pattern of amplification of plex 4

• Mutation identification



	Validation Study 22 patients	Blind Study 32 patients
Point mutations (exons, promoter regions)	17	9
Indel mutations	4	1
Indel mutation in homopolymers	2	
Whole gene deletion	1	
Heterozygous SNPs	122	224
Homozygous SNPs	88	205

✓ **Validation study** All known mutations located in amplicons with a coverage >30X were identified. Sequence analysis was based on two independent bioinformatic analysis (AVA2 software and Seqnext software). The use of two bioinformatical tools was important in our experience for analyzing mutations located in GC-rich amplicons (Fig 5) or in homopolymers (Fig. 6) since differences of read counts between forward and reverse stands were observed.

✓ **Blind study** 24 Patients with MODY and 8 with CHI of unknown origin were analysed. 5 mutations were identified in MODY patients and 5 in CHI patients. All detected mutations and negative genotypes identified were confirmed by Sanger sequencing.

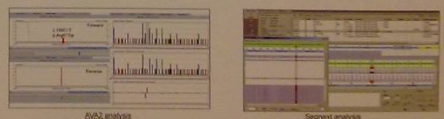


Fig. 5: Example of AVA2 and Seqnext analysis of the *HNF4A* mutation c.199C>T, p.Arg67Trp located in the high GC-rich exon 2 of *HNF4A* and only detected in the reverse strand.

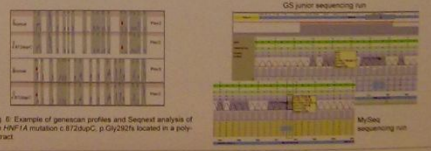


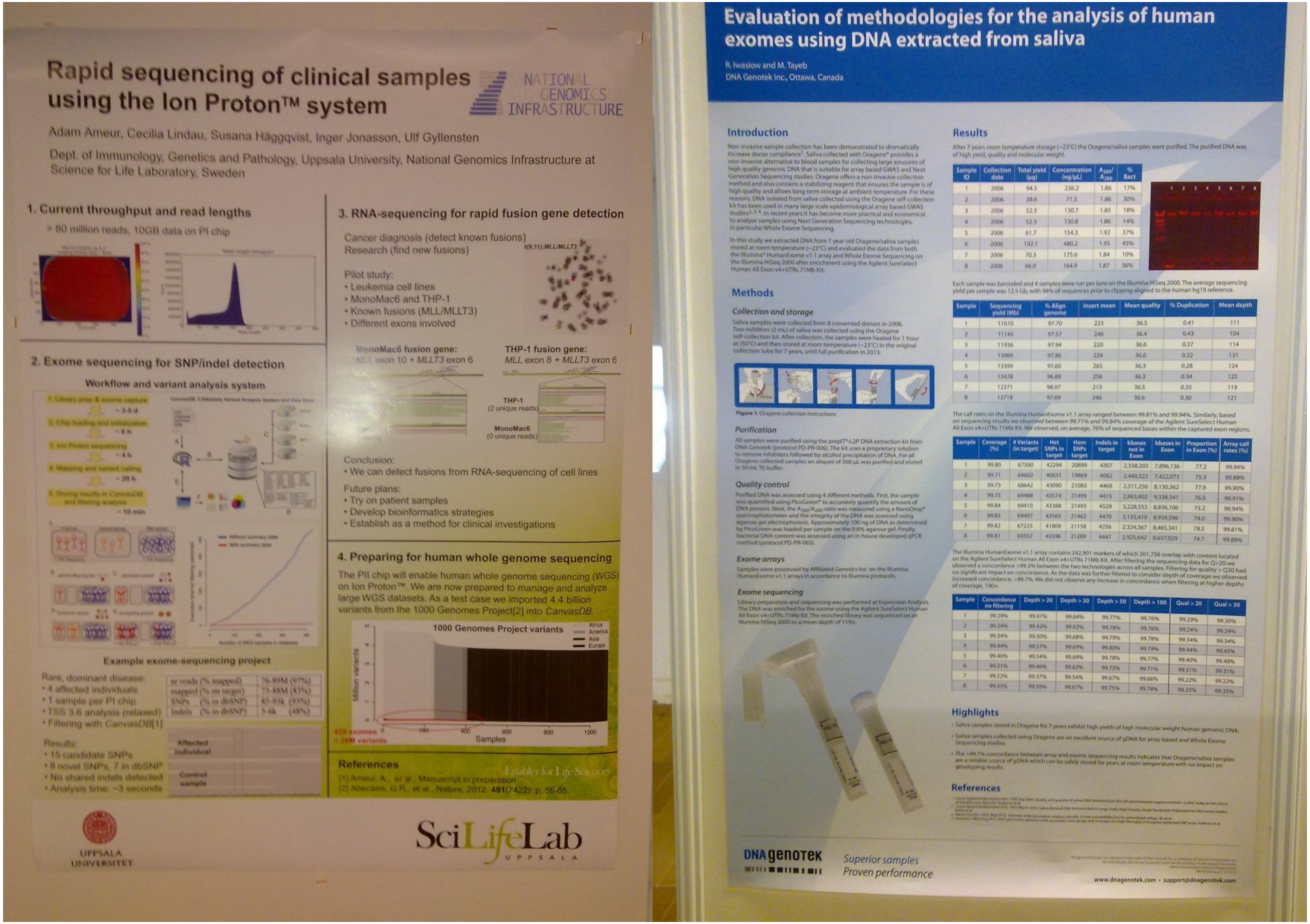
Fig. 6: Example of genescan profiles and Seqnext analysis of the *HNF1A* mutation c.872dupC, p.Gly292fs located in a poly-C tract.

