Meeting Minutes

61st EDNAP Meeting, Barcelona, Spain, Catalunia Plaza Hotel

12.11.2024

Agenda

09:00	**Welcome**	
	Welcome by the host organization	
	Welcome by EDNAP board	EDNAP board
09:20	**Current EDNAP projects**	
	Methylated DNA and age exercise	David Ballard
	MPS RNA cSNPs exercises 3&4	Cordula Haas
	mtDNA heteroplasmy exercise	Walther Parson
	Discussion on the current EDNAP projects	all
11:00	**Coffee/Tea**	
11:30	**Current EDNAP projects**	
	Paper Exercise on Estimating Biogeographic Ancestry from DNA	Marta Diepenbroek, Chris Phillips & W Parson
	Future EDNAP activities	
	part 2 - Paper Exercise on Estimating Biogeographic Ancestry from DNA	Marta Diepenbroek, Chris Phillips & W Parson
13:00	**Lunch**	
14:00	**Future EDNAP activities**	
	Funding for projects with EDNAP participation	EDNAP Board
	CapCell: EU-project initiative on single cell analysis	Walther Parson, Bo Thisted Simonsen
	Brief round: Publications by projects with EDNAP-participation	All
14:45	**EDNAP topics**	
	Member management	EDNAP Board
	Applications for membership	All
	Online meetings	All
	EDNAP Homepage	EDNAP Board
	New logo for EDNAP? (include the use of ISFG-logo?)	EDNAP Board
15:30	**Coffee/Tea**	
16:00	**Updates from other reserch**	
	The ENFSI ReAct project	Peter Gill
	MitoMetrics	Vania Pereira
	Other ISFG projects, that can be adopted by EDNAP	All
16:45	**Any other business**	
	next EDNAP meeting: Luxemburg 6 May 2025 followed by the ENFSI DNA Expert Working Group meeting 7-9 May 2025	EDNAP Board
	other	
17:00	**Closure of the meeting**	
	Closure of the meeting	EDNAP Board

Summary of the presentations/discussions

Current EDNAP projects

1. **Methylated DNA and age exercise** (David Ballard)

David discussed the age methylation collaborative exercise that was carried out in two parts and has been deferred since the pandemic. He presented a summary of both the data and the challenges associated with turning this data into a publication. There was a discussion regarding whether the PGM results were relevant given the instrument is not in common use anymore, where points were raised that the underlying technology is the same as the S5 and that the difference between laboratories is interesting irrespective of the difference between technologies. It was decided to explore what could be done with the data we have and see how a story could be written with the paper. David said he could not look at this before Christmas but would work on it In January.

2. **MPS RNA cSNPs exercises 3&4** (Cordula Haas)

Cordula summarized the results of parts 3 and 4 of the MPS RNA cSNPs exercise, on body fluid identification and donor association. A draft manuscript was shared with the participants, who provided valuable feedback. The manuscript will be submitted to FSI Genetics before Christmas 2024.

3. **mtDNA heteroplasmy exercise** (Walther Parson)

Walther presented the preliminary results of the mtDNA heteroplasmy exercise, which had been delayed due to the pandemic. A total of 24 participants submitted results for the first part of the exercise, which focused on comparing the detection of point and length heteroplasmy across Sanger, Ion Torrent, and Illumina technologies. The exercise utilized five DNA extracts provided by the organizing laboratory in Innsbruck. While the detection and reporting of point heteroplasmy were generally consistent across the technologies, notable differences were observed in the reporting of length heteroplasmy, both within and between the technologies. These findings may necessitate revisions to the current interpretation guidelines. Walther will share the result tables with the participants, inviting comments and corrections. A manuscript summarizing and presenting the findings will be prepared.

4. **Paper Exercise on Estimating Biogeographic Ancestry from DNA – part 1** (Marta Diepenbroek)

Marta delivered the final report on the joint exercise on biogeographic ancestry (BGA) estimation.

The slides focused on reports provided by the participants of the exercise, by presenting data in a structure aligned with the planned publication. The nine samples used in the study were divided into three groups with varying levels of difficulty: “easy” samples, samples with

similar genetic patterns, and “advanced” samples. Marta provided a summary of how the labs reported the biogeographical ancestry (BGA) of the samples, emphasizing the different approaches participants used. Additionally, she discussed how labs reported phenotype predictions based on the provided HIrisPlex-S p-values.

It was concluded that simple, unequivocal ancestries received consistent and clear reporting across all labs, while complex samples did not. This highlighted a need for guidelines on data interpretation and the use of software such as STRUCTURE. It was also noted that phenotype predictions lacked consistency between labs, despite most using the available HPS guidelines, indicating a need for updated guidelines on phenotype reporting. Marta also discussed the use of uniparental markers for BGA analysis, concluding that the interpretation of both maternal and paternal markers lacks a clear pathway.

Marta presented the proposed structure of the publication summarizing the study, which will include results from the analyses, lab reports, and feedback from questionnaires completed by participants. It was concluded that this publication should play an observational role. The manuscript is planned for submission before the next EDNAP meeting in May 2025.

Future EDNAP activities

1. **Paper Exercise on Estimating Biogeographic Ancestry from DNA – part 2** (Marta Diepenbroek)

Marta introduced the idea of a follow-up study, primarily planned based on feedback gathered during the first exercise. In this study, a smaller number of samples (five) will be used, and based on an updated ethics committee agreement, raw data (genotypes and haplotypes) will be sent to the participants. Marta shared that guidelines on using STRUCTURE, Snipper, Genogeographer, as well as on interpreting X-SNPs and uniparental markers, will be provided to the participants. In the study, labs will be able to choose which set of markers they will receive and decide how they want to conduct the data analysis. The proposal was positively received by the meeting participants. The follow-up study is planned to begin in 2025.

2. **CapCell** (Bo Thisted Simonsen and Walther Parson)

The EDNAP Board facilitated discussions on potential future activities. Bo and Walther introduced the CapCell consortium, which has applied for EU funding under the current Open Call HORIZON-CL3-2024-FCT-01. If successful, the consortium will focus on developing methods, software, and interpretive frameworks for single-cell forensic DNA analysis. EDNAP, represented by Bo, will support the consortium by reviewing research plans and achievements.

EDNAP Topics

1. **Member management**: Bo Thisted Simonsen summarized the Board's considerations regarding member management. The Board recommended retaining all current members who have attended meetings within the past five years. Applications for new memberships will be reviewed and voted for during (online) meetings. Former members who have not attended meetings in the past five years may reapply for membership. The boards considerations on member management was approved.
2. **Applications for membership**: There were two applications for membership, Natalie de Jong-Weiler, The Netherlands (presented by Walther Parson) and Andreas Tillmar, Sweden (presented by Bo Thisted Simonsen). Both applications were unanimously approved.
3. **Online meetings**: The group discussed the option of online meetings. It was decided that online meetings can be held on an *ad hoc* basis in addition to the annual personal meetings.
4. **EDNAP homepage**: The EDNAP homepage, which is part of the ISFG website, is scheduled for an update. As part of this process, the EDNAP homepage will also undergo revisions.
5. **New EDNAP logo**: There was a brief discussion, no decision was made.

Updates from other research

1. **The ENFSI ReAct project** (Peter Gill)

Peter provided a summary for the ENFSI supported ReAct project, which is an inter-laboratory study to investigate direct and indirect transfer using an agreed experimental design. The first paper is now published as a pre-print and is complemented by databases and software to calculate likelihood ratios. There were wide differences in DNA recovery which affects the likelihood ratios. The ReAct II project is designed to address the issues of reproducibility between labs, by introducing a method to measure DNA recovery. These data will be used to update ReAct I project results.

2. **MitoMetrics** (Vania Pereira)

Vania Pereira presented MitoMetrics - an initiative that aims to address challenges when interpreting forensic mtDNA evidence. The work has a particular focus on heteroplasmy and mtDNA profile discrepancies when comparing different tissues from the same donor. The study introduced a preliminary model to calculate the weight of mtDNA-based evidence using a likelihood ratio approach that accounts for profile discrepancies between tissues.

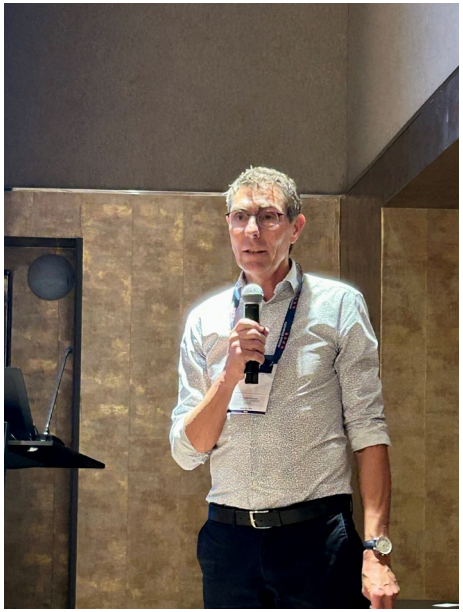
Participants

Last Name	First Name	Contact E-mail	Institution
Aili Fagerholm	Siri	siri.aili-fagerholm@polisen.se	National Forensic Centre, Swedish Police Authority
Ansell	Ricky	ricky.ansell@polisen.se	Swedish Police Authority, National Forensic
Ballard	David	david.ballard@kcl.ac.uk	King's College London
Banemann	Regine	regine.banemann@bka.bund.de	BKA, Federal Criminal Police Office
Bastisch	Ingo	ingo.bastisch@icmp.int	ICMP
Brągoszewska	Anna	anna.bragoszewska@policja.gov.pl	Central Forensic Laboratory of the Police, Departament of Genetics
Courts	Cornelius	cornelius.courts@uk-koeln.de	University Hospital of Cologne - Institute of Legal Medicine
Diepenbroek	Marta	marta.diepenbroek@med.uni-muenchen.de	Institute of Legal Medicine LMU Munich
Fonneløp	Ane Elida	rmanfo@ous-hf.no	Oslo University Hospital
Forsberg	Christina	christina-u.forsberg@polisen.se	Swedish National Forensic Centre
GAMONAL SIMON	Marina	marina.gamonal@gmail.com	POLICIA NACIONAL
Gentile	Fabiano	fabiano.gentile@carabinieri.it	Arma dei Carabinieri
Gill	Peter	peterd.gill@gmail.com	Oslo University Hospital
Guiness	June	june.guiness@homeoffice.gov.uk	Office of the Forensic Science Regulator
Haas	Cordula	cordula.haas@irm.uzh.ch	University of Zurich
Kal	Arnoud	a.kal@nfi.nl	Netherlands Forensic Institute
Kneppers	Sander	s.kneppers@nfi.nl	Netherlands Forensic Institute
Kondili	Aikaterina	k.kondili@astynomia.gr	Forensic Sciences Division, Hellenic Police
Mariiko	Viacheslav	viacheslavmariiko@gmail.com	The State Scientific Research Forensic Center (SSRFC) of the Ministry of Internal Affairs of Ukraine
NOEL	Fabrice	fabrice.noel@just.fgov.be	NICC
O Donnell	Geraldine	gaodonnell@fsi.gov.ie	Forensic Science Ireland
Palencia Madrid	Leire	leire.palencia@seg.euskadi.eus	Forensic Genetics Section, Basque Country Police - Ertzaintza
Parson	Walther	walther.parson@i-med.ac.at	Medical University of Innsbruck
Pereira	Vania	Vania.Pereira@sund.ku.dk	Department of Forensic Medicine, University of Copenhagen
Pietikäinen	Sanna	sanna.pietikainen@poliisi.fi	National Bureau of Investigation Forensic Laboratory
Rodriguez Cuadrado	Irene	irodriguez0061@policia.es	POLICIA NACIONAL
Simonsen	Bo Thisted	bo.simonsen@sund.ku.dk	Dept. Forensic Medicine, University of Copenhagen
Syn	Christopher Kiu Choong	christopher_syn@hsa.gov.sg	Health Sciences Authority Singapore
Syndercombe Court	Denise	denise.syndercombe=court@kcl.ac.uk	King's College London
Wong	Hang Yee	wong_hang_yee@hsa.gov.sg	Health Sciences Authority, Singapore
Zatkalikova	Livia	livia.zatkalikova@minv.sk	Institute of Forensic Science, Ministry of Interior, Slovakia


mossos d'esquadra
Police of Catalonia
Generalitat
de Catalunya
Government
of Catalonia


University of Zurich
Zurich Institute of Forensic Medicine
Acknowledgements


EDNAP Meeting, Barcelona, Spain, Nov 12 2024
mtDNA heteroplasmy exercise



Universität Zürich[UZH]
Zurich Institute of Forensic Medicine

Evolution of EDNAP

EDNAP meeting 12.11.2024

Cordula Haas, Bo Simonsen, Walther Parson

History of EDNAP

1988: Foundation of EDNAP with the aim of harmonizing DNA technology for crime investigation by organizing collaborative intercomparison exercises

1991: EDNAP was accepted as a working group of the International Society for Forensic Genetics (ISFG)
Each European country is represented by one laboratory

First collaborative exercises on:
single-locus DNA probes
STR typing > led to the selection of the "European standard set of loci"

1997-2000: STADNAP network > exercises on autosomal and Y-STRs, mtDNA

1999: The EDNAP Forensic mtDNA Population Database EMPOP was established

Since 2004: joint meetings with the DNA Expert Working Group of the European Network of Forensic Science Institutes (ENFSI)

Recent collaborative exercises on:
single nucleotide polymorphism (SNP) typing
identification of body fluids in forensic stain samples using mRNA analysis
forensic ancestry analysis by SNP typing
forensic phenotyping of eye colour
age estimation by measuring DNA methylation of informative CpG DNA nucleotides

33 EDNAP publications

Here the 5 most recent ones:

Ingold S, Dørum G, Hanson E, Ballard D, Berti A, Gettings KB, Giangasparo F, Kampmann ML, Laurent FX, Morling N, Parson W, Steffen CR, Ulus A, van den Berge M, van der Gaag KJ, Verdoliva V, Xavier C, Ballantyne J, Haas C. Body fluid identification and assignment to donors using a targeted mRNA massively parallel sequencing approach - results of a second EUROFORGEN / EDNAP collaborative exercise. Forensic Sci Int Genet. 2020; 45:102208.

Ingold S, Dørum G, Hanson E, Berti A, Branicki W, Brito P, Elsmore P, Gettings KB, Giangasparo F, Gross TE, Hansen S, Hanssen EN, Kampmann ML, Kayser M, Laurent FX, Morling N, Mosquera-Miguel A, Parson W, Phillips C, Porto MJ, Pośpiech E, Roeder AD, Schneider PM, Schulze Johann K, Steffen CR, Syndercombe-Court D, Trautmann M, van den Berge M, van der Gaag KJ, Vannier J, Verdoliva V, Vidaki A, Xavier C, Ballantyne J, Haas C. Body fluid identification using a targeted mRNA massively parallel sequencing approach - results of a EUROFORGEN/EDNAP collaborative exercise. Forensic Sci Int Genet. 2018; 34:105-115.

Weiler NE, Baca K, Ballard D, Balsa F, Bogus M, Børsting C, Brisighelli F, Červenáková J, Chaitanya L, Coble M, Decroyer V, Desmyter S, van der Gaag KJ, Gettings K, Haas C, Heinrich J, João Porto M, Kal AJ, Kayser M, Kúdelová A, Morling N, Mosquera-Miguel A, Noel F, Parson W, Pereira V, Phillips C, Schneider PM, Syndercombe Court D, Turanska M, Vidaki A, Woliński P, Zatkalíková L, Sijen T. A collaborative EDNAP exercise on SNaPshot™-based mtDNA control region typing. Forensic Sci Int Genet 2017; 26: 77-84.

Santos C, Fondevila M, Ballard D, Baneman R, Bentod AM, Børsting C, Branicki W, Brisighelli F, Burrington M, Capal T, Chaitanya N, Daniel R, Decroyer V, England R, Gettings KB, Gross TE, Haas C, Harteveld PJ, Hoff-Oisen P, Hoffmann A, Kayseri M, Linacre A, Kohler P, Mayr-Eduardoffu M, McGovern C, Morling N, Noel F, O'Donnell G,Parson W, Pascali VL, Porto MJ, Roset A, SchneiderPM, Sijen T, Sten V, Syndercombe Court D, Templeton J, Turanska M, Vallone PM, van Oorschot PAV, Zatkalikova L, The EUROFORGEN-NoE Consortium, Carracedo A, Phillips C. Forensic ancestry analysis with two simple capillary electrophoresis AlMs panels: Results of a collaborative EDNAP exercise. Forensic Sci Int Genet 2015; 19: 56-67.

Chaitanya L, Walsh S, Andersen JD, Ansell R, Ballantyne K, Ballard D, Banemann R, Bauer CM, Bento AM, Brisighelli F, Capal T, ClarisseL, Groß T, Haas C, Hoff-Olsen P, Hollard C, Keyser C, Kiesler CM, Kohler P, Linacre A, Minawi A, Morling N, Nilsson H, Norén L, Ottens R, Parson W, Pascali VL, Phillips C, Porto MJ, Sajantila A, Schneider P, Sijen T, Söchtig J, Syndercombe-Court D, Tilmar A, Turanska M, Vallone PM, Zatkalíková L, Zidkova A, Branicki W and Kayser M. Collaborative EDNAP Exercise on the IrisPlex system for DNA based prediction of human eye colour. Forensic Sci Int Genet 2014; 11: 241-51.

New directions of EDNAP

Today, forensic genetic typing methods are well-established and harmonized in Europe and other parts of the World. The EDNAP laboratories perform accredited DNA analyses. Most EDNAP laboratories are members of the DNA Expert Working Group of ENFSI, which serves as a platform for practical scientific collaboration among European forensic genetic laboratories. The ENFSI group competently addresses many of the issues initially dealt with by EDNAP.
> The need for EDNAP's role in harmonizing standard DNA typing methods no longer exists, a lot of overlap with ENFSI

EDNAP is undergoing a transformation and in this context has adopted Terms of Reference and Statutes for EDNAP. The main points are:

Aims:
- share information, explore new research areas, and drive the development of forensic genetics
- organize collaborative exercises, workshops, in-depth discussions

Members:
- change from laboratory/country membership to individual/personal membership
- the General Assembly decides on membership with a simple majority
- invite guests (temporary)

Meetings:
- at least one annual personal scientific meeting, together with ENFSI
- additional personal, internet-based, or combined meetings

Projects:
- explore new forensic genetic research areas suitable for research projects
- form research groups
- apply for funding

New directions of EDNAP

New structure and organization of EDNAP:
- is a working group with its own statutes under the International Society for Forensic Genetics
- is organized with an elected board with a Chairman, a Deputy Chairman, a Secretary/Treasurer

At the last EDNAP meeting (29. May 2024 in Copenhagen)
- the Statutes and the terms of references (ToR) were adopted
- new EDNAP board members were elected
  Chair: Cordula Haas
  Deputy Chair: Bo Simonsen
  Secretary and Treasurer: Walther Parson

New directions of EDNAP

Change of collaboration style:

- Colleagues that have new research questions/suggestions can get in contact with the EDNAP group directly/immediately
- Colleagues that plan to present ongoing exercise results should send results well ahead of the meeting to allow for more meaningful discussions
- Particularly problems, limitations that arise during the analyses should be shared when they arise to allow for better troubleshooting

Focus more on EDNAP members' practical experiences (at court)

On-going collaborative exercises:

- Paper Exercise on Estimating Biogeographic Ancestry (M. Diepenbroek, C. Phillips & W. Parson)
- mtDNA heteroplasmy exercise (W. Parson)
- Second exercise on methylated DNA and age (D. Syndercombe Court, D. Ballard)
- Exercise 4 on mRNA typing with MPS (C. Haas)
- The series of exercises relating to DNA transfer (R. van Oorschot, B. Kokshoorn)

→ New topics/ideas are welcome ☺!

New agenda structure

Preliminary EDNAP Agenda
12. Nov 24
Catalonia Barcelona Plaza Hotel

Vs. 3.0 (04.11.2024)

	09:00	**Welcome**	
10 min		Welcome by the host organization	
10 min		Welcome by EDNAP board	EDNAP board
	09:20	**Current EDNAP projects**	
30 min		Methylated DNA and age exercise	David Ballard
20 min		MPS RNA cSNPs exercises 3&4	Cordula Haas
30 min		mtDNA heteroplasmy exercise	Walther Parson
20 min		Discussion on the current EDNAP projects	all
	11:00	**Coffee/Tea**	
	11:30	**Current EDNAP projects**	
45 min		Paper Exercise on Estimating Biogeographic Ancestry from DNA	Marta Diepenbroek, Chris Phillips & W Parson
		Future EDNAP activities	
30 min		part 2 - Paper Exercise on Estimating Biogeographic Ancestry from DNA	Marta Diepenbroek, Chris Phillips & W Parson
	13:00	**Lunch**	
	14:00	**Future EDNAP activities**	
10 min		Funding for projects with EDNAP participation	EDNAP Board
20 min		CapCell: EU-project initiative on single cell analysis	Walther Parson, Bo Thisted Simonsen
10 min		Brief round: Publications by projects with EDNAP-participation	All
	14:45	**EDNAP topics**	
45 min		Member management	EDNAP Board
		Applications for membership	All
		Online meetings	All
		EDNAP Homepage	EDNAP Board
		New logo for EDNAP? (include the use of ISFG-logo?)	EDNAP Board
	15:30	**Coffee/Tea**	
	16:00	**Updates from other reserch**	
15 min		The ENFSI ReAct project	Peter Gill
15 min		MitoMetrics	Vania Pereira
5 min		Other ISFG projects, that can be adopted by EDNAP	All
	16:45	**Any other business**	
		next EDNAP meeting: Luxemburg 6 May 2025 followed by the ENFSI DNA Expert Working Group meeting 7-9 May 2025	EDNAP Board
		other	
	17:00	**Closure of the meeting**	
		Closure of the meeting	EDNAP Board

Contact
Cordula Haas: cordula.haas@irm.uzh.ch
Bo Simonsen: bo.simonsen@sund.ku.dk
Walther Parson: walther.parson@i-med.ac.at

https://www.isfg.org/EDNAP

EDNAP/ENFSI Rome 2018

Methylated DNA & Age Exercise

David Ballard

EDNAP, Barcelona 2024

KING'S College LONDON

EDNAP Exercise

- 15 laboratories participated
  - 8 MiSeq only
  - 5 PGM only
  - 2 MiSeq and PGM/S5
- Part 1 - 7 Methylation standards between 0-100% sent out to all labs
- Part 2 – 7 blood stains sent out to laboratories to test. Also optional submission of extra blood samples. Age prediction by ANN from methylation values at 12 markers.

Introduction

The development of methods that can accurately estimate an individual's chronological age from trace evidence is an ongoing quest in the field of forensic DNA intelligence. The retrieval of this information, as well as information regarding externally visible characteristics, like eye, hair or skin colour and hair morphology [1-4], from DNA samples recovered from crime scenes, can significantly aid police investigations, especially in cases lacking eye witness testimonies and/or intact human remains.

While multiple biomarkers for chronological age have been suggested over the years [5-19], the quantification of DNA methylation, an epigenetic modification that mainly affects cytosines when these are followed by guanines in a 5'->3' direction and is a known modulator of genetic expression [20], has been the focus of recent research. The main reasons behind this choice are the strong and specific correlation of multiple methylation sites with chronological age, as shown in multiple studies [20-24], as well as methylation's biological stability over time [24-28].

Several different approaches have been established for the quantification of DNA methylation, with the four main ones being (i) massively parallel sequencing (MPS), (ii) pyrosequencing, (iii) methylation SNaPshot and (iv) MALDI-TOF mass spectrometry (EpiTYPER). Massively parallel sequencing offers high sensitivity as well as single-base resolution and is able to cope with large scale multiplexing, characteristics that place it to the top of the choices for DNA methylation quantification for forensic purposes. Furthermore, forensic laboratories worldwide are becoming increasing familiar with this technology as it has been applied to multiple aspects of forensic analysis [29-33]. The most common MPS instrumentations currently being used in forensic laboratories include the MiSeq (Illumina), MiSeq FGx (Verogen) and the ION Personal Genome Machine (ION PGM) and ION S5 systems (Life Technologies).

Recent publications reveal significant scientific leaps towards making age estimation through DNA methylation a reality for forensic casework, with the developed methods showcasing promising results in terms of accuracy, robustness and sensitivity [34-37]. However, even though few published methods have been successfully reproduced across the forensic community [38, 39] little research has been conducted on the transferability of the proposed methods between different laboratories as well as different instruments. While the stages of identification of promising markers and optimisation of the potential methods are vital to the development of new forensic tools, transferability between different forensic facilities is also an important factor that needs to be investigated, especially when the proposed methods involve high cost equipment like the MPS instrumentation.

In order to investigate further into this matter, this exercise focuses on the transferability of a previously described DNA methylation-based age prediction method originally developed on the MiSeq FGx platform [40]. The same protocol, with minor instrument-related alterations was performed in 14 different labs using different types of MPS technology including the MiSeq, MiSeq FGx, ION PGM and ION S5 systems and the results were compared both for standards of known methylation and real samples.

Materials and Methods

DNA methylation standards

For the first part of this study 7 pre-mixed methylation standards ranging from 0% to 100% methylation were purchased from EpigenDx (Massachusetts, USA) at a concentration of 50ng/µL. Standards were diluted and delivered to the participating laboratories at a final concentration of 2.5ng/µL.

Sample Collection

For the leading research group of this exercise in King's College London, sample collection for this study was performed under ethical approval granted by the Biomedical Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Research Ethics Subcommittee (BDM/13/14-30). A total of 7 donors aged between 27.7 and 79.7 years were recruited for the collection of whole blood samples (samples A-G) via venepuncture following the acquisition of full informed consent. Samples were stored at 4°C.

Sample Shipping

Methylation standards were shipped to the participating laboratories in sealed 1.5mL snap-top containers, while samples A-G were shipped in the form of blood stain cards.

DNA Extraction and Quantification

Genomic DNA was extracted by different methods depending on the laboratory, with the BioRobot®EZ1 automated purification instrument (Qiagen, Hilden, Germany) in combination with the EZ1 Blood and Investigator kits being the most popular choice. Other methods included the QIAamp DNA mini kit (Qiagen, Hilden, Germany), QIAamp DNA Investigator kit (Qiagen, Hilden, Germany), DNA IQ system (Promega Corporation, Wisconsin USA), Wizard® Genomic DNA Purification kit (Promega Corporation, Wisconsin USA), PrepFiler Forensic DNA Extraction kit (Thermo Fisher Scientific, Massachusetts, USA), Chelex and organic extraction (Supplementary File 1a).

Similarly, several different methods were employed for the quantification of the DNA extracts with the most common ones being a fluorometric quantitation using Qubit for double stranded DNA high sensitivity (Thermo Fisher Scientific, Massachusetts, USA) and a real-time PCR quantitation with Quantifiler® Trio DNA Quantification kit (Thermo Fisher Scientific, Massachusetts, USA) in full or half volumes. Additional methods included the Quantifiler® Human DNA Quantification (Thermo Fisher Scientific, Massachusetts, USA), Quantifiler® Duo (Thermo Fisher Scientific, Massachusetts, USA), Quantifiler® Human Plus (Thermo Fisher Scientific, Massachusetts, USA), AluQuant™ Human DNA Quantitation (Promega Corporation, Wisconsin USA), Quantus™ Fluorometer in combination with the QuantiFluor™ dsDNA Dyes (Promega Corporation, Wisconsin USA) and NanoDrop Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) (Supplementary File 1b).

	A	B	C	D	E
8		Illumina			
9	7. NFI	KAPA Hyper Prep kit for Illumina®			
10	8. Oslo	Agencourt AMPure XP; Ion Plus Fragment Library Kit (Cat. no. 4471252); Ion PGM Hi-Q Chef Kit; Ion Xpress™ Barcode Adapters 1–16 Kit; Ion 314™ Chip Kit v2 BC	ampliseq protocol?		
11	9. Zurich	IonXpress Plus gDNA Fragment Library Kit, Thermofisher	Ion PGM Hi-Q OT2 200 Kit, Ion PGM Hi-Q Sequencing Kit, Ion 318 Chip v2		
12	10.Florida	KAPA hyper prep kit for illumina; agencourt AMPpure XP beads; Roche SeqCap A/B adapters			
13	11. Singapore MiSeq	KAPA Hyper Prep for Illumina platforms, NimbleGen SeqCap Adapter Kit A and B Used 1.87 µl of 10µM of the index adapters	KAPA Hyper Prep for Illumina platforms, NimbleGen SeqCap Adapter Kit A and B Agencourt AMPure XP reagent, Beckman Coulter Genomics		
14	11. Singapore S5		Ion XpressPlus gDNA Fragment Library Preparation Ion Xpress Barcode Adapters Used 50ng PCR product for end-repair instead of 200ng, and performed library amp before library quan. Also have tried using 100ng without library amp following suggested PGM protocol, but S5 run failed. Agencourt AMPure XP reagent, Beckman Coulter Genomics		
15	12. Lyon	KAPA Hyper Prep kit for Illumina®, KAPABiosystems (Cat.No.: KK8502, 48 reactions) Agencourt® AMPure® XP reagent, Beckman Coulter Genomics (Cat.No.: A63881, 60 ml SeqCap Adapter Kit, Roche, Cat.No.: 07141530001 for set A/07141548001 for set B			
16	13. NIST	KAPA Hyper Prep for Illumina, with adapters from Illumina TruSeq 96plex Adapter Plate			
17	14. Victoria				

Bisulphite conversion | PCR | Clean up | Post Clean up quant | Library Prep

3. Results

3.1 *DNA standards of known methylation*

In the first part of this collaborative exercise, the different laboratories were provided with the same pre-mixed primers as well as the same pre-mixed DNA methylation standards ranging from 0 to 100% methylation. All standards were analysed in duplicate by each participant laboratory and the detected methylation values were averaged across the labs using MiSeq technology and those using Ion PGM or Ion S5 sequencers (Fig.1). While results showed good correlation between the two different sequencing technologies for 7/12 markers, significant differences ($p<0.05$) were observed for cg04528819, cg22736354 and cg06493994.

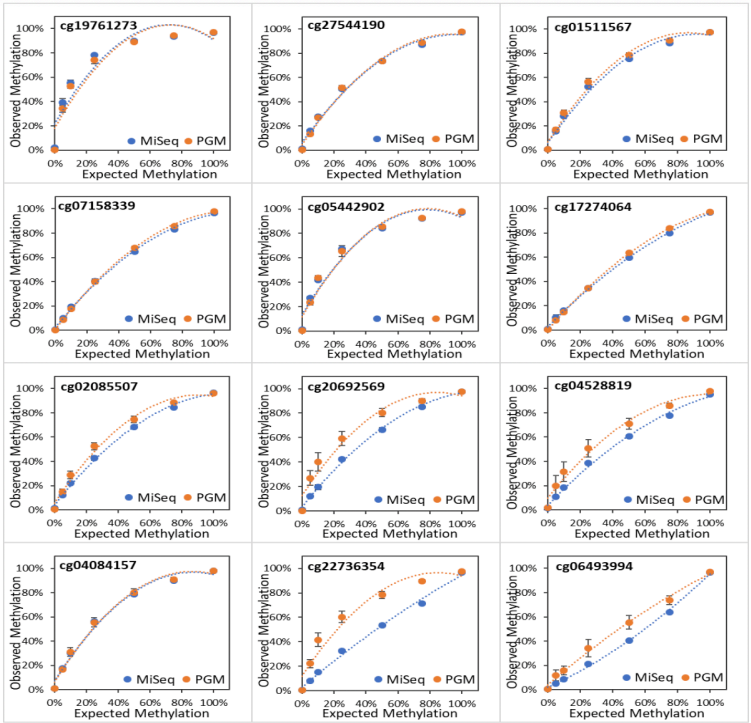


Figure 1. Averaged detected DNA methylation values for the 7 standards of known methylation for the laboratories using MiSeq (blue) and PGM/S5 (orange) technology. Significant differences were observed for cg04528819,

- Average sequencing reads for the different markers in the MiSeq and the PGM/S5 platforms. Error bars represent the standard deviation.

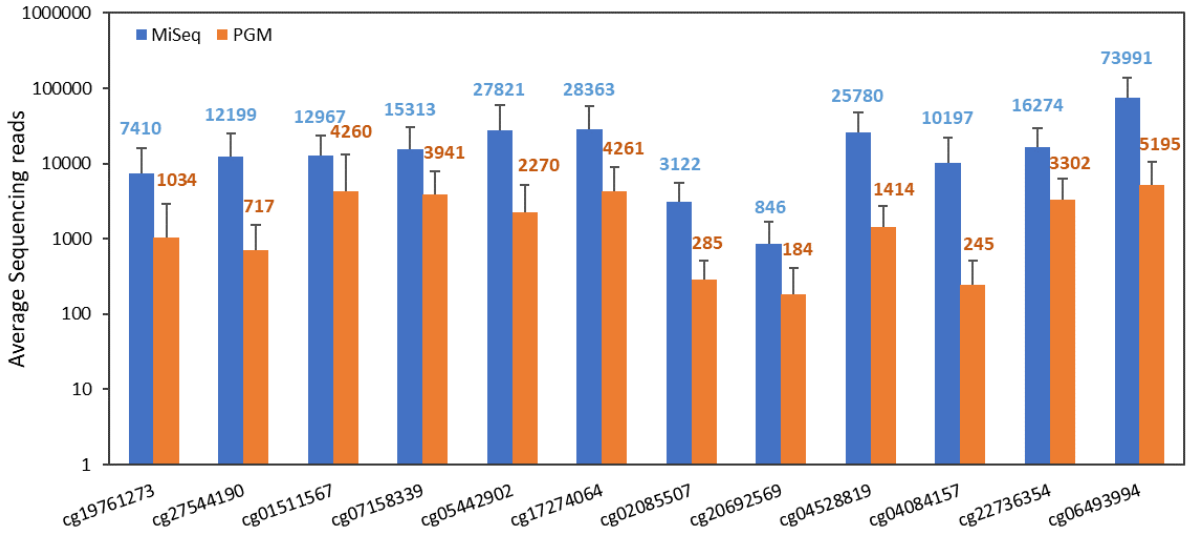


Figure 1 – Laboratory prediction of samples A-F

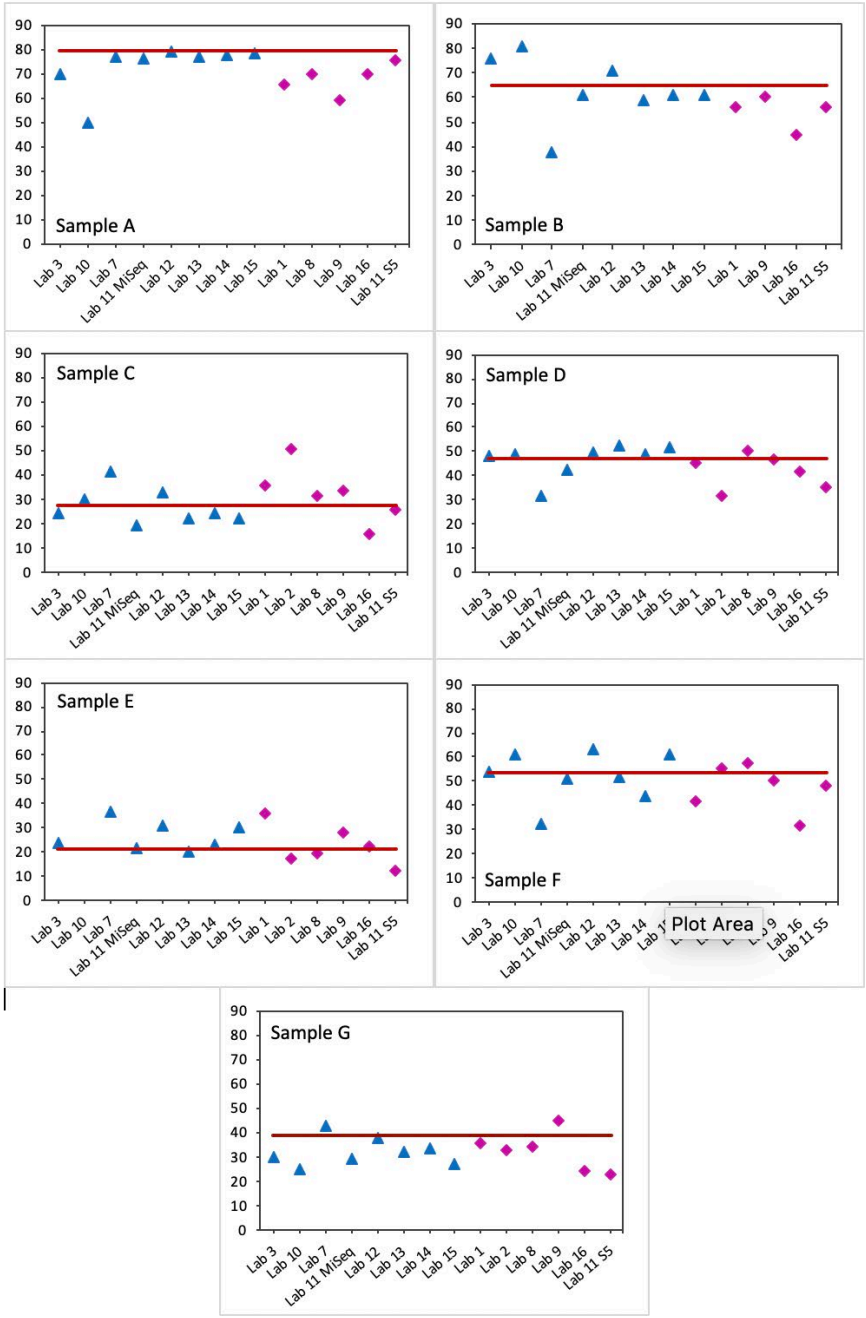
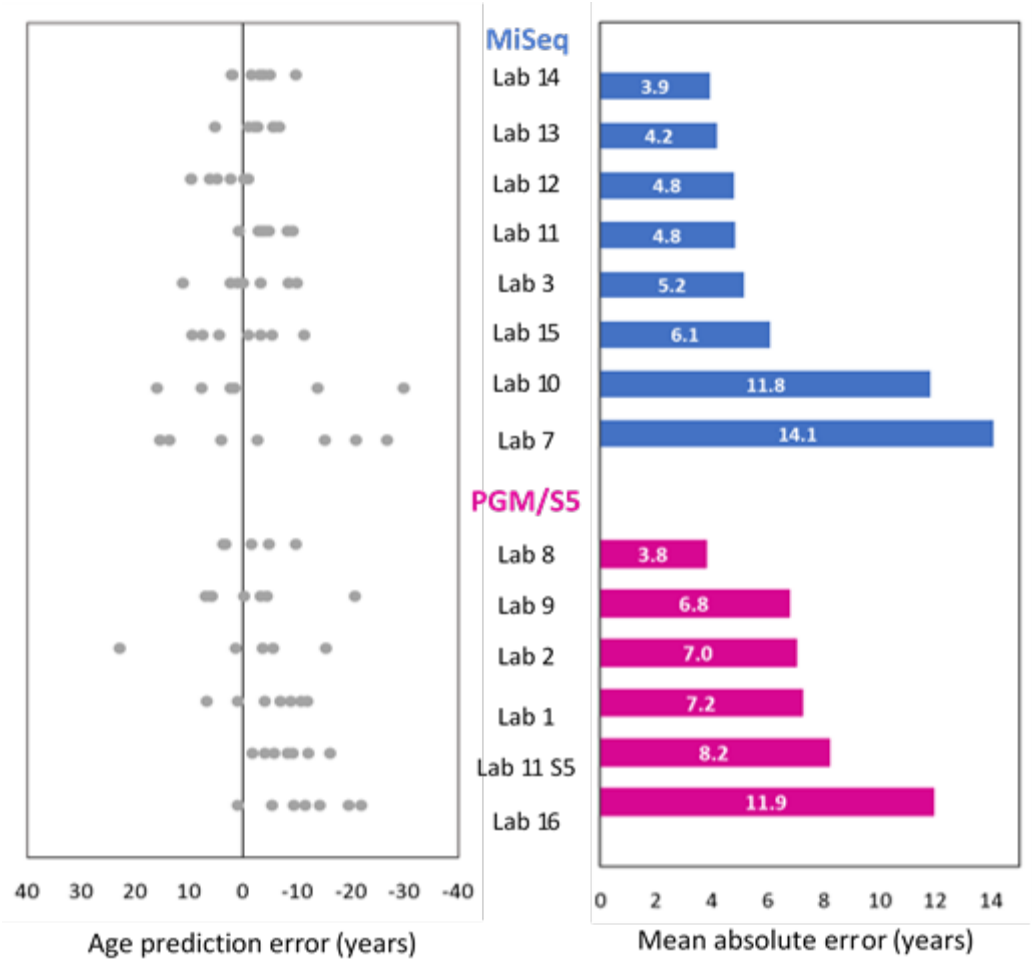


