

The EDNAP collaborative study on mtDNA heteroplasmy detection using Sanger and Massively Parallel Sequencing technologies[☆]

Lena Ewers^a, Gabriela Huber^a, Katja Anslinger^b, David Ballard^c, Pedro A. Barrio Caballero^d, Magdalena Bogus^e, Claus Børsting^f, Michael D. Brandhagen^g, Francesca Brisighelli^{h,i}, Lauren Canale^j, Gianluca Casamassima^k, Vincent Castella^l, Jennifer Daniels-Higginbotham^m, Stijn Desmyterⁿ, Marta Diepenbroek^b, Sophie Dognauxⁿ, Bethany Forsythe^{o,p}, Kristiaan J. van der Gaag^q, Oscar Garcia^r, Federica Giangasparo^c, Nair Gouveia^s, Mario Gysi^t, Cordula Haas^t, SallyAnn Harbison^{o,u}, Johannes Hedman^{v,w}, Mitchel M. Holland^{x,y}, Beatriz Heinrich Roson^d, Kevin M. Kiesler^z, Odile Loreille^g, Sabine Lutz-Bonengel^{aa}, Alberto Marino^k, Charla Marshall^{m,ab}, Patrik Matejov^{ac}, Jennifer A. McElhoe^x, Marta Montesino^{ad}, Marie-Laure Morerod^l, Ana Mosquera-Miguel^{ae}, Jana Naue^z, Leire Palencia-Madrid^r, Vincenzo Pascali^{af,ag}, Vania Pereira^f, Maria de la Puente^{ae}, Erica L. Romsos^y, Lisa Sampaio^s, Peter Schneider^{e,1}, Cibeles Serna-Menor^{ad}, Bing Hong Shue^{ah}, Maja Sidstedt^v, Titia Sijen^q, Kimberly Sturk-Andreaggi^m, Christopher Kiu-Choong Syn^{ah}, Claudia Wöss^a, Nicole Huber^a, Arne Dür^{ai}, Walther Parson^{a,x,*}

^a Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck 6020, Austria

^b Institute of Legal Medicine, the Ludwig Maximilian University Munich, Munich 80336, Germany

^c King's Forensics, King's College London, London SE1 9NH, UK

^d National Institute of Toxicology and Forensic Sciences, Madrid Department, Spain

^e Institute of Legal Medicine, University of Cologne, Cologne 50823, Germany

^f Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, Copenhagen DK-2100, Denmark

^g FBI Laboratory, DNA Support Unit, Quantico, VA, USA

^h Departmental Unit of Molecular and Genomic Diagnostics, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome 00168, Italy

ⁱ Genomics Research Core Facility, Gemelli Science and Technology Park (GSTeP), Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome 00168, Italy

^j California Department of Justice, Jan Bashinski DNA Laboratory, Richmond, CA 94804, USA

^k Reparto Carabinieri Investigazioni Scientifiche di Parma, Sezione di Biologia, Strada della Fonderie 10, Parma 43100, Italy

^l Unit of Forensic Genetics, University Centre of Legal Medicine, Lausanne - Geneva, Lausanne University Hospital and University of Lausanne, Chemin de la Vulliette 4, Lausanne CH - 1000, Switzerland

^m SNA International, LLC, Contractor Supporting the Armed Forces DNA Identification Laboratory (AFDIL), Alexandria, VA 22314, USA

ⁿ National Institute for Criminalistics and Criminology (NICC), Brussels B-1120, Belgium

^o Forensic Science Programme, School of Chemical Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

^p International Commission on Missing Persons, The Hague 2514 AA, the Netherlands

^q Department of Human Biological Traces, Netherlands Forensic Institute, Laan van Ypenburg 6, The Hague 2497 GB, the Netherlands

^r Forensic Science Unit, Forensic Genetics Section, Basque Country Police, Erandio, Bizkaia, Spain

^s Serviço de Genética e Biologia Forenses, Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Delegação do Centro, Largo da Sé Nova, Coimbra 3000-213, Portugal

^t Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland

^u New Zealand Institute for Public Health and Forensic Science, Private Bag 92021, Auckland 1142, New Zealand

^v National Forensic Centre, Swedish Police Authority, Linköping SE-581 94, Sweden

^w Division of Biotechnology and Applied Microbiology, Department of Process and Life Science Engineering, Lund University, Lund SE-221 00, Sweden

^x Forensic Science Program, Department of Biochemistry & Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA

[☆] Given their role as Editor-in-Chief, Walther Parson had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to another journal editor

* Corresponding author at: Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck 6020, Austria.

E-mail address: walther.parson@i-med.ac.at (W. Parson).

<https://doi.org/10.1016/j.fsigen.2026.103593>

Received 29 May 2026; Received in revised form 21 July 2026; Accepted 21 July 2026

Available online 23 July 2026

1872-4973/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

^y Mitotyping Technologies, an Actuated Medical Company, Bellefonte, PA, USA^z US National Institute of Standards and Technology, Gaithersburg, MD, USA^{aa} Institute of Forensic Medicine, Medical Center, University of Freiburg, Medical Faculty, University of Freiburg, Germany^{ab} present address: DNAID Services LLC, Delaware, USA^{ac} Institute of Forensic Science, Ministry of Interior Slovakia, Slovakia^{ad} DNA Laboratory, Comisaría General de Policía Científica, Madrid, Spain^{ae} Forensic Genetics Unit, Institute of Forensic Science, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain^{af} Department of Healthcare Surveillance and Bioethics, Section of Legal Medicine, Università Cattolica del Sacro Cuore, Rome, Italy^{ag} Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy^{ah} DNA Profiling Laboratory, Biology Division, Health Sciences Authority, 11 Outram Road, Singapore 169078, Singapore^{ai} Institute of Mathematics, University of Innsbruck, Austria

ARTICLE INFO

Keywords:

Mitochondrial DNA

Next Generation Sequencing

Homopolymeric poly(C)-tracts

Interpretation

Forensic Genetics

ABSTRACT

This EDNAP (European DNA Profiling Group) collaborative study evaluated the detection and reporting of mitochondrial DNA (mtDNA) point heteroplasmy (PHP) and length heteroplasmy (LHP) across forensic laboratories using Sanger, Ion Torrent, and Illumina sequencing platforms. Standardized DNA extracts were centrally prepared and distributed to participating laboratories to assess inter-laboratory concordance and platform-specific (platform = instrument + software) effects. Raw sequence data showed high overall consistency across platforms, and observed mitotype discrepancies were due to transcription errors. PHP detection was generally concordant across laboratories and technologies. MPS data provided relatively stable mixture ratios across platforms. Low-level PHPs were often not reported in this study when they fell below laboratory-specific minor allele frequency thresholds, although raw data review confirmed their presence. In contrast, LHP interpretation showed substantially greater variation. Sanger-derived LHP patterns could be harmonized through review under current ISFG guidelines, whereas MPS-based LHP results differed both between and within sequencing technologies. Standardization of the alignment workflow improved concordance to some extent, but important discrepancies remained, particularly in Ion Torrent HVS-I (Hypervariable Segment I) poly(C)-tract data. These differences were attributable to sequencing technology, analysis software, and human interpretation.