CONCLUSIONS

These results showed that this customized MASTR assay coupled to Next-Generation sequencing workflow is adapted to the routine molecular diagnosis of MODY diabetes and congenital hyperinsulinism. However, further optimizations are required for GC-rich regions and amplicons with homopolymers. Further experiments should also be done to confirm the usefulness of MASTR assay for large exonic deletion/insertion identification.

The bioinformatical analysis is now a major step of the diagnosis workflow based on Next-Generation Sequencing. The use of two distinct bioinformatical algorithms was an added value in our experience for an accurate sequence analysis.

In conclusion, combining MASTR assays with massive parallel sequencing technology offers high throughput and cost effective molecular diagnosis of genetically heterogeneous disorders.


Rapid sequencing of clinical samples using the Ion Proton™ system
NATIONAL GENOMICS INFRASTRUCTURE
Adam Ameur, Cecilia Lindau, Susana Häggqvist, Inger Jonasson, Ulf Gyllensten
Dept. of Immunology, Genetics and Pathology, Uppsala University, National Genomics Infrastructure at Science for Life Laboratory, Sweden
1. Current throughput and read lengths
> 80 million reads, 10GB data on PI chip
2. Exome sequencing for SNP/indel detection
Workflow and variant analysis system
Example exome-sequencing project
Rare, dominant disease:
• 4 affected individuals
• 1 sample per PI chip
• TSS 3.6 analysis (relaxed)
• Filtering with CanvasDB[1]
Results:
• 15 candidate SNPs
• 8 novel SNPs, 7 in dbSNP
• No shared indels detected
• Analysis time: ~3 seconds
Affected individual
Control sample
3. RNA-sequencing for rapid fusion gene detection
Cancer diagnosis (detect known fusions)
Research (find new fusions)
Pilot study:
• Leukemia cell lines
• MonoMac6 and THP-1
• Known fusions (MLL/MLLT3)
• Different exons involved
t(9;11);MLL/MLLT3
MonoMac6 fusion gene:
MLL exon 10 + MLLT3 exon 6
THP-1 fusion gene:
MLL exon 8 + MLLT3 exon 6
THP-1
(2 unique reads)
MonoMac6
(0 unique reads)
Conclusion:
• We can detect fusions from RNA-sequencing of cell lines
Future plans:
• Try on patient samples
• Develop bioinformatics strategies
• Establish as a method for clinical investigations
4. Preparing for human whole genome sequencing
The PII chip will enable human whole genome sequencing (WGS) on Ion Proton™. We are now prepared to manage and analyze large WGS datasets. As a test case we imported 4.4 billion variants from the 1000 Genomes Project[2] into CanvasDB.
1000 Genomes Project variants
Million variants
Samples
References
[1] Ameur, A., et al., Manuscript in preparation
[2] Abecasis, G.R., et al., Nature, 2012. 491(7422): p. 56-65.
Enabler for Life Sciences
UPPSALA UNIVERSITET
SciLifeLab
UPPSALA
Evaluation of methodologies for the analysis of human exomes using DNA extracted from saliva
R. Iwasiow and M. Tayeb
DNA Genotek Inc., Ottawa, Canada
Introduction
Non-invasive sample collection has been demonstrated to dramatically increase donor compliance[1]. Saliva collected with Oragene® provides a non-invasive alternative to blood samples for collecting large amounts of high quality genomic DNA that is suitable for array based GWAS and Next Generation Sequencing studies. Oragene offers a non-invasive collection method and also contains a stabilizing reagent that ensures the sample is of high quality and allows long-term storage at ambient temperature. For these reasons, DNA isolated from saliva collected using the Oragene self-collection kit has been used in many large scale epidemiological array based GWAS studies[2, 3, 4]. In recent years it has become more practical and economical to analyze samples using Next Generation Sequencing technologies, in particular Whole Exome Sequencing.
In this study we extracted DNA from 7 year old Oragene/saliva samples stored at room temperature (~23°C) and evaluated the data from both the Illumina® HumanExome v1.1 array and Whole Exome Sequencing on the Illumina HiSeq 2000 after enrichment using the Agilent SureSelect Human All Exon v4+UTRs 71Mb Kit.
Methods
Collection and storage
Saliva samples were collected from 8 consented donors in 2006. Two millilitres (2 mL) of saliva was collected using the Oragene self-collection kit. After collection, the samples were heated for 1 hour at (50°C) and then stored at room temperature (~23°C) in the original collection tube for 7 years, until full purification in 2013.
Figure 1: Oragene collection instructions
Purification
All samples were purified using the prepIT®-L2P DNA extraction kit from DNA Genotek (protocol PD-PR-006). The kit uses a proprietary solution to remove inhibitors followed by alcohol precipitation of DNA. For all Oragene collected samples an aliquot of 500 µL was purified and eluted in 50 mL TE buffer.
Quality control
Purified DNA was assessed using 4 different methods. First, the sample was quantified using PicoGreen® to accurately quantify the amount of DNA present. Next, the A260/A280 ratio was measured using a NanoDrop® spectrophotometer and the integrity of the DNA was assessed using agarose gel electrophoresis. Approximately 100 ng of DNA as determined by PicoGreen was loaded per sample on the 0.8% agarose gel. Finally, bacterial DNA content was assessed using an in-house developed qPCR method (protocol PD-PR-065).
Exome arrays
Samples were processed by Affiliated Genetics Inc. on the Illumina HumanExome v1.1 arrays in accordance to Illumina protocols.
Exome sequencing
Library preparation and sequencing was performed at Expression Analysis. The DNA was enriched for the exome using the Agilent SureSelect Human All Exon v4+UTRs 71Mb Kit. The enriched library was sequenced on an Illumina HiSeq 2000 to a mean depth of 119×.
Results
After 7 years room temperature storage (~23°C) the Oragene/saliva samples were purified. The purified DNA was of high yield, quality and molecular weight.
Sample ID | Collection date | Total yield (µg) | Concentration (ng/µL) | A260/A280 | % Bact
Each sample was barcoded and 4 samples were run per lane on the Illumina HiSeq 2000. The average sequencing yield per sample was 12.5 Gb, with 98% of sequences prior to clipping aligned to the human hg19 reference.
Sample | Sequencing yield (Mb) | % Align genome | Insert mean | Mean quality | % Duplication | Mean depth
The call rates on the Illumina HumanExome v1.1 array ranged between 99.81% and 99.94%. Similarly, based on sequencing results we observed between 99.71% and 99.84% coverage of the Agilent SureSelect Human All Exon v4+UTRs 71Mb Kit. We observed, on average, 76% of sequenced bases within the captured exon regions.
Sample | Coverage (%) | # Variants (in target) | Het SNPs in target | Hom SNPs target | Indels in target | kbases not in Exon | kbases in Exon | Proportion in Exon (%) | Array call rates (%)
The Illumina HumanExome v1.1 array contains 242,901 markers of which 201,756 overlap with content located on the Agilent SureSelect Human All Exon v4+UTRs 71Mb Kit. After filtering the sequencing data for Q>20 we observed a concordance >99.2% between the two technologies across all samples. Filtering for quality > Q30 had no significant impact on concordance. As the data was further filtered to consider depth of coverage we observed increased concordance, >99.7%. We did not observe any increase in concordance when filtering at higher depths of coverage, 100×.
Sample | Concordance no filtering | Depth > 20 | Depth > 30 | Depth > 50 | Depth > 100 | Qual > 20 | Qual > 30
Highlights
• Saliva samples stored in Oragene for 7 years exhibit high yields of high molecular weight human genomic DNA.
• Saliva samples collected using Oragene are an excellent source of gDNA for array-based and Whole Exome Sequencing studies.
• The >99.7% concordance between array and exome sequencing results indicates that Oragene/saliva samples are a reliable source of gDNA which can be safely stored for years at room temperature with no impact on genotyping results.
References
DNAgenotek
Superior samples
Proven performance
www.dnagenotek.com • support@dnagenotek.com