Figure 2 – Laboratory prediction of samples A-F



Manuscript

- Introduction
- Methods
- Results Phase 1
  - Methylation control reproducibility per lab and marker
  - Methylation control differences MiSeq vs PGM/S5
- Results Phase 2
  - Age prediction reproducibility per lab
  - Blind prediction samples
- Discussion/conclusion

Acknowledgments

- Anastasia Aliferi
- Athina Vidaki
- Leon Barron
- Denise Syndercombe Court

DAVID BALLARD
DNA ANALYSIS AT KING'S
KING'S COLLEGE LONDON
LONDON
UK

DAVID.BALLARD@KCL.AC.UK

KING'S College LONDON

University of Zurich[UZH]

Zurich Institute of Forensic Medicine

NATIONAL CENTER FOR FORENSIC SCIENCE

EDNAP mRNA MPS collaborative exercises 3 and 4 (BFID-cSNP-BSS and BFID-cSNP-6F)

Cordula Haas, Nadescha Hänggi, Erin Hanson, Jack Ballantyne

EDNAP Meeting, 12. November 2024, Barcelona

EUROFORGEN Network of Excellence

EUROFORGEN-NoE is funded by the European Commission within the 7th Framework Programme

Association of Body Fluids with a Donor: cSNPs

DNA Profile:

- Mixture of 2 persons

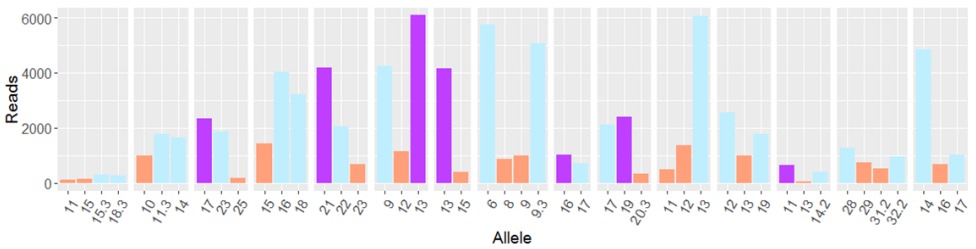
RNA Profile:

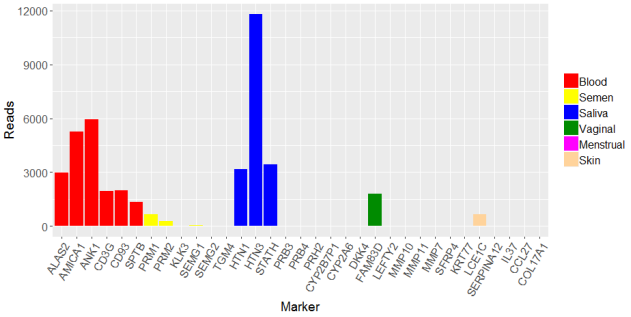
- Body fluid identification (BFID)

cSNPs:

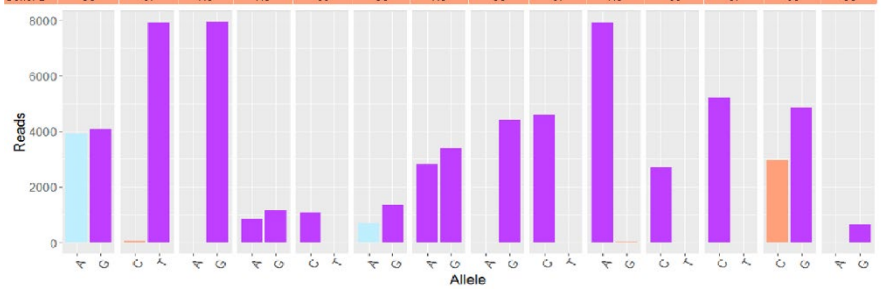
- Association to donors

Donor	D1S1656	D2S441	D2S1338	D3S1358	FGA	D8S1179	D10S1248	TH01	vWA	D12S391	D16S539	D18S51	D19S433	D21S11	D22S1045
Donor 1	15.3;18.3	11.3;14	17;23	16;18	21;22	9;13	13	6;9.3	16;17	17;19	13	12;19	11;14.2	28;32.2	14;17
Donor 2	11;15	10	17;25	15	21;23	12;13	13;15	8;9	16	19;20.3	11;12	13	11;13	29;31.2	16





Donor	Blood markers											Saliva markers		
	AMICA1_1	AMICA1_2	ANK1_1	ANK1_2	ANK1_3	ANK1_4	CD3G	CD93_1	CD93_2	CD93_3	SPTB	MUC7_1	MUC7_2	PRB3
Donor 1	AG	TT	GG	AG	CC	AG	AG	GG	CC	AA	CC	CC	CC	GG
Donor 2	GG	CT	AG	AG	CC	GG	AG	GG	CT	AG	CC	CT	CG	GG



Association of Body Fluids with a Donor: cSNPs

International Journal of Legal Medicine (2023) 137:13–32
https://doi.org/10.1007/s00414-022-02908-9

ORIGINAL ARTICLE

Check for updates

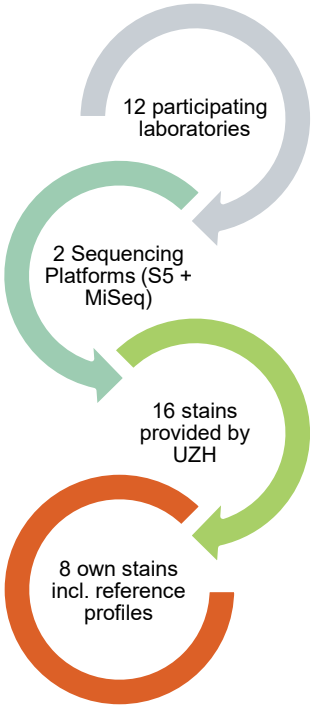
Targeted S5 RNA sequencing assay for the identification and direct association of common body fluids with DNA donors in mixtures

Erin Hanson[1,2] · Guro Dørum[3] · Manuel Zamborlin[3] · Shouyu Wang[3] · Mario Gysi[3] · Sabrina Ingold[3] · Robert Lagace[4] · Chantal Roth[4] · Cordula Haas[3] · Jack Ballantyne[1,2]

BFID-cSNP-BSS blood, semen, saliva

BFID-cSNP-6F 6 fluids/tissue

EDNAP mRNA MPS Exercises 3 and 4



		3rd Exercise BFID-cSNP-BSS assay		4th Exercise BFID-cSNP-6F assay	
Body Fluid	BFID Marker	included	# cSNPs	included	# cSNPs
Blood (BL)	ANK1		2		2
	CD3G		1		1
	SPTB		4		4
Semen (SE)	PRM1		1		1
	SEMG2		1		1
	KLK3		2		2
	TGM4		4		4
Saliva (SA)	HTN3		3		3
	PRB4		1		1
	PRH2		1		1
	MUC7		1		1
	STATH				
Vaginal Secretion (VAG)	CYP2A6				1
	MUC22				7
	CYP2B7P1				
Menstrual Blood (MB)	MMP10				2
	MMP3				1
	COL6A3				5
	COL12A1				3
	LEFTY2				
Skin (SK)	LCE1C				3
	COL17A1				1
	IL37				2

RNA/DNA Co-Extraction → RT → cDNA STR-Analysis → Library Preparation → Sequencing: S5/MiSeq → Data Analysis

EDNAP mRNA MPS Exercises 3 and 4

Stain N°	BF/T	Amount	Stain Provided
1	SE	10 µl	piece of fabric (boxer shorts)
2	BL-MB	1/2 Swab + 25 µl	1/2 swab
3	SE	50 µl	artificial cotton
4	SA-SE	50 µl + 25 µl	part of a T-shirt
5	BL	50 µl	1 swab
6	SK	1 swab	1 swab
7	BL-BL	25 µl + 25 µl	part of a T-shirt
8	SA	Licked plastic spoon	spoon
9	SA-SA	25 µl + 25 µl	1 swab
10	BL-SA	25 µl +25 µl	1 swab
11	SA	50 µl	part of a T-shirt
12	VAG	1/2 swab	1/2 swab
13	BL	Nose bleed on tissue	part of a tissue
14	SA-SE	25 µl + 25 µl	piece of fabric (boxer shorts)
15	MB	1/2 swab	1/2 swab
16	SE-VAG	½ Swab+ 25 µl SE	1/2 swab

Stain N°	BF/T	Amount	Stain Provided
1	SK	1 swab	1 swab
2	BL-MB	1 swab + 25ul	1/4 swab
3	SA-VAG	1 swab + 25ul	1/4 swab
4	SE-MB	1 swab + 25ul	1/4 swab
5	BL-SE	25ul + 25ul	part of T-Shirt
6	SE-SE	25ul + 25ul	1 swab
7	SA-MB	1 swab + 50ul	1/4 swab
8	SA-SK	1 swab + 25ul	1 swab
9	VAG	cotton part of a slip	a piece of it
10	MB	menstrual pad	a part of it
11	SE	50ul	part of a glove (latex)
12	BL	20ul	part of a T-Shirt
13	SA-SE	50ul + 10ul	artificial cotton
14	VAG-BL	1 swab + 25ul	1/4 swab
15	SA	50ul	stockings (nylon)
16	VAG-SE	1 swab + 25ul	1/4 swab

Light blue: single donor, low input
Dark blue: single donor, high input
Orange: mixtures

Participating Laboratories

6x S5
3x MiSeq
2x both sequencing platforms

Netherlands Forensic Institute, Ministry of Justice and Security, Netherlands

National Forensic Center, Swedish Police Authority, Sweden

Department of Analytical, Environmental and Forensic Sciences, King's College London, UK

Institute of Forensic Medicine, University of Zurich, Switzerland

Department of Forensic Medicine, University of Copenhagen, Denmark

Institute of Forensic Medicine, University Medical Center Cologne, University of Cologne, Germany

National Center for Forensic Science, University of Central Florida (UCF), USA

Institute of Forensic Sciences, DNA department, Bavarian State Criminal Police Office, Germany

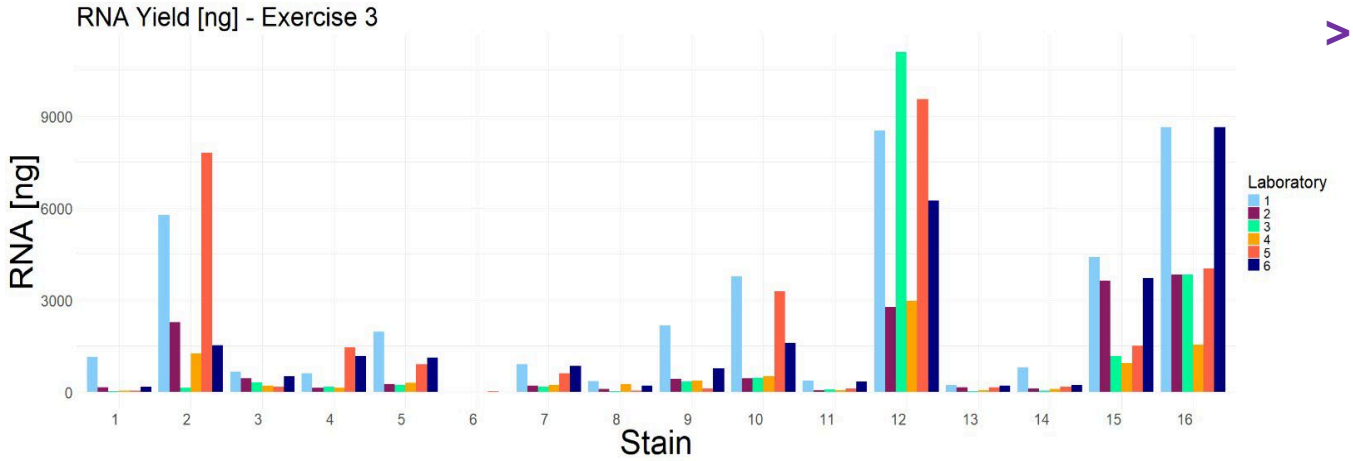
Departement of Forensic Sciences, Oslo University Hospital, Norway

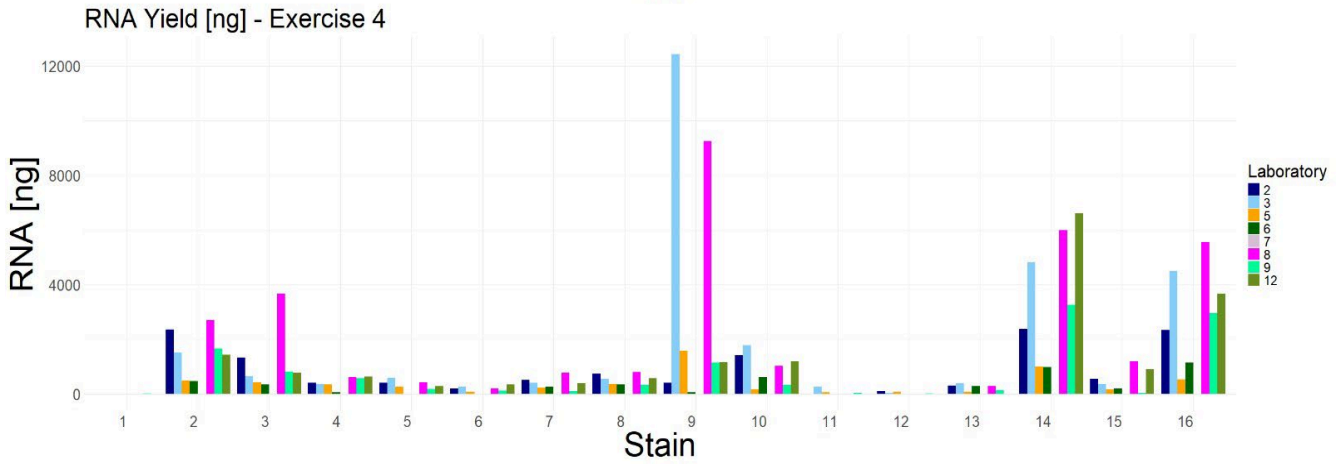
Institute of Legal Medicine, Innsbruck Medical University, Austria

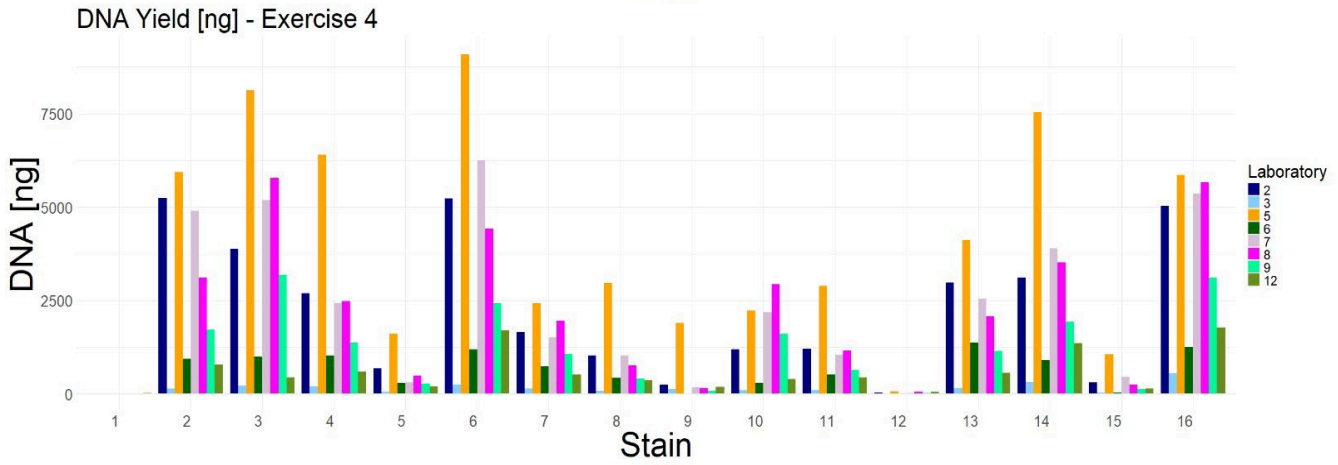
Instituto Nacional de Medicina Legal, I.P., Ministry of Justice, Portugal

RNA / DNA yields

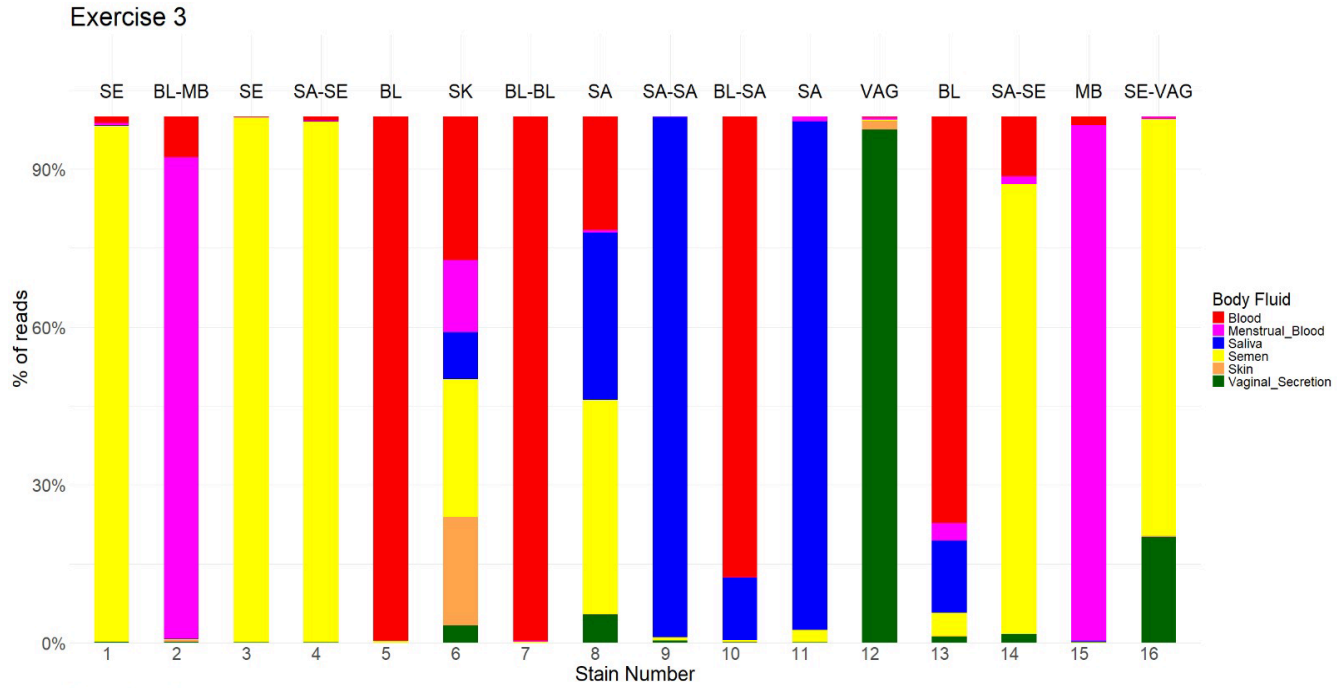
> Suppl

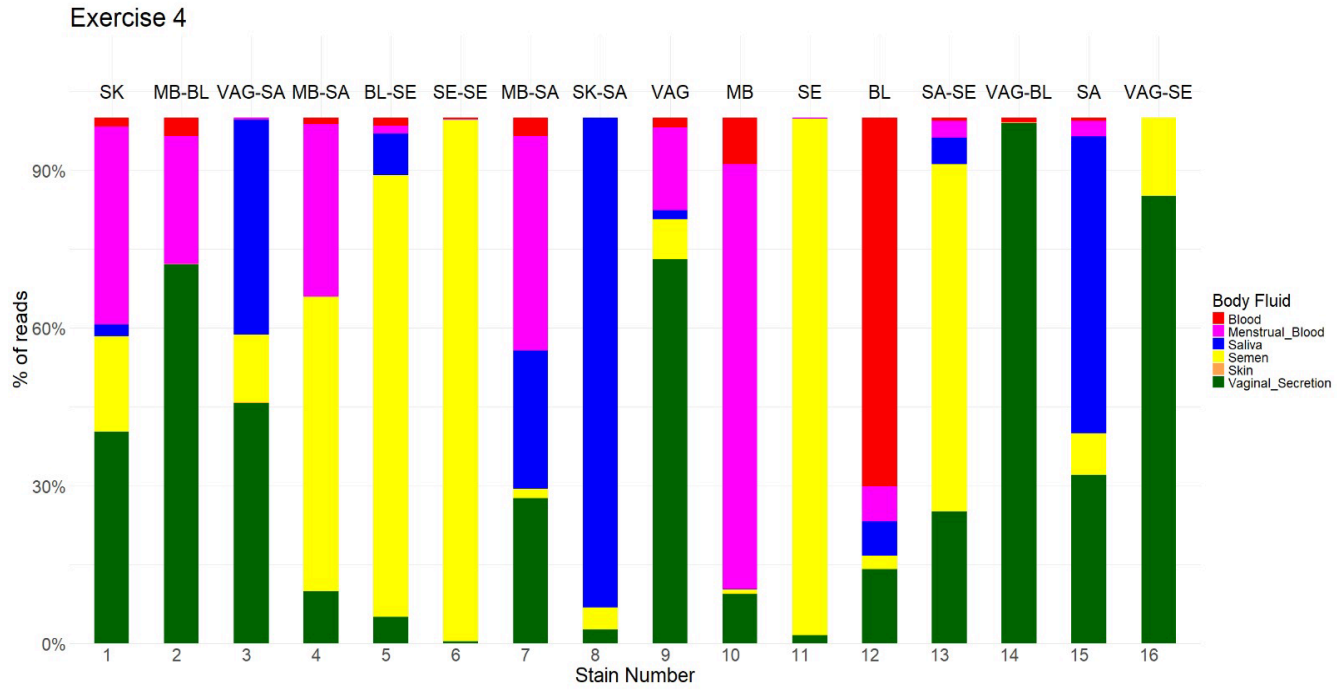






Composition analysis of stains by body fluid percentages



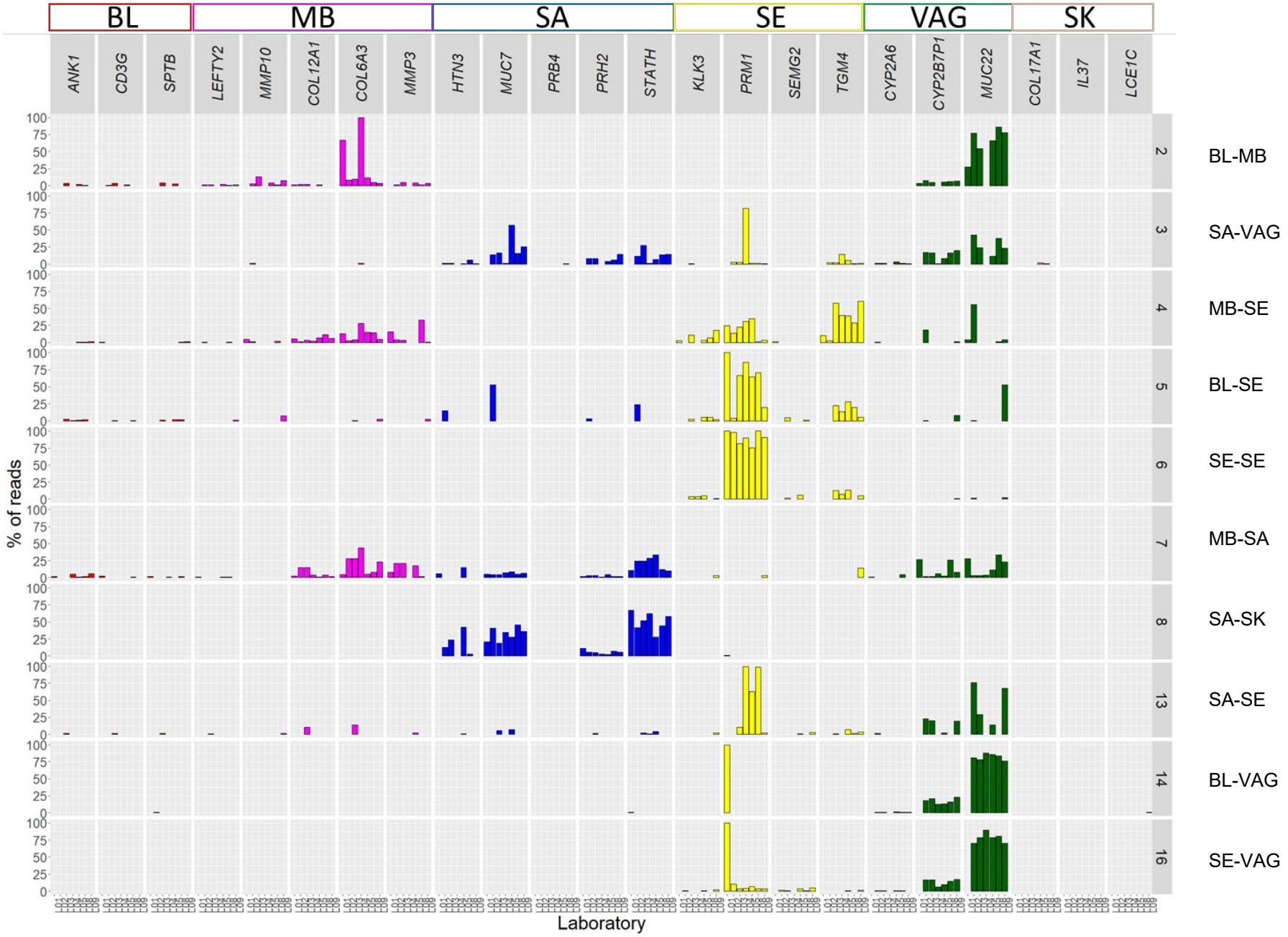


Exercise 4: Single body fluid stains analyzed by S5 laboratories

> Suppl

Exercise 4: Mixed body fluid stains analyzed by S5 laboratories

> Suppl



Association to Donors in Mixed Stains

Stain 3	HTN3	HTN3	HTN3	MUC7	PRB4	PRH2	HTN3	CYP2A6	MUC22	MUC22	MUC22	MUC22	MUC22	MUC22	MUC22	MUC22	MUC22	MUC22
SA-VAG	rs1849937_rs1136515	rs1849937	rs1136515	rs2306948	rs1052808	rs10772391	rs75067954	rs8192721	rs12110470_rs12110785	rs12110470	rs12110785	rs3869098_rs4248153	rs3869098	rs4248153	rs1419664_rs3094672	rs1419664	rs3094672	rs10947121
IonCode_135	TC	T/T	C/C	C/T	G/G	C/C	C/C	C/C	GT	G/G	T/T	AA/GA	A/G	A/A	CT/TA/TC	C/T	T/A	T/T
IonCode_136	CT/CC	C/C	T/C	C/C	G/C	C/C	C/C	C/C	TC/GT	T/G	C/T	AA/GG	A/G	A/G	CA/CC	C/C	A/C	T/C
Lab1 SS - Genotype																		
Lab1 SS - Read Counts																		
Lab2 SS - Genotype	TC/CT	T/C	C/T	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA/ACT	C/C	A/T	C/T
Lab2 SS - Read Counts	204\51	204\51	204\51	2185\1911	15\15	2327\2327	204\204		672\600	672\600	672\600	5858\3628	5858\3628	5858\3628	1600\274	1600\1600	1600\274	672\600
Lab3 SS - Genotype	TC/TC	T/T	C/C	C/T		C/C	C/C		TC/GT	T/G	C/T	CGG/CAA	G/A	G/A	ACA	C/C	A/A	C/T
Lab3 SS - Read Counts	215		215\215	1724\1504		1632\1632	215\215		269\200	269\200	269\200	1700\1548	1700\1548	1700\1548	1065	1065\1065	1065\1065	269\200
Lab4 SS - Genotype																		
Lab4 SS - Read Counts																		
Lab5 SS - Genotype	TC/CC	T/C	C/C	T/C		C/C	C/C					CAA/CGG	A/G	A/G	ACT	C/C	T/T	
Lab5 SS - Read Counts	42\17\15	42\17	42\42	4405\3462		560\560	42\42					950\668	950\668	950\668	19	19\19	19\19	
Lab6 MiSeq - Genotype				C/T		C/C			TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab6 MiSeq - Read Counts				34\19		51\51			76\32	76\32	76\32	103\102	103\102	103\102	250	250\250	250\250	76\32
Lab7 MiSeq - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab7 MiSeq - Read Counts		14464\14464	14464\14464	21108\17201	2418\2418	8824\8824	14464\14464		8858\8081	8858\8081	8858\8081	27716\18812	27716\18812	27716\18812	30115	30115\30115	30115\30115	8858\8081
Lab8 SS - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CGG/CAA	G/A	G/A	ACA	C/C	A/A	C/T
Lab8 SS - Read Counts	33244	33244\33244	33244\33244	47296\36046	3064\3064	33033\33033	33244\33244		11954\10099	11954\10099	11954\10099	57295\51908	57295\51908	57295\51908	70883	70883\70883	70883\70883	11954\10099
Lab9 SS - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab9 SS - Read Counts	2536	2536\2536	2536\2536	68781\41249	298\298	60076\60076	2536\2536		6614\3587	6614\3587	6614\3587	30194\23648	30194\23648	30194\23648	36149	36149\36149	36149\36149	6614\3587
Lab10 MiSeq - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab10 MiSeq - Read Counts	311	311\311	311\311	1017\559	5\5	329\329	311\311		197\71	197\71	197\71	1740\1328	1740\1328	1740\1328	761	761\761	761\761	197\71
Lab11 MiSeq - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		GT/TC	G/T	T/C	CAA/CGG	A/G	A/G	ACA	C/C	A/A	T/C
Lab11 MiSeq - Read Counts	2971	2971\2971	2971\2971	3854\3124	268\268	3107\3107	2971\2971		1978\1777	1978\1777	1978\1777	9956\9835	9956\9835	9956\9835	7131	7131\7131	7131\7131	1978\1777
Lab12 MiSeq - Genotype	TC/TC	T/T	C/C	C/T		C/C	C/C					CAA/CGG	A/G	A/G	ACA	C/C	A/A	
Lab12 MiSeq - Read Counts	36	36\36	36\36	120\110		15\15	36\36					106\69	106\69	106\69	14	14\14	14\14	

Stain 3 (SA-VAG):

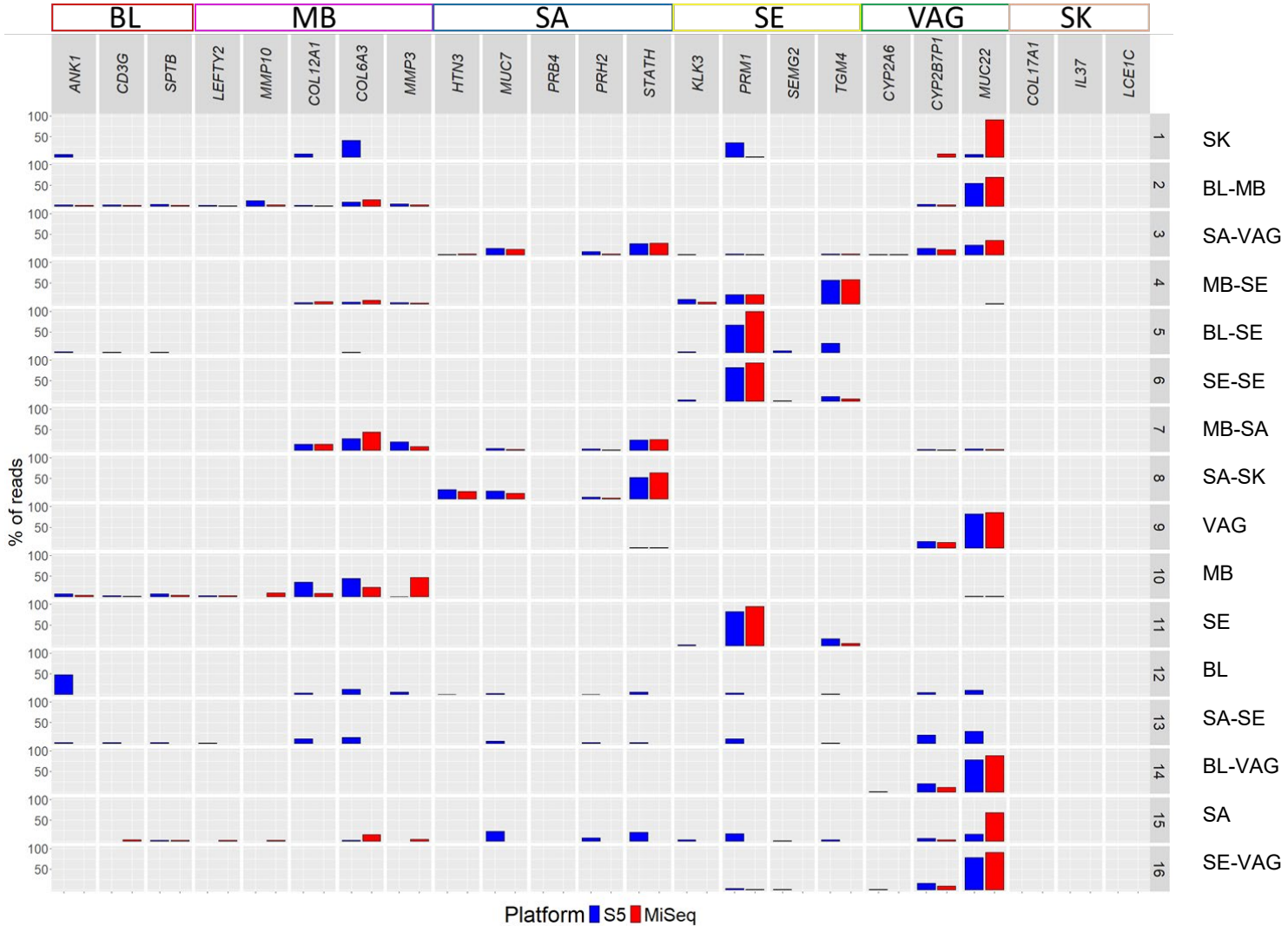
- high number of reads
- RNA cSNP genotype mostly reflects donor genotypes

Stain 14	CYP2A6	MUC22.0	MUC22.1	MUC22.2	MUC22.3	MUC22.4	MUC22.5	MUC22.6	MUC22.7	MUC22.8	MUC22.9	ANK1.0	ANK1.1	CD3G	SPTB.0	SPTB.1	SPTB.2	SPTB.3	SPTB.4
VAG-BL	rs8192721	rs12110470_rs12110785	rs12110470	rs12110785	rs3869098_rs4248153	rs3869098	rs4248153	rs1419664_rs3094672	rs1419664	rs3094672	rs10947121	rs504574	rs7816734	rs3753059	rs1741488_rs1741487	rs1741488	rs1741487	rs229592	rs229586
IonCode_139	C/C	GT	T/T	A/A/GA	A/G	A/A	A/A	CT/CA/CC	C/C	T/A	T/T	C/G	G/G	T/T	CA	C/C	A/A	A/A	C/T
IonCode_147	C/C	TC/TT	T/T	C/T	GG/AG	G/A	G/G	CA/TA/CC/TC	C/T	A/A	C/T	G/C	G/G	T/T	CA/TG	C/T	A/G	A/G	C/C
Lab1 SS - Genotype																			
Lab1 SS - Read Counts																			
Lab2 SS - Genotype		GT/GT	G/G	T/T	CAA/CGA	A/G	A/A	ACT/ACA	C/C	T/A	T/T	C/G	G/G	T/T	ATG/ACA	T/C	G/A	G/A	C/C
Lab2 SS - Read Counts		16600	16600\16600	16600\16600	76026\60510	76026\60510	76026\76026	38149\36457	38149\38149	38149\36457	16600\16600	530\375	868\868	413\413	244\228	244\228	244\228	304\60	1511\1511
Lab3 SS - Genotype		GT/GT	G/G	T/T	CAA/CGA	A/G	A/A	ACA/ACT	C/C	A/T	T/T	C/G	G/G	T/T	ATG/ACA	T/C	G/A	G/A	C/C
Lab3 SS - Read Counts		6370	6370\6370	6370\6370	25976\23531	25976\23531	25976\25976	18803\18118	18803\18803	18803\18118	6370\6370	84\18	421\421	326\326	82\20	82\20	82\20	31\11	312\312
Lab4 SS - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACA/ACT	C/C	A/T	T/T	C/G	G/G	T/T	ATG/ACA/AAG	T/C	G/A		C/C
Lab4 SS - Read Counts		39616	39616\39616	39616\39616	179565\170161	179565\170161	179565\179565	39416\35836	39416\39416	39416\35836	39616\39616	798\682	56\56	1055\1055	99\64\8	99\64	99\64		1604\1604
Lab5 SS - Genotype		GT/GT	G/G	T/T	CAA/CGA	A/G	A/A	ACT/ACA	C/C	T/A	T/T	C/C	G/G	T/T	ACA	C/C	A/A	G/G	C/C
Lab5 SS - Read Counts		39196	39196\39196	39196\39196	200156\191512	200156\191512	200156\200156	79575\73926	79575\79575	79575\73926	39196\39196	932\932	1226\1226	1665\1665	735	735\735	735\735	479\479	2034\2034
Lab6 MiSeq - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACA/ACT	C/C	A/T	T/T	G/C	G/G	T/T	ATG/ACA	T/C	G/A	G/A	C/T
Lab6 MiSeq - Read Counts		18646	18646\18646	18646\18646	30610\12528	30610\12528	30610\30610	17009\8256	17009\8256	17009\8256	18646\18646	7770\7770	1125\1125	14020\14020	1100	1100\1100	1100\1100	24\24	9180\9180
Lab7 MiSeq - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACT/ACA	C/C	T/A	T/T	G/C	G/G	T/T	ACA/ATG	C/T	A/G	G/A	C/C
Lab7 MiSeq - Read Counts		21248	21248\21248	21248\21248	77165\76895	77165\76895	77165\77165	46196\46069	46196\46196	46196\46069	21248\21248	299\175	274\274	485\485	177\22	177\22	177\22	21\19	588\588
Lab8 SS - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACT/ACA	C/C	T/A	T/T	C/G	G/G	T/T	ACA/ATG	C/T	A/G	A/A	C/C
Lab8 SS - Read Counts		36495	36495\36495	36495\36495	187101\175195	187101\175195	187101\187101	158341\153043	158341\158341	158341\153043	36495\36495	1263\261	189\189	2392\2392	248\231	248\231	248\231	253\253	661\661
Lab9 SS - Genotype		GT/GT	G/G	T/T	CAA/CGA	A/G	A/A	ACT/ACA	C/C	T/A	T/T	G/C	G/G	T/T	ACA	C/C	A/A	G/G	C/C
Lab9 SS - Read Counts		57011	57011\57011	57011\57011	154560\153950	154560\153950	154560\154560	157605\149433	157605\157605	157605\149433	57011\57011	378\364	550\550	1372\1372	301	301\301	301\301	181\181	2274\2274
Lab10 MiSeq - Genotype		GT/GT	G/G	T/T	CAA/CGA	A/G	A/A	ACA/ACT	C/C	A/T	T/T	G/C	G/G	T/T	ACA	C/C	A/A		C/C
Lab10 MiSeq - Read Counts		1106	1106\1106	1106\1106	8971\8630	8971\8630	8971\8971	3612\3557	3612\3612	3612\3557	1106\1106	16\16	26\26	49\49	6	6\6	6\6		24\24
Lab11 MiSeq - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACA/ACT	C/C	A/T	T/T	C/G	G/G	T/T	ATG/ACA	T/C	G/A	A/A	C/C
Lab11 MiSeq - Read Counts		8562	8562\8562	8562\8562	45238\45119	45238\45119	45238\45238	26448\24013	26448\26448	26448\24013	8562\8562	102\42	133\133	77\77	35\31	35\31	35\31	13\13	53\53
Lab12 MiSeq - Genotype		GT/GT	G/G	T/T	CGA/CAA	G/A	A/A	ACA/ACT	C/C	A/T	T/T	G/C	G/G	T/T	ACA	C/C	A/A	G/G	C/C
Lab12 MiSeq - Read Counts		1697	1697\1697	1697\1697	10233\9976	10233\9976	10233\10233	6486\4946	6486\6486	6486\4946	1697\1697	59\9	120\120	49\49	13	13\13	13\13	14\14	25\25