1. Introduction

Mitochondrial DNA (mtDNA) heteroplasmy was first described in humans in 1983, when intraspecific sequence and length variability was observed in the non-coding region [1]. Heteroplasmy refers to the coexistence of multiple mtDNA types within a single individual that differ either by sequence (point heteroplasmy, PHP) or by size (length heteroplasmy, LHP). The first forensic observations of PHP emerging in the mid-1990s were found during the identification of the Romanov family remains [2]. A mixture of cytosine (C) and thymine (T) that was observed at nucleotide position (ntp) 16169 in the putative remains of Tsar Nicholas II was confirmed through cloning, providing one of the first unambiguous examples of Control Region (CR) PHP. The 16169Y PHP was also observed in the remains of the Tsar's brother, Grand Duke Georgij Romanov [3] and further confirmed by studies on the Tsar's remains employing more advanced sequencing methods [4,5] demonstrating the stability of PHP detection across technologies.

One of the first U.S. cases in which mtDNA evidence was admitted in court was *State of Tennessee v. Paul Ware* (1996). In this case, mtDNA analysis of hair recovered from the victim's throat and the crime scene revealed PHP in the CR [6]. The complexity of PHP transmission in living individuals was demonstrated by [7] and [8] through the analysis of members of a maternal pedigree. In the first study, blood and buccal swab samples from a mother and her two children exhibited varying levels of C and T at ntp 16355, while hair samples from the same individuals showed either homoplasmy or different mixtures of C and T at that site, which highlighted the dynamic nature of mtDNA heteroplasmy across tissues (see also [9]). The second study [8] reported on the rate and pattern of intergenerational sequence substitutions, including observations of heteroplasmy in 134 lineages across 327 generational events.

The European DNA Profiling (EDNAP) group conducted a collaborative multicenter study in the early 2000s to investigate mtDNA heteroplasmy in hair shafts [10]. Using a single donor with PHP 16234Y in

blood, 55 hair shafts were individually sequenced by ten laboratories. Results showed differential segregation of heteroplasmic variants: some hairs had only C, others only T, and some a mix at 16234, often at different ratios. These findings were confirmed by follow-up studies including one study that analyzed over 450 hair fragments from 21 donors using Sanger sequencing [11], a massively parallel sequencing (MPS) study assessing drift of heteroplasmy at different levels of minor variant frequency [12], and another study using MPS to analyze 475 hair samples along with 26 buccal references [13]. The latter studies demonstrated that MPS could detect low-level heteroplasmic variants often missed by Sanger sequencing, thereby reducing apparent mismatches between hair and buccal samples in some cases.

The largest Sanger-based systematic study on CR heteroplasmy analyzed 5000 high-quality CR sequences from global populations, based on multiple reads of forward and reverse strands, conducted independently in two laboratories, reported that PHP and LHP was found in blood and buccal samples in approximately 6% and 52% of the tested individuals, respectively [14]. The observed LHP patterns were strongly correlated with haplogroup backgrounds, particularly in the poly(C)-tracts within the hypervariable segments HVS-I (around ntp 16193) and HVS-III (around ntp 573), while LHP in HVS-II (around ntp 309) seemed largely independent of the phylogeny. PHPs were typically limited to single instances per individual, with only rare cases showing two or three heteroplasmic sites across the CR.

With the advent of MPS technologies, the Minor Allele Frequency (MAF) thresholds used to define PHP have decreased significantly, from the 10–20% range established for Sanger sequencing to values below 10%, in some cases as low as 1–2% [15,16]. However, the authors cautioned against uncritical interpretation of low-level variants, as artefacts such as post-mortem damage, nuclear mtDNA elements (NUMTs), and miscoding lesions can occur at similar frequencies, potentially leading to false-positive PHP calls. Consequently, as noted in [17], MPS-generated mtDNA data may not always be suitable for confidently calling very low-level PHP. More conservative and forensically robust thresholds, in the order of approximately 10% MAF, are generally recommended, particularly for casework samples, which are more prone to background noise, damage patterns, degradation,

¹ Prof. Peter M. Schneider deceased before publication

contamination and NUMTs (e.g., [18]).

A study comparing Sanger- and MPS-based data (Illumina and Ion Torrent) revealed different LHP patterns in poly(C)-stretches of high-quality DNA reference samples [19]. The authors concluded that these discrepancies were not only due to the sequencing technologies themselves but also to the alignment software used. Such differences in LHP detection between sequencing platforms could impact the interpretation of mtDNA in forensic cases. The results of this study prompted the EDNAP group to conduct a systematic multicenter study using defined control samples to evaluate the detection of PHP and LHP across various sequencing platforms, including both instrumentation and software. The collaborative study on mtDNA heteroplasmy compared how consistently participating laboratories detected PHP and LHP in standardized DNA extracts using their respective platforms.

2. Materials and methods

The study employed a cross-sectional observational design with descriptive analysis. A total of five DNA extracts were taken from two commercially available cell lines, 007 (CS-1; lot 1707017; Applied Biosystems by Thermo Fisher Scientific (TFS), Waltham, Massachusetts, USA) and 9947 A (CS-2; lot 18961603; Promega, Madison, Wisconsin, USA), two buccal swabs (CS-3, CS-5) from two volunteers under informed consent (Ethical Commission of the Medical University of Innsbruck; Nr. 1086/2017) and one blood sample (CS-4) from a previous study [20] by the organizing laboratory (Lab 0). DNA was extracted from the two buccal swab samples using the Chelex method, while DNA from blood sample was extracted using the Qiagen BioRobot M48 Robotic Workstation (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The commercially obtained cell line samples were already provided as DNA extracts. Batches of identical aliquots of these five samples were prepared in total volumes of 20 µL and DNA concentrations were set at 0.1 ng/µL based on SDquant results (for details see [21]). In brief, two mtDNA-specific targets (69 and 143 bp) and a nuclear DNA-specific target (70 bp) were used to simultaneously quantify the DNA in a single multiplex PCR assay on an ABI Prism 7500 instrument (TFS).