P15.01

Investigation of intra individual gene expression variations in venous blood from healthy volunteers

Stefan Lehnert, Martine Jaspers, Harry Cuppens

Abstract:

High-throughput sequencing techniques as RNA-Seq may soon be used to discover diseases with diagnostic tests at the earliest possible onset. RNA-Seq techniques create a high resolution readout of transcript levels. The gene-expression profiles have a near digital character, which exceed sensitivity and dynamic range of the hybridization-based array technologies. In order to discover small changes in transcript abundance that may lead to a disease, within individual variability, but also technical and statistical biases of high-throughput sequencing need to be well described and new methods need to be developed. Here, we investigated intra-individual variations over a time span of 2 years collecting between 18 to 26 time points per volunteer. Pipelines as the Tophat&Cufflinks offer the user a number of custom settings to fine-tune the analyses result. We will investigate the impact of these settings on the prediction of transcript abundance systematically. The reproducibility of differential expression calls were estimated in a custom pipeline applying the "Irreproducible discovery rate" test. Technical issues were studied using a number of replicates. Consolidated guidelines will be generated and applied in subsequent analyses together with further samples from later-time points within the same individual.

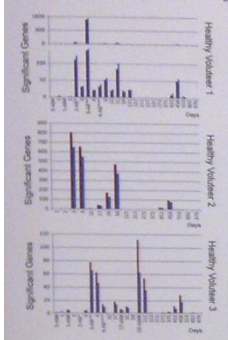


Figure 2. The number of significant differential expressed genes of a 676 day time series is shown for each healthy volunteer (A split panel for volunteer 1). Their reproducibility was scored: red highest average reproducibility, blue lowest average reproducibility (See Figure3 and Methods). Graphs show the number of significant genes and the day of blood collection, where A and B marks comparisons of technical replicates as indicated on Figure 1, B## 2h RT + 2 month storage on 4C°. The table summarizes count differences between the "red" and "blue" dataset of volunteer 1, 2 & 3.