Stain 14 (VAG-BL):

- high number of reads in most markers
- RNA cSNP genotype reflects donor genotypes

Comparison of Sequencing Platforms (Lab1)



Forensic Science International: Genetics 73 (2024) 103125

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Comprehensive body fluid identification and contributor assignment by combining targeted sequencing of mRNA and coding region SNPs

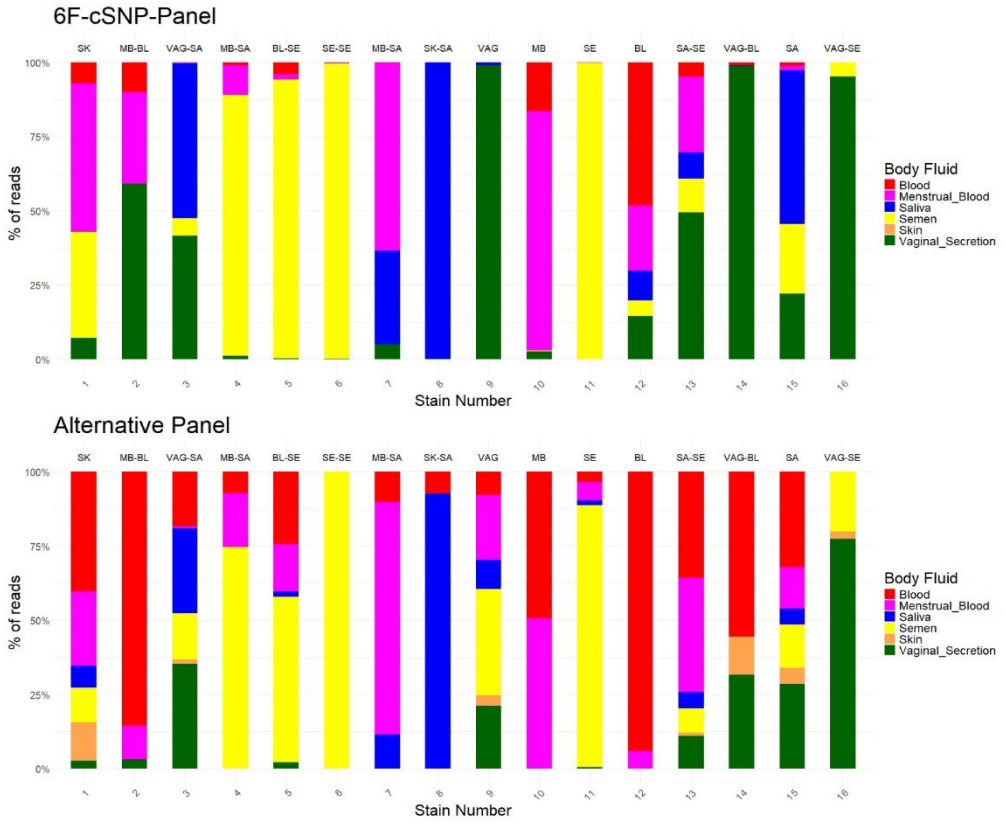
Maximilian Neis [a,*], Theresa Groß [b], Harald Schneider [b], Peter M. Schneider [a,f], Cornelius Courts [a]

Evaluating an Alternative cSNP Panel (Cologne)

- Target amplification from the same cDNA

Stain Nr.	BF/T	6F-cSNP-Panel	alternative cSNP Panel
1	SK	?	?
2	BL-MB	BL-MB	BL-MB
3	SA-VAG	VAG-SA (SE in VAG?)	?
4	MB-SE	MB-SE	MB-SE
5	BL-SE	SE-BL	?
6	SE-SE	SE-SE	SE-SE
7	MB-SA	MB-SA	MB-SA
8	SA-SK	SA	SA-BL
9	VAG	VAG	?
10	MB	MB	MB
11	SE	SE	SE
12	BL	?	MB
13	SA-SE	?	?
14	VAG-BL	VAG	VAG-BL
15	SA	SA(?)	?
16	VAG-SE	VAG-SE	VAG-SE

Composition Analysis of Stains by Body Fluid Percentages

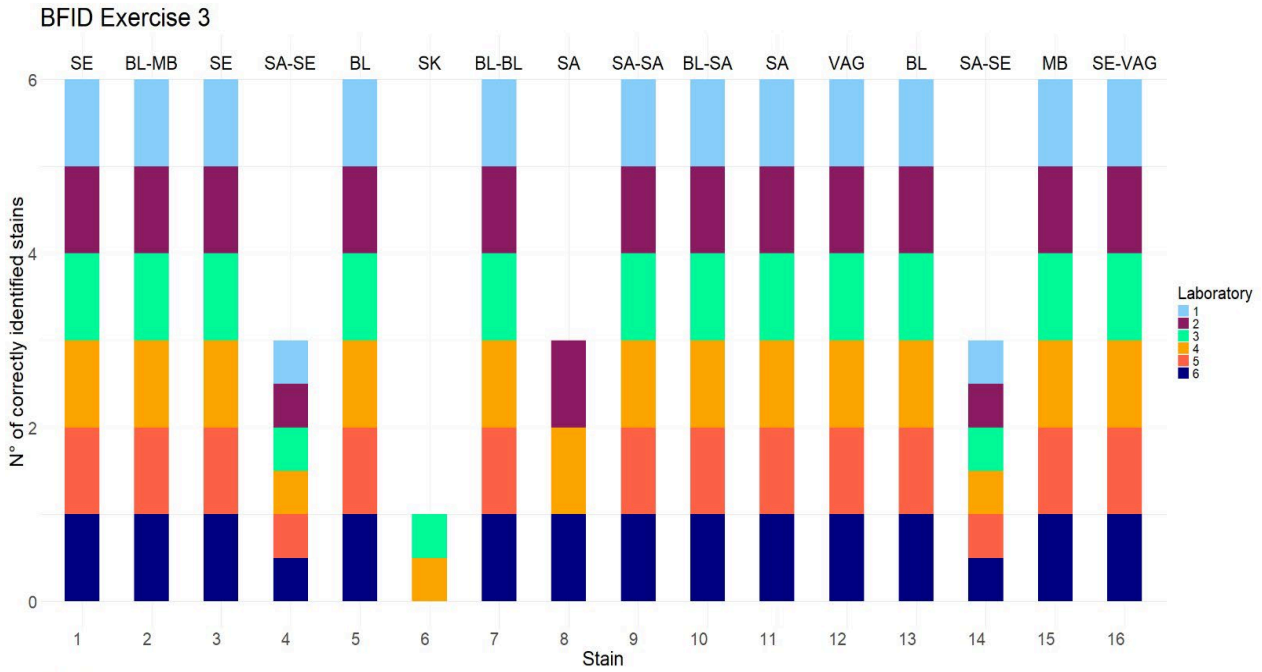


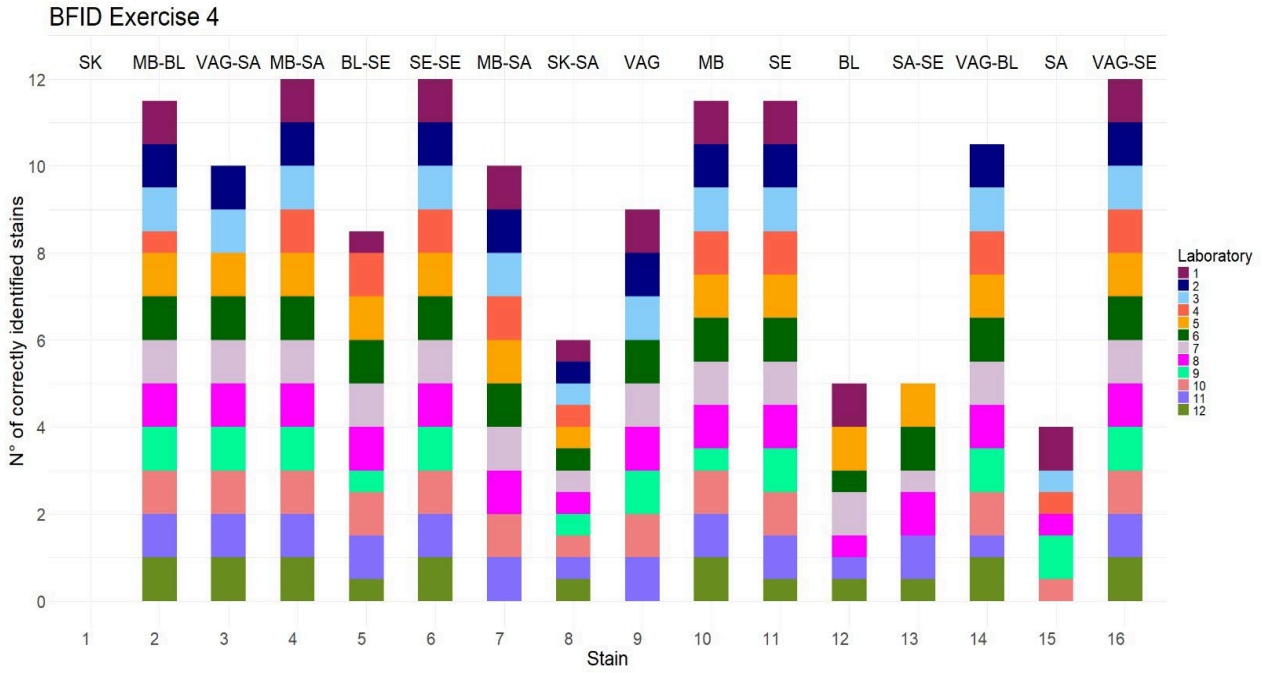

Comparison of the 2 panels
ANK1 BL
CD3G BL
SPTB BL
ALAS2 BL
HBB BL
JAML BL
MMP10 MB
MMP3 MB
LEFTY2 MB
COL6A3 MB
COL12A1 MB
MMP7 MB
STC1 MB
HTN3 SA
PRB4 SA
MUC7 SA
STATH SA
PRH2 SA
PRM1 SE
SEMG2 SE
KLK3 SE
TGM4 SE
SEMG1 SE
PRM2 SE
CYP2B7P VAG
CYP2A6 VAG
MUC22 VAG
ESR1 VAG
MYOZ1 VAG
LCE1C SK
COL17A1 SK
IL37 SK
LORICRIN SK
SERPINA12 SK
% of reads
80
40
1 SK
2 BL-MB
3 SA-VAG
4 MB-SE
5 BL-SE
6 SE-SE
7 MB-SA
8 SA-SK
9 VAG
10 MB
11 SE
12 BL
13 SA-SE
14 BL-VAG
15 SA
16 SE-VAG
Panel 6F Alt.

Feedback from participants

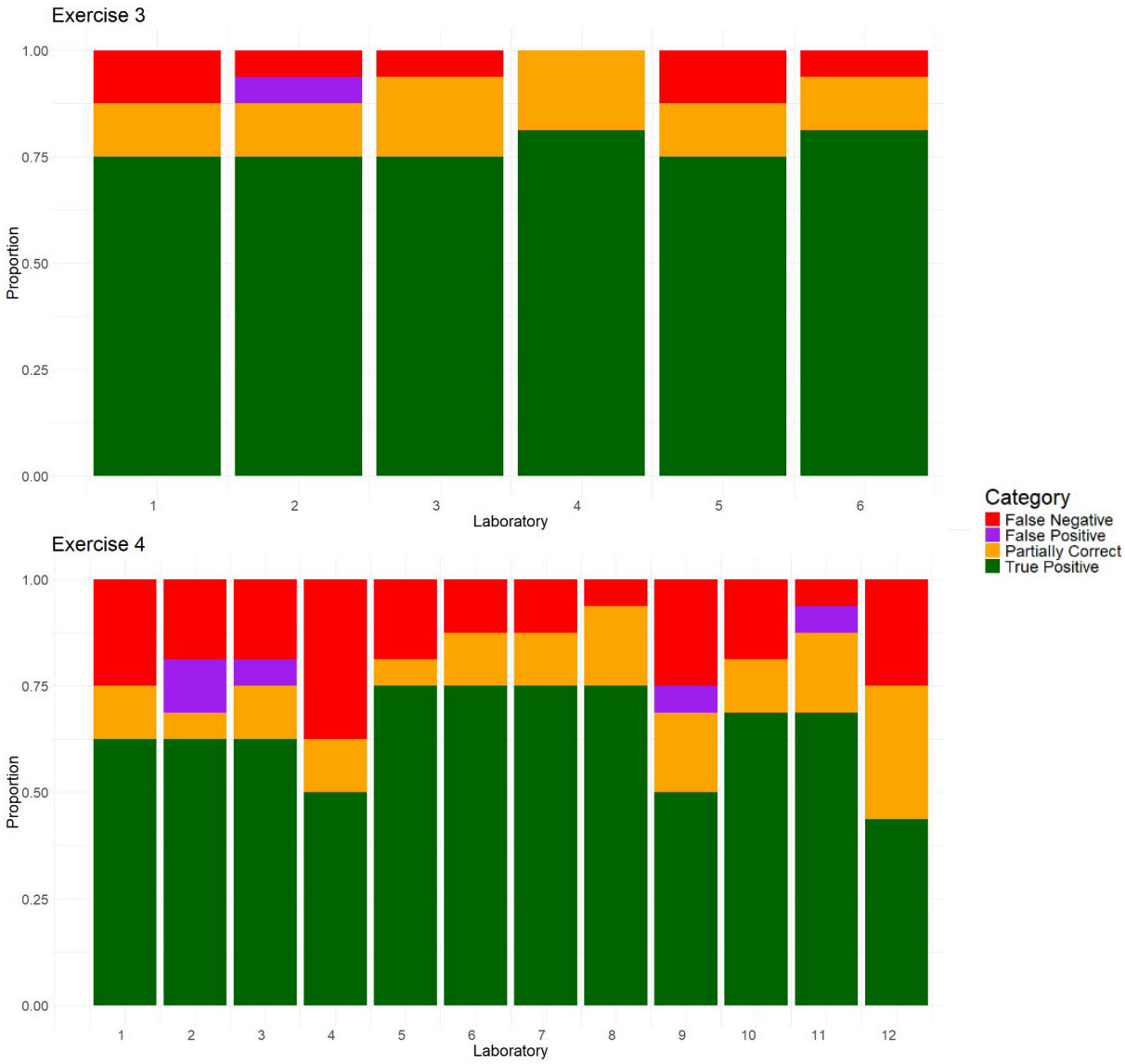
- well organized exercise, great opportunity for us to get some hands on practice on mRNA sequencing, monster task (different labs, different knowledge, different sequencers, different panels, two different exercises - hats off)
- very nice manuscript, really nice work, we are impressed, great job and not an easy task, well structured, super introduction – very nice overview
- too many “detours” > focus on main story
- Main suggestions:
  - performance by laboratory
  - thresholds / cross-reactions / false positives
  - presentation of cSNP results

The number of laboratories that correctly identified one or more body fluid components

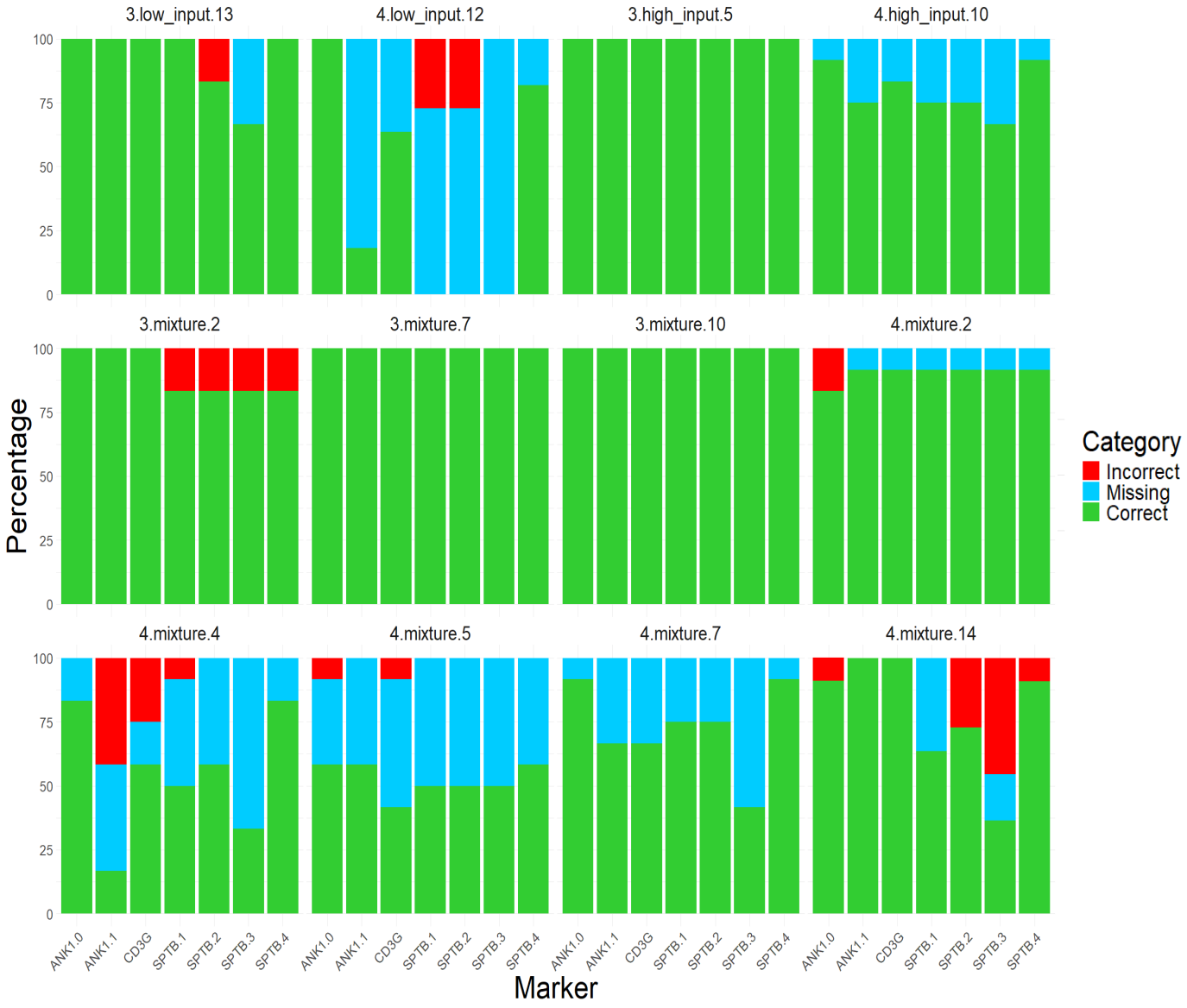




Proportions of false negatives, false positives, partially correctly predicted stains and true positives



Percentage of blood cSNP genotypes found in the respective categories (correct/missing/incorrect) per stain



The number of full cSNP profiles identified by the participating laboratories in each exercise per stain and per body fluid component

		Full cSNP Profile(s)		
Stain Nr.	BF/T	BF/T 1	BF/T 2	Complement
Ex.3-1	SE	3	NA	
Ex.3-2	BL-MB	5	NA	
Ex.3-3	SE	2		
Ex.3-4	SA-SE	0	0	
Ex.3-5	BL	6	NA	
Ex.3-6	SK	NA	NA	
Ex.3-7	BL-BL	6	6	
Ex.3-8	SA	0	NA	
Ex.3-9	SA-SA	3	3	
Ex.3-10	BL-SA	6	0	
Ex.3-11	SA	5	NA	
Ex.3-12	VAG	NA	NA	
Ex.3-13	BL	3	NA	
Ex.3-14	SA-SE			
Ex.3-15	MB	NA	NA	
Ex.3-16	SE-VAG	4	NA	
Ex.4-1	SK	0	NA	
Ex.4-2	BL-MB	9	9	
Ex.4-3	SA-VAG	5	0	
Ex.4-4	MB-SE	6	6	BL: 1
Ex.4-5	BL-SE	5	2	
Ex.4-6	SE-SE	1	1	
Ex.4-7	MB-SA	6	6	VAG:8*, BL: 4
Ex.4-8	SA-SK	7	0	
Ex.4-9	VAG	9*	NA	
Ex.4-10	MB	7	NA	VAG:11*, BL:8
Ex.4-11	SE	0	NA	
Ex.4-12	BL	0	NA	
Ex.4-13	SA-SE	0	0	
Ex.4-14	VAG-BL	11	3	
Ex.4-15	SA	0	NA	
Ex.4-16	VAG-SE	11	6	

Poster presentation at
ISFG congress 2024
in Santiago de Compostela

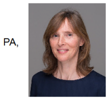
30th Congress of the International Society for Forensic Genetics
SANTIAGO DE COMPOSTELA
9-13 SEPTEMBER 2024

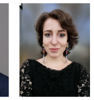
POSTER NR.
145

Body fluid identification and donor association of mock case samples: Results of two EDNAP collaborative exercises

Nadescha Hänggi[1], Antonio Amorim[2,3], Heloisa Afonso Costa[2,4], Jeppe D. Andersen[5], Niels Morling[5], Marie-Louise Kampmann[5], Cornelius Courts[6], Annica Gosch[6], Maximilian Neis[6], Denise Syndercombe Court[7], Federica Giangasparo[7], Ane Elida Fonneløp[8], Helen Johannessen[8], Thorsten Hadrys[9], Walther Parson[10,11], Harald Niederstätter[10], Maja Sidstedt[12], Titia Sijen[13], Margreet van den Berge[13], Erin Hanson[14,15], Jack Ballantyne[14,15], Cordula Haas[1]

[1] Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland
[2] Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal
[3] Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal
[4] Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Lisboa, Portugal
[5] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[6] Institute of Legal Medicine, University Hospital of Cologne, Cologne, Germany
[7] Department of Pharmacy and Forensic Science, King's College London, London, UK
[8] Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway
[9] State Criminal Police Office, Forensic Science Institute, Munich, Germany
[10] Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria
[11] Forensic Science Program, The Pennsylvania State University, University Park, PA, USA
[12] National Forensic Centre, Swedish Police Authority, Linköping, Sweden
[13] Division Biological Traces, Netherlands Forensic Institute, The Hague, The Netherlands
[14] National Center for Forensic Science, Orlando, FL, USA
[15] Department of Chemistry, University of Central Florida, Orlando, FL, USA





Contact
cordula.haas@irm.uzh.ch
nadescha.haenggi@irm.uzh.ch

1. Introduction

- mRNA profiling has emerged as a promising technique for body fluid identification (BFID)
- Coding region SNPs within body fluid specific transcripts facilitate the association between a body fluid and its donor
- Two RNA BFID-cSNP-assays have been developed[1]:
  - the BSS-cSNP-assay for BFID of all forensically relevant body fluids and skin, incl. cSNPs for blood, saliva and semen
  - the 6F-cSNP-assay for BFID incl. cSNPs for blood, saliva, semen, vaginal secretion, menstrual blood, and skin
- Each BFID-cSNP RNA assay comes with a concomitant DNA-cSNP assay for donor genotyping

2. Principle of the Method

Assigning a body fluid to a donor is a three-step-process:

- STR analysis to obtain the number of contributors
- mRNA profiling to determine the body fluid(s) present
- Comparison of the DNA and RNA cSNP genotypes

3. Aims

In two consecutive collaborative exercises organized by the Zurich Institute of Forensic Medicine (ZIFM), both BFID-cSNP assays were evaluated with respect to the following aspects:

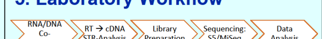
- The robustness and reproducibility across different laboratories using two sets of stains provided by the ZIFM:
  - 6 participating laboratories in exercise 3
  - 12 participating laboratories in exercise 4
- The performance on stains prepared by participants (own stains)
- The effect of different sequencing platforms

4. EDNAP mRNA MPS Exercises

The EDNAP exercises 3 and 4 entailed the analysis of:

- 16 provided stains (single source and/or mixed, Tables 1 + 2)
- 8 own stains (single source and/or mixed) incl. the respective DNA reference profiles (results not shown)
- Participating laboratories were given detailed instructions for stain analysis
- All the data from the participating laboratories was collected and analysed at the ZIFM.

5. Laboratory Workflow



References

1. Hanson, Erin, et al. "Targeted S5 RNA sequencing assay for the identification and direct association of common body fluids with DNA donors in mixtures." *International journal of legal medicine* 137.1 (2023): 13-32.

6. Results

- Sequencing results were compared among participants (Fig. 1)
- Stain compositions were predicted by considering the proportion of body-fluid-specific markers:
  - 13/16 correctly predicted stains in exercise 3
  - 11/16 correctly predicted stains in exercise 4
- Sequencing results were compared across platforms in exercise 4 (Fig. 2)
- The cSNP genotypes were compared to the DNA reference profiles to associate body fluids with their donors (Table 3)

Abbreviations:
BF/T stands for body fluid/tissue, BL for blood, MB for menstrual blood, SA for saliva, SE for semen, VAG for vaginal secretion, and SK for skin.

Stain N°	BF/T	Amount	Stain Provided
1	SE	10 µl	piece of fabric (boxer shorts)
2	BL-MB	1/2 Swab + 25 µl	1/2 swab
3	SE	50 µl	artificial cotton
4	SA-SE	50 µl + 25 µl	part of a T-shirt
5	BL	50 µl	1 swab
6	SK	1 swab	1 swab
7	BL-BL	25 µl + 25 µl	part of a T-shirt
8	SA	Licked plastic spoon	spoon
9	SA-SA	25 µl + 25 µl	1 swab
10	BL-SA	25 µl +25 µl	1 swab
11	SA	50 µl	part of a T-shirt
12	VAG	1/2 swab	1/2 swab
13	BL	Nose bleed on tissue	part of a tissue
14	SA-SE	25 µl + 25 µl	piece of fabric (boxer shorts)
15	MB	1/2 swab	1/2 swab
16	SE-VAG	½ Swab+ 25 µl SE	1/2 swab

Table 1: Composition of the provided stains in exercise 3. The colour code indicates high input stains (dark blue), low input stains (light blue), and mixtures (orange).

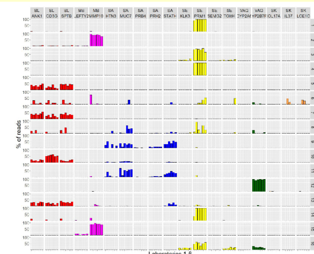


Fig. 1: Sequencing results of all laboratories participating in exercise 3. Markers are depicted as proportions of the total number of reads, and colored in the body-fluid-specific colors (red for BL, pink for MB, blue for SA, yellow for SE, green for VAG, and orange for SK).

Stain N°	BF/T	Amount	Stain Provided
1	SK	1 swab	1 swab
2	MB-BL	1 swab + 25 µl	1/4 swab
3	VAG-SA	1 swab + 25 µl	1/4 swab
4	MB-SA	1 swab + 25 µl	1/4 swab
5	BL-SE	25 µl + 25 µl	part of T-Shirt
6	SE-SE	25 µl + 25 µl	1 swab
7	MB-SA	1 swab + 50 µl	1/4 swab
8	SK-SA	1 swab + 25 µl	1 swab
9	VAG	cotton part of a slip	a piece of it
10	MB	menstrual pad	a part of it
11	SE	50 µl	part of a glove (latex)
12	BL	20 µl	part of a T-Shirt
13	SA-SE	50 µl + 10 µl	artificial cotton
14	VAG-BL	1 swab + 25 µl	1/4 swab
15	SA	50ul	piece of stockings (nylon)
16	VAG-SE	1 swab + 25 µl	1/4 swab

Table 2: Composition of the provided stains in exercise 4. The colour code indicates high input stains (dark blue), low input stains (light blue), and mixtures (orange).

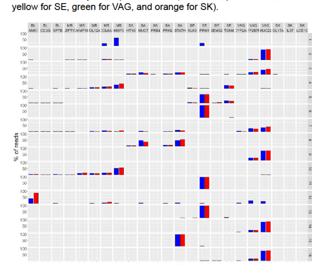


Fig. 2: Sequencing results of one laboratory participating in exercise 4 after stain analysis starting from the same cDNA. Markers are depicted as proportions of the total number of reads.

Ex4_Stain 3	HTN3.0	HTN3.1	HTN3.2	MUC7	PRB4	PRH2	HTN3.4	CYP2A6	MUC22.0	MUC22.1	MUC22.2	MUC22.3	MUC22.4	MUC22.5	MUC22.6	MUC22.7	MUC22.8	MUC22.9
Donor genotype 1	TC	T/T	C/C	C/T	G/G	C/C	C/C	C/C	GT	G/G	T/T	AA/GA	A/G	A/A	CT/TA/TC	C/T	T/A	T/T
Donor genotype 2	CT/CC	C/C	T/C	C/C	G/C	C/C	C/C	C/C	TC/GT	T/G	C/T	AA/GG	A/G	A/G	CA/CC	C/C	A/C	T/C
Lab7 MiSeq - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab7 MiSeq - Read Counts	14464	14464\14464	14464\14464	21108\17201	2418\2418	8824\8824	14464\14464		8858\8081	8858\8081	8858\8081	27716\18812	27716\18812	27716\18812	30115	30115\30115	30115\30115	8858\8081
Lab8 S5 - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CGG/CAA	G/A	G/A	ACA	C/C	A/A	C/T
Lab8 S5 - Read Counts	33244	33244\33244	33244\33244	47296\36046	3064\3064	33033\33033	33244\33244		11954\10099	11954\10099	11954\10099	57295\51908	57295\51908	57295\51908	70883	70883\70883	70883\70883	11954\10099
Lab9 S5 - Genotype	TC/TC	T/T	C/C	C/T	G/G	C/C	C/C		TC/GT	T/G	C/T	CAA/CGG	A/G	A/G	ACA	C/C	A/A	C/T
Lab9 S5 - Read Counts	2536	2536\2536	2536\2536	68781\41249	298\298	60076\60076	2536\2536		6614\3587	6614\3587	6614\3587	30194\23648	30194\23648	30194\23648	36149	36149\36149	36149\36149	6614\3587

Table 3: Example for donor association of a provided stain (stain 3, exercise 4, SA-VAG). DNA cSNP genotypes (donors 1 and 2) in the markers of interest are compared to the RNA cSNP genotypes (results of three exemplary laboratories, no results reported for CYP2A6). The matching donors for the identified body fluids are marked in light blue.

7. Conclusion

- Very promising results, i.e. the majority of the stains in both exercises were typed correctly
- Results were quite consistent across different laboratories
- The 6F-BFID-cSNP assay performed well on both sequencing platforms
- Laboratories with limited RNA experience also reported good results
- For body fluid identification difficulties arise because of various misleading reads:
  - If a stain has low reads in general, reads in target markers are low as well → harder to interpret
- Performance of the association of a body fluid and a donor is dependent on how many markers are detected per body fluid:
  - Not every marker includes a cSNP with the power to exclude a donor

Summary

- Overall promising results with both assays
- Some participants did not (fully) follow the recommendations
- Laboratories with limited RNA experience also achieved good results
- Results were quite consistent across different laboratories
- The cSNP panels performed well on both sequencing platforms
- Comparison with Cologne cSNP panel (30 body fluid markers, 70 cSNPs)

➢ Manuscript will be submitted to FSI Genetics soon

Acknowledgements

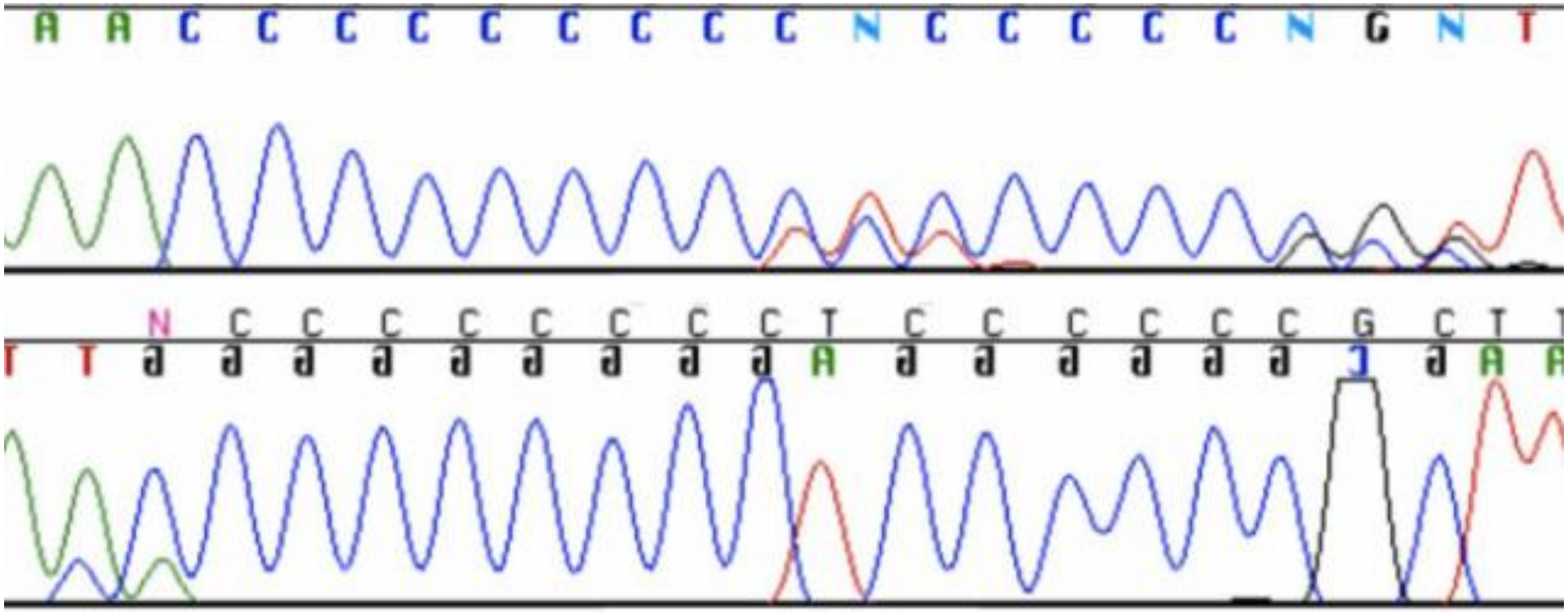
University of Zurich:
Research team 2024

Nadescha Hänggi

University of Central Florida:
Jack Ballantyne, Erin Hanson

Thermofisher:
Robert Lagace, Chantal Roth

EDNAP Meeting, Barcelona, Spain, Nov 12 2024



mtDNA heteroplasmy exercise

Walther Parson[1,2], Lena Ewers[1], Gabriela Huber[1], Nicole Huber[1], Claudia Wöss[1], Arne Dür[3]
[1] Institute of Legal Medicine, Medical University of Innsbruck, Austria
[2] Forensic Science Program, Penn State University, PA, USA
[3] Institute of Mathematics, University of Innsbruck, Austria

EMPOP

Mitotypes from DNA extracts provided by Innsbruck

CS-1: 007 (TFS 1707017)

73G 152C 199C 204C 207A 250C 263G 315.1C 460C **573.XC** 750G 1438G 1719A **2413Y** 2706G 4529T 4769G 6293C 7028T 8251A 8860G 9438A 10034C 10238C 10398G 11719A 12501A 12705T 13780G 14766T 15043A 15326G 15758G 15924G 16129A 16223T 16391A 16519C

CS-2: 9947A (Promega 18961603)

93G 195C 214G 263G **309.XC** 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C

CS-3: reference sample volunteer 1 (U8b1b1)

73G 152C 195C 263G **309.XC** 315.1C **573.XC** 750G 1438G 2706G 3480G 4769G 5165T 7028T 8860G 9055A 9698C 11467G 11719A 12308G 12372A 14053G 14167T 14766T 15326G 15924G 16183C 16189C **16193.XC** 16234T 16324C 16519C

CS-4: reference sample volunteer 2 (H*)

16093Y 16311C 16519C 73G 263G 315.1C 750G 1438G 4769G 8860G **12483Y** 15326G

CS-5: reference sample volunteer 3 (H1b1)

16189C **16193.XC** 16356C 16362C 16519C **234R** 263G 315.1C 523del 524del **573.XC** 750G 1438G 3010A 3796G 4769G 8860G 15326G

Challenges

Some laboratories sent results back with significant delay (covid, etc …)

Reporting of heteroplasmy not uniform

Huge diversity of individual platforms (= instrumentation + software) introduced additional variability

Developed agnostic software “MPSAligner” to remove variability due to different software (versions) used