Samples were distributed to 25 forensic laboratories in sterile 1.5 mL microtubes (Sarstedt AG & Co. KG, Nümbrecht, Germany) and shipped at ambient temperature. Six laboratories quantified DNA of the shipped extracts upon arrival using SD quant ([21]; Lab 2), Quantifiler Trio (TFS; Lab 1), a published protocol according to [22] (Lab 11 and Lab 14) and a home-made, unpublished assay (Lab 16). Where feasible, degradation indexes were evaluated by dividing the quantification values of the smaller target by those of the larger target.

Participating laboratories were invited to sequence mtDNA from the five DNA extracts following their established methods and protocols and to provide the results according to their interpretation guidelines (see Supplementary Table S1a-c).

The reported mitotypes, as well as the sequence raw data, were sent back to Lab 0 for comparative analyses. Lab 0 collated the reported mitotypes and instances of heteroplasmy and, in an additional step, all submitted raw sequencing data were independently reviewed by two experienced analysts following the revised mtDNA guidelines of the International Society for Forensic Genetics (ISFG; [23]). The reviewers were not involved in the initial sequence interpretation and were blinded to the respective sequence variants prior to assessment. Final interpretations were based on consensus between both reviewers. This procedure was used to minimize potential ascertainment bias when evaluating whether reported or non-reported variants were supported by the underlying raw sequencing data. We note that the analysis by Lab 0 was not considered "ground truth", and the purpose of this analysis was not to classify individual laboratory results as correct or incorrect. Rather, the aim was to identify and describe discrepancies in variant reporting and interpretation between laboratories. Due to its semi-quantitative nature PHP was called for Sanger-data by visual

inspection of the peaks relative to the background noise rather than actual peak height values. For MPS-derived data, binary alignment (BAM) files were scrutinized and also re-analyzed in Lab 0 using in-house software that follows the phylogenetic alignment principles for human mtDNA, published in [24]. Minimal sequence threshold was set at 20 reads and the following settings were applied: for Ion Torrent, total read depth per position ≥ 20, PHP call ≥ 5%, insertion call ≥ 20%, deletion call ≥ 30%, factor for forward/reverse counts ≥ 20%; for Illumina, total read depth per position ≥ 20, PHP call ≥ 5%, insertion call ≥ 5%, deletion call ≥ 5%, factor for forward/reverse counts ≥ 20%. Length heteroplasmic regions were inspected in the following sequence ranges: HVS-I (16179–16196), HVS-II (300–316) and HVS-III (564–577) and the dominant variant (most abundant molecule) was called with a MAF of 20 reads (strand bias disregarded). Thus, the mitotypes reported by the participating laboratories could be compared to those derived from the in-house software to account for possible variation due to different platforms used by the participants. Analysis and comparison of tabular data were performed in Microsoft Excel and graphical illustrations were generated in RStudio (RStudio 2023.06.0 +421 "Mountain Hydrangea" Release).

3. Results

3.1. DNA quantification

The five DNA extracts produced in Lab 0 were diluted to a concentration of approx. 0.1 ng/µL nuclear DNA (nuDNA) and yielded 90 (CS-2) to 147 pg/µL (CS-1) after quantifying the 70 bp target ([21]; Table 1a). Mitochondrial DNA quantification in Lab 0 using the same protocol yielded between 4298 (CS-4) and 26,155 copies mtDNA/µL (CS-1) with the 69 bp target and between 3402 (CS-4) and 25,714 copies (CS-1) using the 143 bp. Six participating laboratories also

Table 1a
Tabular summary of nuDNA and mtDNA quantification results of the five DNA extracts shared for the collaborative exercise with participating laboratories. The amount of DNA in the set of samples was quantified in Lab 0 prior to the shipment to the participating laboratories (a) and upon receipt of a batch of samples that was shipped to one participant and sent back (b) to evaluate shipping conditions.

Labs	Quant. target	CS-1	CS-2	CS-3	CS-4	CS-5
Lab 0 ^a	mt-69/143	26,155/ [copies/µL]	7093/ 7832	8656/ 6266	4298/ 3402	8694/ 6756
	Xavier 2019					
Lab 0 ^a	nu-70 [pg/µL]	125	90	115	147	108
	Xavier 2019					
Lab 0 ^b	mt-69/143	18,506/ [copies/µL]	2345/ 2471	3814/ 3184	1002/ 956	4014/ 3292
	Xavier 2019					
Lab 0 ^b	nu-70 [pg/µL]	82	45	77	47	54
	Xavier 2019					
Lab 1	nu-80/214	34/30	29/23	29/23	25/22	-
	[pg/µL]					
	Quant Trio					
Lab 2	mt-69 [copies/µL]	63,341	17,768	21,842	6685	27,287
	Xavier 2019					
Lab 4	mtGE/µL gDNA (mito copies estimated based on 150 per cell)	11,957	1574	3865	1316	2455
Lab 11	mt-105 [copies/µL]	21,477	2946	4050	1359	3903
	Kavlic 2011					
Lab 14	mt-105 [copies/µL]	65,736	9858	11,510	8562	10,889
	Kavlic 2011					
Lab 16	mt-70/217 [copies/µL]	2750/ 1918	1560/ 1031	1052/ 1144	864/ 770	1175/ 1248
	unpubl.					

quantified nuDNA/mtDNA of the shipped extracts with target sizes between 69 and 217 bp and yielded results that fell between approx. 10% (CS-1; Lab 16) and approx. 314% (CS-5; Lab 2) of the quantities observed in Lab 0 (Table 1a).