Healthy Volunteer	Red/Blue	Significant genes	Total gainloss	Median gainloss
1	red	2267	7,2%	9,8%
1	blue	2219	18,8%	17,6%
3	red	282	30,8%	16,0%

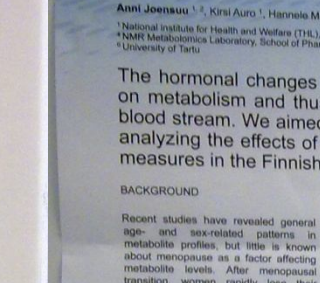


Figure 1. Count vs dispersion plot by sample for all genes. As a quality control measure, biological variation of samples is visualized by estimated overdispersion plots. Exemplary dispersion-levels are shown for samples from healthy volunteer 2. Samples are numbered day 0 to 676 and technical replicates indicated by B. A and B# samples were processed following the PAXgene Blood RNA Tube standard protocol. There a 2h RT step subsequent to the collection of whole blood was mandatory. This step was replaced for sample B* by 48h RT, B** by 4C° for 1 day + 1.5h on RT, B*** by 2h RT + 4 days 4C° together their referred to as "impaired" Samples.

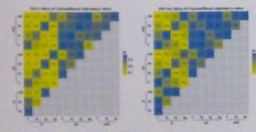


Figure 3. Heatmap summarizing distinctive "Irreproducible discovery rate" rho and p-Values of custom Cufflinks and Cuffdiff parameters. Parameters that resulted identical results were omitted.

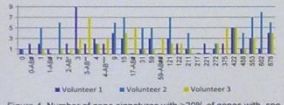


Figure 4. Number of gene signatures with >20% of genes with specificity >0.6 are shown (1 being uniquely up-regulated in one Sample). No bias is seen towards detection of gene signatures in the impaired "B" samples that hold the majority of differential expressed genes. Black truncated columns indicate missing data points.

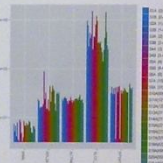


Figure 5. Gene expression plot of 5 selected genes of the volunteer 1 time series. Time points and technical replicates as indicated for Figure 1. Constant expression is shown by the housekeeping gene actinin (ACTN4).

Conclusion:

All samples showed a comparable biological variation within each time series (Figure 1). However, changes in the number of differential expressed genes are observed. Impaired samples showed the majority of differential expressed genes. As expected, period of storage, temperature and incubation periods are essential for RNA quality. Custom Cufflinks and Cuffdiff parameters can also influence the number of significant differential regulated genes (7 to 30% total (Figure1, table). Reproducibility testing (Figure 3) identified the custom parameters (Volunteer 1: ucG/ubc & G/bc; V. 2: NucG/ubc & G/b; V. 3: ucG/bc & G/b). No bias was detected towards the identification of gene signatures that are upregulated uniquely in impaired samples by hard clustering (Figure 4). This was of interest as these samples hold the majority of differential regulated genes. Hard clustering may lack sensitivity. Boundaries between temporal patterns may be not well defined, and changes may occur in multiple directions (Figure 5). Looking at healthy individuals the question for a specific function may lack. Soft clustering methods may help to overcome these problems as it allows to assign genes in degree of membership to different clusters and as the number of clusters are not fixed but depends to the resolution scale.

TopHat: • G

Cufflinks: • G • uG • ucG • NucG

Cuffdiff: • b • bc • bM • Nb • Nbc • NbM • Nub • Nubc • NubM • ub • ubc • ubM

D1 vs D2 -b, -bc, -bM, -Nb, -Nbc, -NbM, -ub, -ubc, -ubM

Method: Samples were analyzed with the "Tuxedo suit" pipeline composed of Bowtie, Tophat, Cufflinks, Cuffdiff using custom parameters as indicated. Custom parameters of Cufflinks and Cuffdiff were used to modulate differential expression calls for pairwise comparisons as indicated in the scheme to the left. The variation of these modulated differential expression calls were estimated in a custom pipeline centered around the "Irreproducible discovery rate" R-package. The reproducibility of differential expression calls was estimated. Subsequent, the samples were analyzed with the Cuffdiff –T option applying the identified parameter sets. The gene expression sets were screened against 3500 gene signatures from GeneSigDB: a manually curated database and resource for analysis of gene expression signatures (NAR 2012).