Staff change at GMI that required a new person to pick up the work

Platforms and settings

Overview Sanger

Table S1a: Sanger data						
Lab	**Range**	**Kit/Protocol**	**Instrument**	**Seq. chemistry**	**Anal. software**	**AT/IT***
0	CR	Advantage 2 Polymerase Mix PCR, CS	3500XL	BDT CS v1.1	Sequencher v5.1	10%*
5	HVS-I/II	Taq Gold PCR, CS	3500	BDT CS v1.1	SeqScape v3	n.a.
6	HVS-I/II	Preformulated PCR Optimization Kit Buffer B PCR, CS	3130xL	BDT CS v1.1	Sequencher v5.1	PHP: 10%, 10-79.9%
7	CR	Advantage 2 Polymerase Mix Mini PCR or Midi PCR, CS	3130Avant	BDT CS v1.1	Sequencher v5.4.6	10% 10-20%
21	CR	AmpliTaq Gold Polymerase PCR, CS	3500XL	BDT CS v1.1	Sequencher v3	?
22	CR	Qiagen Multiplex PCR Kit PCR, CS	3130	BDT CS v3.1	Sequencing Analysis v6.0	?
23	HVS-I/II	Taq DNA Polymerase or Platinum® Taq DNA Polymerase PCR, CS	3130	BDT CS v3.1	Sequencing Analysis v? BioEdit MEGA6	?
24	CR	Qiagen Multiplex PCR Kit PCR, CS	3130xL	BDT CS v3.1	SeqScape v2.6	?
* Note that Sanger Sequencing is semi-quantitative						

Platforms and settings

Overview Ion Torrent

Table S1b: Ion Torrent data							
Lab	**Range**	**Kit/Protocol**	**Instrument**	**Seq. chemistry**	**Anal. software**	**Alignment**	**AT/IT**
0	MtG	PID mtDNA WGP AB MAN0017771	Ion S5	Ion 530	TSS v5.10.0 Con. v2.1/2.3 MVC v1.09b	HID GT v2.1/2.3	10%
1	MtG	PID mtDNA WGP AB MAN0017771	Ion S5	Ion 530	TSS v5.10 Con. v2.1 IGV/MVC	HID GT? or TMAP?	n.a.
2	MtG	PID mtDNA WGP MAN0015831	Ion S5	Ion 530	TSS v5.10.1 Con. v2.2 IGV	HID GT v2.1	20 reads 96% confirmed call variant strand bias ≤ 0.6 30% del, 20% ins, 10% PHP
3	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10.1 Con. v2.1 IGV	TMAP	20 reads, 20 reads PHP: 20 reads, 10%
4	MtG	PID mtDNA WGP AB MAN0017771	Ion S5	Ion 530	TSS v5.10.1 Con. v2.2 IGV v?	TMAP	96 PHP: 10
5	CR	PID mtDNA WGP AB MAN0015831	Ion S5	Ion 520	TSS v5.12.2 TVC v5.2.2.41 IGV	TMAP	n.a.
6	MtG	PID mtDNA WGP AB MAN0017770	Ion S5	Ion 530	TSS v5.10 IGV v2.3.94/MVC	HID GT v2.1	20 reads, 20 reads PHP: 10%, 10-79.9%
10	MtG	Precision ID Ion S5 XL AB MAN0015831	Ion S5 XL	Ion 530	TSS v5.10.0 TVC v5.10.0.18 Con. v2.1 MVC v1.09b	TMAP? HID GT v2.1	20 reads, 20 reads 30% del, 20% ins, 10% PHP (default thresholds Converge)
12	CR	PID mtDNA CRP AB MAN0017772	Ion S5 XL	Ion 530	TSS v5.10.0 Con. v2.2 IGV v.2.7.1	HID GT v2.2	20 reads 96% confirmed call 30% del, 20% ins PHP: 5%[a]
13	MtG	PID mtDNA WGP MAN0015830	Ion PGM	Ion 318	TSS v5.2.2 TVC v5.2.1.38 GM-HTS (demo v1.0.17.1266)[b] IGV v2.3.72[b]	TMAP?	20 reads, 10% 40% del, 30% ins, 10% PHP
14	MtG	Precision ID Ion S5 AB MAN0017770 B.0	Ion S5 XL	Ion 520	TSS v5.10.1 TVC v5.10.1.19 IGV coustom Excel workbook Con. v? (LHP levels only)	TMAP?	20 reads, 100 reads PHP: 5%
15	MtG	Precision ID Ion Chef & Ion S5 MAN0017770 Rev B	Ion S5 XL	Ion 530	TSS v5.12 Con. v2.2	HID GT v2.3	PHP: 20 reads, 10%
17	MtG	Early Access Mito Kit v.1 AB MAN0017771	Ion S5	Ion 530	TSS v5.10.1 IGV	TMAP	50 reads 8% (minor variant > 15 reads)

[a] ... for this exercise only

[b] ... only positions with frequency < 90% and allele call as "heterozygous" and some positions usually not reported (e.g. 309, 315, 573, 16183-16189) were reanalyzed using IGV and GeneMarker HTS

AB MAN0017770 ... detailed Pcrotocol S5 for both manuel ans automatic approach (2018-2023)

AB MAN0017771 ... quick reference for manual approach based on AB MAN0017770

AB MAN0017772 ... quick reference for automatic approach based on AB MAN0017770

MAN0015830 ... PGM protocol

AB MAN0015831 ... detailed Pcrotocol S5 for both manuel ans automatic approach (2016)

Platforms and settings

Overview Ion Torrent

Table S1b: Ion Torrent data							
Lab	**Range**	**Kit/Protocol**	**Instrument**	**Seq. chemistry**	**Anal. software**	**Alignment**	**AT/IT**
0	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10.0	HID GT v2.1/2.3	10%
		AB MAN0017771			Con. v2.1/2.3		
					MVC v1.09b		
1	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10	HID GT?	n.a.
		AB MAN0017771			Con. v2.1	or	
					IGV/MVC	TMAP?	
2	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10.1	HID GT v2.1	20 reads
		MAN0015831			Con. v2.2		96% confirmed call
					IGV		variant strand bias ≤ 0.6
							30% del, 20% ins, 10% PHP
3	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10.1	TMAP	20 reads, 20 reads
					Con. v2.1		PHP: 20 reads, 10%
					IGV		
4	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10.1	TMAP	96
		AB MAN0017771			Con. v2.2		PHP: 10
					IGV v?		
5	CR	PID mtDNA WGP	Ion S5	Ion 520	TSS v5.12.2	TMAP	n.a.
		AB MAN0015831			TVC v5.2.2.41		
					IGV		
6	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10	HID GT v2.1	20 reads, 20 reads
		AB MAN0017770			IGV v2.3.94/MVC		PHP: 10%, 10-79.9%

Platforms and settings

Overview Ion Torrent

Table S1b: Ion Torrent data							
Lab	**Range**	**Kit/Protocol**	**Instrument**	**Seq. chemistry**	**Anal. software**	**Alignment**	**AT/IT**
6	MtG	PID mtDNA WGP	Ion S5	Ion 530	TSS v5.10	HID GT v2.1	20 reads, 20 reads
		AB MAN0017770			IGV v2.3.94/MVC		PHP: 10%, 10-79.9%
10	MtG	Precision ID Ion S5 XL	Ion S5 XL	Ion 530	TSS v5.10.0	TMAP?	20 reads, 20 reads
		AB MAN0015831			TVC v5.10.0.18	HID GT v2.1	30% del, 20% ins, 10% PHP
					Con. v2.1		(default thresholds Converge)
					MVC v1.09b		
12	CR	PID mtDNA CRP	Ion S5 XL	Ion 530	TSS v5.10.0	HID GT v2.2	20 reads
		AB MAN0017772			Con. v2.2		96% confirmed call
					IGV v.2.7.1		30% del, 20% ins
							PHP: 5%[a]
13	MtG	PID mtDNA WGP	Ion PGM	Ion 318	TSS v5.2.2	TMAP?	20 reads, 10%
		MAN0015830			TVC v5.2.1.38		40% del, 30% ins, 10% PHP
					GM-HTS (demo v1.0.17.1266)[b]		
					IGV v2.3.72[b]		
14	MtG	Precision ID Ion S5	Ion S5 XL	Ion 520	TSS v5.10.1	TMAP?	20 reads, 100 reads
		AB MAN0017770 B.0			TVC v5.10.1.19		PHP: 5%
					IGV		
					coustom Excel workbook		
					Con. v? (LHP levels only)		
15	MtG	Precision ID Ion Chef & Ion S5	Ion S5 XL	Ion 530	TSS v5.12	HID GT v2.3	PHP: 20 reads, 10%
		MAN0017770 Rev B			Con. v2.2		
17	MtG	Early Access Mito Kit v.1	Ion S5	Ion 530	TSS v5.10.1	TMAP	50 reads
		AB MAN0017771			IGV		8% (minor variant >15 reads)

Platforms and settings

Overview Illumina

Table S1c: Illumina data							
Lab	**Range**	**Kit/Protocol**	**Instrument**	**Seq. chemistry**	**Anal. software**	**Alignment**	**AT/IT**
7	CR	custom kit based on Mini-PCR with Nextera XT	MiSeq FGx	v2 (2x150 bp)	MRS v2.5.1.3	BWA-MEM	1%
		Mini PCR			BS-mtDNA Var. Proc. v1.0		5%
					mtDNA Variant Analyzer IGV		Q-score ≥ 30; 30 reads
8	MtG	myBaits Mito and NEBNext Ultra II	MiSeq FGx	v2 (2x150 bp)	IGV	BWA-MEM	>10%[a]
		Capture					
9	MtG	KAPA HyperPlus and myBaits Custom RNA-Seq	MiSeq FGx	v3 (2x300 bp)	MRS v2.5.1.3	CLC GW v12	100 reads
		Mito Capture			CLC GW v12/AQME		5%
11	MtG	KAPA HyperPlus (KR1145 – v5.19)	MiSeq FGx	v3 (2x300 bp)	MRS v2.5.1	BWA-Hash	45 reads
		with PID Primer Panels (A+B)			GM-HTS		20% (minor variant ≥ 40 reads)
16	CR	MPS mito-mini,	MiSeq FGx	v3 (2x300 bp)	FDSTools v1.2.11	FDSTools v1.2.11	30 reads
		in-house design (adjusted version of Sanger Mitominis)					3%[b]
18	CR	ForenSeq mtDNA Control Region Kit	MiSeq FGx	v3 (2x300 bp)	MRS v2.5.1.3	UAS v2.5.0	10%
		VD2019001-A			UAS v2.5.0		Q-score ≥ 30; 64 reads
19	CR	Promega PowerSeq CRM Nested System	MiSeq FGx	v3 (2x300 bp)	GM-HTS v1.2.2	CLC GW/CLC Bio v12.0.3	10 reads
							10%
							Q-score ≥ 30
20	MtG	Promega PowerSeq WGM	MiSeq	na (1101)	MRS v2.6.2.3	BWA-Hash	2%
		PowerSeq Systems Prototype			GM-HTS v?		

[a] … >10% MAF for PHP for this exercise only

[b] … 3% of highest within locus allele (based on per fragment haplotype frequency and on position frequency) ignoring the read-counts of singletons for the calculation

Results DNA extracts – Point Heteroplasmy

Sanger

method	lab	2413Y (%T)	7861Y (%C)	16093Y (%C)	12483Y (%T)	234R (%G)
Sanger Sequencing	lab 0	n.a.	n.a.	Y	n.a.	R
	lab 5	n.a.	n.a.	Y	n.a.	R
	lab 6	n.a.	n.a.	Y	n.a.	R
	lab 7	n.a.	n.a.	Y	n.a.	R
	lab 21	n.a.	n.a.	Y	n.a.	R
	lab 22	n.a.	n.a.	C	n.a.	R
	lab 23	n.a.	n.a.	C	n.a.	R
	lab 24	n.a.	n.a.	Y	n.a.	R

Results DNA extracts - Heteroplasmy

Evaluated data in two forms

1) reported results

2) reviewed results

Review: MPSAligner accepts bam files and performs phylogenetic alignment according to ISFG guidelines

=> alignment specific to human mtDNA

MPSAligner is based on the EMPOP query engine SAM2 (Huber et al 2018) that uses rCRS-coded mitotypes/fasta-like data as source

Results DNA extracts – Point Heteroplasmy

Conclusions

Reported and reviewed levels of PHP comparable to each other

Little variation between platforms and instruments

Keep in mind that PHP levels can be affected by primer sequences, which needs to be taken into account in data analysis

Results DNA extracts

Length heteroplasmy (LHP)

Participants were asked to determine and report LHP according to their established guidelines

Most labs reported the **dominant variant** (= major molecule) as recommended by revised ISFG guidelines (Parson et al 2014)

Results – DNA extracts

Reporting of LHP in DNA extracts not concordant between technologies and labs

In total we discern three sources of variation: **technological, software** and **interpretational**

MPSAligner removes individual user settings and variation in software but still results in **LHP differences** between labs

It maybe difficult if not impossible to achieve consistent results across labs (see also Sturk-Andreaggi et al 2020)

Needs to be reflected in interpretation guidelines

Will guide interpretation of results in hair shafts (follow up study)

Way forward

Share tables with participants to confirm data and add missing information

work on manuscript on DNA extracts

evaluate hair results based on platform-specific findings of this study

Quo vadis, BGA?

A collaborative EDNAP exercise on estimating biogeographic ancestry from the DNA of unknown samples

Final update

Organized by:

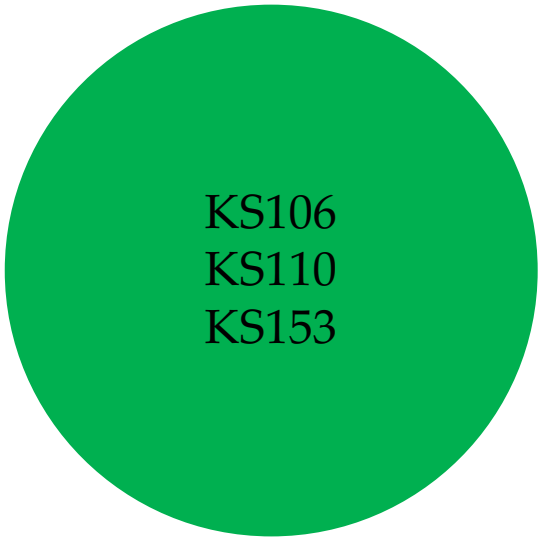
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

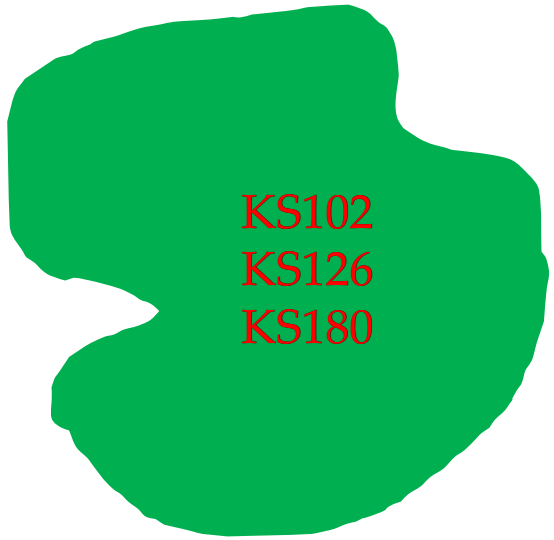
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Samples

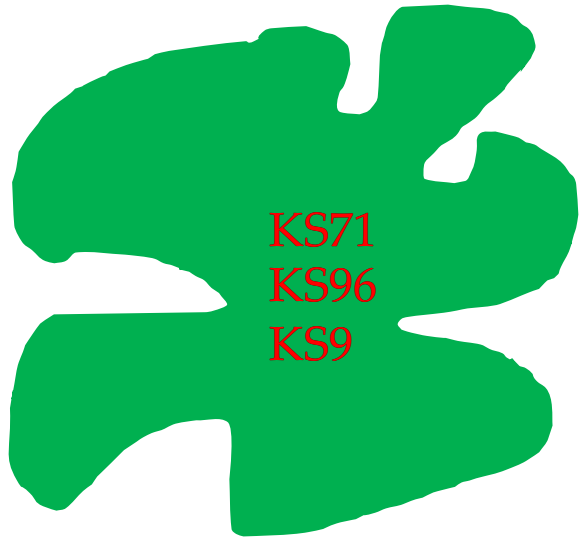
Level 1 – „easy“ samples



Level 2 – samples with similar genetic pattern

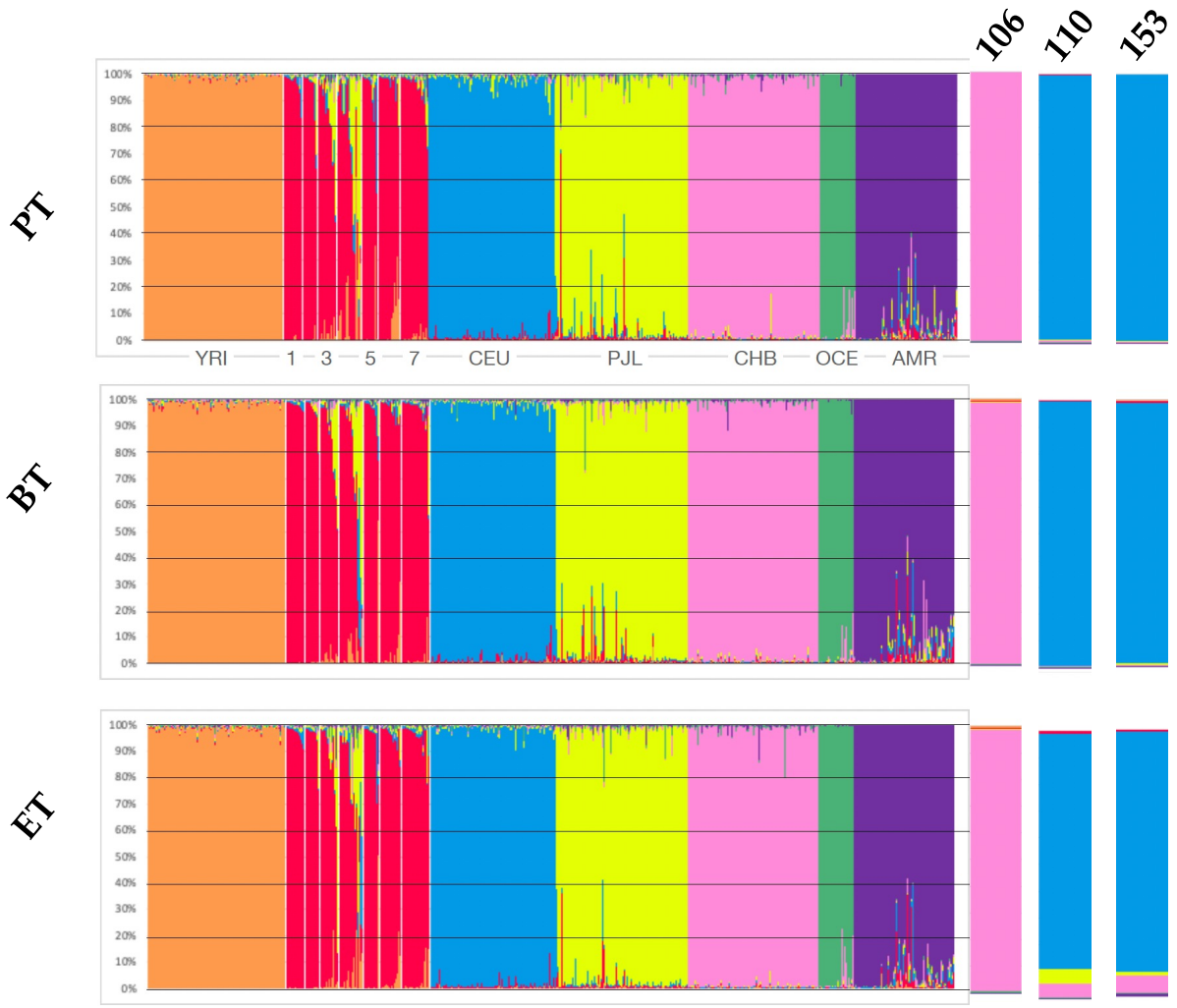


Level 3 – „advanced“ samples



Which individual's data did you find the most difficult to analyse?

KS106, KS110, KS153 - admixture analysis - STRUCTURE



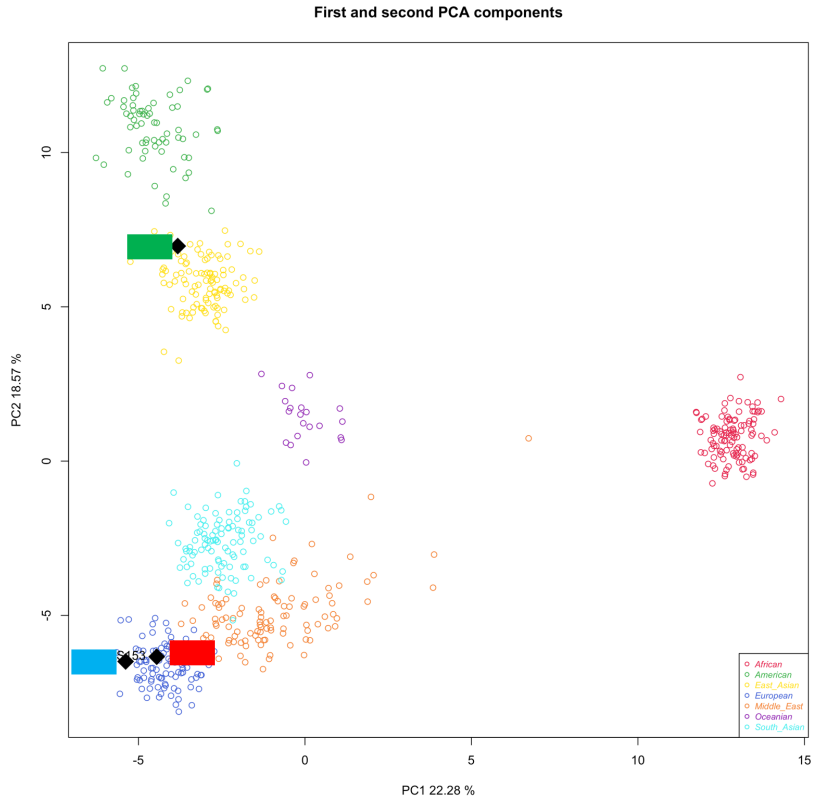
	106			110			153		
	PT	BT	ET	PT	BT	ET	PT	BT	ET
AFR	0.001	0.009	0.007	0.001	0.001	0.001	0.001	0.005	0.001
ME	0.001	0.002	0.003	0.006	0.005	0.009	0.003	0.007	0.008
EUR	0.001	0.001	0.001	0.977	0.987	0.880	0.988	0.975	0.900
SAS	0.001	0.002	0.002	0.005	0.002	0.056	0.004	0.007	0.013
EAS	0.991	0.980	0.982	0.006	0.004	0.051	0.001	0.002	0.066
OCE	0.003	0.003	0.005	0.002	0.001	0.002	0.002	0.002	0.003
AME	0.003	0.003	0.001	0.002	0.001	0.001	0.001	0.002	0.009

KS106, KS110, KS153 - admixture analysis - CONVERGE

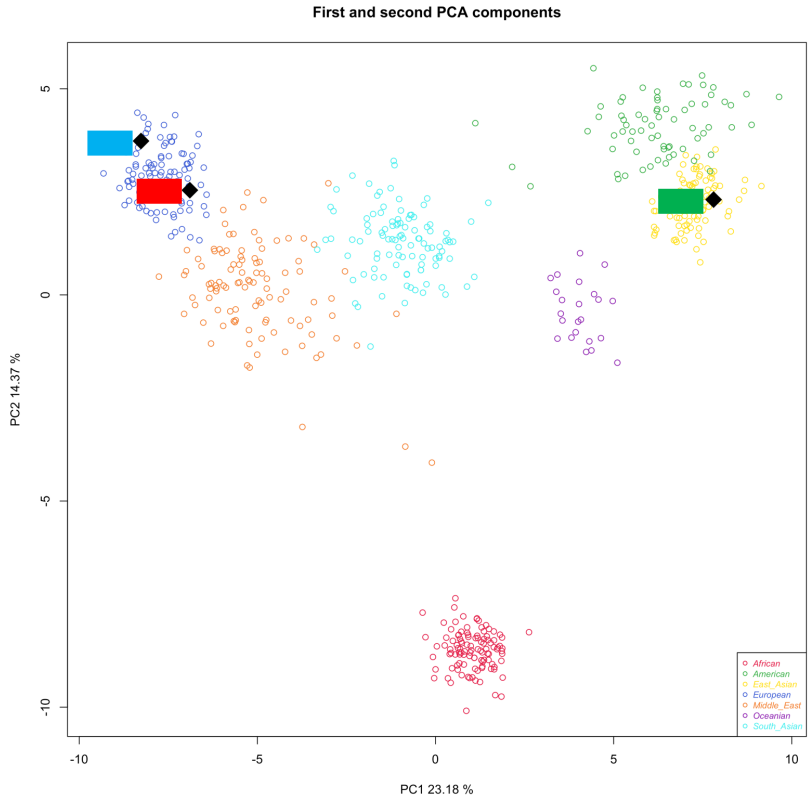
PT only

	106		110		153	
	TFS	MAC	TFS	MAC	TFS	MAC
AFR	0.000	0.000	0.000	0.000	0.000	0.000
ME	0.000	0.000	0.001	0.001	0.001	0.000
EUR	0.000	0.000	0.999	0.999	0.999	0.913
SAS	0.000	0.000	0.000	0.000	0.000	0.087
EAS	0.995	0.968	0.000	0.000	0.000	0.000
OCE	0.005	0.032	0.000	0.000	0.000	0.000
AME	0.000	0.000	0.000	0.000	0.000	0.000

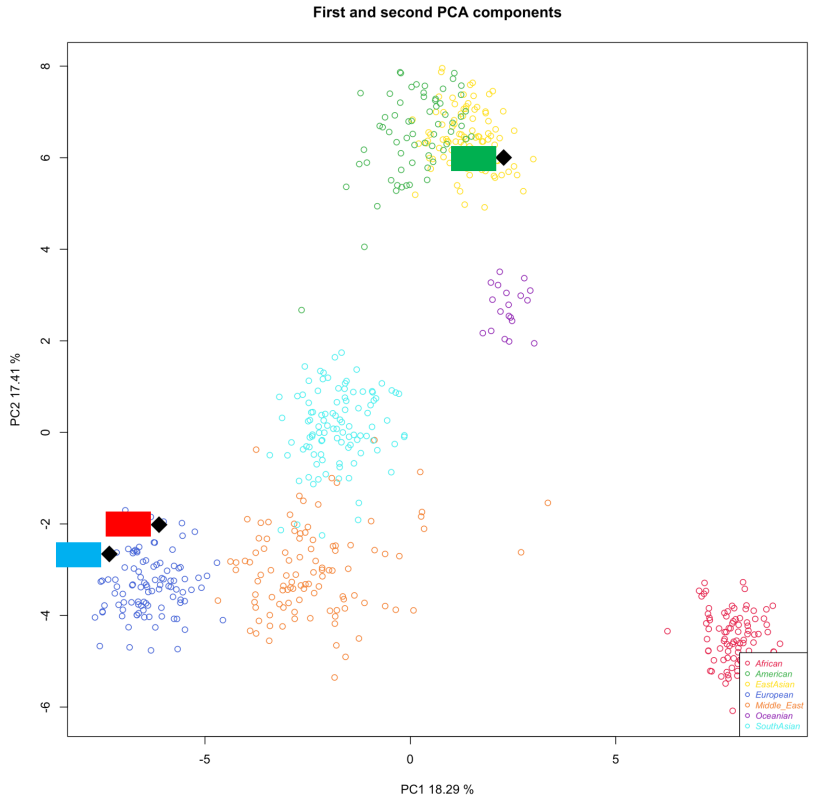
KS106, KS110, KS153 - PCA



PT



BT



ET

KS106, KS110, KS153 - GenoGeographer

106		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
All populations are rejected		
BT		
population	*z-score*	*p-value*
East Asia	-1.33	0.908
ET		
population	*z-score*	*p-value*
East Asia	0.736	0.231

110		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
Europe	-0.353	0.638
BT		
population	*z-score*	*p-value*
Europe	0.509	0.305
ET		
population	*z-score*	*p-value*
All populations rejected		

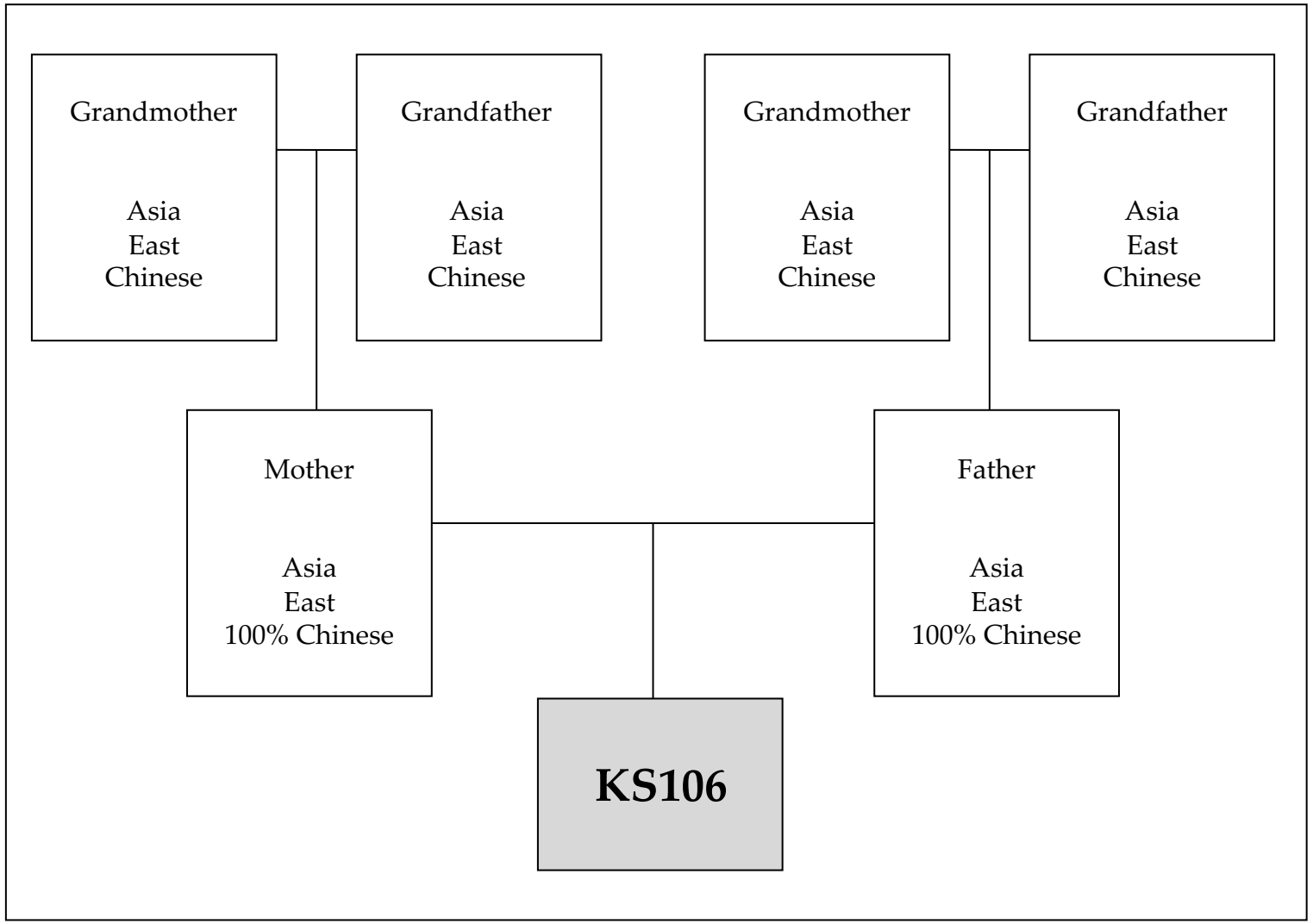
153		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
Europe	-1.159	0.877
BT		
population	*z-score*	*p-value*
Europe	-0.283	0.611
ET		
population	*z-score*	*p-value*
Europe	1.633	0.051

KS106, KS110, KS153 - extra markers

			106	110	153
			p-value		
eye colour		*blue*	0,000	0,099	**0,903**
		intermediate	0,002	0,134	0,074
		brown	**0,998**	**0,767**	0,023
hair	colour	*blond*	0,001	**0,362**	**0,319**
		brown	0,122	**0,521**	**0,605**
		red	0,000	0,002	0,007
		black	**0,877**	0,116	0,068
	shade	*light*	0,002	**0,742**	**0,812**
		dark	**0,998**	0,258	0,188
skin colour		*very pale*	0,000	0,008	0,078
		pale	0,000	**0,304**	**0,563**
		intermediate	**0,965**	**0,681**	**0,369**
		dark	0,035	0,007	0,000
		dark to black	0,000	0,000	0,000

	106	110	153
mtDNA	F1e3	H2a1c	U4c1
87 ET Y-SNPs	O-M119	♀	R-P312
116 PT Y-SNPs	O-M119	♀	R-U152
16 ET X-SNPs	East Asian specific X-chromosome	European specific X-chromosomes	European specific X-chromosome

KS106 - ancestry results



STRUCTURE

	PT	BT	ET
AFR	**0.001**	**0.009**	**0.007**
ME	**0.001**	**0.002**	**0.003**
EUR	**0.001**	**0.001**	**0.001**
SAS	**0.001**	**0.002**	**0.002**
EAS	**0.991**	**0.980**	**0.982**
OCE	**0.003**	**0.003**	**0.005**
AME	**0.003**	**0.003**	**0.001**

mt: F1e3
Y: O-M119
X: EA

KS106 - ancestry results

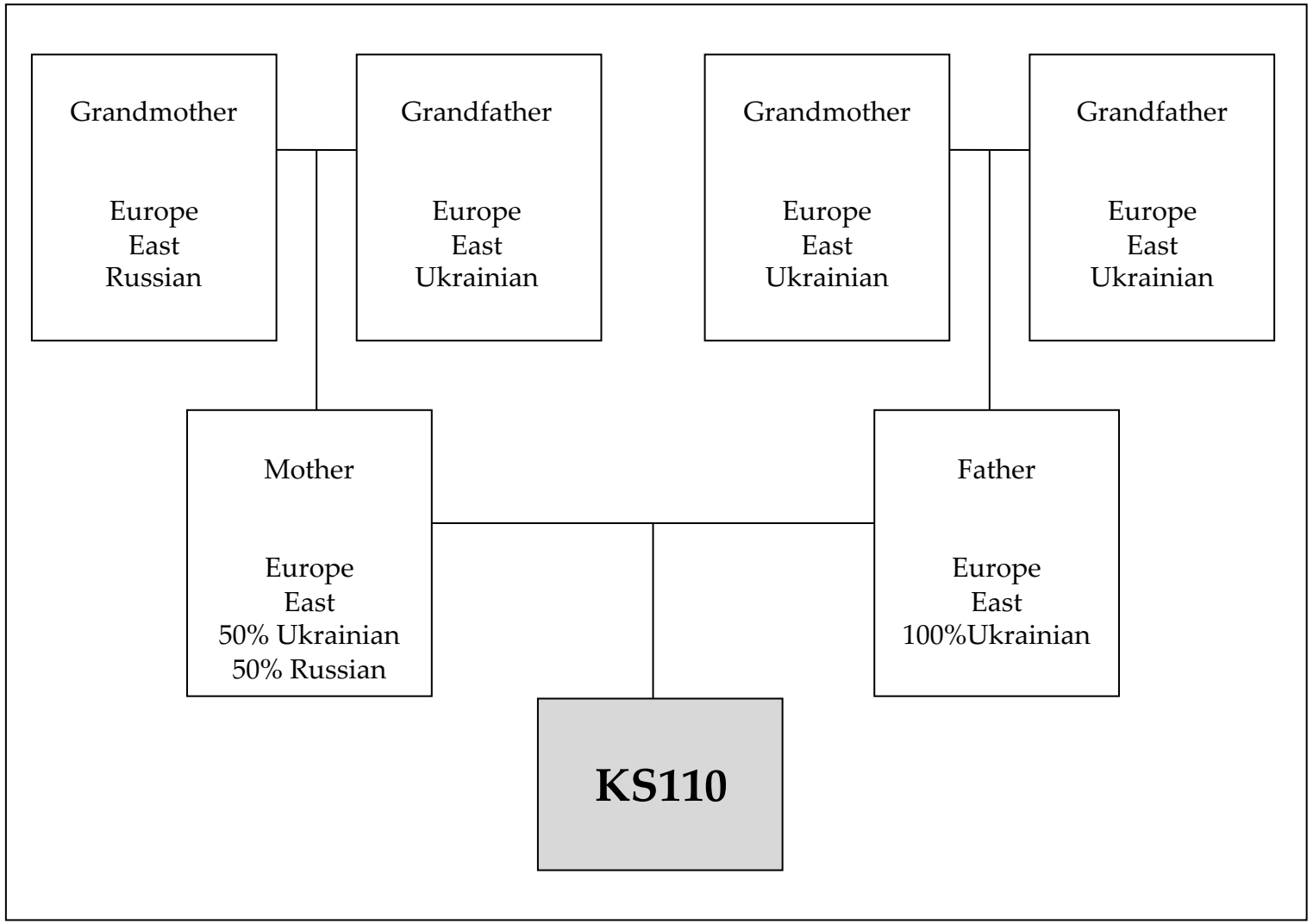
colour code more "direct" less "direct"

All assigned to East Asia (8 labs used all markers, *4 only biparental)*, examples:

- „individual of EAS origin"
- „most likely an East Asian biogeographic origin"
- *„ancestry from East Asia"*
- *„is most likely of East Asian ancestry"*

- *„biogeographical ancestry being East Asia, or a population with East Asian ancestry"*
- *„more likely if the investigated DNA comes from an individual from the East Asian reference population than if the investigated DNA comes from an individual from the Native American, South Asian, Oceanian, Middle Eastern, European, or African reference populations"*
- „biogeographical ancestry results are in line with ancestry in East Asia. The regions considered to have low probability are ME, America, Europe, Africa and South Asia"