To assess the impact of shipment on DNA quantity and quality, one participating laboratory received two batches of samples, one of which was returned to Lab 0 within 4 days (Table 1a, b). The quantification analysis of the returned samples showed degradation indexes close to 1, indicating no detectable mtDNA degradation at fragment sizes of 69 bp and 143 bp. However, a notable loss of DNA quantity was observed for both targets in the returned samples, ranging from approx. 26% (CS-1) to approx. 77% (CS-4) of the original mtDNA amount (Table 1b).

3.2. Mitotypes of the selected samples

Before shipping the five DNA extracts to the participating laboratories, their CRs were sequenced using the Sanger method according to [25], and their complete mitogenome sequences were established using the Precision ID mtDNA Whole Genome Panel (PID; TFS), following the protocol of [26] in Lab 0. The mitotypes showed instances of PHP at ntps 2413 (CS-1; PID), 7861 (CS-2; PID), 12483 and 16093 (CS-4; PID, Sanger and PID, respectively) and 234 (CS-5; Sanger and PID; Table 2). Instances of LHP were observed in the HVS-I poly(C)-tract (16193.XC; CS-3, CS-5), in the HVS-II poly(C)-tract (309.XC; CS-2, CS-3) and in the HVS-III poly(C)-tract (573.XC; CS-1, CS-3, CS-5, Table 2).

3.3. Comparison of reported mitotypes and evaluation of the raw data

A total of 24 laboratories returned results in this collaborative study. The final dataset comprised Sanger-based results from eight laboratories, Ion Torrent-based sequences from 13 laboratories, and Illumina-based sequences from eight laboratories (Supplementary Tables S1a-c).

3.3.1. Sanger-based sequencing

Of the eight submitted Sanger-based sequencing data, five reported the entire CR and three analyzed the two hypervariable segments HVS-I and HVS-II only. Sequencing instruments included the 3130 (XL) and the 3500 (XL) systems (both TFS) and a total of five different analysis software packages were used (Supplementary Table S1a). All reported haplotypes were concordant for non-heteroplasmic variants except for three differences to the rCRS that were omitted by one laboratory at ntps 152, 207 (both CS-1) and ntp 16234 (CS-3) (data not shown). Their presence was however confirmed by inspecting the raw data in Lab 0.

3.3.2. Ion Torrent-based sequencing

Eleven of 13 laboratories reported entire mitogenome sequences and two laboratories typed the CR using Ion S5 instrumentation (12 labs) and the Personal Genome Machine (PGM; 1 lab). There was variability in sequence analysis and alignment software employed, including different versions of the Torrent Suite Software (TSS; Supplementary Table S1b). Single non-heteroplasmic discrepancies between laboratories were observed at ntps 250, 460 (both CS-1), 16311, 16519, 73 (all three CS-4) and 3010 (CS-5; data not shown) across three participating laboratories; one of these laboratories was involved in multiple

Table 2

Full mitogenome sequences of five control samples, which are DNAs from two cell lines, 007 (= CS-1) and 9947 A (= CS-2), and DNA extracted from two buccal swabs (CS-3, CS-5) and one blood sample (CS-4) provided by volunteers. Instances of point and length heteroplasmy are highlighted in bold letters. Note, that point and length heteroplasmy are known to vary between different batches (lot numbers) of cell lines (CS-1 and CS-2; [27]).

Samples	Mitotypes
CS-1	16129 A 16223 T 16391 A 16519 C 73 G 152 C 199 C 204 C 207 A 250 C 263 G 315.1 C 460 C 573.XC 750 G 1438 G 1719 A 2413Y 2706 G 4529 T 4769 G 6293 C 7028 T 8251 A 8860 G 9438 A 10034 C 10238 C 10398 G 11719 A 12501 A 12705 T 13780 G 14766 T 15043 A 15326 G 15758 G 15924 G
CS-2	16311 C 16519 C 93 G 195 C 214 G 263 G 309.XC 315.1 C 750 G 1438 G 4135 C 4769 G 7645 C 7861Y 8448 C 8860 G 9315 C 13572 C 13759 A 15326 G
CS-3	16183 C 16189 C 16193.XC 16234 T 16324 C 16519 C 73 G 152 C 195 C 263 G 309.XC 315.1 C 573.XC 750 G 1438 G 2706 G 3480 G 4769 G 5165 T 7028 T 8860 G 9055 A 9698 C 11467 G 11719 A 12308 G 12372 A 14053 G 14167 T 14766 T 15326 G 15924 G
CS-4	16093Y 16311 C 16519 C 73 G 263 G 315.1 C 750 G 1438 G 4769 G 8860 G 12483Y 15326 G
CS-5	16189 C 16193.XC 16356 C 16362 C 16519 C 234 R 263 G 315.1 C 523del 524del 573.XC 750 G 1438 G 3010 A 3796 G 4769 G 8860 G 15326 G

observations. The presence of variants at these positions was confirmed by further review of the sequencing raw data in Lab 0.

3.3.3. Illumina-based sequencing

Four of the eight laboratories using Illumina sequencing technology reported entire mitogenomes and four laboratories analyzed the CR using the MiSeq FGx (7 labs) or the MiSeq (1 lab) instrumentation. Also here, a variety of analysis and alignment software packages and versions was observed (Supplementary Table S1c). Single non-heteroplasmic discrepancies were reported at ntps 16223 (CS-1), 16189 (CS-3) and 73, 263, 16519 (all three CS-4; data not shown). These discrepancies involved four laboratories in total, including one laboratory that contributed to two observations. Again, raw data inspection confirmed their presence.

3.4. Evaluation of point heteroplasmy (PHP)

This evaluation focuses on five instances of PHP detected in the control samples: CS-1 (2413Y), CS-2 (7861Y), CS-4 (12483Y, 16093Y), and CS-5 (234 R). Additional, systematic, and individual observations of PHP, mostly at low MAF, are summarized further below.

3.4.1. Sanger-based PHP detection

All Sanger-based sequencing data encompassed (the hypervariable regions of) the CR, thus no information was recorded for the coding region PHPs present in samples CS-1, CS-2 and CS-4. All eight participating laboratories reported 234 R in CS-5 and six laboratories reported 16093Y in CS-4 (Supplementary Table S2a); the remaining two called the major molecule 16093 C, but the presence of 16093Y was confirmed in the sequence raw data in Lab 0.