KATHOLIEKE UNIVERSITEIT LEUVEN

Contact: Prof Harry Cuppens, Harry.Cuppens@med.kuleuven.be

Center for Human Genetics, Campus Gasthuisberg, Katholieke Universiteit Leuven, Belgium

MENOPAUSAL EFFECTS ON METABOLIC SIGNATURES IN A FINNISH POPULATION COHORT

FIMM Institute for Molecular Medicine Finland

Anni Joensuu [1, 2], Kirsi Auro [1], Hannele Mattsson [1, 2], Johannes Kettunen [1, 2], Aki S. Havulinna [1], Pasi Soininen [3, 4], Antti J Kangas [3], Veikko Salomaa [1], Mika Ala-Korpela [3, 4, 5], Markus Perola [1, 2, 6]

[1] National Institute for Health and Welfare (THL), Finland; [2] Institute for Molecular Medicine Finland (FIMM); [3] Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland; [4] NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; [5] Computational Medicine, School of Social and Community Medicine, University of Bristol, UK; [6] University of Tartu

The hormonal changes in menopause are known to have a profound effect on metabolism and thus on the levels of circulating small molecules in the blood stream. We aimed to study which metabolites are affected the most by analyzing the effects of age and menopause on the levels of 135 metabolite measures in the Finnish population sample FINRISK 1997.

BACKGROUND

Recent studies have revealed general age- and sex-related patterns in metabolite profiles, but little is known about menopause as a factor affecting metabolite levels. After menopausal transition, women rapidly lose their advantage in cardiovascular disease risk compared to age-matched men, which has partly been explained with pro-atherogenic alterations in serum lipids.

METHODS

Our original study sample consisted of 7981 Finnish individuals aged 24-74. After exclusions (see Fig. 2) the levels of 135 NMR-measured metabolites or metabolite ratios were individually treated as dependent variables in linear regression analysis with the effect of menopause status in women and the cutoff age of 51-years (the mean age of menopause) in men as the variable of interest. The analyses were adjusted for BMI, smoking and hours fasted before sampling.

The changes of the expression levels of 18 transcripts (measured from leukocytes) from genes in the statin pathway (Fig. 3) were analyzed in a similar linear analysis of 430 individuals of FINRISK 2007, adjusting for BMI and smoking.

RESULTS

Our results highlight a systematic, consistent shift towards a more atherogenic lipid profile by age. In women, the steepest slope of change strictly overlaps the time of menopausal transition, whereas in men the pattern shows a gradual change beginning already at early middle age (Fig. 4). Menopause associated significantly with altogether 94 metabolites (26 correlated <0.8) and in men the young/old status associated with 46 metabolites (8 correlated < 0.8) in the analysis of 24-74-year-olds. Figure 1 shows that especially VLDL-measures and fatty acid measures associate with menopause, but not with age in <40y women. In men the 51-year cutoff was non-significant, but several metabolite measures show strong correlations with age in young men.

The expression levels of LDLR, APOB and LRP1 were observed to vary with menopause in women (P<0.0028). When limiting the analysis to women in the menopausal transition period, LDLR, APOB and APOA2 showed suggestive associations of P<0.05. (Fig. 3)

Fig. 4: The deviation of the mean metabolite level in one year age groups from the mean of under 35-year-olds in units of standard deviation in a) women and b) men of the FINRISK 1997 cohort. The average menopause age of 51 years is marked with a horizontal line.

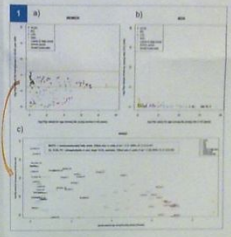


Fig. 1: P-values of a) menopause and b) age limit of 51 years (mimicking menopause) in 40-55-year olds (N=1096 in women and 1279 in men) against P-value of age in <40-year olds (N=1183 in women and 1014 in men). c) the metabolites associated with menopause in 40-55-year-old women. The gray lines mark the 0.05 significance level adjusted for 21 PCs accounting for 95% of the variance in the metabolites.

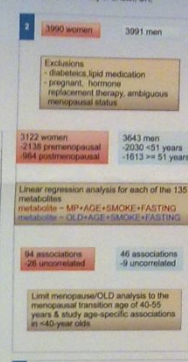


Fig. 2: A flow chart of the metabolic analysis.

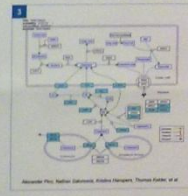
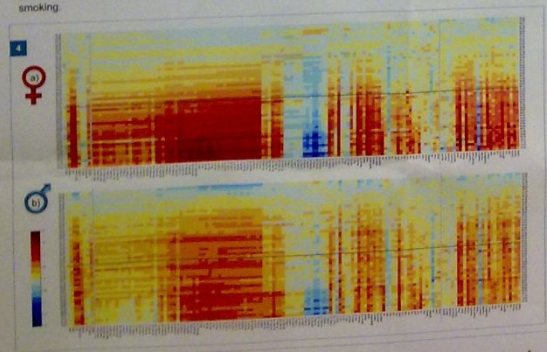
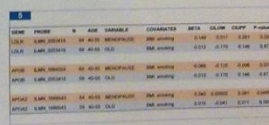


Fig. 3: The statin pathway. Genes analyzed in transcriptomic analysis marked with turquoise.

Fig. 5: Suggestive associations of statin pathway transcripts with menopause, with corresponding analysis in men.