KS110 - ancestry results

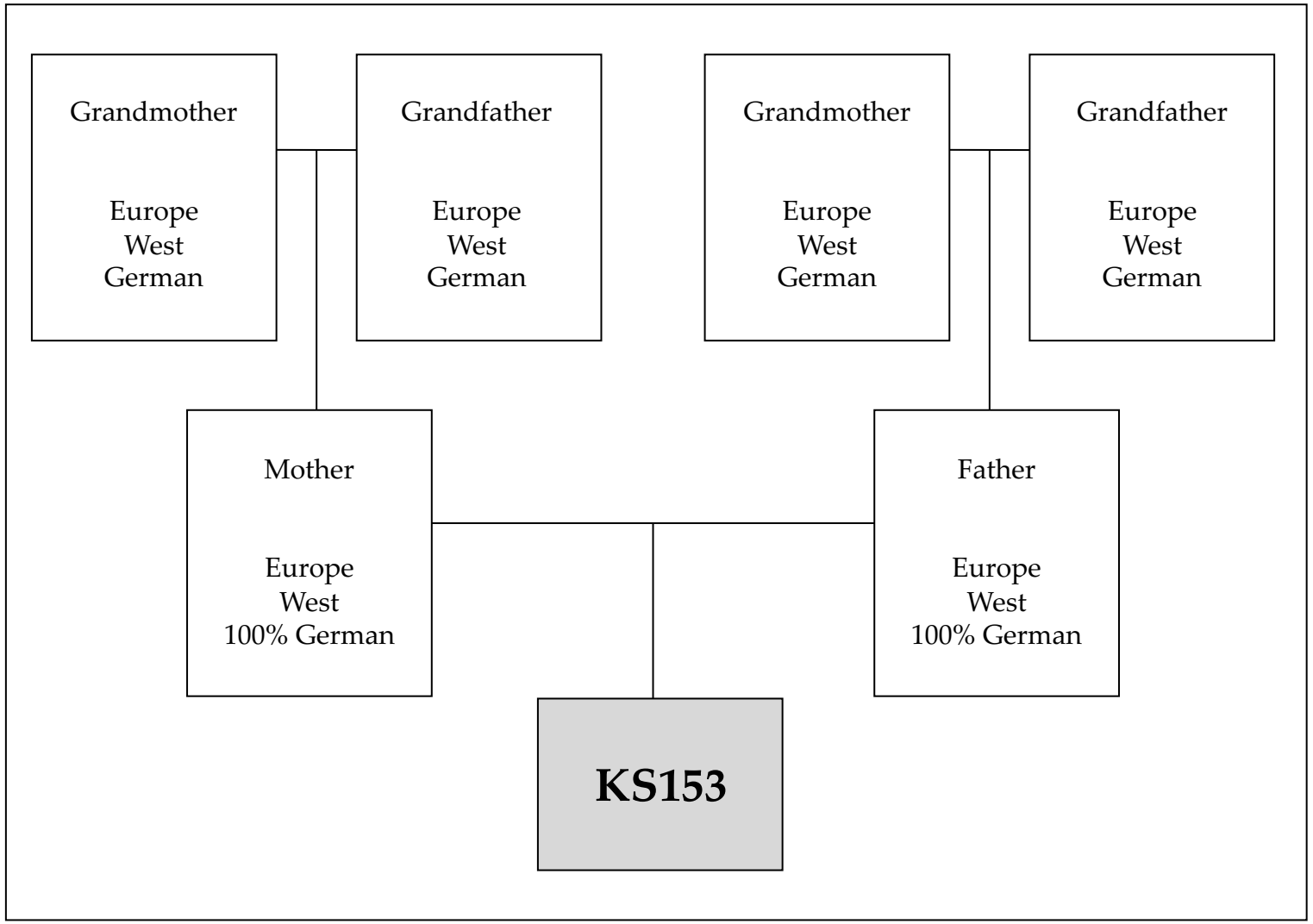


STRUCTURE

	PT	BT	ET
AFR	**0.001**	**0.001**	**0.001**
ME	**0.006**	**0.005**	**0.009**
EUR	**0.977**	**0.987**	**0.880**
SAS	**0.005**	**0.002**	**0.056**
EAS	**0.006**	**0.004**	**0.051**
OCE	**0.002**	**0.001**	**0.002**
AME	**0.002**	**0.001**	**0.001**

mt: H2a1c
X: EU & EU

KS153 - ancestry results



STRUCTURE

	PT	BT	ET
AFR	0.001	0.005	0.001
ME	0.003	0.007	0.008
EUR	0.988	0.975	0.900
SAS	0.004	0.007	0.013
EAS	0.001	0.002	0.066
OCE	0.002	0.002	0.003
AME	0.001	0.002	0.009

mt: U4c1
Y: R-U152
X: EU

KS110 and KS153 - ancestry results

colour code more "direct" less "direct"

All assigned to Europe (8 labs used all markers, *4 labs only biparental SNPs*), examples:

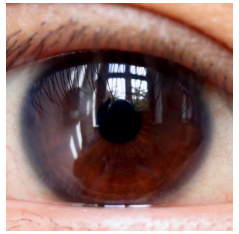
- „individual of European origin"
- „most likely has a European biogeographic background"
- *„ancestry from Europe"*
- *„is most likely of European ancestry"*

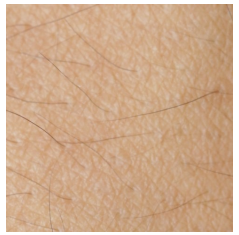
- *„biogeographical ancestry being Europe, or a population with European ancestry"*
- *„more likely if the investigated DNA comes from an individual from the European reference population than if the investigated DNA comes from the South Asian, Middle Eastern, Oceanian, East Asian, Native American, or African reference populations"*
- „results are in line with ancestry in Europe. low probability are ME, Europe, Africa, America, East Asia, Oceania and South Asia"

KS106 - phenotype results

		106
		p-value
eye colour	*blue*	0,000
	intermediate	0,002
	brown	**0,998**
hair colour	*blond*	0,001
	brown	0,122
	red	0,000
	black	**0,877**
hair shade	*light*	0,002
	dark	**0,998**
skin colour	*very pale*	0,000
	pale	0,000
	intermediate	**0,965**
	dark	0,035
	dark to black	0,000

reference photos



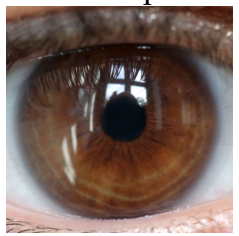


„brown eyes, black hair, intermediate skin“

KS110 - phenotype results

		110
		p-value
eye colour	*blue*	0,099
	intermediate	0,134
	brown	**0,767**
hair **colour**	*blond*	**0,362**
	brown	**0,521**
	red	0,002
	black	0,116
hair **shade**	*light*	**0,742**
	dark	0,258
skin colour	*very pale*	0,008
	pale	**0,304**
	intermediate	**0,681**
	dark	0,007
	dark to black	0,000

reference photos



„brown eyes, dark-brown to black hair and intermediate skin“

„brown eyes, light hair shade and brown to blond hair, intermediate to pale skin“

„brown eyes, dark brown hair and intermediate skin“

„brown eyes, blond or brown hair, and intermediate skin“

„brown eyes, brown hair and intermediate skin“

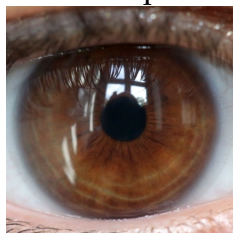
„brown eyes, dark brown or black hair and a pale to intermediate skin“

„brown eyes, dark blond or brown hair and intermediate skin“

KS110 - phenotype results

			110
			p-value
eye colour		*blue*	0,099
		intermediate	0,134
		brown	**0,767**
hair	colour	*blond*	**0,362**
		brown	**0,521**
		red	0,002
		black	0,116
	shade	*light*	**0,742**
		dark	0,258
skin colour		*very pale*	0,008
		pale	**0,304**
		intermediate	**0,681**
		dark	0,007
		dark to black	0,000

reference photos



HPS guidelines: dark brown/black hair (natural: brown)

„dark-brown/black hair "
– 4 labs, 3 used HPS and *1 updated guidelines*

„dark brown hair"
– 1 lab, HPS guidelines

„light hair shade and brown to blond hair "
– 1 lab, HPS guidelines

„brown hair"
– 1 lab, updated HPS guidelines

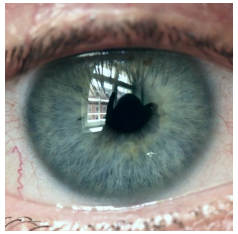
„blond or brown hair "
– 1 lab, updated HPS guidelines

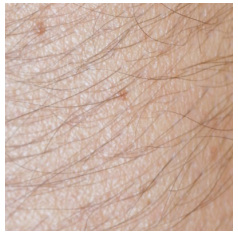
„dark blond or brown hair"
– 1 lab, updated HPS guidelines

KS153 - phenotype results

			153
			p-value
eye colour		*blue*	**0,903**
		intermediate	0,074
		brown	0,023
hair	colour	*blond*	**0,319**
		brown	**0,605**
		red	0,007
		black	0,068
	shade	*light*	**0,812**
		dark	0,188
skin colour		*very pale*	0,078
		pale	**0,563**
		intermediate	**0,369**
		dark	0,000
		dark to black	0,000

reference photos





„blue eyes, dark blond/brown hair and pale/intermediate skin“

„blue eyes, brown/dark brown hair and pale to intermediate skin“

„blue eyes, light hair shade and brown to blond hair and pale to intermediate skin“

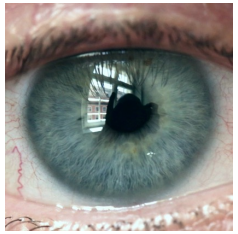
„blue eyes, blond or brown hair and pale to intermediate skin“

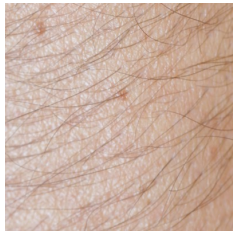
„blue eyes, brown hair and pale/intermediate skin“

KS153 - phenotype results

		153
		p-value
eye colour	*blue*	**0,903**
	intermediate	0,074
	brown	0,023
hair colour	*blond*	**0,319**
	brown	**0,605**
	red	0,007
	black	0,068
hair shade	*light*	**0,812**
	dark	0,188
skin colour	*very pale*	0,078
	pale	**0,563**
	intermediate	**0,369**
	dark	0,000
	dark to black	0,000

reference photos





HPS guidelines: brown/dark brown hair (natural: dark brown)

„brown/dark brown hair“
– 5 labs, 4 with HPS and *1 with updated guidelines*

„brown hair“
- 1 lab, HPS guidelines

„light hair shade and brown to blond hair “
– 1 lab, HPS guidelines

„dark blond or brown hair“
– 1 lab, updated HPS guidelines

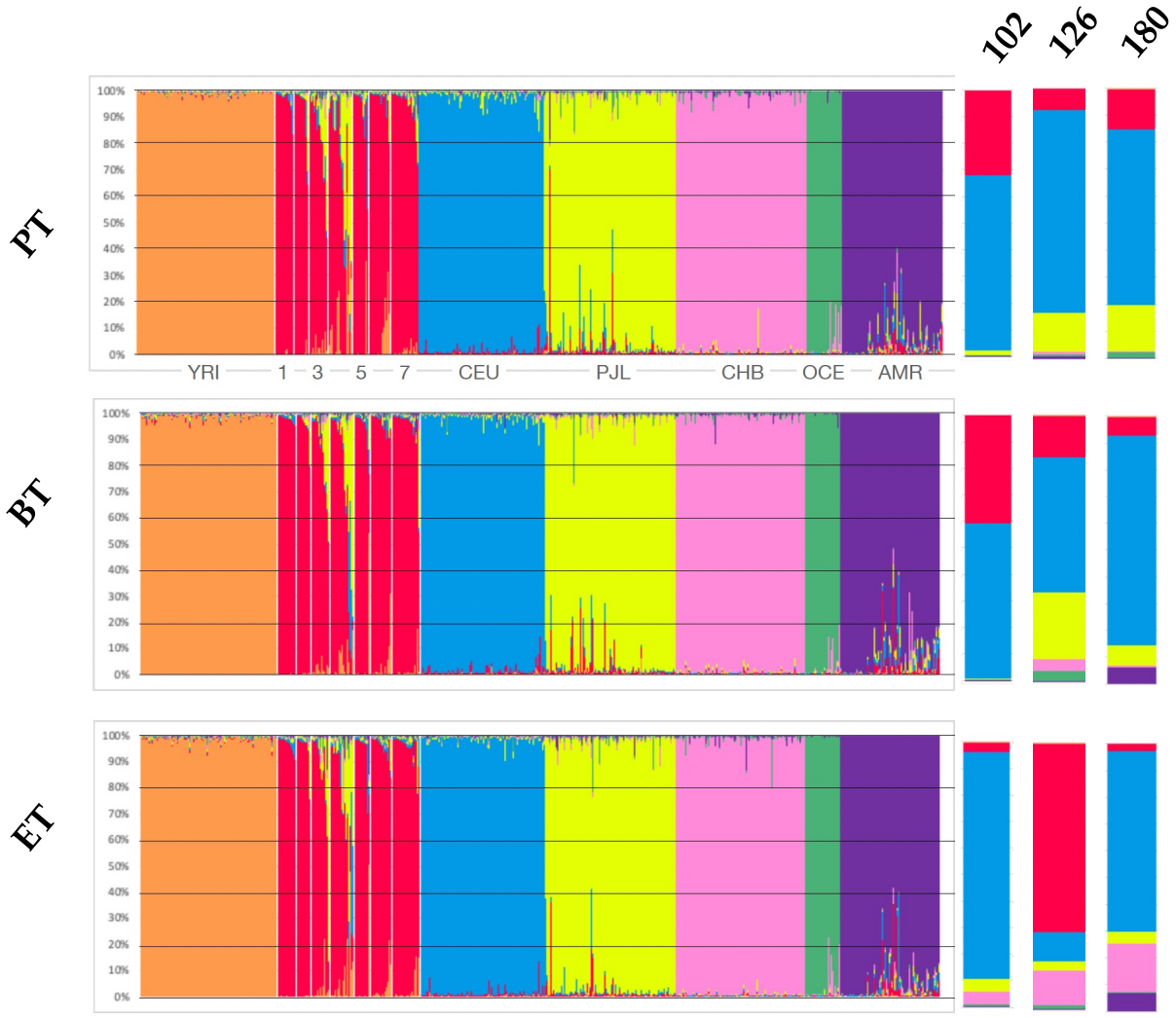
„blond or brown hair“
– 1 lab, updated HPS guidelines

Summary

- simple unequivocal ancestries – consistent clear reporting among all labs

- eye colour based on highest p-value consistent among all labs: predictions tend towards the more secure binary classification approach: i.e., blue vs brown, or neither
- hair and skin colour more ambiguous when close and below 0.7
- HPS guidelines tend to ‚darken' European hair colour predictions in comparison to updated guidelines used by some labs
- no consistency between labs that claimed to use HPS guidelines

KS102, KS126, KS180 - admixture analysis - STRUCTURE



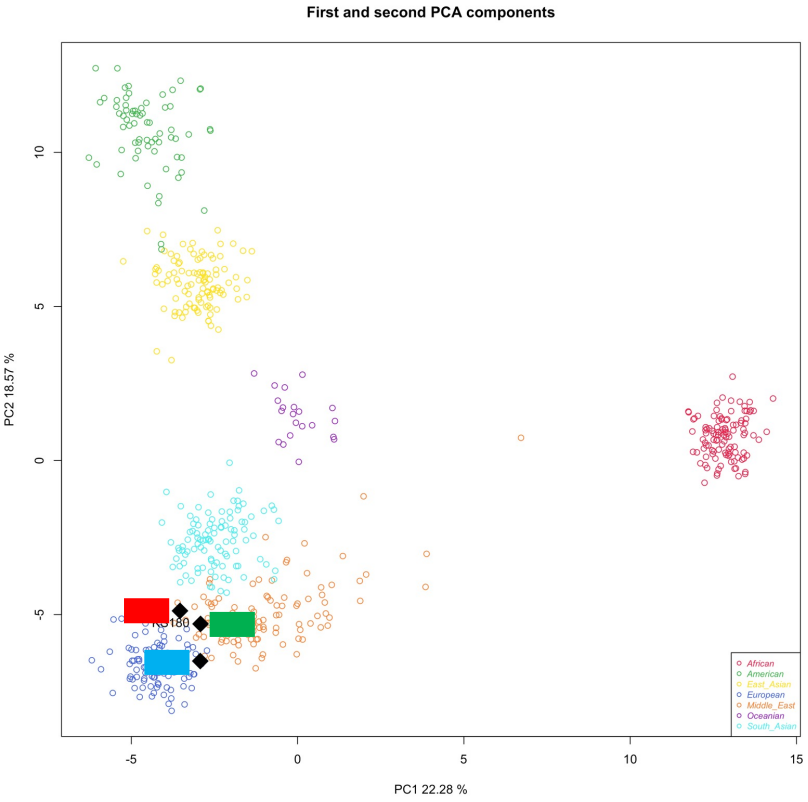
	102			126			180		
	PT	BT	ET	PT	BT	ET	PT	BT	ET
AFR	0.003	0.002	0.003	0.003	0.003	0.004	0.005	0.003	0.002
ME	0.317	0.406	0.036	0.079	0.156	0.704	0.149	0.069	0.025
EUR	0.658	0.585	0.855	0.751	0.505	0.112	0.648	0.786	0.674
SAS	0.017	0.004	0.048	0.141	0.250	0.035	0.171	0.074	0.047
EAS	0.002	0.001	0.048	0.011	0.046	0.130	0.005	0.007	0.179
OCE	0.004	0.002	0.004	0.007	0.036	0.012	0.019	0.001	0.006
AME	0.001	0.001	0.005	0.008	0.005	0.003	0.003	0.061	0.067

KS102, KS126, KS180 - admixture analysis - CONVERGE

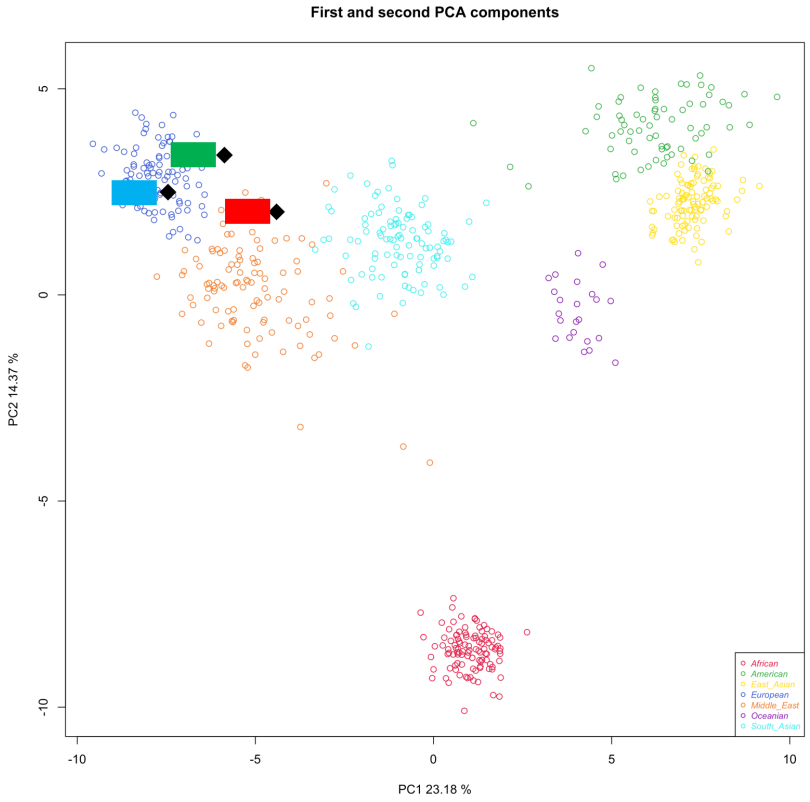
PT only

	102		126		180	
	TFS	MAC	TFS	MAC	TFS	MAC
AFR	0.000	0.000	0.000	0.000	0.007	0.005
ME	0.445	0.439	0.445	0.439	0.030	0.031
EUR	0.549	0.525	0.549	0.525	0.657	0.633
SAS	0.003	0.036	0.003	0.036	0.306	0.331
EAS	0.001	0.000	0.001	0.000	0.000	0.000
OCE	0.001	0.000	0.001	0.000	0.000	0.000
AME	0.001	0.000	0.001	0.000	0.000	0.000

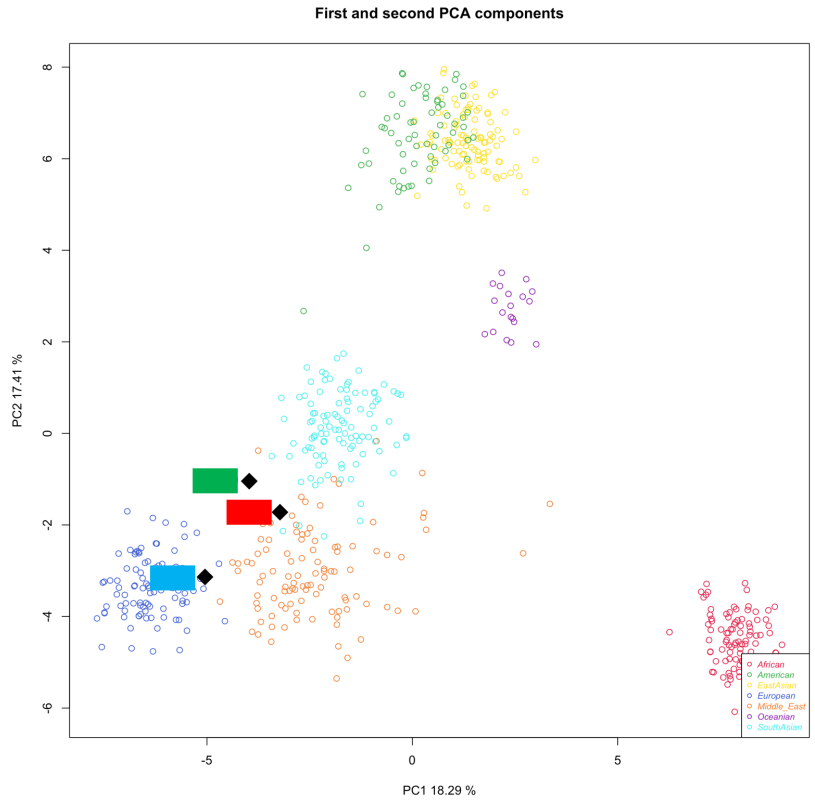
KS102, KS126, KS180 - PCA



PT



BT



ET

KS102, KS126, KS180 - GenoGeographer

102		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
Middle East	-0.956	0.830
BT		
population	*z-score*	*p-value*
Europe & ME	-1.546	0.939
ET		
population	*z-score*	*p-value*
Europe & ME	1.356	0.088

126		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
ME & S. Asia	0.306	0.380
BT		
population	*z-score*	*p-value*
Europe & S. Asia	1.525	0.064
ET		
population	*z-score*	*p-value*
Middle East	0.347	0.364

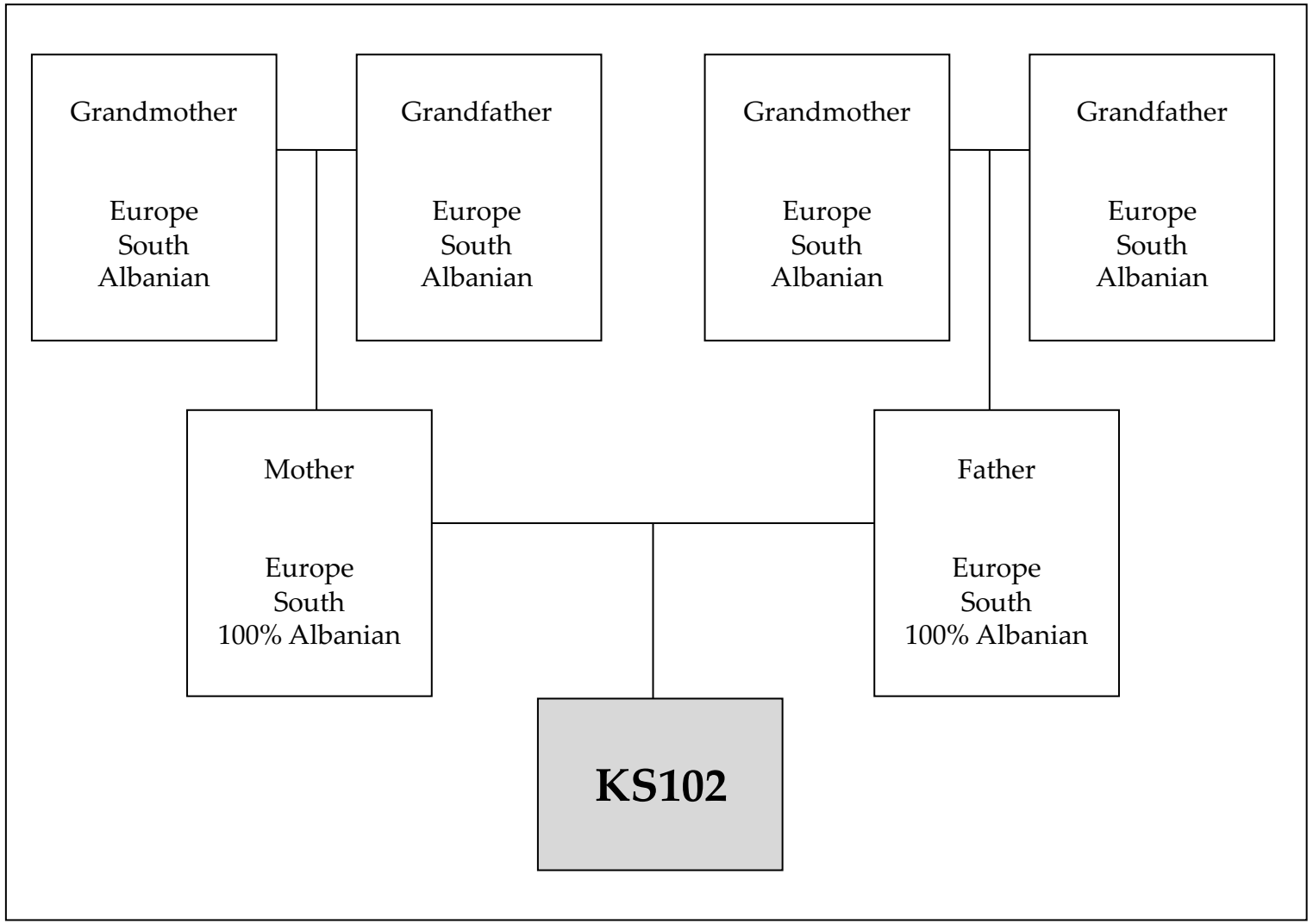
180		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
Europe & S. Asia	-0.865	0.806
BT		
population	*z-score*	*p-value*
Europe & S. Asia	0.257	0.398
ET		
population	*z-score*	*p-value*
Middle East	-0.087	0.535

KS102, KS126, KS180 - extra markers

			102	126	180
			p-value		
eye colour		*blue*	0,001	0,000	0,082
		intermediate	0,032	0,013	0,142
		brown	**0,967**	**0,986**	**0,776**
hair	colour	*blond*	0,027	0,101	0,256
		brown	**0,604**	**0,657**	**0,617**
		red	0,000	0,000	0,058
		black	**0,369**	**0,241**	0,069
	shade	*light*	0,060	0,246	**0,864**
		dark	**0,940**	**0,754**	0,136
skin colour		*very pale*	0,006	0,004	0,138
		pale	**0,159**	0,103	**0,458**
		intermediate	**0,807**	**0,755**	**0,383**
		dark	0,026	**0,136**	0,021
		dark to black	0,002	0,002	0,000

	102	126	180
mtDNA	W1	U7a	U2e2a1a
87 ET Y-SNPs	R-CTS1078	R-CTS1078	J-M172
116 PT Y-SNPs	R-M269	R-M269	J-M67
16 ET X-SNPs	European specific X-chromosome	European specific X-chromosome	European specific X-chromosome

KS102 - ancestry results



STRUCTURE

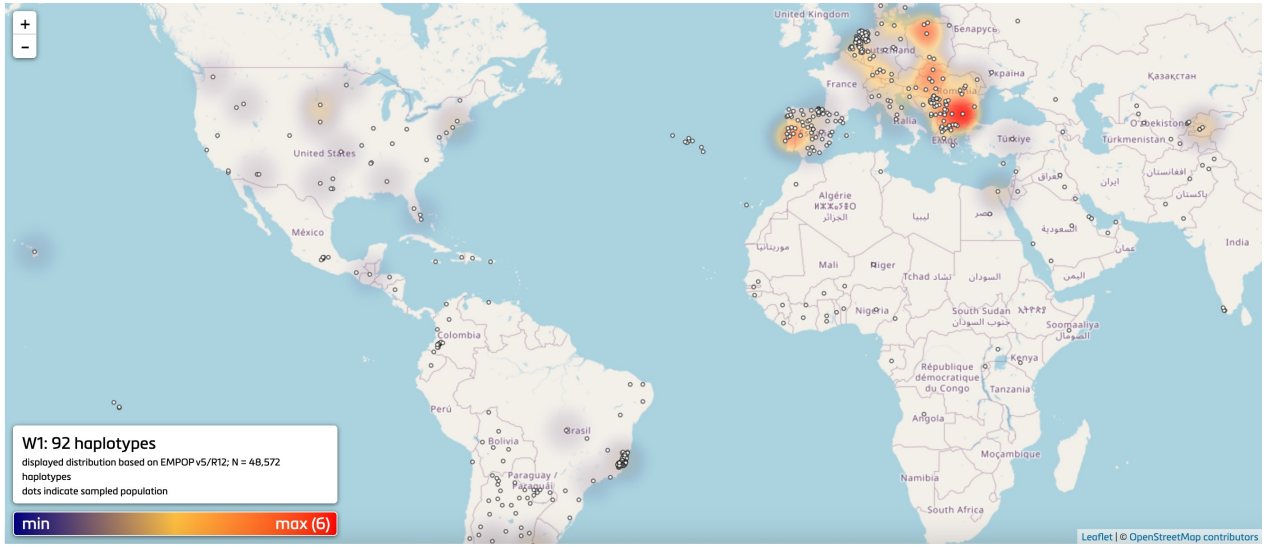
	PT	BT	ET
AFR	0.003	0.002	0.003
ME	0.317	0.406	0.036
EUR	0.658	0.585	0.855
SAS	0.017	0.004	0.048
EAS	0.002	0.001	0.048
OCE	0.004	0.002	0.004
AME	0.001	0.001	0.005

mt: W1
Y: R-CTS1078
X: EU

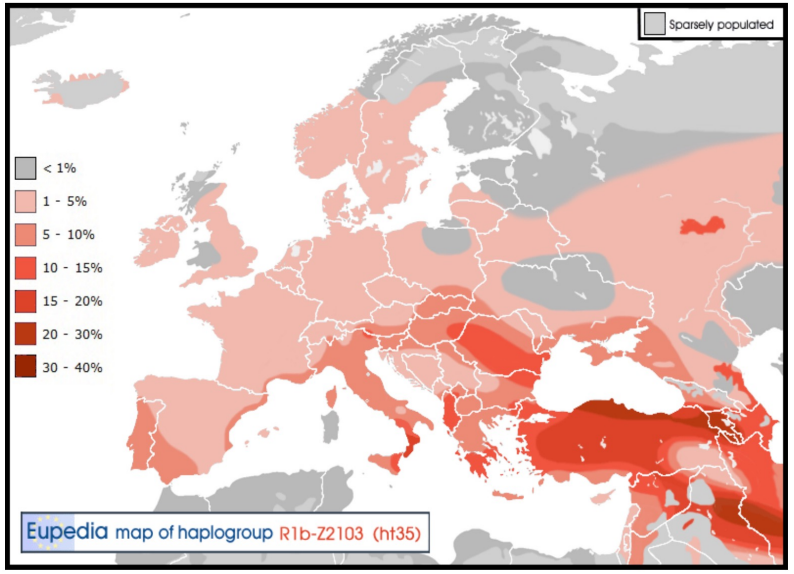
KS102 - ancestry results

- **W1**: found in most of Europe, Central Asia, Iran, Pakistan and northwest India / found in Neolithic Anatolia and Europe

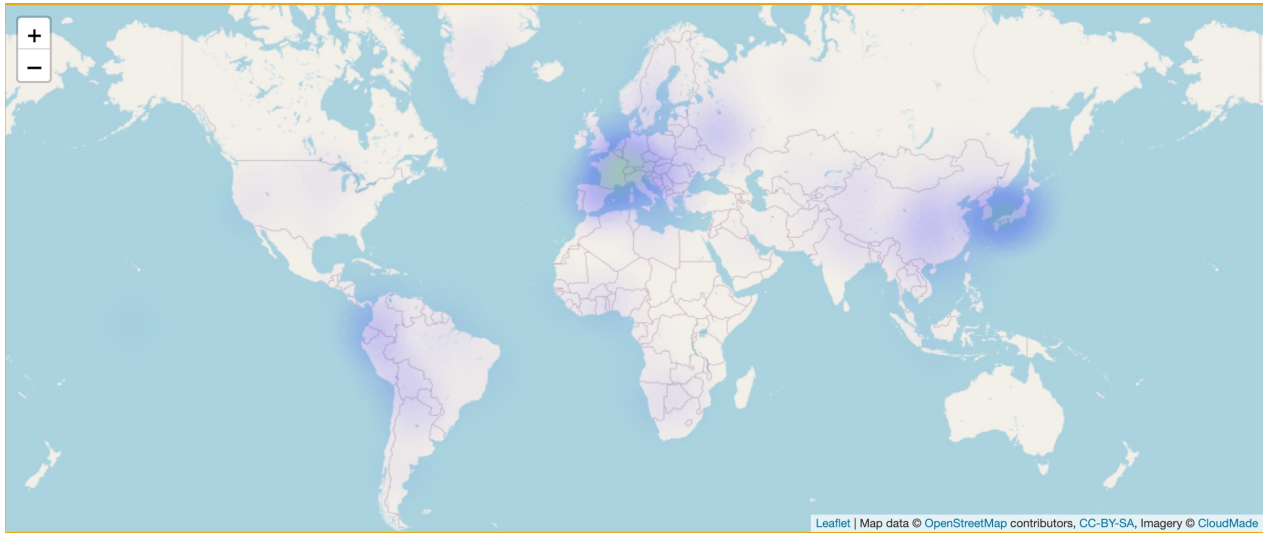
mt: W1



Distribution of haplogroup R1b-ht35 (Z2103) in Europe



Y: R-CTS1078



KS102 - ancestry results

colour code EU OR/AND ME EU/ME admixture EU

Assigned as of European or/and Middle East ancestry (3 labs used all markers, *4 only biparental*), examples:

- „ancestry in Europe, or a location between Europe and the ME regions; probable that one parent has European ancestry while the other could have ME ancestry“
- „predominant biogeographical origin is European and/or Middle Eastern (indication of admixture)
- *„either of European ancestry or of Middle-East and European co-ancestry“*
- *„biogeographical ancestry being Europe or Middle East, or a population with European or Middle East ancestry, or an admixture of the aforementioned meta-populations“*

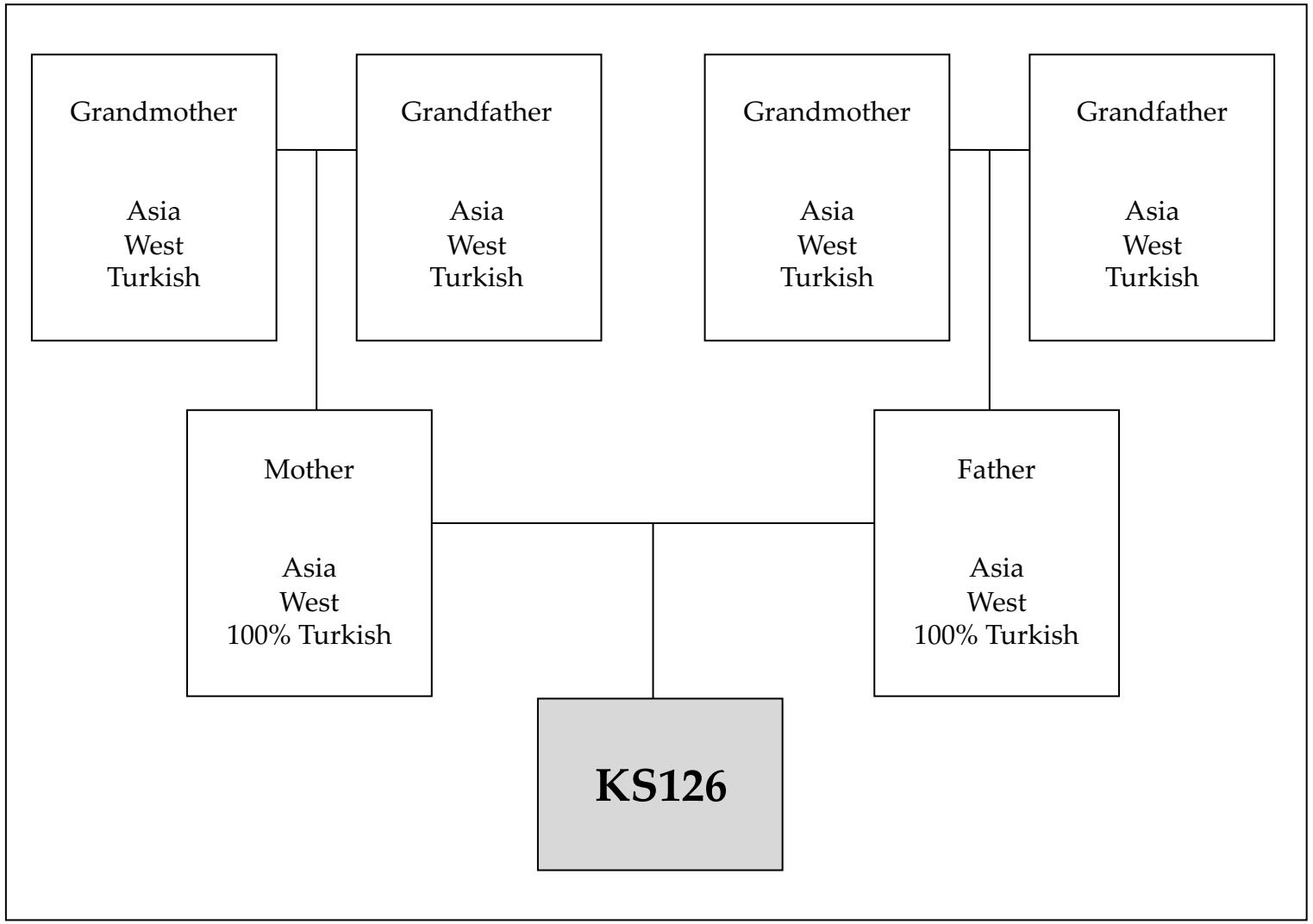
Assigned as European/Middle East admixture (3 labs used all markers):

- „a double inclusion biogeographic ancestry with two root populations, 53.80% European and 46.20% Southwest Asia“
- „either an admixed individual of European and Middle East origin or stems from an admixed population “
- „admixed (Europe & Middle East)“

Assigned to Europe (2 labs used all markers):

- „probably has European biogeographical background“
- „most likely ancestry is from Europe“

KS126 - ancestry results



STRUCTURE

	PT	BT	ET
AFR	0.003	0.003	0.004
ME	0.079	0.156	0.704
EUR	0.751	0.505	0.112
SAS	0.141	0.250	0.035
EAS	0.011	0.046	0.130
OCE	0.007	0.036	0.012
AME	0.008	0.005	0.003