Table 1b

Tabular summary of mtDNA quantification results (copies mtDNA/μl) of the five DNA extracts as shipment controls. These samples were sent to a study participant (sent) and quantified on return (received, rec.). The loss of mtDNA quantity (%) is shown for both fragment sizes 69 and 143 bp. D.I. (degradation index) = small target quant./large target quant).

Samples	mt69 _{sent}	mt143 _{sent}	D.I. _{sent}	mt69 _{rec.}	mt143 _{rec.}	D.I. _{rec.}	loss _{mt69}	loss _{mt143}
CS-1	26,155	25,714	1.02	18,506	19,139	0.97	29%	26%
CS-2	7093	7832	0.91	2345	2471	0.95	67%	68%
CS-3	8656	6266	1.38	3814	3184	1.20	56%	49%
CS-4	4298	3402	1.26	1002	956	1.05	77%	72%
CS-5	8694	6756	1.29	4014	3292	1.22	54%	51%

3.4.2. Ion Torrent-based PHP detection

All 13 Ion Torrent-based sequencing results reported PHP in CS-4 (16093Y: median allele frequency (medAF)_C= 58%, mean absolute deviation (MAD)= 1.3) and CS-5 (234 R: medAF_G=50.5%, MAD=1.3). PHP in CS-4 (12483Y: medAF_T=33.3%, MAD=4.9) was recorded by eleven laboratories (Table 3; Supplementary Table S2a). PHPs in CS-1 (2413Y) and CS-2 (7861Y) were present at low minor allele frequencies of about or below 10%, which corresponded to the MAF settings of many participants, which is why only few laboratories reported PHP for CS-1 (2413Y: medAF_T=9.8%, MAD=1.2) and CS-2 (7861Y: medAF_C=8.0%; single observation, Table 3). Review of PHP in all Ion Torrent data in Lab 0 with MAF > 5% resulted in similar results for CS-4 and CS-5, namely 16093Y: medAF_C= 57.9%, MAD= 3.1; 234 R: medAF_G= 50.9%, MAD= 1.8 and 12483Y: medAF_T= 33.4%, MAD= 4.6, respectively. Due to using a lower threshold for re-analysis, more PHP calls were feasible in CS-1 (2413Y: medAF_T=8.1%, MAD=1.9) and CS-2 (7861Y: medAF_C=8.2%, MAD=1.2; (Table 4; Supplementary Table S2b)).

3.4.3. Illumina-based PHP detection

Six of the eight Illumina-based sequencing results reported PHP in CS-4 (16093Y: medAF_C=58.1%, MAD=1.6) and seven reported PHP in CS-5 (234 R: medAF_G=48.3%, MAD=3.4). PHP in CS-4 (12483Y: medAF_T=28.2%, MAD=0.7) was recorded by two laboratories (Table 3, Supplementary Table S2a). PHP in CS-1 and CS-2 was reported by two participants each with values of 2413Y: medAF_T= 7.4%, MAD= 0.6 and 7861Y: medAF_C= 8.5%, MAD= 0.7. Review of PHP in the Illumina data in Lab 0 with MAF > 5% resulted in values for CS-4 and CS-5; 16093Y: medAF_C= 57.0%, MAD= 1.2; 234 R: medAF_G= 46.4%, MAD= 6.8 and 12483Y: medAF_T= 29.7%, MAD= 0.1, respectively (Table 4). We note that the re-analysis of 234 R in CS-5 for the raw data of Labs 7 and 20 yielded different PHP ratios, with MAF differences between reported and reviewed data of -12.1% and -18.8%, respectively (data not shown), as primers remained in the raw data covering ntp 234 and therefore contributing reads to the PHP in the re-analysis by Lab 0. Other labs removed the primer sequences from their raw data (e.g. Lab 16), which made the re-analysis possible. The PHP review in CS-1 and CS-2 resulted in the following values: 2413Y: medAF_T= 7.4%, MAD= 0.2 and 7861Y: medAF_C= 8.4%, MAD= 1.9 (Table 4; Supplementary Table S2b).

3.4.4. Comparing MPS-based PHP detection

Both MPS-based platforms yielded comparable ratios of the alternative variant for both the reported (Fig. 1, Table 3, Supplementary Table S2c and S2d) and the reviewed raw data (Fig. 1, Table 4, Supplementary Table S2c and S2d), except for a few outliers (Ion Torrent: N = 7; Illumina: N = 3), defined as observations below Q1 - 1.5 × IQR or above Q3 + 1.5 × interquartile range (IQR).

The two PHPs 16093Y (CS-4) and 234 R (CS-5) were analyzed by

Table 3

Summary of reported point heteroplasmies at ntps 2413 (CS-1), 7861 (CS-2), 16093 and 12483 (CS-4) and 234 (CS-5) for Ion Torrent (Ion) and Illumina (ILMN) sequence data. Median allele frequency (medAF) and mean absolute deviation (MAD) values are compared. Number of observations are shown in brackets. Mean allele frequency (meanAF) values are provided for both MPS-based methods (sd=standard deviation).

Method	CS-1 2413Y (%) T)	CS-2 7861Y (%) C)	CS-4 16093Y (%) C)	CS-4 12483Y (%) T)	CS-5 234 R (%) G)
medAF (Ion)	9.8 (n = 4)	8.0 (n = 1)	58.0 (n = 13)	33.3 (n = 11)	50.5 (n = 13)
MAD	1.2	-	1.3	4.9	1.3
medAF (ILMN)	7.4 (n = 2)	8.5 (n = 2)	58.1 (n = 6)	28.2 (n = 2)	48.3 (n = 7)
MAD	0.6	0.7	1.6	0.7	3.4
meanAF (MPS)	8.6 (n = 6)	8.3 (n = 3)	56.8 (19)	33.3 (n = 13)	49.6 (n = 20)
sd	1.76	0.6	3.9	3.9	3.2

Table 4

Summary of reviewed (re-analyzed) point heteroplasmies at ntps 2413 (CS-1), 7861 (CS-2), 16093 (CS-4), 12483 (CS-4) and 234 (CS-5) based on MPS raw data provided by the participating laboratories. For this purpose, the minor allele frequency threshold was set to 5% for all data. Median allele frequency (medAF) and mean absolute deviation (MAD) values are compared. Mean allele frequency (meanAF) values are provided for both MPS-based methods (sd=standard deviation).