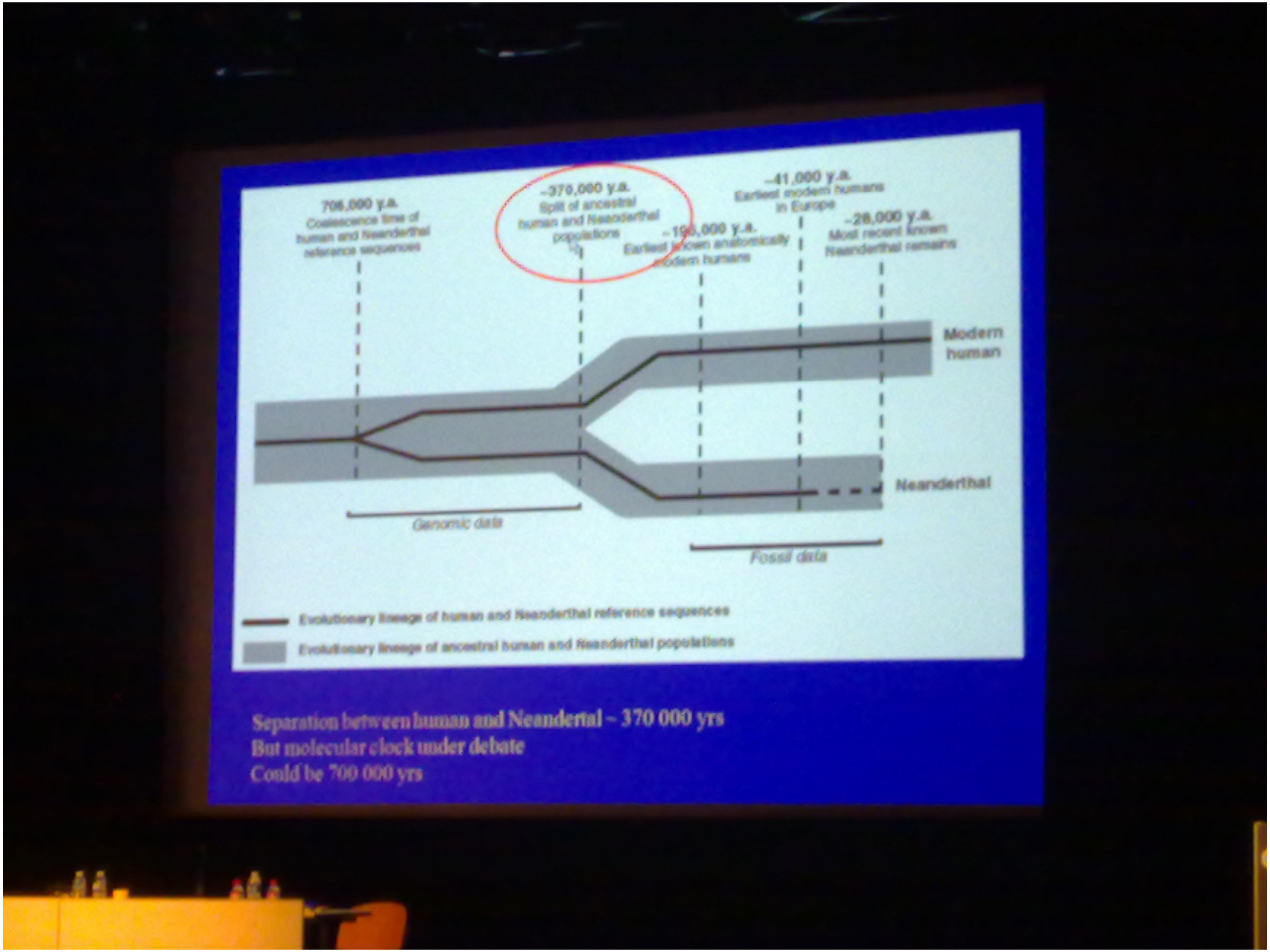


CONCLUSIONS

Our results provide population-based knowledge on the effects of menopause on more than 100 metabolite measures. A vast number of the metabolites show non-linear age associations: menopause affects especially measures regarding VLDL-particles and fatty acids, but also the levels of some amino acids. Our results add valuable insights to CVD-risk assessment especially in women, emphasizing the need to stratify all analysis by sex and the importance of menopausal status alongside age in all CVD- and lipid-related risk analysis.

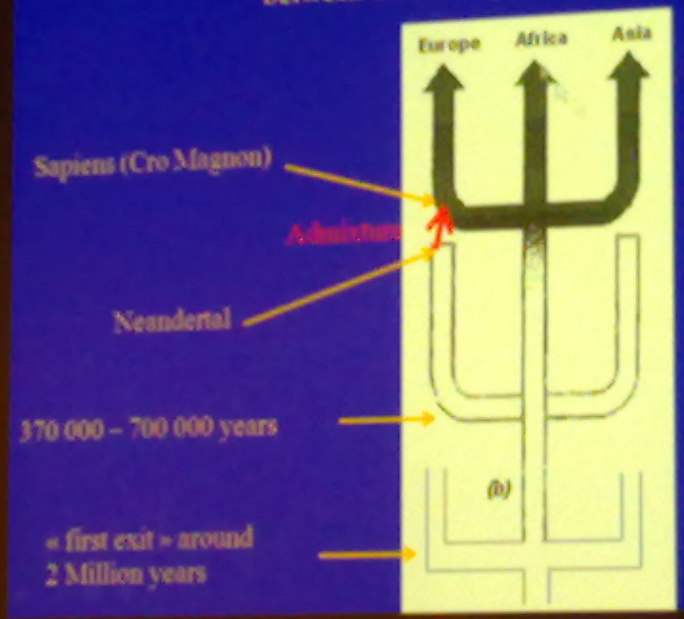
FIMM - Institute for Molecular Medicine Finland
P.O. Box 20, FI-00014 University of Helsinki, Finland
Biomedicum Helsinki 2U, Tukholmankatu 8, 00290 Helsinki, Finland