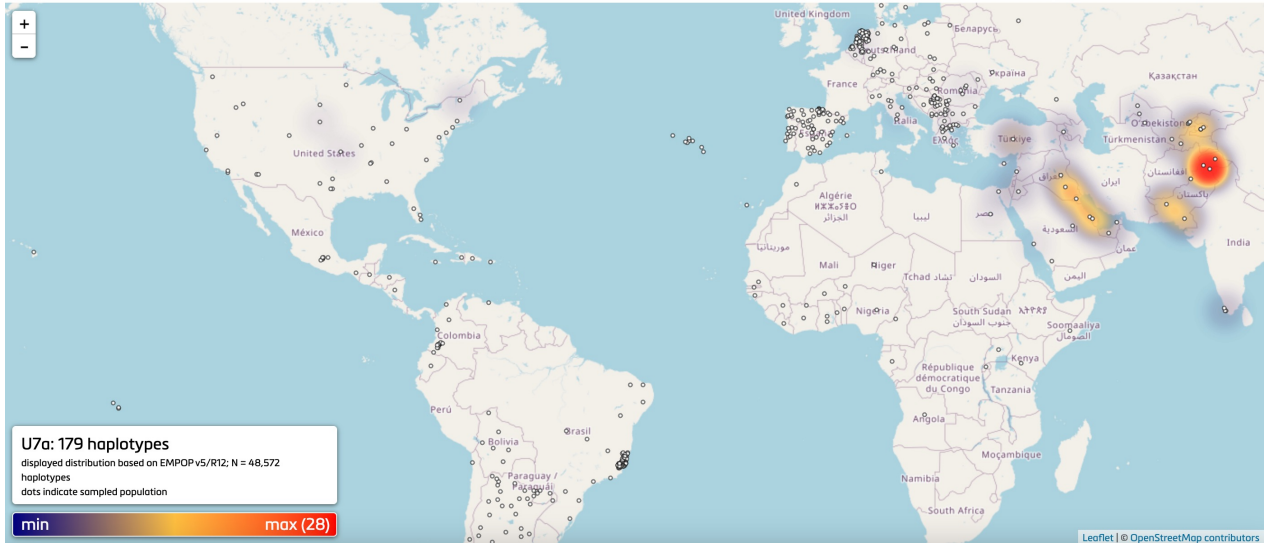
mt: U7a
Y: R-CTS1078
X: EU

KS126 - ancestry results

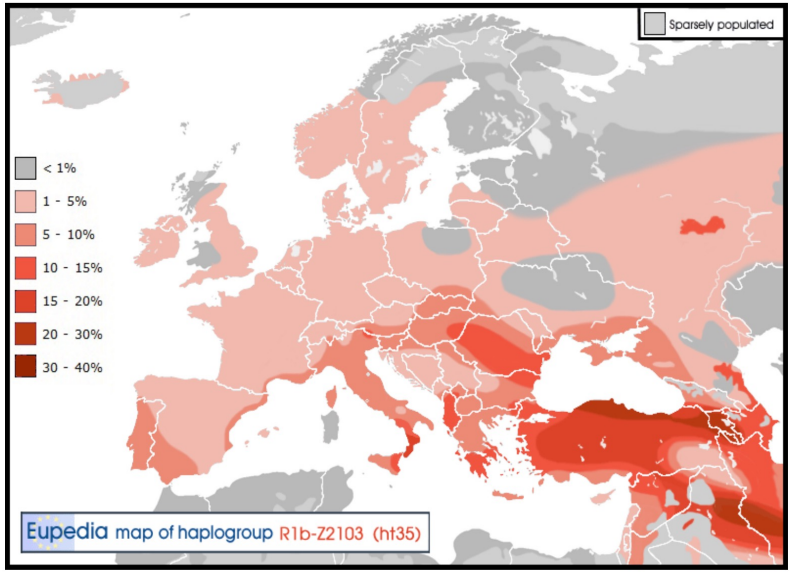
Haplogroup U7 [edit]

Haplogroup U7 is considered a West Eurasian–specific mtDNA haplogroup, believed to have originated in the Black Sea area approximately 30,000 years ago.[67][114][115] In modern populations, U7 occurs at low frequency in the Caucasus,[115] the western Siberian tribes,[116] West Asia (about 4% in the Near East, while peaking with 10% in Iranians),[67] South Asia (about 12% in Gujarat, the westernmost state of India, while for the whole of India its frequency stays around 2%, and 5% in Pakistan),[67] and the Vedda people of Sri Lanka where it reaches it highest frequency of 13.33% (subclade U7a).[117] One third of the West Eurasian-specific mtDNAs found in India are in haplogroups U7, R2 and W. It is speculated that large-scale immigration carried these mitochondrial haplogroups into India.[67]

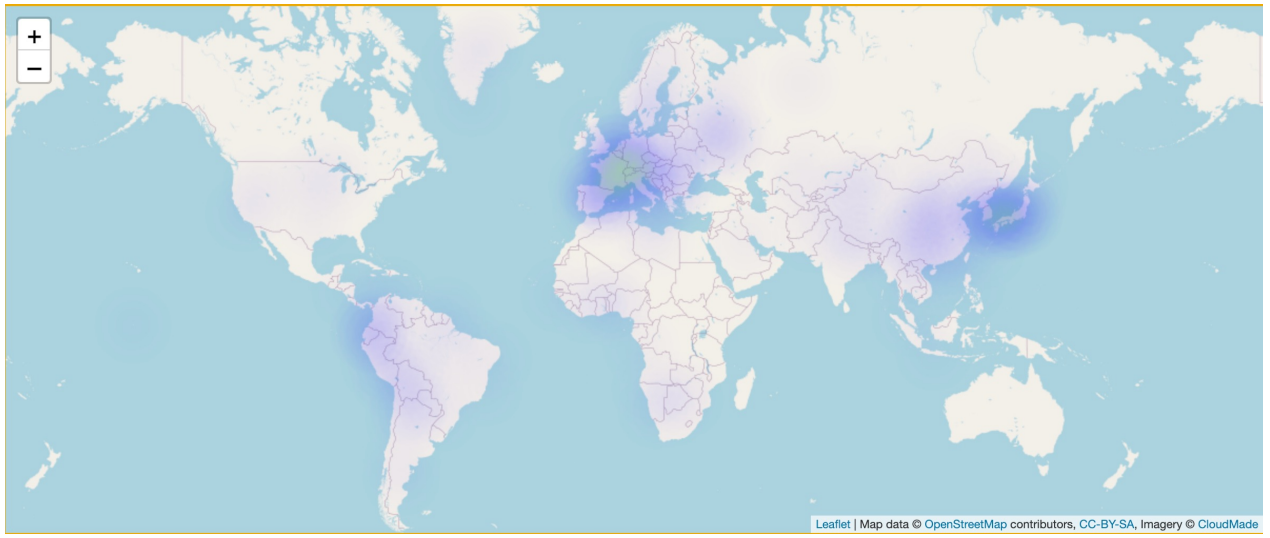
mt: U7a



Distribution of haplogroup R1b-ht35 (Z2103) in Europe



Y: R-CTS1078



KS126 - ancestry results

colour code EU/ME/SA admixture EU/ME admixture co-parentage

Assigned as of European, Middle East and South Asian admixture (4 labs used all markers, *4 only biparental*), examples:

- „admixed (Europe & Middle East & South Asia)“
- „either an admixed individual of European, Middle East and South Asian origin or stems from an admixed population“
- *„more likely if the investigated DNA comes from an individual of European and Middle Eastern genetic admixture, of European and South Asian genetic admixture or of Middle Eastern and South Asian genetic admixture than if the investigated DNA comes from an individual from the European, Middle Eastern, East Asian, South Asian, Oceanian, Native American, or African reference populations“*
- *„biogeographical ancestry being Europe or Middle East or South Asia, or a population with European or Middle East or South Asian ancestry, or an admixture of the aforementioned meta-populations“*

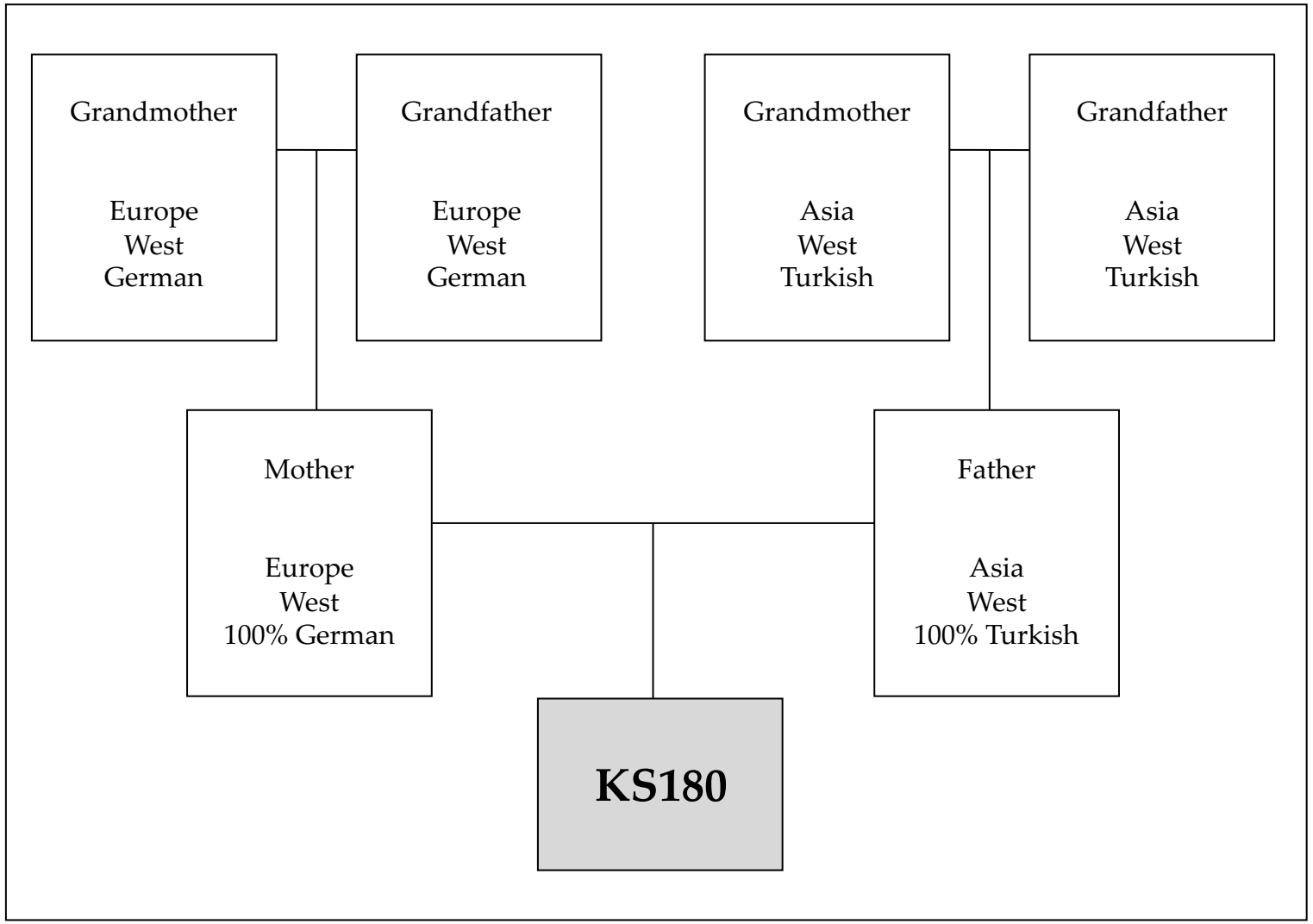
Assigned as of European and Middle East admixture (2 labs used all markers):

- „ancestry is either Middle Eastern, or admixed with a major European component“
- „double inclusion biogeographic ancestry with two root populations, 54.87% European and 44.45% Southwest Asia“

Referred to co-parentage (2 labs used all markers):

- „probably has a mixed biogeographical background a European (probably paternal) - Middle Eastern (probably maternal) origin“
- „under the unadmixed scenario, ancestry in the Middle East or Europe. under the admixed scenario, the results are suggestive of paternal ancestry in Eurasia and maternal ancestry that could be in Europe, Middle East or South Asia “

KS180 - ancestry results



STRUCTURE

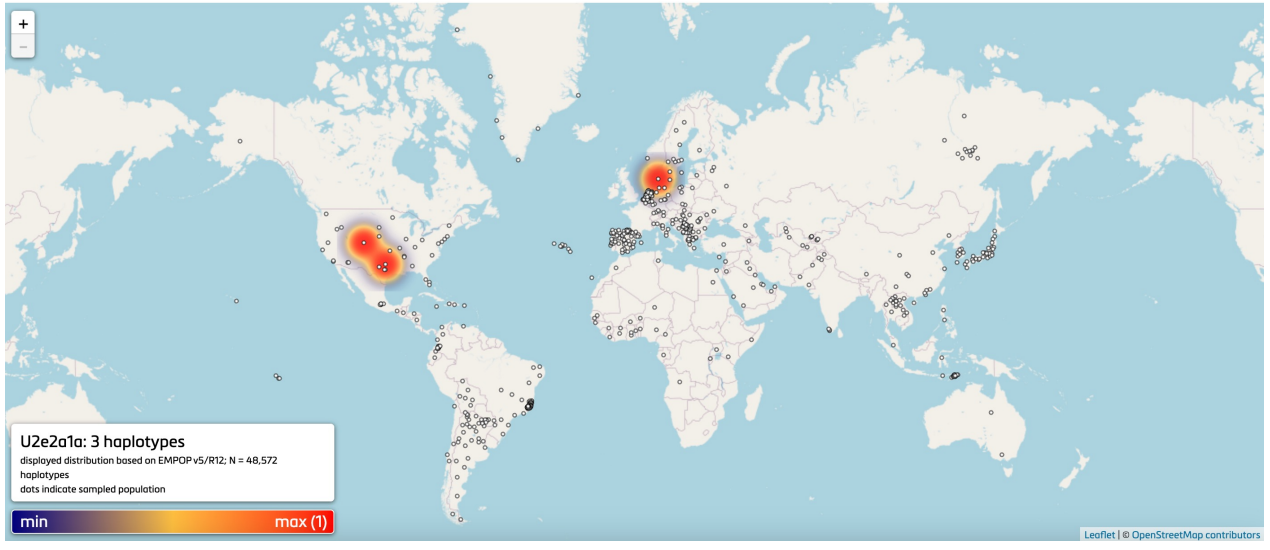
	PT	BT	ET
AFR	0.005	0.003	0.002
ME	0.149	0.069	0.025
EUR	0.648	0.786	0.674
SAS	0.171	0.074	0.047
EAS	0.005	0.007	0.179
OCE	0.019	0.001	0.006
AME	0.003	0.061	0.067

mt: U2e2a1a
Y: J-M67
X: EU

KS180 - ancestry results

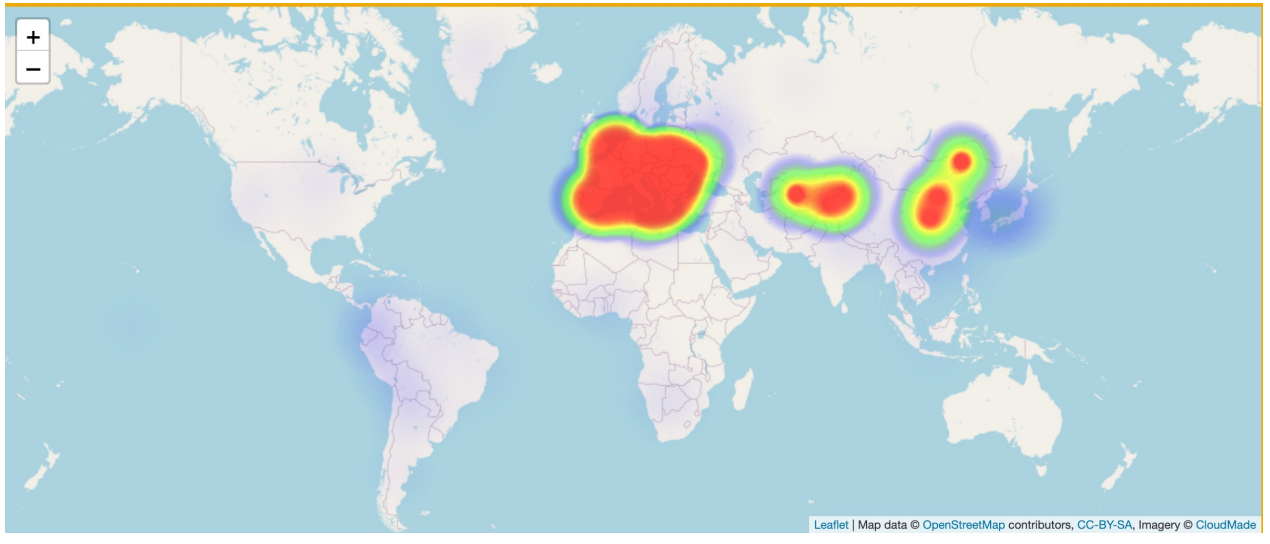
- U2e2a
  - U2e2a1 : found in Germany and Scandinavia

mt: U2e2a1a



- **J2a1-M67** is the most common subclade in the Caucasus (Vainakhs, Ingushs, Chechens, Georgians, Ossetians, Balkars) and in the Levant (Lebanese, Jews). It is also common in western India, the Arabian Peninsula, Anatolia (esp. north-west), Greece (esp. Crete), Italy (esp. Marche and Abruzzo) and Iberia. M67 was probably a major Bronze Age lineage expanding from the Caucasus to Greece to the west and the Indus valley to the east.

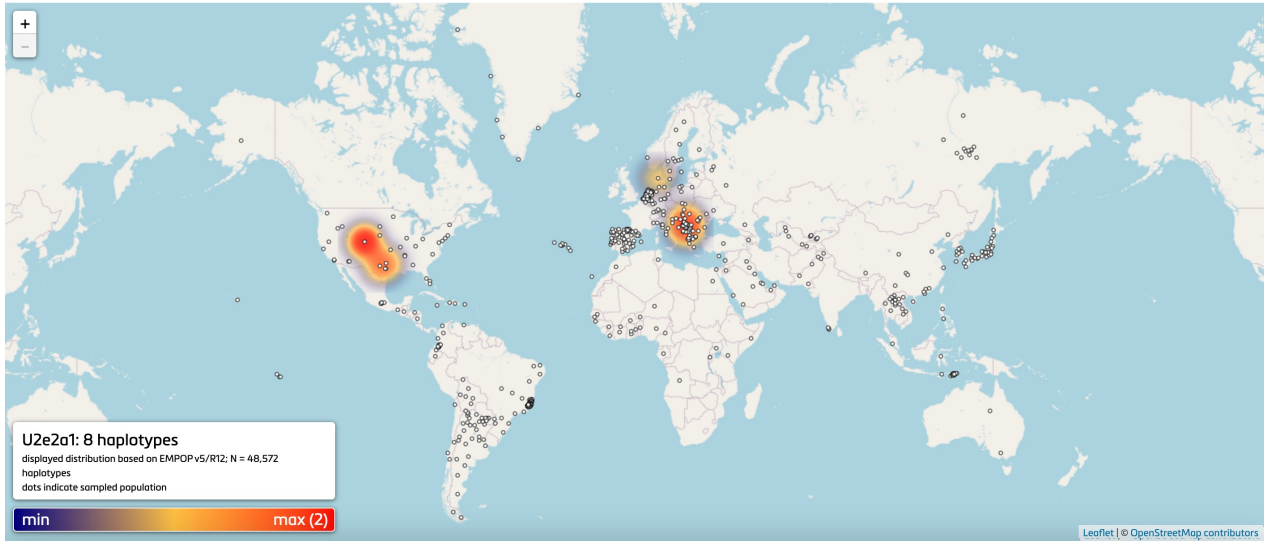
Y: J-M67



KS180 - ancestry results

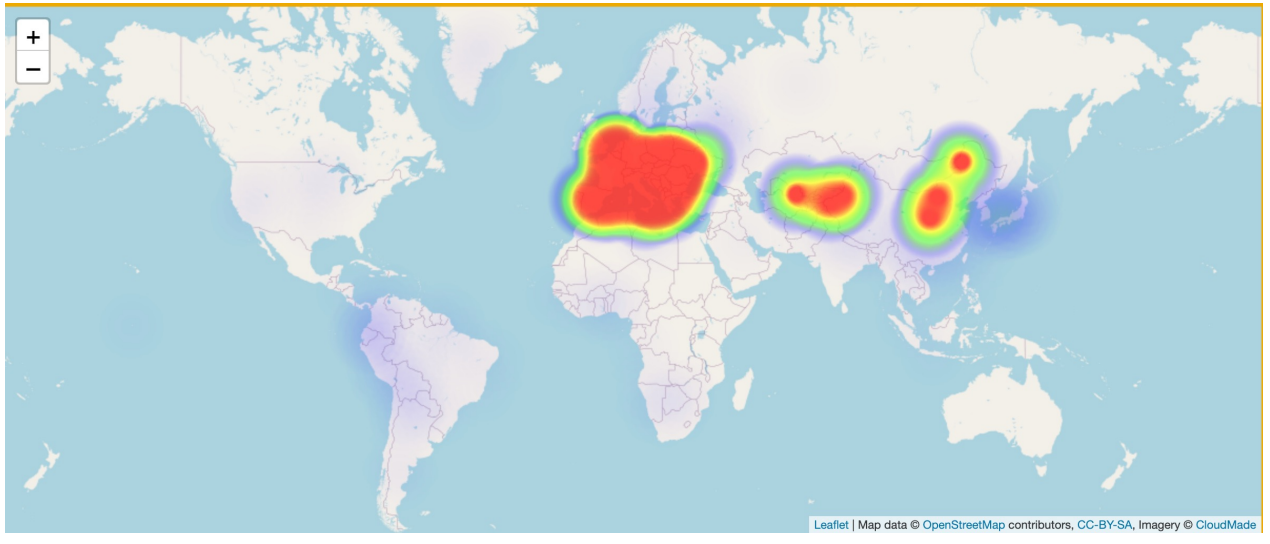
- U2e2a
  - U2e2a1 : found in Germany and Scandinavia

mt: U2e2a1a



- **J2a1-M67** is the most common subclade in the Caucasus (Vainakhs, Ingushs, Chechens, Georgians, Ossetians, Balkars) and in the Levant (Lebanese, Jews). It is also common in western India, the Arabian Peninsula, Anatolia (esp. north-west), Greece (esp. Crete), Italy (esp. Marche and Abruzzo) and Iberia. M67 was probably a major Bronze Age lineage expanding from the Caucasus to Greece to the west and the Indus valley to the east.

Y: J-M67



KS180 - ancestry results

colour code EU/ME/SA admixture co-parentage

Assigned as of European, Middle East and South Asian admixture (4 labs used all markers, *4 only biparental*), examples:

- „admixed (Europe & South Asia & Middle East)“
- „is either an admixed individual of EUR, ME and/or SAS origin or stems from an admixed population“
- *„biogeographical ancestry being Europe or Middle East or South Asia, or a population with European or Middle East or South Asian ancestry, or an admixture of the aforementioned meta-populations“*
- *„more likely if the investigated DNA comes from an individual of European and Middel Eastern genetic admixture, of European and South Asian genetic admixture or of Middle Eastern and South Asian genetic admixture than if the investigated DNA comes from an individual from the European, Middle Eastern, East Asian, South Asian, Oceanian, Native American, or African reference populations“*

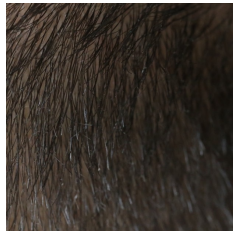
Referred to co-parentage (4 labs used all markers), examples:

- „is likely to have a mixed biogeographic background of European (probably maternal) – Middle Eastern/ Caucasus (probably paternal) origin“
- „results are in line with some degree of European ancestry. Under the hypothesis of admixture, it is possible that one parent has European ancestry and the other parent has ME or South Asian ancestry“

KS102 - phenotype results

		102
		p-value
eye colour	*blue*	0,001
	intermediate	0,032
	brown	**0,967**
hair colour	*blond*	0,027
	brown	**0,604**
	red	0,000
	black	**0,369**
hair shade	*light*	0,060
	dark	**0,940**
skin colour	*very pale*	0,006
	pale	0,159
	intermediate	**0,807**
	dark	0,026
	dark to black	0,002

reference photos



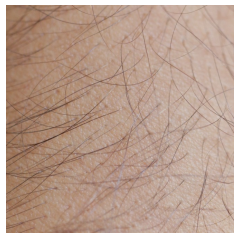
„brown eyes, dark-brown/black hair and intermediate/pale skin“

„brown eyes, brown or black hair, intermediate skin“

KS126 - phenotype results

		126
		p-value
eye colour	*blue*	0,000
	intermediate	0,013
	brown	**0,986**
hair colour	*blond*	0,101
	brown	**0,657**
	red	0,000
	black	**0,241**
hair shade	*light*	0,246
	dark	**0,754**
skin colour	*very pale*	0,004
	pale	0,103
	intermediate	**0,755**
	dark	**0,136**
	dark to black	0,002

reference photos



„brown eyes, dark-brown to black hair and intermediate skin“

„brown eyes, dark hair shade and brown to black hair, intermediate to dark skin“

KS180 - phenotype results

		180
		p-value
eye colour	*blue*	0,082
	intermediate	0,142
	brown	**0,776**
hair colour	*blond*	**0,256**
	brown	**0,617**
	red	0,058
	black	0,069
hair shade	*light*	**0,864**
	dark	0,136
skin colour	*very pale*	0,138
	pale	**0,458**
	intermediate	**0,383**
	dark	0,021
	dark to black	0,000

reference photos

„brown eyes, blond or brown hair, pale to intermediate skin“

„brown eyes, light hair shade and brown to blond hair, pale to intermediate skin“

„brown eyes, brown hair and pale to intermediate skin“

“brown eyes, brown to dark brown hair and pale to intermediate skin“

„brown eyes, dark blond or brown hair and pale / intermediate skin“

KS180 - phenotype results

		180
		p-value
eye colour	*blue*	0,082
	intermediate	0,142
	brown	**0,776**
hair **colour**	*blond*	**0,256**
	brown	**0,617**
	red	0,058
	black	0,069
hair **shade**	*light*	**0,864**
	dark	0,136
skin colour	*very pale*	0,138
	pale	**0,458**
	intermediate	**0,383**
	dark	0,021
	dark to black	0,000

reference photos

HPS guidelines: brown/dark brown (natural: brown)

“brown to dark brown hair “
– 5 labs, 4 with HPS and 1 with updated guidelines

„brown hair“
– 1 lab, HPS guidelines

„light hair shade and brown to blond hair “
– 1 lab, HPS guidelines

„blond or brown hair“
- 1 lab, updated HPS guidelines

„dark blond or brown hair“
– 1 lab, updated HPS guidelines

Summary

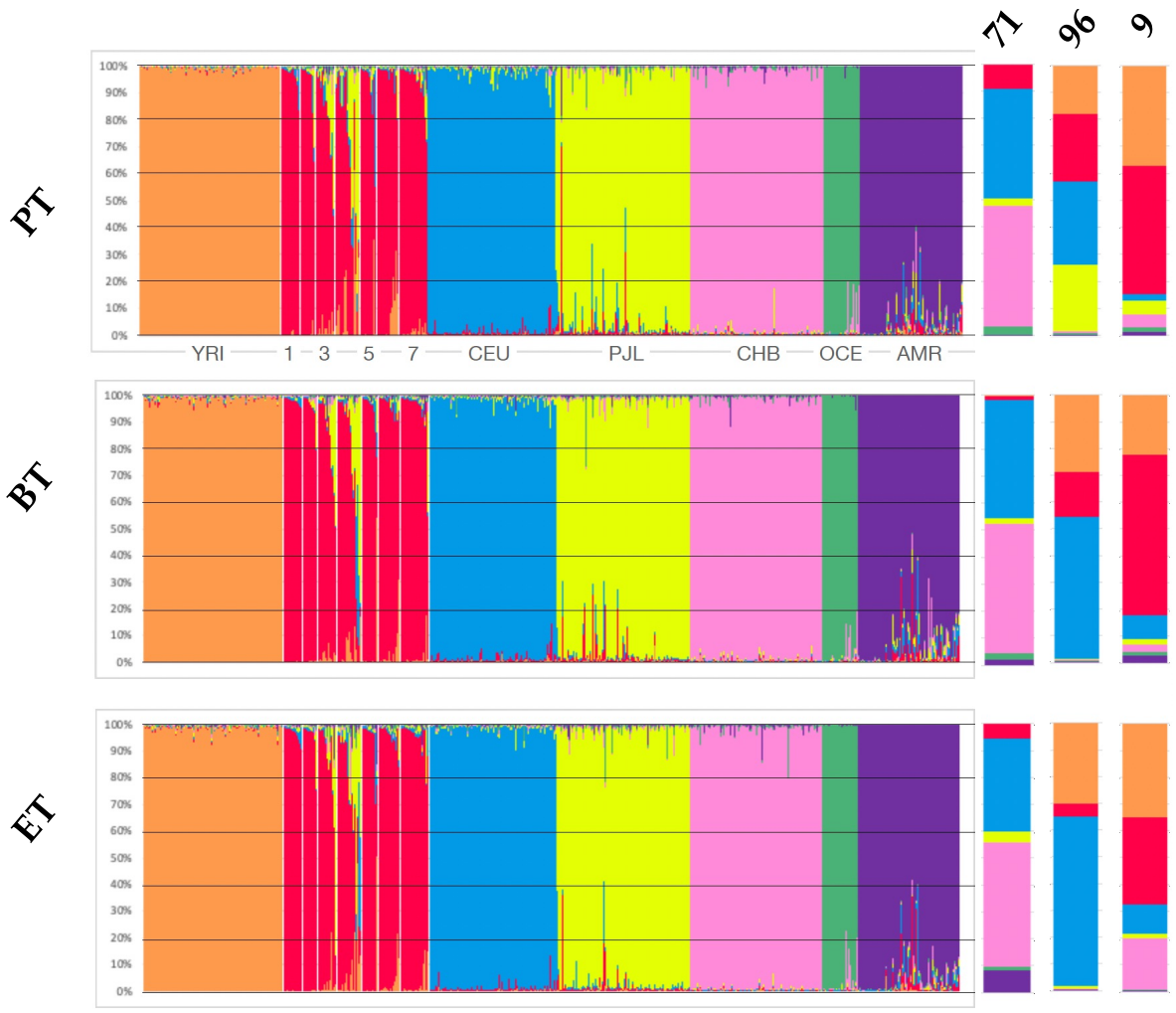
- ET gave the clearest STRUCTURE patterns for KS102 (Albanian) and K126 (Turkish) samples in terms of relative EUR-ME proportions
- KS180 (German-Turkish) had incorrect EAS cluster with ET
- KS180 and KS126 respectively, gave PT and BT >0.15 SAS clusters

- differences appeared between KS126 (Turkish) and KS102 (Albanian) in how interpretations were made and how they were reported
- KS126 with stronger ME-SAS description in most reports: consistent in saying the individual was not completely European
- KS180 (German-Turkish) is the first sample where more labs distinguishing the paternal and maternal lineages and commenting on co-parentage

- tendency to over-emphasise relatively small co-ancestry proportions - no guidelines were given by us for the interpretation of minor co-ancestry cluster membership proportions

- the lack of any differentiation of ME-SAS with EUR for the ET X-chromosome SNP set was not explained to participants – data might have been misread

KS71, KS96, KS9 - admixture analysis - STRUCTURE



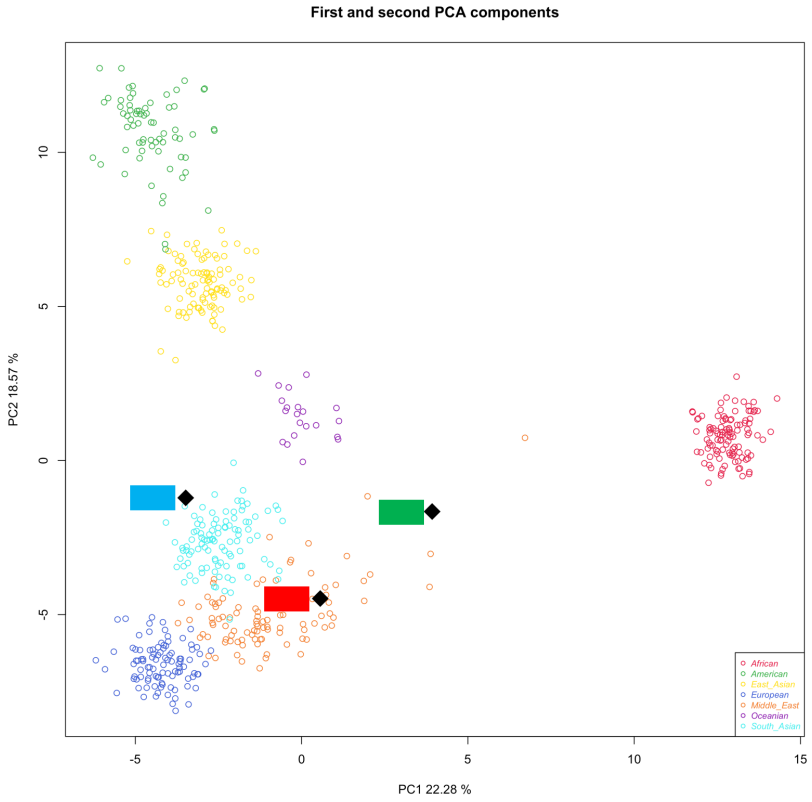
	71			96			9		
	PT	BT	ET	PT	BT	ET	PT	BT	ET
AFR	0.002	0.003	0.002	0.180	0.287	0.302	0.372	0.222	0.349
ME	0.089	0.014	0.054	0.249	0.170	0.048	0.475	0.600	0.326
EUR	0.405	0.438	0.344	0.308	0.531	0.633	0.026	0.090	0.111
SAS	0.026	0.021	0.042	0.247	0.004	0.011	0.049	0.019	0.016
EAS	0.447	0.479	0.463	0.006	0.002	0.002	0.047	0.029	0.192
OCE	0.029	0.025	0.012	0.007	0.003	0.002	0.018	0.011	0.003
AME	0.003	0.020	0.083	0.003	0.003	0.002	0.012	0.029	0.004

KS71, KS96, KS9 - admixture analysis - CONVERGE

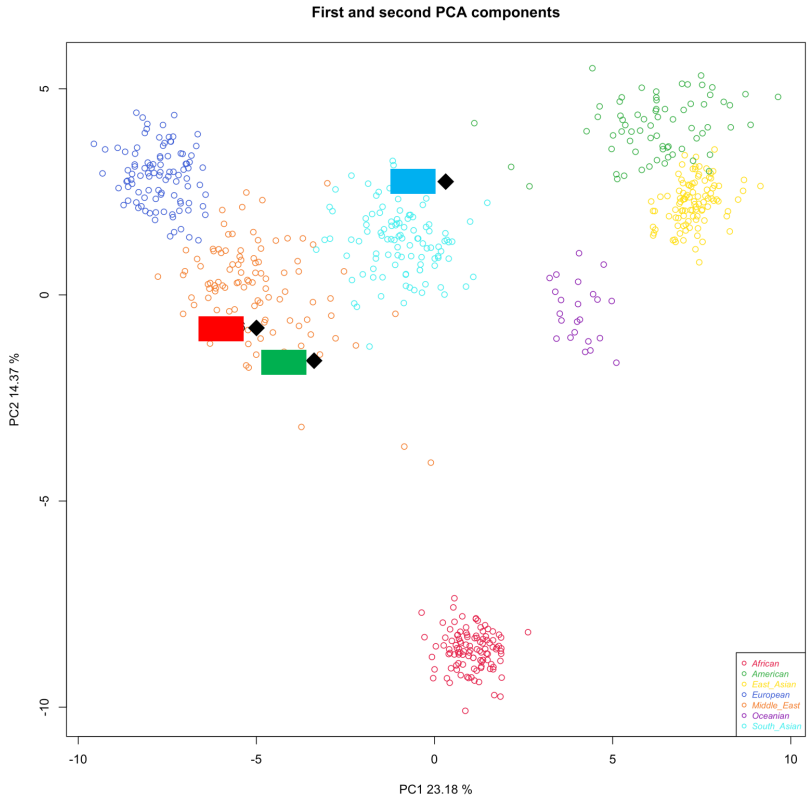
PT only

	71		96		9	
	TFS	MAC	TFS	MAC	TFS	MAC
AFR	0.000	0.000	0.283	0.250	0.463	0.115
ME	0.012	0.242	0.115	0.137	0.236	0.732
EUR	0.479	0.291	0.582	0.551	0.007	0.000
SAS	0.033	0.000	0.020	0.062	0.294	0.002
EAS	0.471	0.467	0.000	0.000	0.000	0.151
OCE	0.005	0.000	0.000	0.000	0.000	0.000
AME	0.000	0.000	0.000	0.000	0.000	0.000

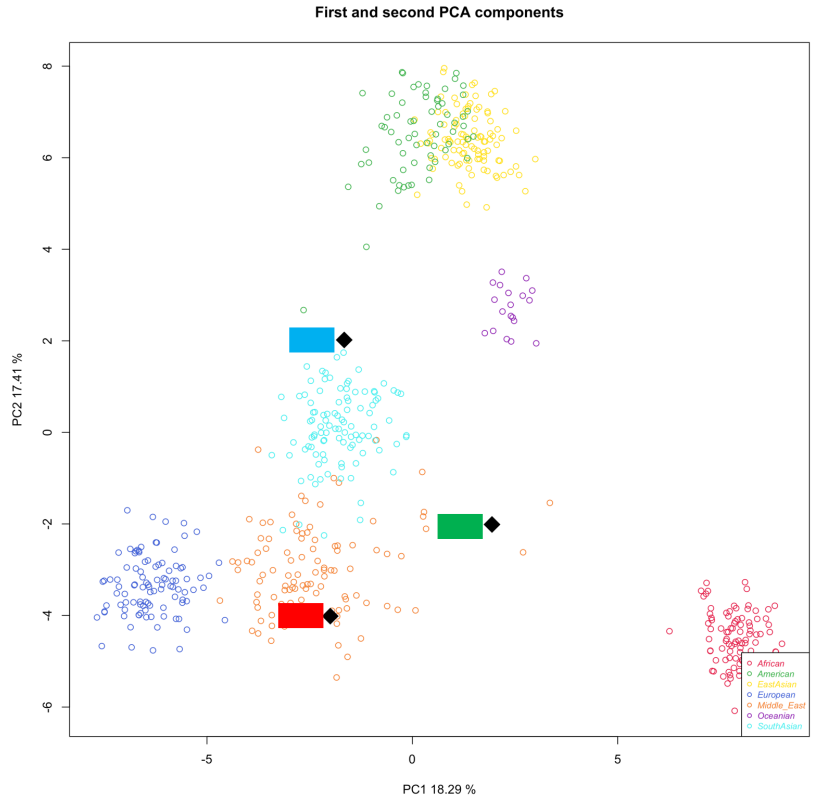
KS71, KS96, KS9 - PCA



PT



BT



ET

KS71, KS96, KS9 - GenoGeographer

71		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
E. Asia & ME	-0.153	0.561
BT		
population	*z-score*	*p-value*
South Asia	1.241	0.107
ET		
population	*z-score*	*p-value*
All populations are rejected		

96		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
Middle East	0.823	0.205
BT		
population	*z-score*	*p-value*
All populations are rejected		
ET		
population	*z-score*	*p-value*
Middle East	0.957	0.169