Method	CS-1 2413Y (%) T)	CS-2 7861Y (%) C)	CS-4 16093Y (%C)	CS-4 12483Y (%T)	CS-5 234 R (%) G)
medAF (Ion)	8.1 (n = 9)	8.2 (n = 11)	57.9 (n = 12)	33.4 (n = 11)	50.9 (n = 12)
MAD	1.9	1.2	3.1	4.6	1.8
medAF (ILMN)	7.4 (n = 4)	8.4 (n = 4)	57.0 (n = 7)	29.7 (n = 3)	46.4 (n = 5)
MAD	0.2	1.9	1.2	0.1	6.8
meanAF (MPS)	8.1 (n = 13)	8.1 (n = 15)	56.7 (19)	33.0 (n = 14)	49.3 (n = 17)
sd	1.4	1.2	3.9	4.2	3.6

the majority of Ion Torrent and Illumina laboratories, since they are located in the CR, thus allowing a more detailed comparison.

For Ion Torrent data, the relative contribution of 16093 C to the heteroplasmic mixture was reported between 49.0% and 61.7% (median=58.0%; interquartile range (IQR) = 1.8%; N = 13). The data review resulted in comparable ranges, 49.1% and 62.0% (median=57.9%; N = 12) and a slightly increased IQR (3.8%). The relative contribution of 16093 C reported for the Illumina PHP ranged between 48.3% and 59.5% (median=58.1%; IQR=1.8%; N = 6) and the reviewed values varied between 48.3% and 58.7% (median=57.0%; IQR=1.4%; N = 7). The IQR was smaller for the Illumina data for that PHP, albeit the number of datapoints was also smaller compared to Ion Torrent. However, we note that both MPS-based technologies arrived at similar median values for the contribution of 16093 C in that PHP, both for reported and reviewed data.

PHP 234 R was reported by Ion Torrent laboratories with a median of 50.5% G contribution (min=48.0%; max=53.9%; IQR=1.6%; N = 13) and similar values for the reviewed data (median=50.9%; min=48.5%; max=53.5%; IQR=2.2%; N = 12). Reported values for the 234 G contribution for Illumina-based data were: median= 48.3%, min= 39.7%, max= 54.6%, and IQR= 5.1%. Data review resulted in median= 46.4%, min= 40.1%, max= 52.1%, and IQR= 7.7%.

The IQR was higher for Illumina than for Ion Torrent data, which may be related to the fact that one of the primers used by some laboratories covered position 234. Consequently, sequences generated with this primer may, if not completely removed during the analysis and alignment process, have contributed additional sequence information to the base call at position 234. Because the primer sequences of both commercial providers were confidential, the exact interpretation of the mixture observed at this position was limited. This phenomenon may also reflect primer-binding sites located within the concatemer of two adjacent amplicons when the single-amplification approach of the PowerSeq kit was used. During data review, clear outliers attributable to this issue were excluded; however, we cannot rule out whether, and to what extent, this primer-related effect also influenced the results presented here.

3.4.5. Additional mixed positions

In addition to the PHP observations described above, the review of the MPS raw data brought additional instances of mixtures at low MAF level (>5%). Some of these might represent platform-specific artefacts, such as those located around the poly(C)-tracts (HVS-I for Ion Torrent data and HVS-II for Illumina data; Supplementary Table S3, Supplementary Figure S1). Ion Torrent sequencing technology is known to show greater variation in homopolymeric tracts, such as the one around ntp 16189 [19], which is known to also affect the preceding poly