www.fimm.fi


706,000 y.a.
Coalescence time of human and Neanderthal reference sequences
~370,000 y.a.
Split of ancestral human and Neanderthal populations
~195,000 y.a.
Earliest known anatomically modern humans
~41,000 y.a.
Earliest modern humans in Europe
~28,000 y.a.
Most recent known Neanderthal remains
Modern human
Neanderthal
Genomic data
Fossil data
Evolutionary lineage of human and Neanderthal reference sequences
Evolutionary lineage of ancestral human and Neanderthal populations
Separation between human and Neandertal ~ 370 000 yrs
But molecular clock under debate
Could be 700 000 yrs

1 - 4% of curasian genome comes from Neandertal

Recent African Origin with admixture
between Homo Sapiens and Neandertal



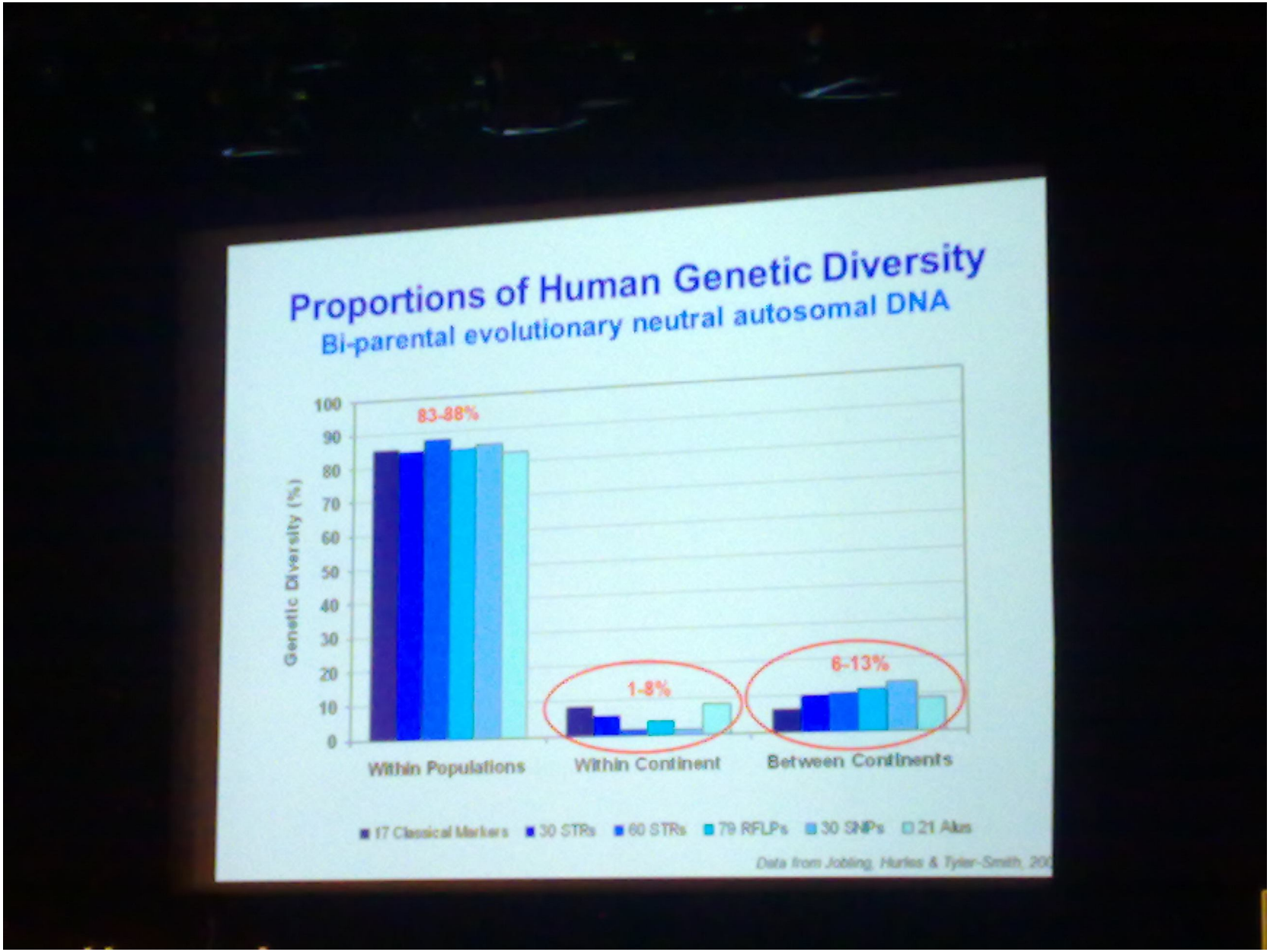
Sapiens (Cro Magnon)