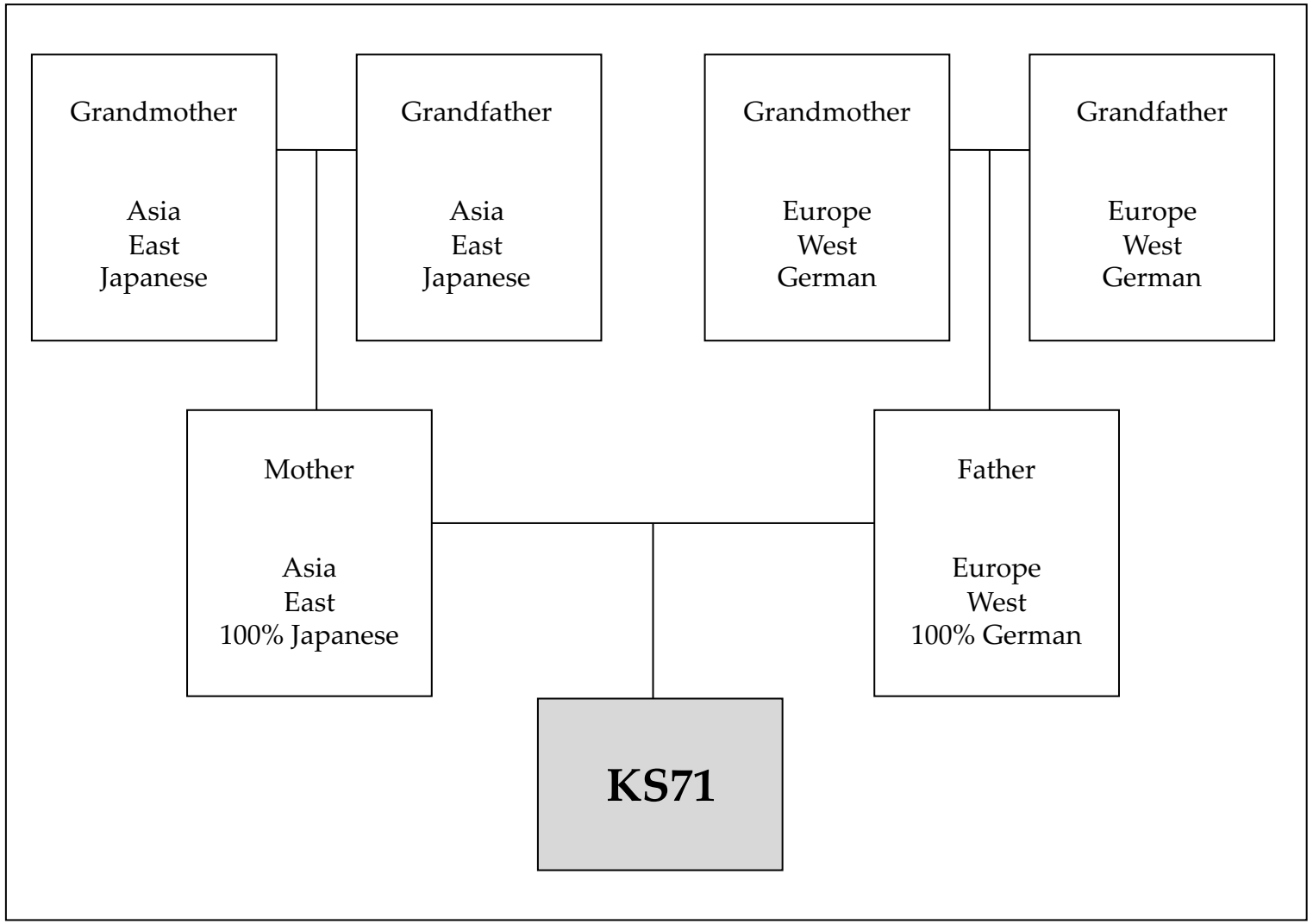
9		
z-score ≤ 1.64; P ≥ 0.05		
PT		
population	*z-score*	*p-value*
All populations are rejected		
BT		
population	*z-score*	*p-value*
All populations are rejected		
ET		
population	*z-score*	*p-value*
All populations are rejected		

KS71, KS96, KS9 - extra markers

			71	96	9
			p-value		
eye colour		*blue*	0,000	0,003	0,000
		intermediate	0,007	0,020	0,002
		brown	**0,993**	**0,977**	**0,998**
hair	colour	*blond*	0,004	0,038	0,003
		brown	**0,347**	**0,725**	**0,474**
		red	0,000	0,000	0,000
		black	**0,649**	**0,236**	**0,523**
	shade	*light*	0,005	0,103	0,005
		dark	**0,995**	**0,897**	**0,995**
skin colour		*very pale*	0,000	0,001	0,000
		pale	0,000	0,028	0,000
		intermediate	**0,705**	**0,262**	0,000
		dark	**0,264**	**0,685**	0,009
		dark to black	0,030	0,024	**0,991**

	71	96	9
mtDNA	M7b1a1a1	L3e2a1a	E1a1a1
87 ET Y-SNPs	♀	R-M343	♀
116 PT Y-SNPs	♀	R-M343	♀
16 ET X-SNPs	European and E.Asian specific X-chromosomes	African specific X-chromosome	African specific X-chromosomes

KS71 - ancestry results

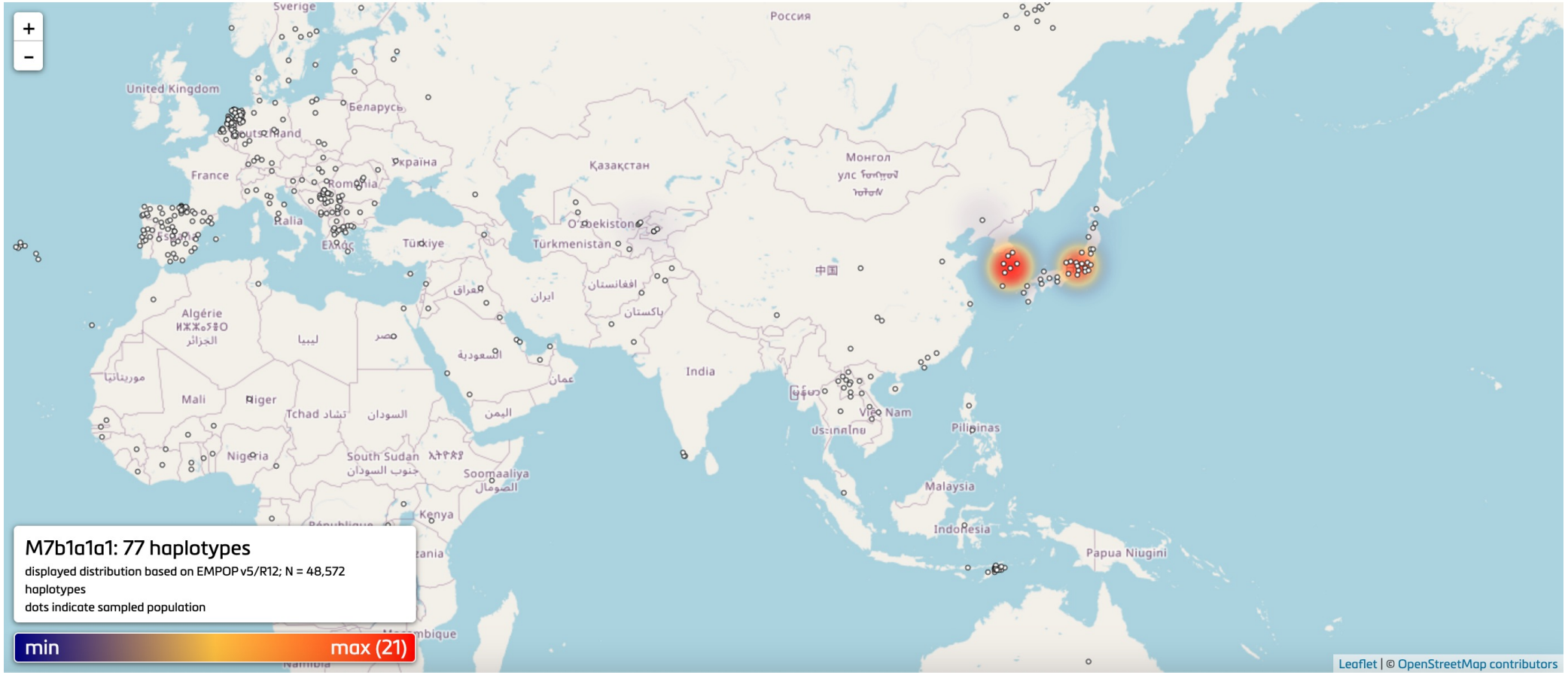


STRUCTURE

	PT	BT	ET
AFR	0.002	0.003	0.002
ME	0.089	0.014	0.054
EUR	0.405	0.438	0.344
SAS	0.026	0.021	0.042
EAS	0.447	0.479	0.463
OCE	0.029	0.025	0.012
AME	0.003	0.020	0.083

mt: M7b1a1a1
X: EU & EA

KS71 - ancestry results



KS71 - ancestry results

colour code co-parentage EU/EA admixture no report

Referred to co-parentage (5 labs used all markers), examples:

- „likely to have a recent mixed biogeographical background of East Asian (probably maternal) – European (probably paternal) origin“
- „admixed ancestry with one parent with European ancestry and one with East Asian. Lineage markers suggest that the East Asian ancestry is maternal while the European ancestry is from the male line“
- „results are in line with one parent having East Asian ancestry. Under the hypothesis of admixture, it is possible the other parent has European ancestry – although the ME and South Asia cannot be excluded as origins for the second parent“

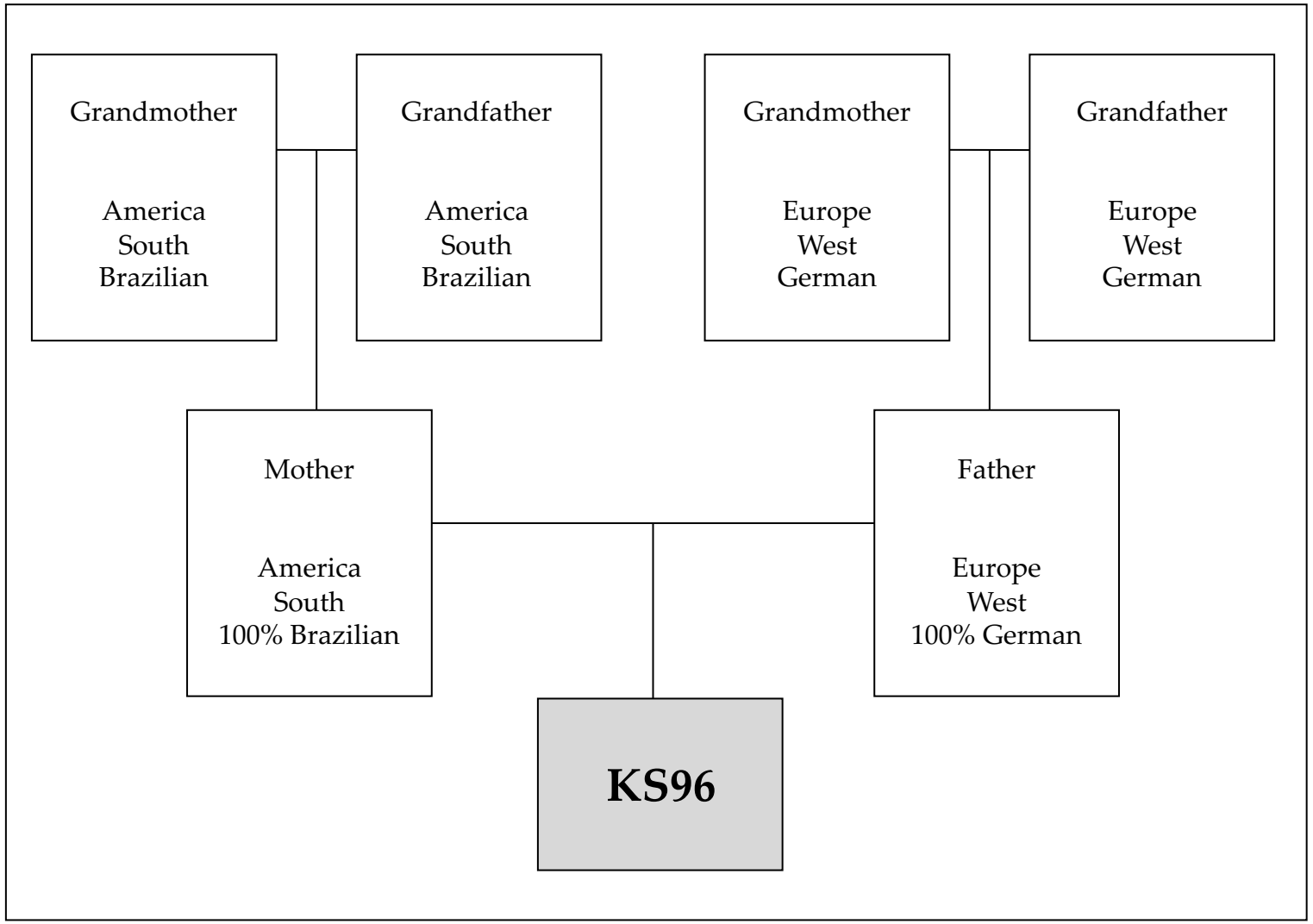
Assigned as of European and East Asian admixture (3 labs used all markers, *3 only biparental*) examples:

- „admixed (Europe & East Asia)“
- *„biogeographical ancestry being an admixture of European and East Asian“*
- *„more likely if the investigated DNA comes from an individual of European and East Asian genetic admixture, of Middle Eastern and East Asian genetic admixture or of South Asian and East Asian genetic admixture than if the investigated DNA comes from an individual from the European, Middle Eastern, East Asian, South Asian, Oceanian, Native American, or African reference populations“*

Didn‘t assign to any reference population (*1 lab used only biparental*):

- *„this individual doesn’t present an African ancestry. The biogeographical origin of this individual could not be predicted more precisely“*

KS96 - ancestry results



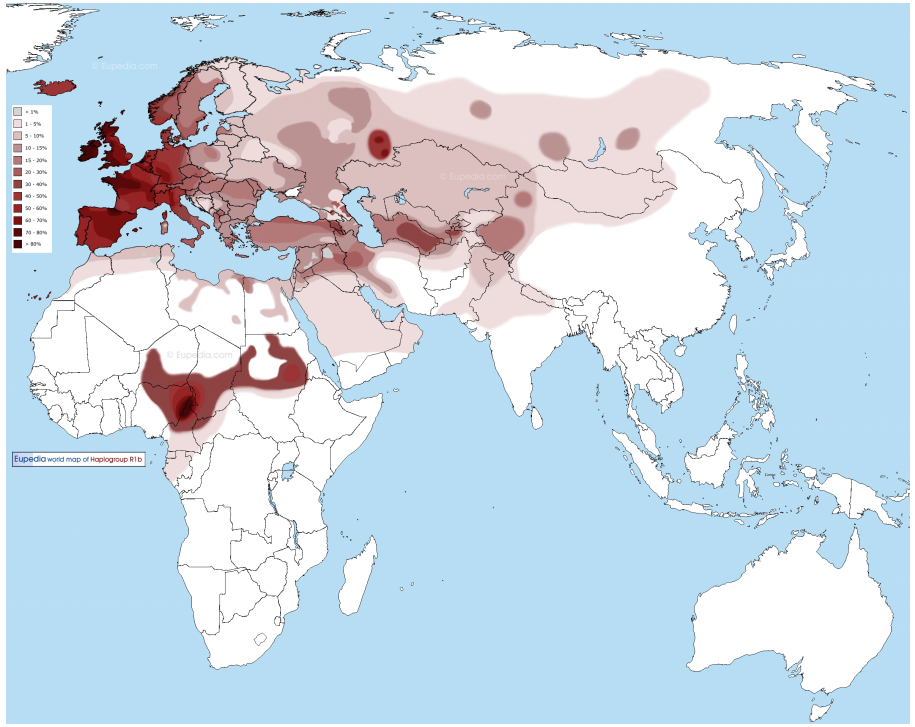
STRUCTURE

	PT	BT	ET
AFR	0.180	0.287	0.302
ME	0.249	0.170	0.048
EUR	0.308	0.531	0.633
SAS	0.247	0.004	0.011
EAS	0.006	0.002	0.002
OCE	0.007	0.003	0.002
AME	0.003	0.003	0.002

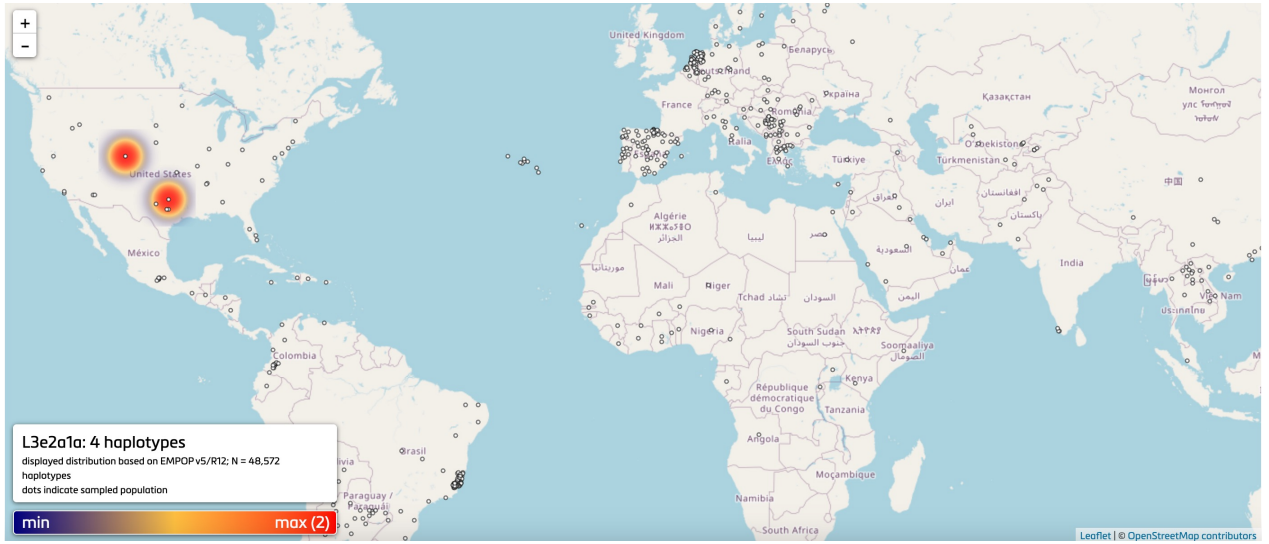
mt: L3e2a1a
Y: R-M343
X: AFR

KS96 - ancestry results

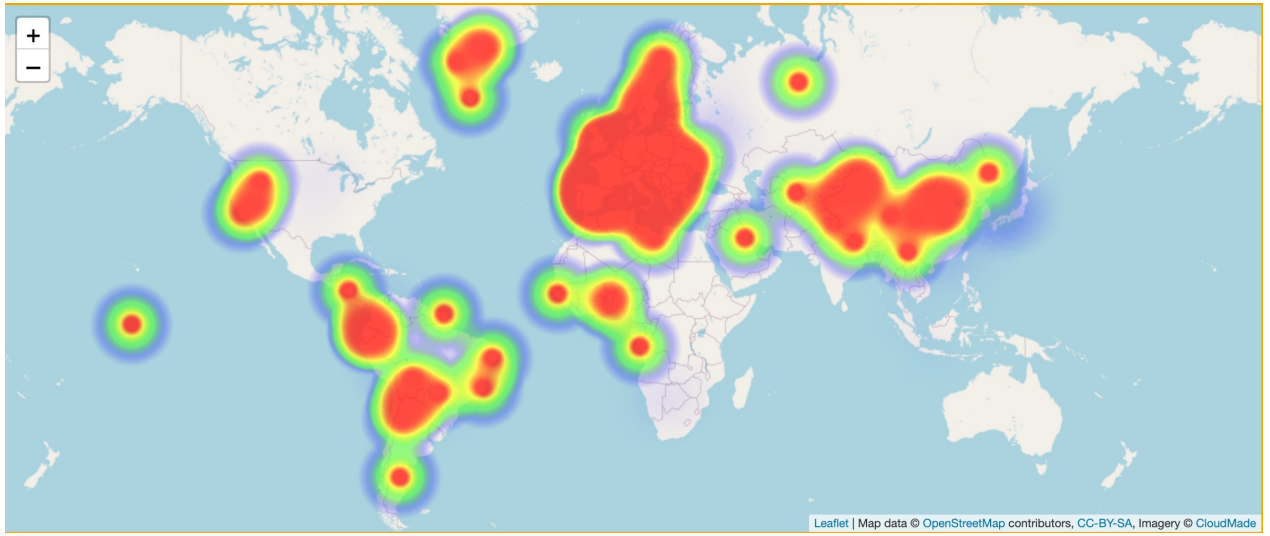
- **L3e** – Suggested to have originated in the Central Africa/Sudan region about 45,000 years ago during the upper paleolithic period.[27] It is the most common L3 sub-clade in Bantu-speaking populations.[28] L3e is also the most common L3 subclade amongst African Americans and Afro-Brazilians.[29]



mt: L3e2a1a

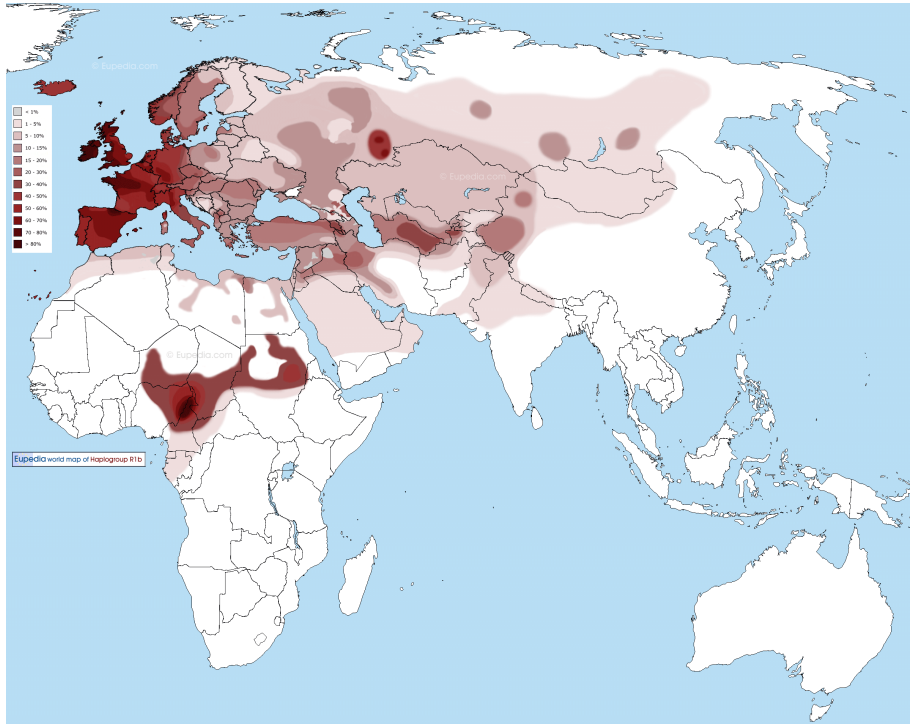


Y: R-M343

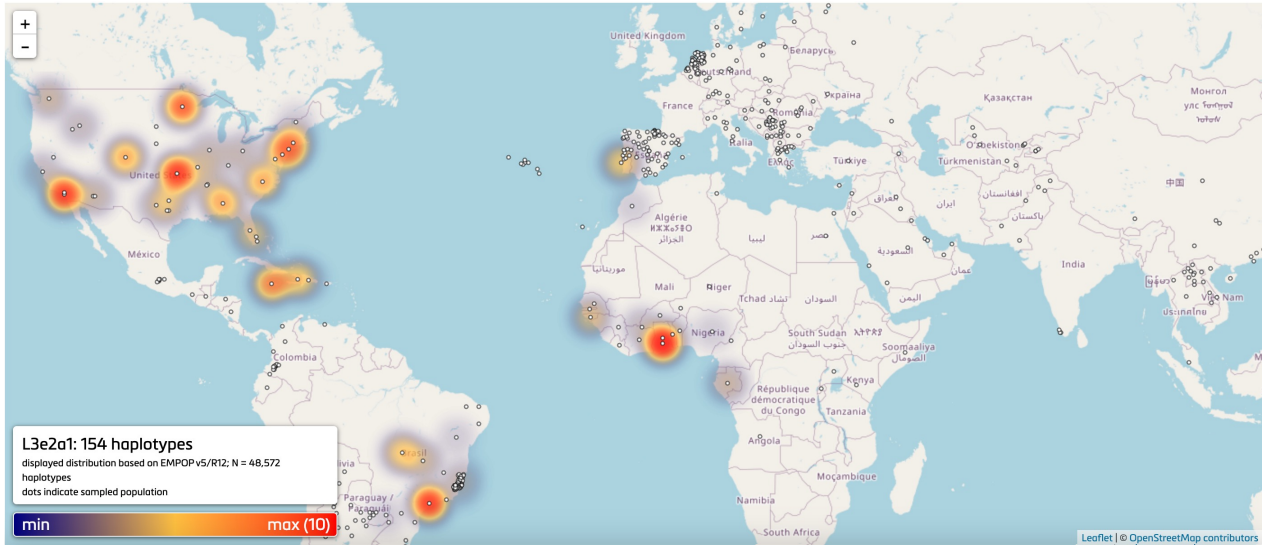


KS96 - ancestry results

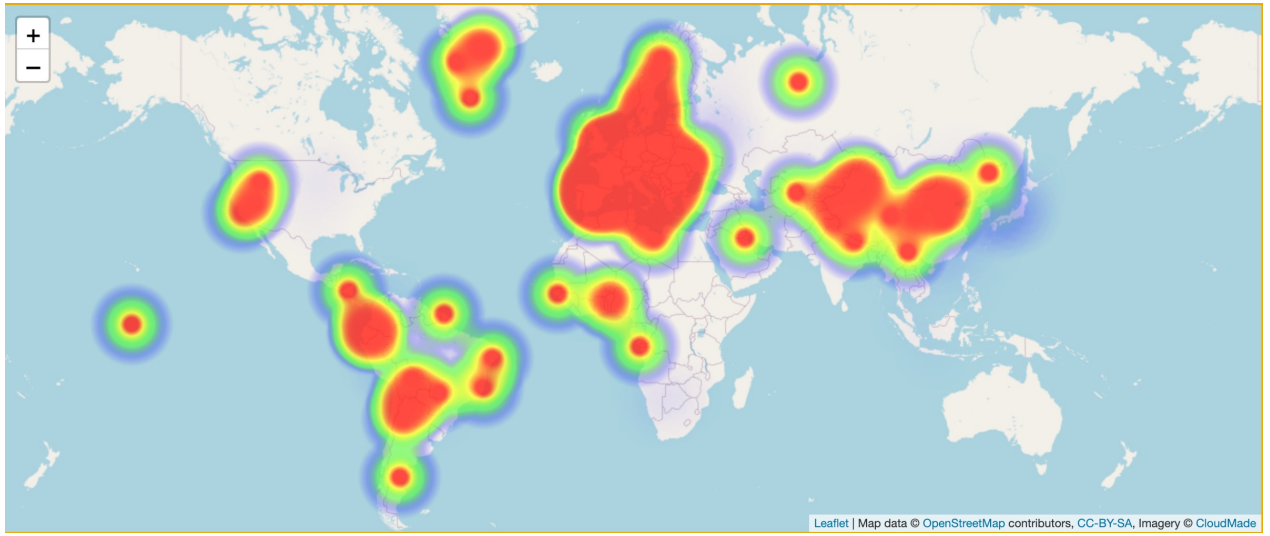
- **L3e** – Suggested to have originated in the Central Africa/Sudan region about 45,000 years ago during the upper paleolithic period.[27] It is the most common L3 sub-clade in Bantu-speaking populations.[28] L3e is also the most common L3 subclade amongst African Americans and Afro-Brazilians.[29]



mt: L3e2a1a



Y: R-M343



KS96 - ancestry results

colour code co-parentage EU/AFR admixture ME/admixed no report

Referred to co-parentage (4 labs used all markers), examples:

- „results are in line with one parent having African ancestry. Under the hypothesis of admixture, it is possible the other parent has European or ME ancestry“
- „either admixed with the male lineage likely European and the female lineage having an African/North-east African component, or from some area on the edge of Europe where there are genetic influences from various regions, e.g. the Middle East or North-east Africa“ – lab 9
- „mixed biogeographical background an African (probably maternal) - European (probably paternal) origin. The origin of the maternal line is presumably in Africa“

Assigned as of European and African admixture (2 labs used all markers):

- „either an admixed individual of AFR and EUR origin or stems from an admixed population“
- „double inclusion biogeographic ancestry with two root populations, 58.21% European and 28.29% African“

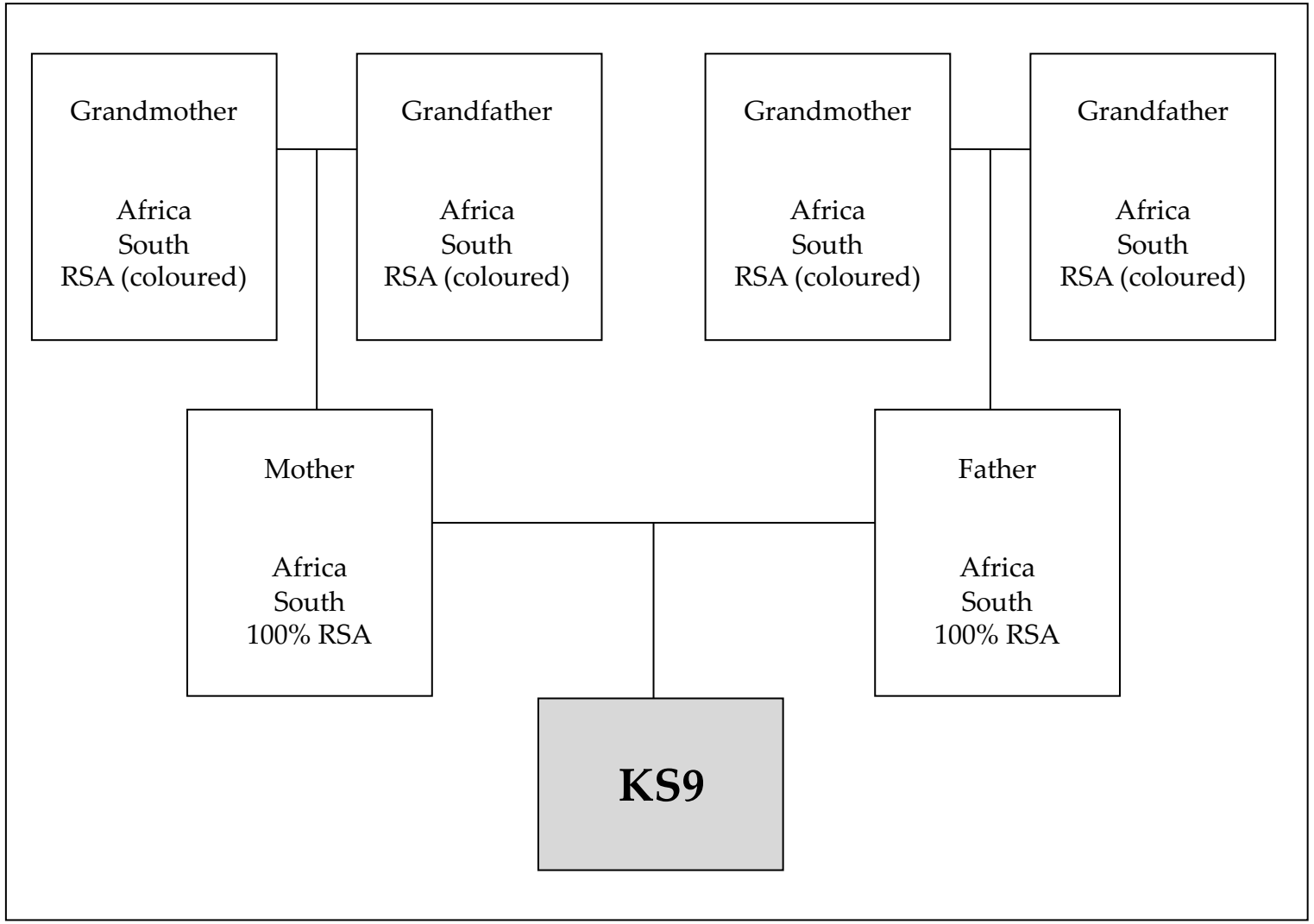
Assigined as of Middle East or admixed (*3 labs used only biparental*, 1 used all markers), examples:

- *„could be either of Middle-East ancestry or of Middle-East and European co-ancestry“*
- *„more likely if the investigated DNA comes from an individual from the Middle Eastern reference population than if the investigated DNA comes from an individual from the European, East Asian, South Asian, Oceanian, Native American, or African reference populations“*

Didn‘t assign to any reference population (1 lab used all markers, *1 used only biparental*):

- „most likely of admixed origin although it cannot be excluded that the ancestry is from a population not represented by the references database. East Asian or Oceania ancestry is not likely, but any other ancestry component cannot be excluded“
- *„prediction of biogeographic ancestry provide support against the unidentified individual's biogeographical ancestry being East Asia, America or Oceania. It cannot be clarified whether the ancestry is Africa, Europe, Middle East, South Asia or an admixture of these meta-populations“*

KS9 - ancestry results

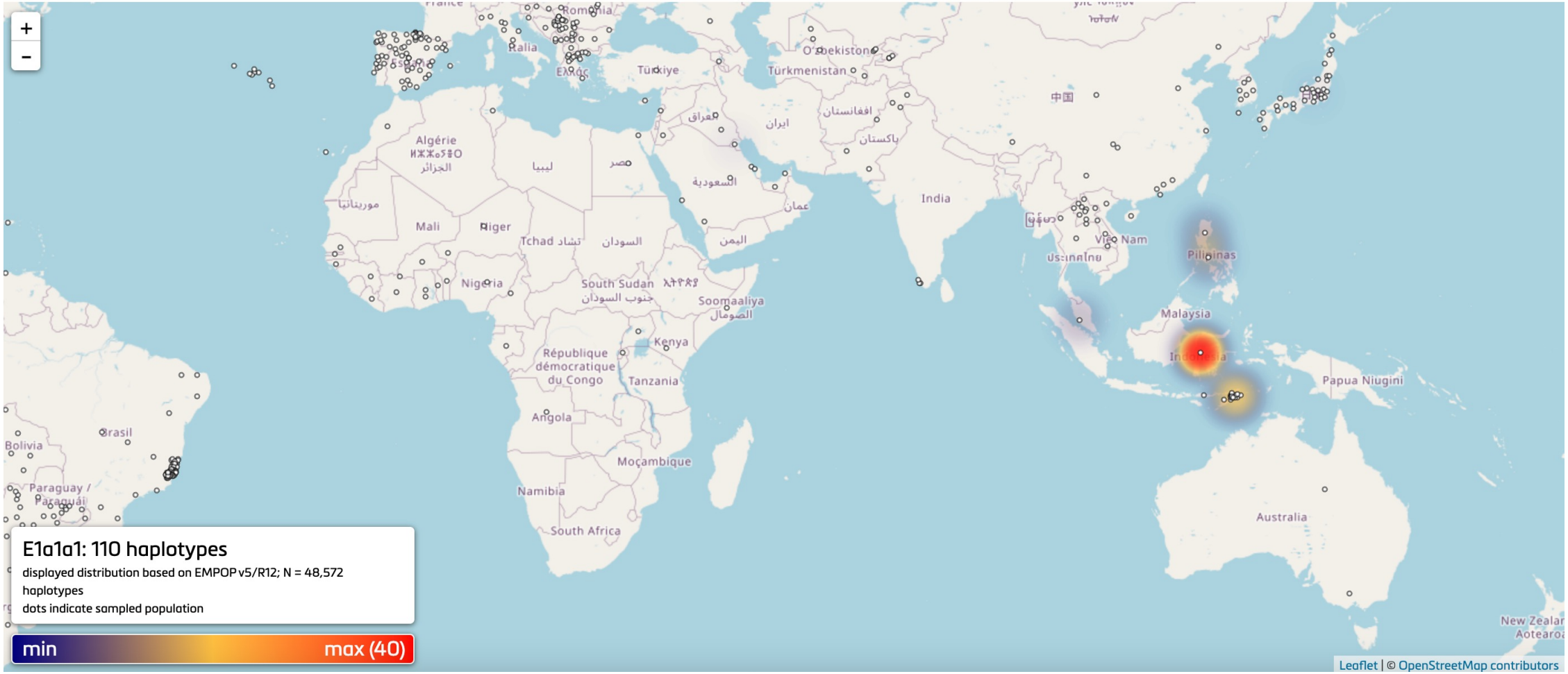


STRUCTURE

	PT	BT	ET
AFR	0.372	0.222	0.349
ME	0.475	0.600	0.326
EUR	0.026	0.090	0.111
SAS	0.049	0.019	0.016
EAS	0.047	0.029	0.192
OCE	0.018	0.011	0.003
AME	0.012	0.029	0.004

mt: E1a1a1
X: AFR & AFR

KS9 - ancestry results



KS9 - ancestry results

colour code: no report | AFR/ME/SEA admixture | AFR/ME admixture

Didn't assign to any reference population (*3 labs used only biparental*, 1 all markers), examples:

- *„the biogeographical origin of this individual could not be predicted"*
- *„not possible to find a most likely reference population or most likely genetic admixture for this sample"*

Assigned as African, Middle Eastern and SouthEast Asian admixture (4 labs used all markers), examples:

- „likely that the sample belongs genetically to the African, Middle Eastern and SouthEast Asian admixture population. It is likely that the maternal linage of the sample comes from SouthEast Asia"
- „predominant biogeographical origin is Middle Eastern or African, but in the maternal line an South East Asian origin is most likely (since autosomally East Asian is not strongly represented, this may be a deeper maternal ancestry). There is indication of admixture although a population in between that is not represented in the reference samples cannot be excluded"

Referred to phenotype prediction (1 lab used all markers):

- „either ancestry from the North-east African/Middle-East area or admixed ancestry with male ancestry from Africa and ancestry on the female side from somewhere in the Middle East/South Asia/South-East Asia region. Given the dark-to-black skin tone, the admixed ancestry option is the most likely"

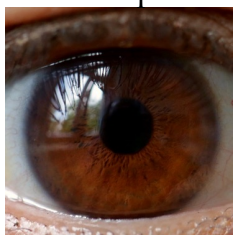
Assigned as of African and Middle East admixture (2 labs used all markers, *1 only biparental*), examples:

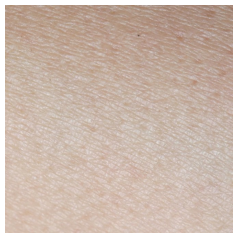
- *„biogeographical ancestry being Middle East or Africa, or a population with Middle East or African ancestry, or an admixture of the aforementioned meta-populations"*
- „likely to have a mixed biogeographical background with a dominant African - Middle Eastern origin"

KS71 - phenotype results

		71
		p-value
eye colour	*blue*	0,000
	intermediate	0,007
	brown	**0,993**
hair **colour**	*blond*	0,004
	brown	0,347
	red	0,000
	black	**0,649**
hair **shade**	*light*	0,005
	dark	**0,995**
skin colour	*very pale*	0,000
	pale	0,000
	intermediate	**0,705**
	dark	**0,264**
	dark to black	0,030

reference photos



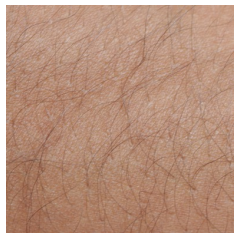


„brown eyes, brown to black / black hair and intermediate to dark skin“

KS96 - phenotype results

			96
			p-value
eye colour		*blue*	0,003
		intermediate	0,020
		brown	**0,977**
hair	**colour**	*blond*	0,038
		brown	**0,725**
		red	0,000
		black	**0,236**
	shade	*light*	0,103
		dark	**0,897**
skin colour		*very pale*	0,001
		pale	0,028
		intermediate	**0,262**
		dark	**0,685**
		dark to black	0,024

reference photos



„brown eyes, brown hair and dark skin“

„brown eyes, brown to dark brown hair and dark to intermediate skin“

„brown eyes, dark hair shade and brown to black hair, dark to intermediate skin“

„brown eyes, brown to dark brown hair and dark skin“

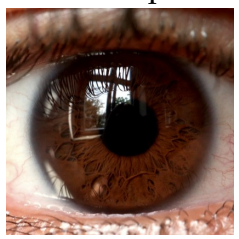
„brown eyes, dark brown hair, dark to intermediate skin“

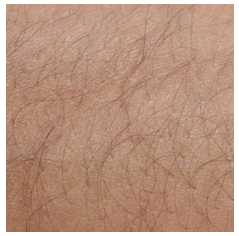
„brown eyes, dark brown hair, dark skin“

KS96 - phenotype results

		96
		p-value
eye colour	*blue*	0,003
	intermediate	0,020
	brown	**0,977**
hair colour	*blond*	0,038
	brown	**0,725**
	red	0,000
	black	**0,236**
hair shade	*light*	0,103
	dark	**0,897**
skin colour	*very pale*	0,001
	pale	0,028
	intermediate	**0,262**
	dark	**0,685**
	dark to black	0,024

reference photos





HPS guidelines:
brown/dark brown hair and dark skin
(natural: dark brown hair and dark skin)

„brown to dark brown hair and dark skin“
– 2 labs, HPS guidelines

„brown hair and dark skin“
– 3 labs, 1 with HPS and *2 with updated guidelines*

„dark brown hair, dark skin“
– 1 lab, HPS guidelines

„brown to dark brown hair and dark to intermediate skin“
– 1 lab, updated HPS guidelines

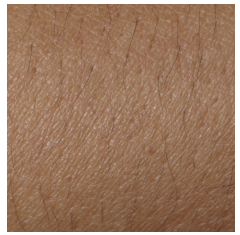
„dark hair shade and brown to black hair, dark to intermediate skin“
– 1 lab, HPS guidelines

„dark brown hair, dark to intermediate skin“
– 1 lab, HPS guidelines

KS9 - phenotype results

		9
		p-value
eye colour	*blue*	0,000
	intermediate	0,002
	brown	**0,998**
hair colour	*blond*	0,003
	brown	**0,474**
	red	0,000
	black	**0,523**
hair shade	*light*	0,005
	dark	**0,995**
skin colour	*very pale*	0,000
	pale	0,000
	intermediate	0,000
	dark	0,009
	dark to black	**0,991**

reference photos



„brown eyes, dark brown / black hair and dark to black skin“

Summary

- KS71: parental admixture STRUCTURE patterns were straightforward for all
- 50% reported uniparental maternal EAS, paternal EUR and 50% reported just admixture
- HIrisPlex-S skin colour test is focussed on EUR, so EAS components produced a dark prediction, highlighting the European-centric SNP selection in this SNP assay

- KS96: produced the most variation in reporting, from the specific: paternal parent EUR, maternal parent AFR, to "only can exclude EAS-OCE-AMR"
- Mito reporting suggests labs are taking EMPOP output without including the qualification that much of the 'global diaspora' distribution of AFR variation which should be necessary to include

- KS9: only sample where many labs refused to make any ancestry prediction
- the most complex patterns of genetic ancestry and consequent interpretations based on which panel, reference sets and software is used

- KS71 and KS9: FDP fully consistent - all probabilities over 0.7
- KS96: first sample where FDP had variable interpretation of the skin colour probabilities

What we learned from you

Which method of biparental SNP analysis did you find the most informative?