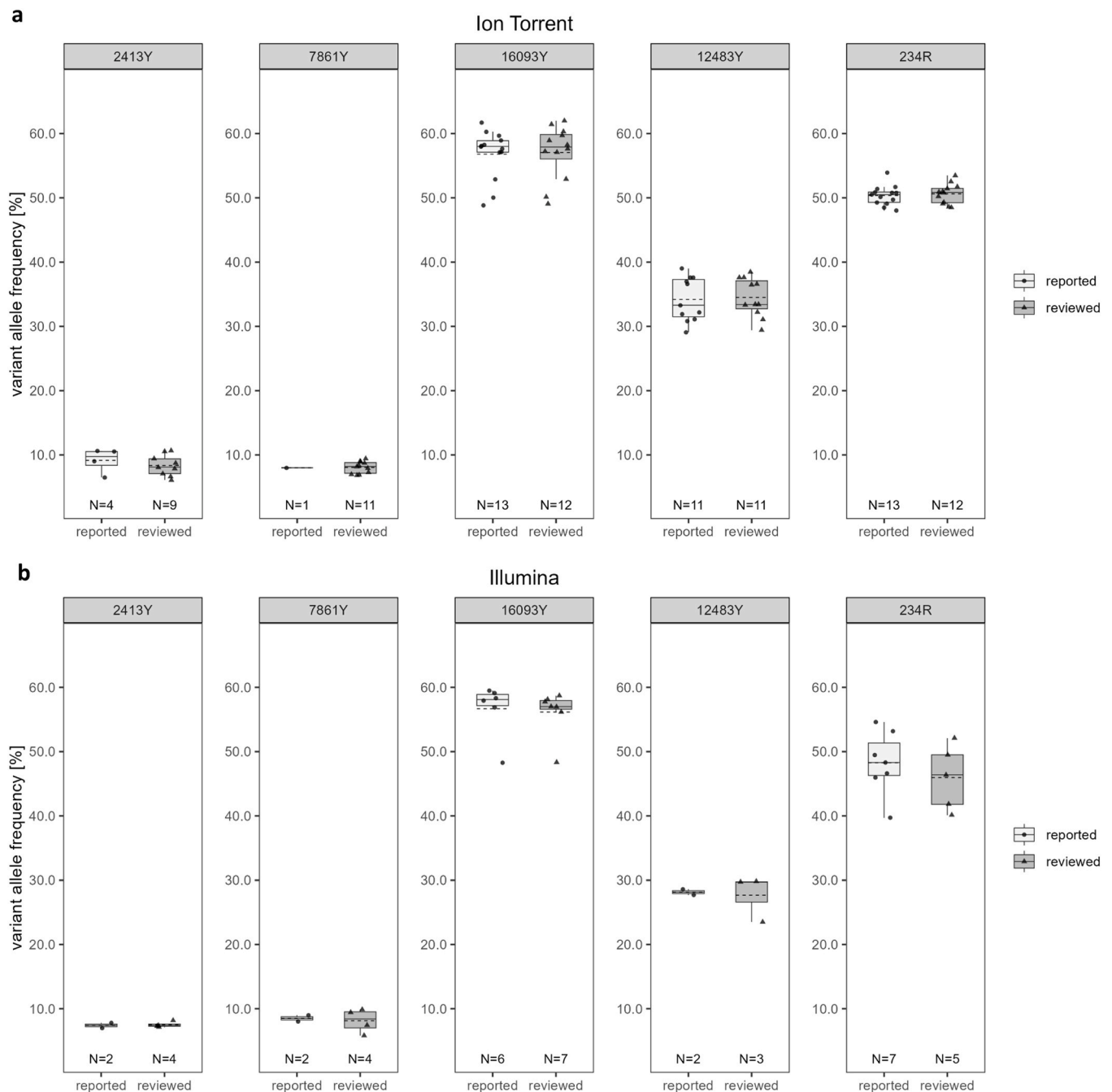


Fig. 1. Distribution of variant allele frequencies at point heteroplasmy positions 2413Y (CS–1), 7861Y (CS–2), 16093Y (CS–4), 12483Y (CS–4) and 234 R (CS–5). Reported and reviewed variation is contrasted in neighboring plots: **1a)** Ion Torrent data, **1b)** Illumina data. Each boxplot represents the median (solid line), mean (dotted line) and interquartile range (IQR; box edges). Whiskers extend to 1.5 x IQR from the quartiles. Individual data points are plotted, with the dataset size indicated along the x-axis for each boxplot ($N = X$).

(A)-tract between ntps 16180–16183 [11,28]. Other mixtures seem to represent sample-specific artefacts or low-level background signal that is not uncommon for mtDNA sequence analysis.

3.5. Evaluation of length heteroplasmy (LHP)

For the evaluation of LHP, the ISFG guidelines [23] recommend reporting the dominant variant (DV), i.e., the most abundant length variant in a mixture. For example, if the predominant molecule in the poly(C)-tract spanning ntps 302–310 contains eight cytosines, the corresponding DV call would be 309.1 C, given that the rCRS contains seven cytosines in this region. Some software tools alternatively denote this as

309.1c to indicate the presence of LHP. While we present the observations from participating laboratories in their original form, we applied ISFG nomenclature when reviewing and interpreting the results.

The LHP results reported by the participating laboratories and the analysis of the accompanying raw data is presented per region in detail below.

3.5.1. LHP in HVS-I (around ntp 16193)

Five of eight Sanger data reported LHP in HVS-I of CS–3, four presented 16193.1 C and one 16193.2 C. The remaining data that did not report LHP in this region correspond to the rCRS (Supplementary Table S4e, Fig. 2a). The review of the raw data following ISFG

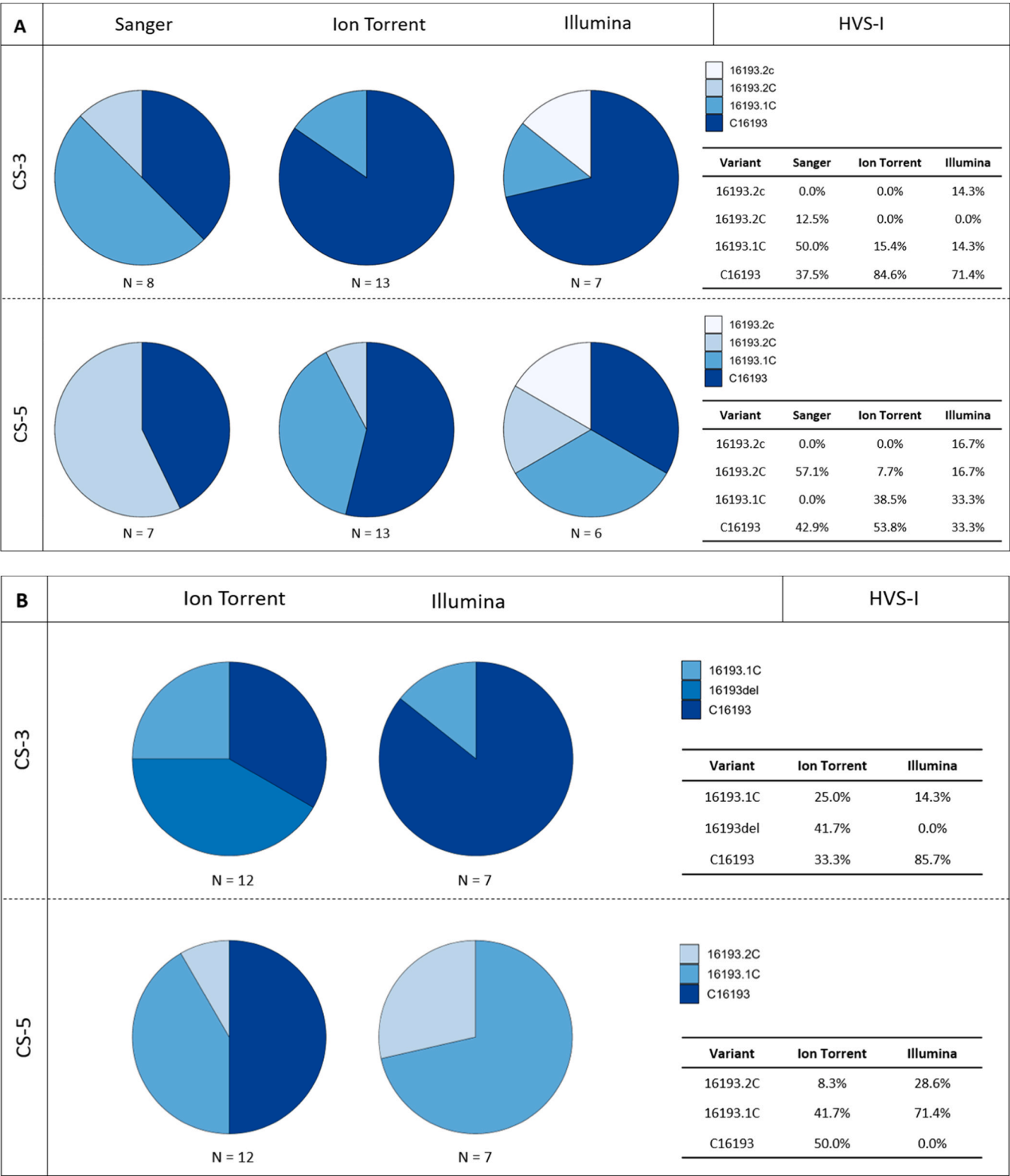


Fig. 2. Summary of a) reported and b) reviewed length heteroplasmy in the HVS-I poly(C)-tract of CS–3 and CS–5.