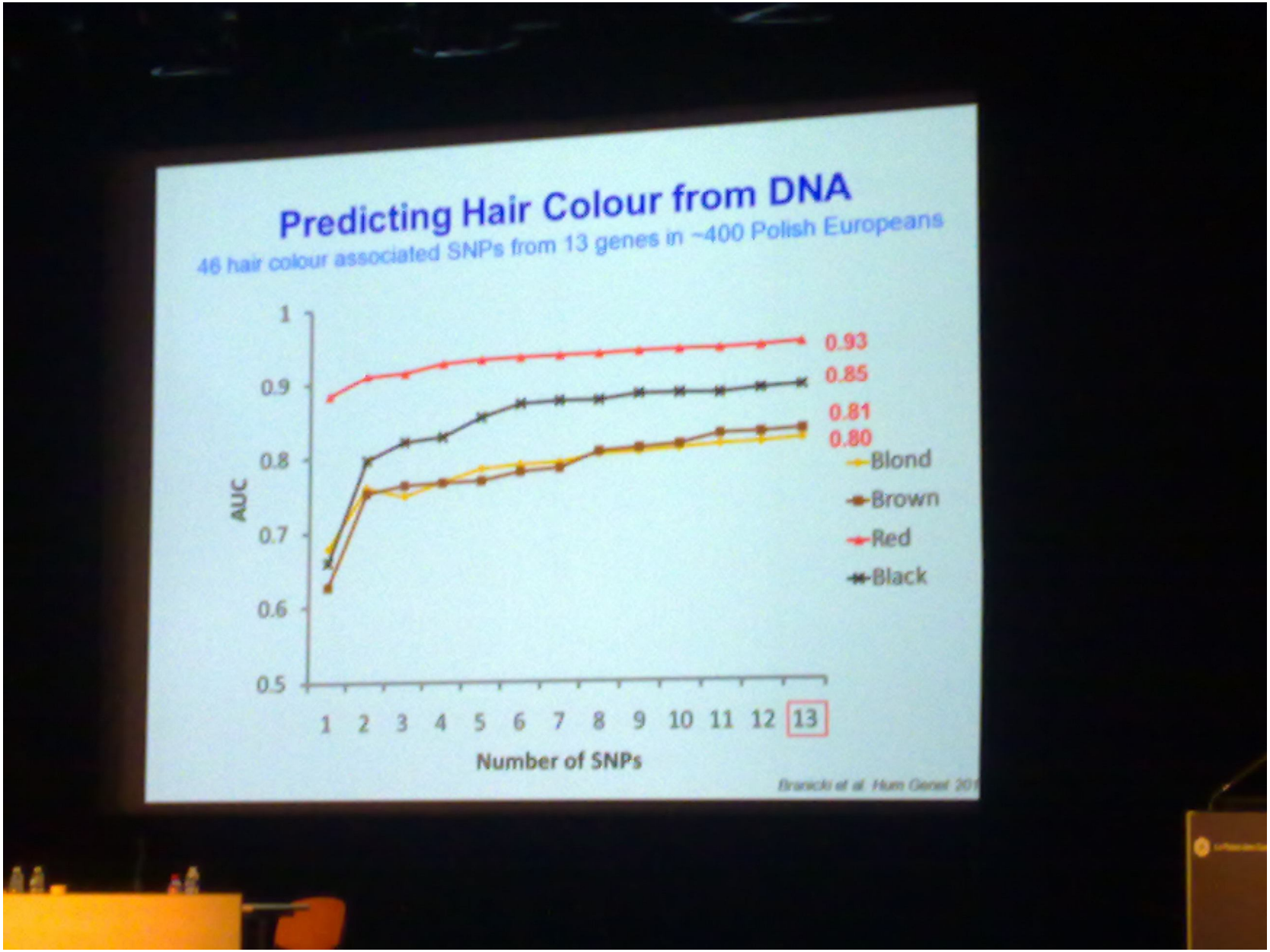
Admixture

Neandertal

370 000 – 700 000 years

« first exit » around
2 Million years


Proportions of Human Genetic Diversity
Bi-parental evolutionary neutral autosomal DNA
Genetic Diversity (%)
100
90
80
70
60
50
40
30
20
10
0
83-88%
1-8%
6-13%
Within Populations
Within Continent
Between Continents
17 Classical Markers
30 STRs
60 STRs
79 RFLPs
30 SNPs
21 Alus
Data from Jobling, Hurles & Tyler-Smith, 20


Predicting Hair Colour from DNA
46 hair colour associated SNPs from 13 genes in ~400 Polish Europeans
AUC
1
0.9
0.8
0.7
0.6
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13
Number of SNPs
0.93
0.85
0.81
0.80
Blond
Brown
Red
Black


Взгляд в NGS будущее!!!