STRUCTURE
PCA
GenoGeographer
Converge

Did you find any of the data provided confusing and/or difficult to interpret?

PCA for admixed
BGA more difficult than phenotype
contradictory results
admixed samples
GenoGeographer
STRUCTURE
PCA lacking quantitative metrics

Which individual's data did you find the most difficult to analyse?

71
126
102
9
96
180

Did you refer to the HPS authors' guidelines or did you interpret the provided p-values using your own criteria?

HPS guidelines
own or both

Which additional markers besides the biparental SNPs were most useful for interpreting the data?

mtDNA
X-SNPs
Y-SNPs

Did the provided phenotype predictions have an impact on your conclusions regarding the biogeographic origin of the studied individuals?

no
somewhat

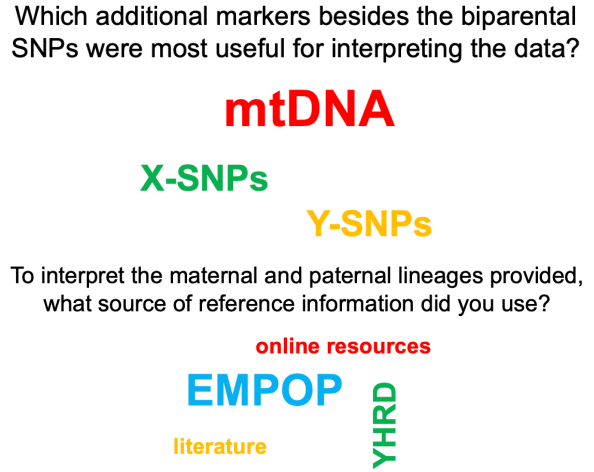
To interpret the maternal and paternal lineages provided, what source of reference information did you use?

online resources
EMPOP
YHRD
literature

Legend: the bigger words reflect the more common replies

General summary

- consistency of genetic patterns between the three autosomal MPS panel's data is important for participants confidence in reporting ancestry, especially when complex
- updated guidelines concerning phenotype reporting are needed
- guidelines concerning usage of common tools like STRUCTURE, SNIPPER and GenoGeographer are needed
- guidelines concerning interpretation of X-SNPs from ET are needed
- Y interpretation lacks a clear path - YHRD was the most common Y search system used, but it is not particularly informative for Y-SNP variation
- similar with EMPOP – cannot be used as only source of information
- uniparental markers significantly aid complex ancestry reporting but guidelines on how to use them are needed



React project

EDNAP

Summary of project so far

- Large consortium of more than 23 laboratories, where the primary purpose was to design experiments simulating typical casework circumstances; collect data and to implement Bayesian networks to assess the value (i.e., likelihood ratio) of DNA results given activity level propositions.

The ReAct project: Analysis of data from 23 different laboratories to characterise DNA recovery given two sets of activity level propositions

Peter Gill[a,b], Ane Elida Fonneløp[a,c], Tacha Hicks[d,e], Stavroulla Xenophontos[f], Marios Cariolou[f], Roland van Oorschot[g,h], Iris Buckel[i], Viktorija Sukser[j], Sunčica Papić[j], Siniša Merkaš[j], Ana Kostic[k], Angela Marques Pereira[l], Christina Teutsch[m], Christina Forsberg[n], Cordula Haas[o], Elizabet Petkovski[l], Fabian Hass[p], Jan Masek[q], Jelena Stosic[k], Yong Sheng Lee[r], Christopher Kiu-Choong Syn[r], Linda Groombridge[s], Marc Trimborn[p], Marilena Hadjivassiliou[f], Michelle Breathnach[t], Jana Novackova[q], Walther Parson[u,v], Petra Hatzer-Grubwieser[u], Sanna Pietikäinen[w], Simone Joas[m], Sascha Willuweit[p], Stefanie Grethe[x], Tamara Milićević[k], Therese Hasselqvist[n], Venus Kallupurackal[o], Vlastimil Stenzl[q], Staffan Jansson[n], Ingrun Glocker[x], Sarah Brunck[x], Karoline Nyhagen[a], Anne Berit Dyve Lingelem[a], Heli Autere[w], Devon Thornbury[y], Natalie Pedersen[y], Stephanie Fox[z], David Moore[aa], Gemma Escott[ab], Cathrine Bie Petersen[ac], Hans Jakob Larsen[ac], Rebecca Giles[ad], Paul Stafford Allen[ad], Ingo Bastisch[i,ae]

Key findings

- Can laboratory A share data with laboratory B?

Recovery varies between
200pg – 5ng

- Our latest work suggests around 10ng of DNA is deposited on a screwdriver when handled
- At best, around half is recovered
- At worst, around 98% is lost
- Median is 10% recovery

Laboratory	LastH Ex1	FirstH Ex2	FirstH Ex3.0	FirstH Ex3.1	Background Ex1
lab_10	202	1	2	1	1
lab_12	222	20	22	16	35
lab_6	301	117	23	5	56
lab_2_ESI	352	558	145	98	165
lab_15	383	81	1	1	1
lab_1_ESS	483	83	42	7	35
lab_1_NGM	488	90	41	9	30
lab_4_ESI	750	314	90	79	40
lab_4_NGM	750	275	111	68	45
lab_21	954	34	64	1	88
lab_9	1073	252	76	1	7
lab_14	1088	1440	87	13	82
lab_8	1921	279	148	25	35
lab_3_ESX	2077	218	114	28	59
lab_3_NGM	2087	310	183	34	6
lab_16	2169	1073	1	1	1
lab_18_F6C	2301	344	73	1	77
lab_18_GOF	2326	289	56	19	58
lab_5	2713	143	1	1	1
lab_13	3143	787	193	91	564
lab_2_NGM	4012	233	41	1	1
lab_17	4959	942	169	113	10
Medians	**1080**	**263**	**68**	**11**	**35**

Table 3: Tabulated raw medians of selected experiments and LastH, FirstH and background recoveries in pg. Ranked by LastH, experiment 1. The median value of medians is shown in the final row.

Back to the future (basics)

- Obviously, if we lose all the DNA from an extraction, we are not going to get much useful information
- In our pursuit of new methods and ever-increasing complexity of technology, have we forgotten something important?

Standardisation

- React project is the first to generate a large database of results: more than 2,700 samples at present
- New data are being accepted
- There are no restrictions on who joins in – i.e open to EDNAP members are everyone else
- However, there is more to do with standardisation, and this is also linked with normalisation of data
- React II will investigate a method where a lab can determine its recovery efficiency. This will be used as a factor which will help normalise data so that lab A can share data with lab B

The problem with publications

- Everyone has assumed that Transfer, Persistence and Recovery are lab independent:

$$p(TPR) = \underbrace{p(T) \times p(P)}_{lab\ \ independent} \times \underbrace{p(R)}_{lab\ \ dependent}$$

- But they are not
- So this has implications

Updated formula

$$p(TPR) = \underbrace{p(TP)}_{lab\ \ independent} \times \underbrace{p(R|TP)}_{lab\ \ dependent}$$

The problem with publications

- Data cannot be easily shared between labs to inform probabilities (because recovery rates are different)
- There is no overall standard for kind of data collected, or the method used to analyse
- Usually the data are not published, or are insufficient
- This also makes peer review very difficult, because we are effectively accepting data on trust
- Important to note that with the ReAct project, all data are published and furthermore all the results in the paper were created from the published data
- ReAct provides a model for the community to follow

Publications – bottom line

- Current publications represent the experience of the lab generating the paper
- The results may be interesting, but they cannot be directly used by other laboratories, as we have insufficient information available, lack of standardisation, and we work on the false premise that the probability of TPR is the same across all labs.

What is to be done

- Let’s recognise there is a problem
- Consensus view needed on standardisation of datasets
- Also, a method is needed to help normalise data between labs so data can be shared (in progress with ReAct II)
- Central databases needed – validation question of course – but we should follow example of STRidER
- In addition, we need to identify those Bayesian networks most used in casework
- So BNs, extraction efficiency and databases and the data used in them are closely linked

Summary

- In summary, the aim is to standardise data collection; define the raw data to be included as open-access data; define a series of open source Bayesian networks that can be used to capture the majority of casework. The ReAct project has published an extensive database which will hopefully kick-start the transition; there is an open invitation for any laboratory to contribute. It is envisaged that the ReAct database can follow the example already provided by STRider for collection of frequency databases. Ideally, it should eventually be placed under the auspices of a scientific society such as ISFG.
- Open to EDNAP members

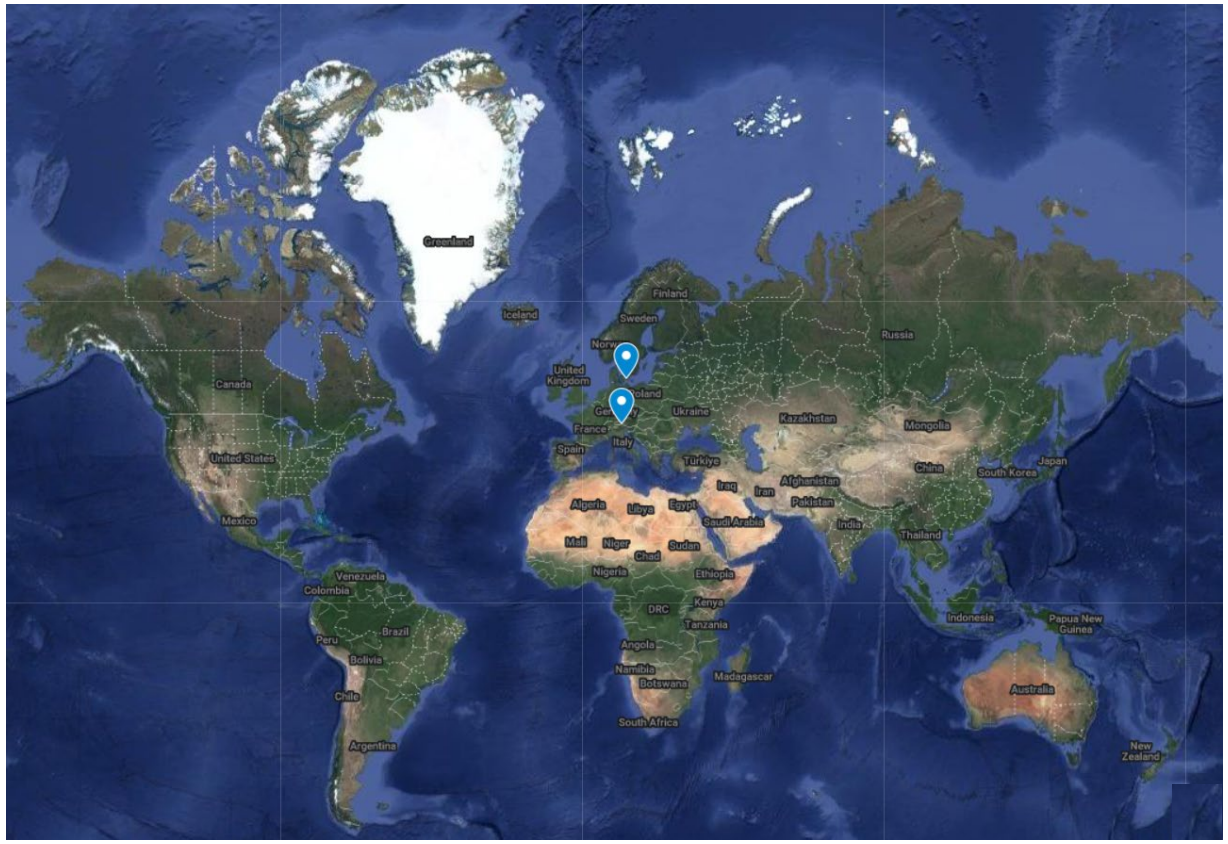
MITOMETRICS:

Characterisation of mitochondrial DNA heteroplasmy in hair shafts and its incorporation on likelihood ratio calculations

Vânia Pereira
Section of Forensic Genetics
Department of Forensic Medicine
University of Copenhagen

KØBENHAVNS UNIVERSITET

About the MitoMetrics initiative



To bring together group of scientists and practitioners working with mitochondrial DNA in the field of forensic genetics to investigate research questions dealing with the analysis and interpretation of mtDNA in forensic genetic casework.

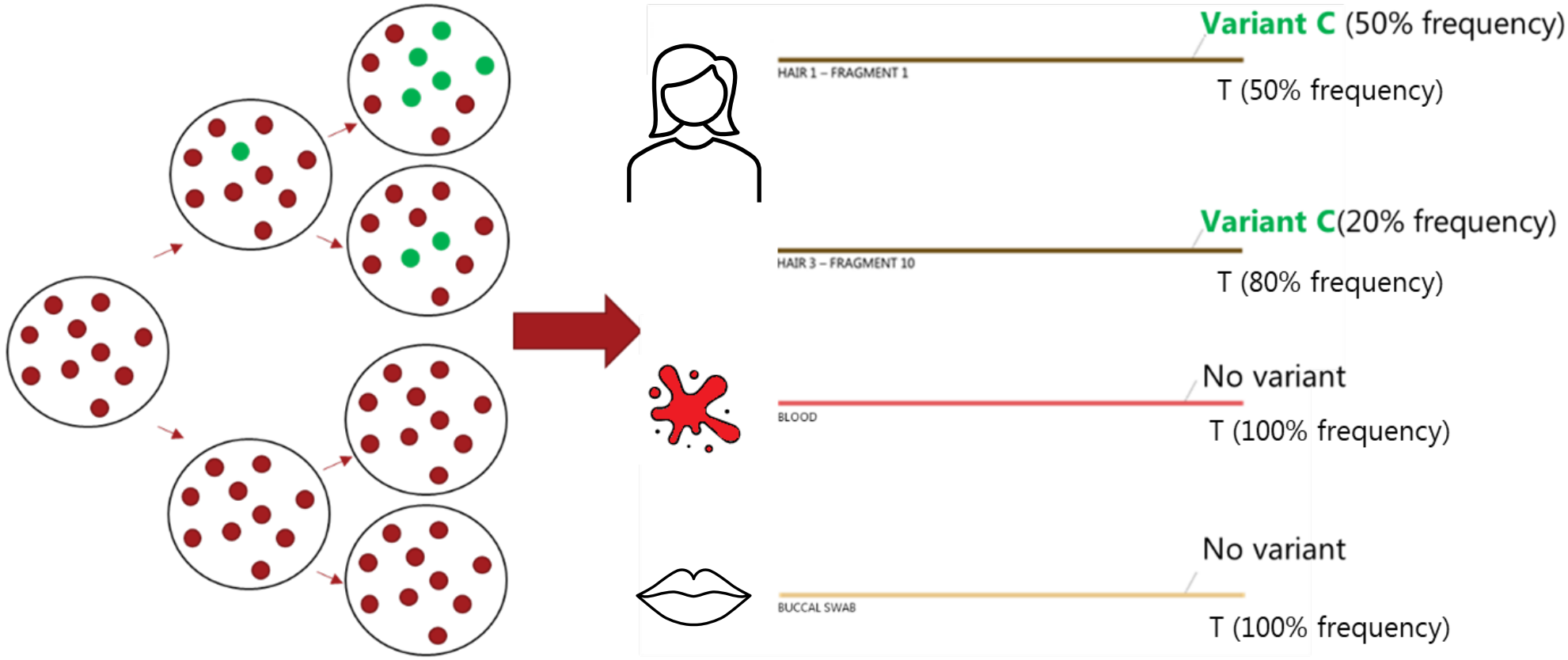
About the MitoMetrics initiative

Goal:
Improve the interpretation of mtDNA forensic evidence

Current forensic interpretation guidelines for mtDNA evidence are ruled-based conventions

Phylogenetic knowledge suggests that this approach is outdated and should instead be based on ground truth data.

mtDNA heteroplasmy

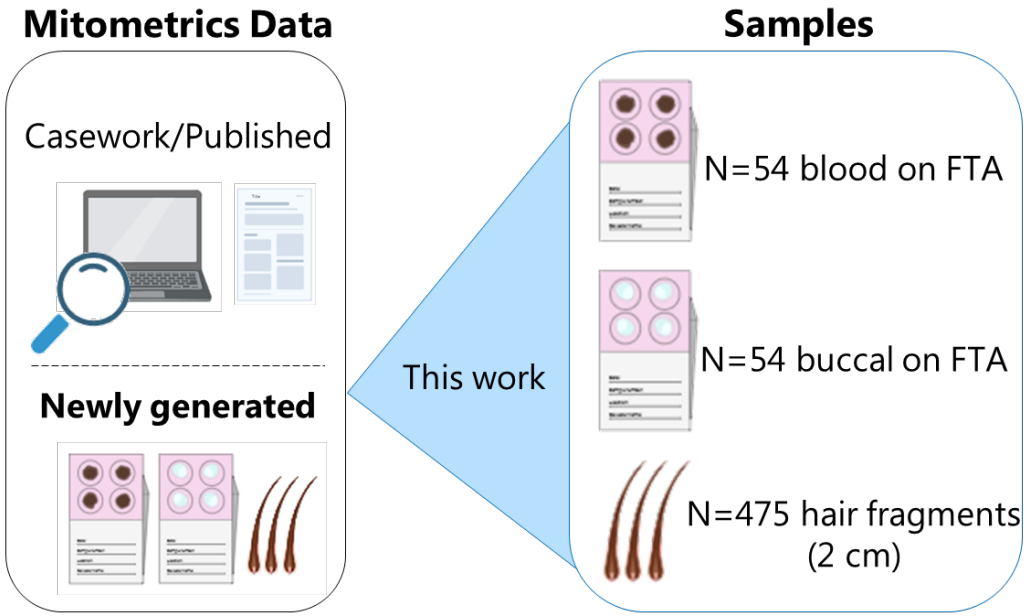


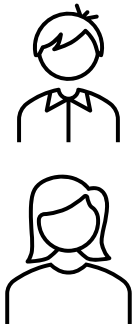
Aims

- Investigate heteroplasmy and the discrepancies in mtDNA profiles from hair shafts and reference samples from the same donor.

- Develop an initial model for calculating the weight of mtDNA-based evidence that incorporates the number of discrepancies observed.

Methods





- Blood drop on FTA card
- Buccal swab on FTA card
- 3 hairs, min. 3 fragments/hair

Methods

Mitometrics Data

Casework/Published

Newly generated

This work

Samples

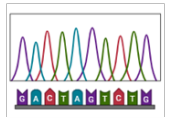
N=54 blood on FTA

N=54 buccal on FTA

N=475 hair fragments (2 cm)

Sequencing

N=120 Sanger (HVI+HVII)



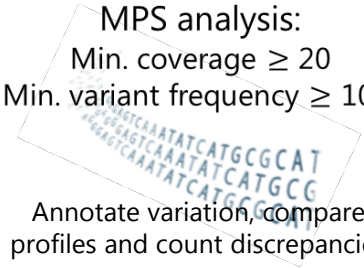
N=76 also MPS (HVI+HVII)

N=463 MPS (whole mtDNA)

N=55 iSeq N=410 Ion S5

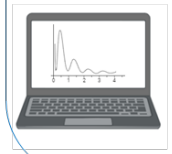
Data Analysis

MPS analysis:
Min. coverage ≥ 20
Min. variant frequency ≥ 10%



Annotate variation, compare profiles and count discrepancies

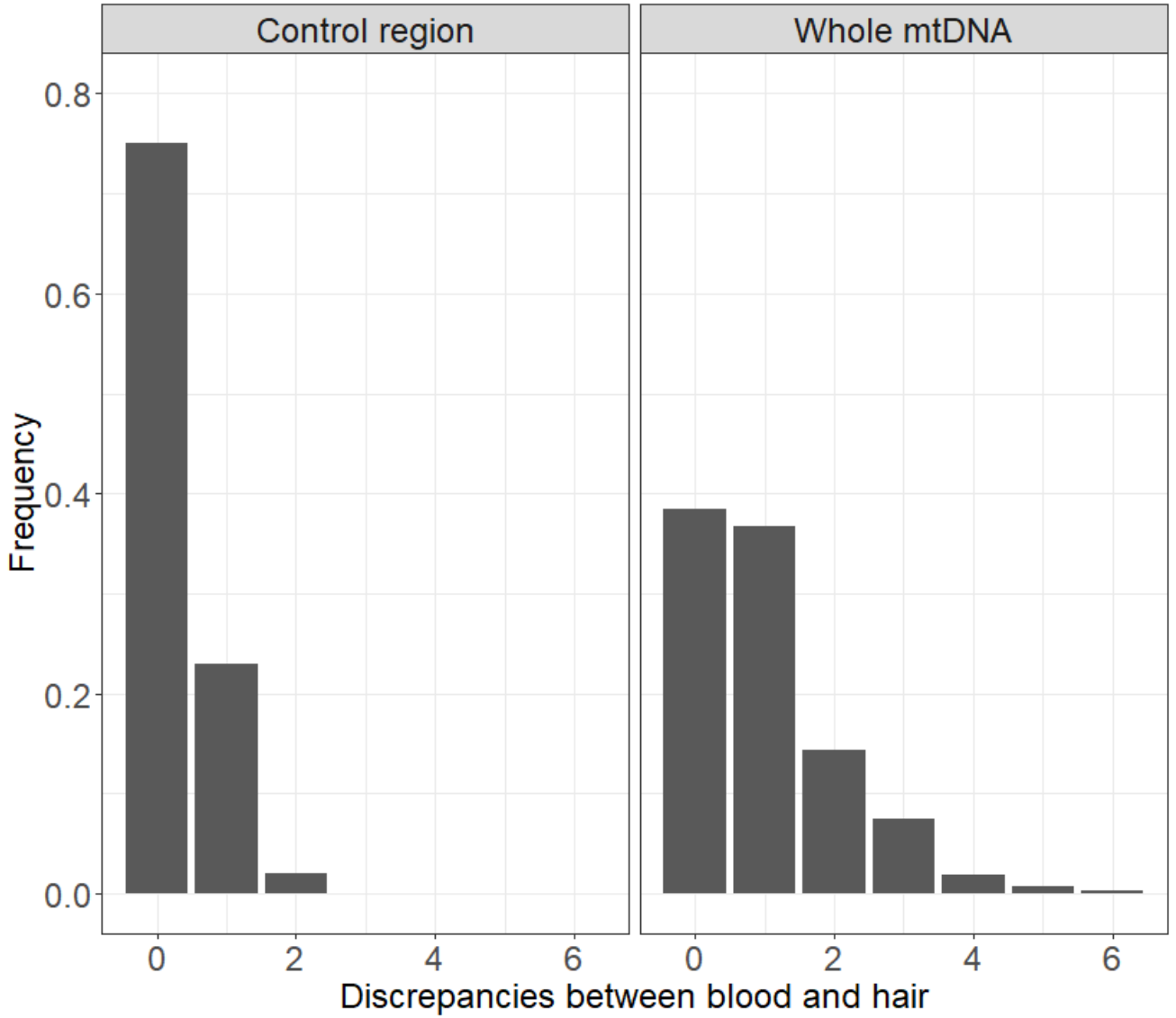
Statistical modelling:



Estimate Poisson for use in LR

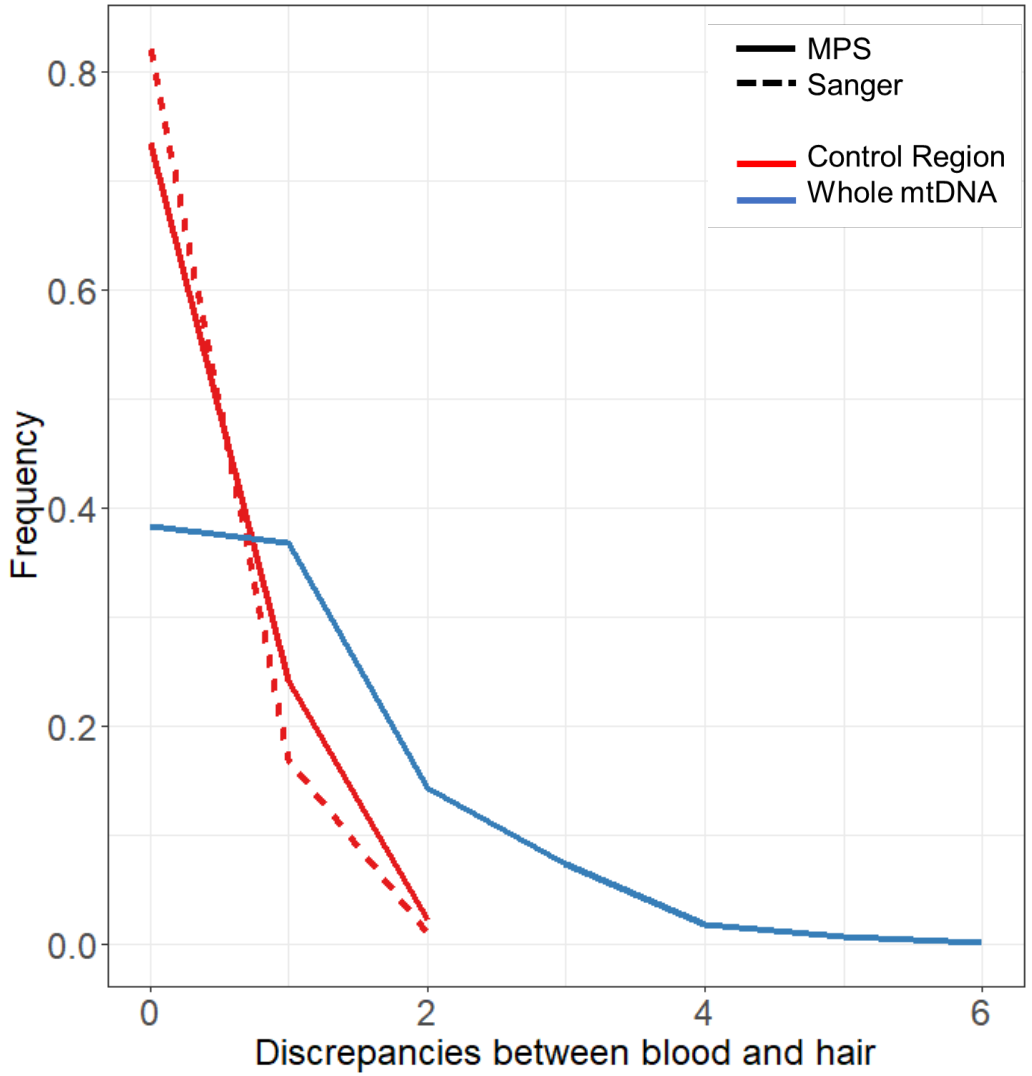
Heteroplasmic discrepancies between blood and hair

- In the control region, most comparisons between blood and hair had no discrepancies reported
- When considering the whole mtDNA molecule, some comparisons revealed 4 or more discrepancies, although these were not very common.



Statistical modelling of heteroplasmy

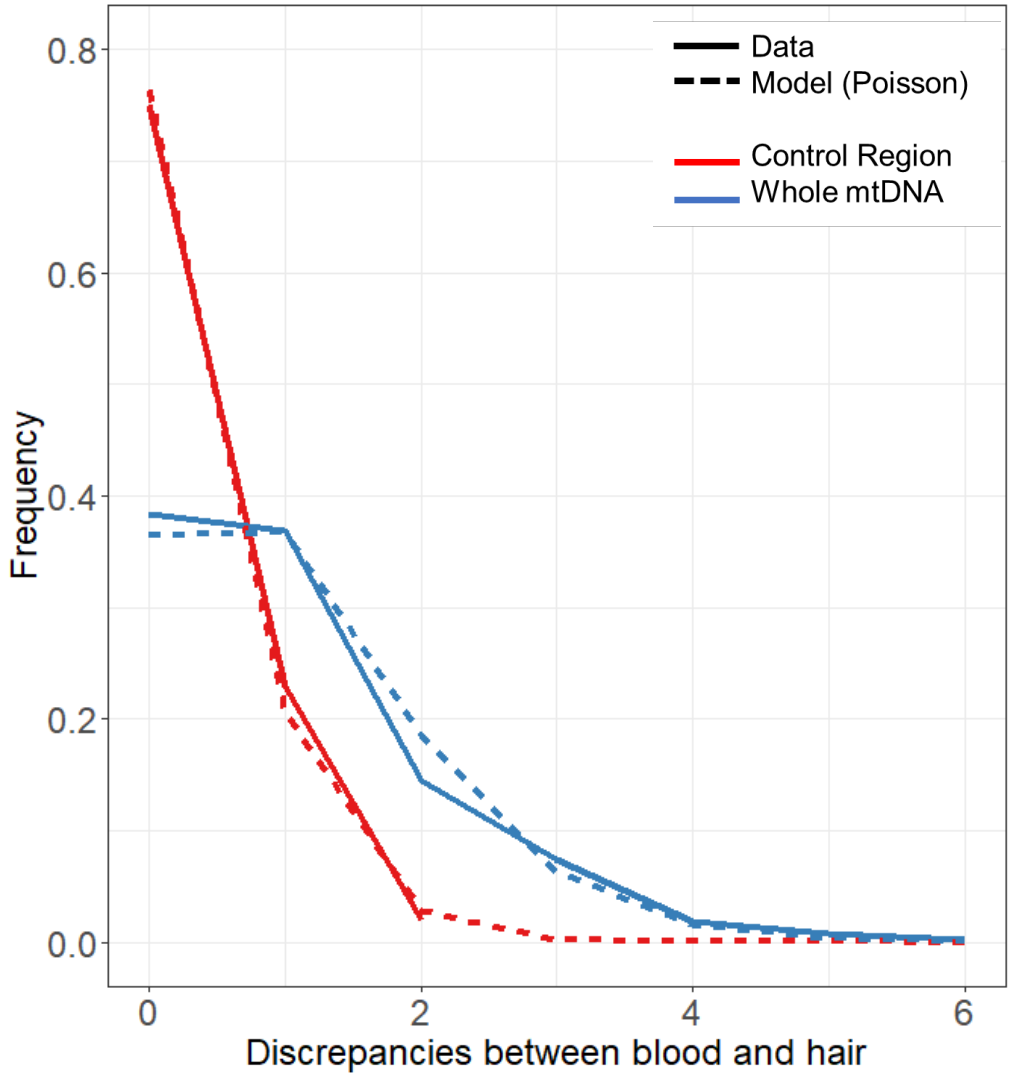
- Similar results were observed for Sanger sequencing and MPS data



Statistical modelling of heteroplasmy

- The frequency of observed discrepancies between reference and trace samples due to heteroplasmy resembled a Poisson distribution

- We can now model these events, and this information can be incorporated into a likelihood ratio (LR) calculation.



Incorporation of heteroplasmy in likelihood ratio calculations

- If the reference and the crime scene samples originate from different biological material from the same individual (blood reference and trace sample from hair, for example) we can then consider:

$$LR = \frac{P(D)}{P(X_Q)}$$

Where:

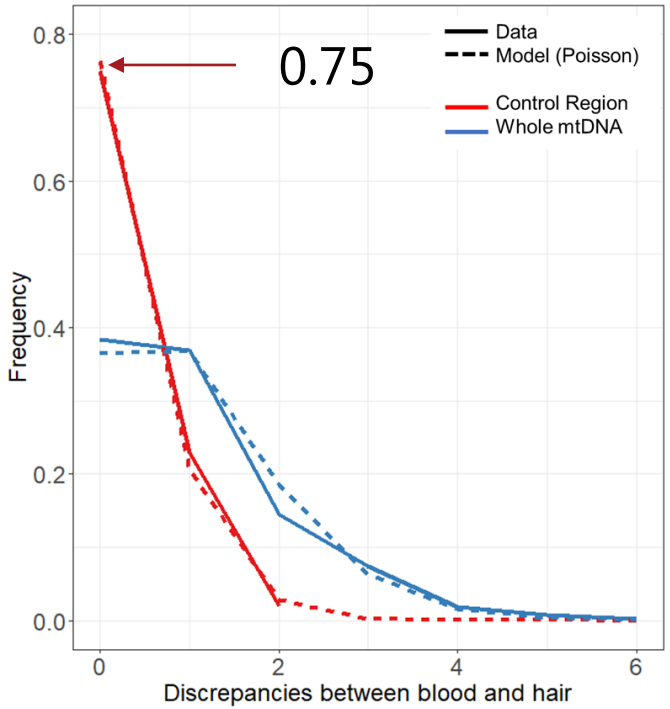
P(D): the frequency of heteroplasmic discrepancies between blood and hairs in a population

$P(X_Q)$: the frequency of the mtDNA profile of the trace in a population

Incorporation of heteroplasmy in likelihood ratio calculations

Assuming that:

- *P(D)* follows the model considering the mtDNA control region
- $P(X_Q)$ = 1/15782 = 6.34×10^{-5}, based on the size of the Westeurasian database of control region mtDNA profiles in EMPOP



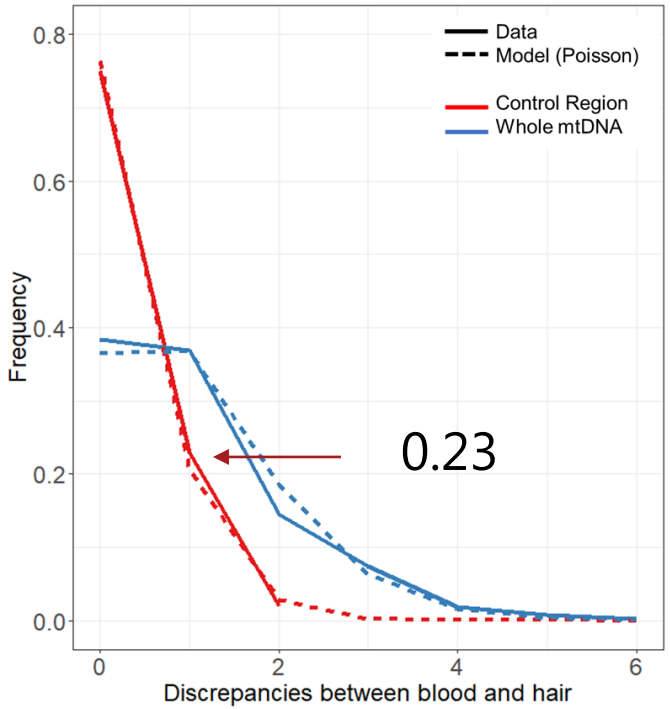
If D = 0 (no observed discrepancies between the profiles):

$$LR = \frac{P(D = 0)}{P(X_Q)}$$

Incorporation of heteroplasmy in likelihood ratio calculations

Assuming that:

- *P(D)* follows the model considering the mtDNA control region
- $P(X_Q)$ = 1/15782 = 6.34×10^{-5}, based on the size of the Westeurasian database of control region mtDNA profiles in EMPOP



If D = 1 (1 observed discrepancy between the profiles):

$$LR = \frac{P(D = 1)}{P(X_Q)}$$

Conclusions and Considerations

- Attempt to calculate the value of mtDNA-based evidence using LR, that considers the possibility of heteroplasmy in different tissues from the same donor.

- Other possible causes of discrepancies such as degradation, sequencing errors, and the effect of methodology used need to be further explored.

Conclusions and Considerations

Future work and further models should address:

→ How common/rare heteroplasmic positions (and the relative frequencies of the alleles) affect the LR

→ How heteroplasmy frequencies vary in and within different phylogenetic lineages

→ If mtDNA reference databases from different biological sources can be used to estimate the weight of the evidence in other tissues, or if any correction factors are needed

Acknowledgements

Floor Claessens

Christina Amory
Gabriela Huber
Walther Parson

Mikkel Meyer Andersen
Tóra Olsen

Cristina Arévalo
Cristina Pie
Lourdes Prieto

Stijn Desmyter

Kris J van der Gaag

Jodie Ward
Kelly Grisedale

Katarzyna Skonieczna
Tomasz Grzybowski

Chiara Turchi
Filomena Melchionda

Hiroaki Nakanishi