guidelines [23] in Lab 0 resulted in concordant LHP calls in all instances (16193.1 C). Ion Torrent data identified no insertions at 16193 (=rCRS variant) in eleven of thirteen data, two reported 16193.1 C. Illumina labs reported no insertions in five of seven cases and 16193.1 C and 16193.2c in two instances. The review of Ion Torrent raw data (N = 12) yielded five deletions at 16193, four times the rCRS variant and three times a single C insertion (16193.1 C, Supplementary Table S4f, Fig. 2b).

The review of the Illumina data (N = 7) resulted in the rCRS variant in six cases and in a single C-insertion (16193.1 C) once.

The other sample that displayed LHP in HVS-I, CS–5, yielded similar variation (Supplementary Table S4i, Fig. 2a). Four Sanger data reported two insertions (16193.2 C) and three identified no insertions; however, data review in Lab 0 resulted in 16193.2 C in all cases. Ion Torrent data were described with seven rCRS variants, five times single C-insertions

and one time two C-insertions at 16193, respectively. Two Illumina-based datasets reported no insertions, two called single C-insertions, one called 16193.2 C and another one reported 16193.2c. The review of Ion Torrent data resulted six times in no indels, five times in single C-insertions and one time in two C-insertions (16193.2 C). The review of Illumina data resulted in five single C-insertions at 16193 and the

16193.2 C variant in two instances (Supplementary Table S4j, Fig. 2b). LHP in the HVS-I poly(C)-tract is often associated with variation in the poly(A)-tract preceding the C-tract, i.e., ntps 16182 and 16183. Due to the complexity of the data in this region, the observed variation was not further described in detail. The remaining samples did not show LHP in the HVS-I poly(C)-tract.

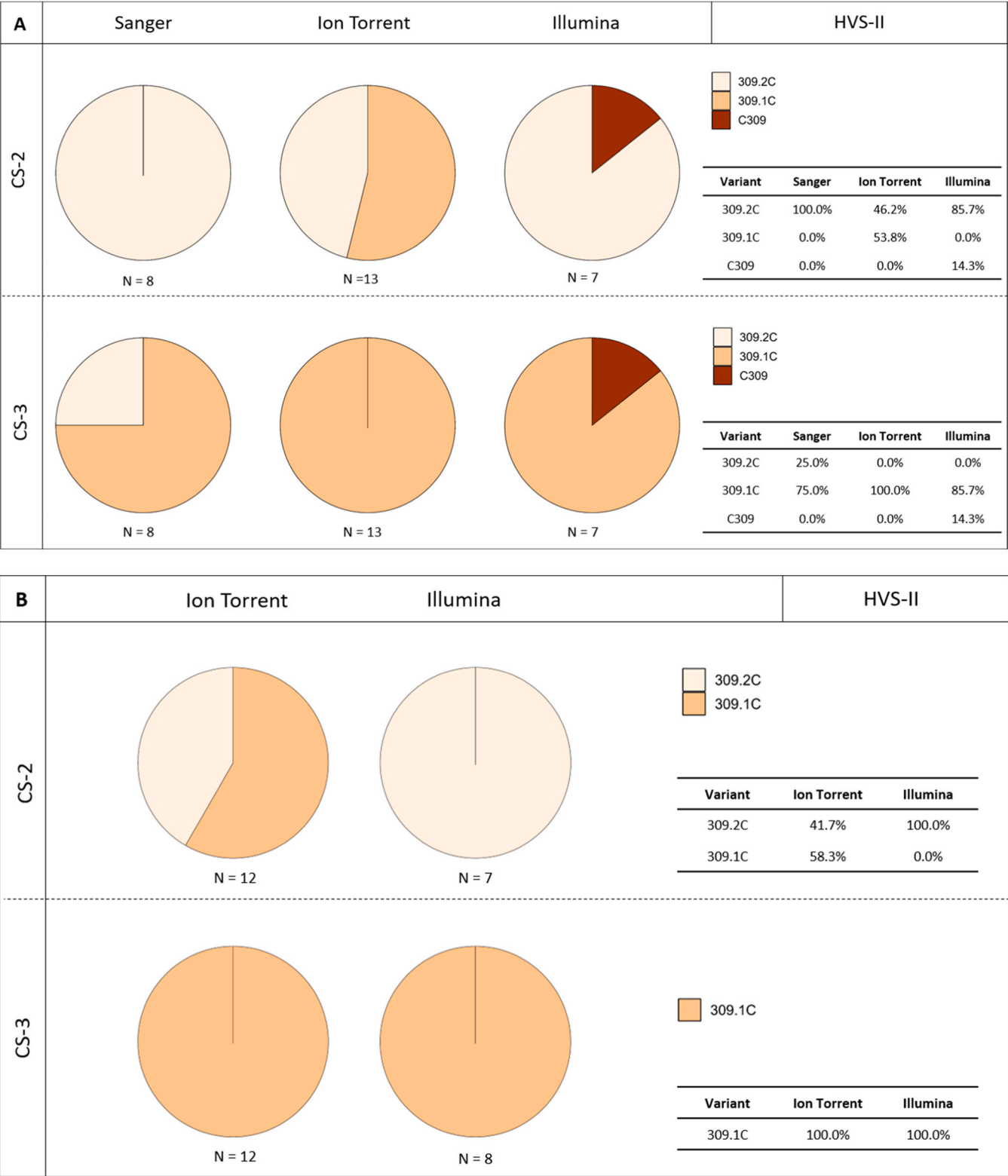


Fig. 3. Summary of a) reported and b) reviewed length heteroplasmy in the HVS-II poly(C)-tract of CS-2 and CS-3.

They were all reported accordingly and brought the same results also after data review (Supplementary Table S4a, b, c, d, g, h).

3.5.2. LHP in HVS-II (around ntp 309)

LHP in the poly(C)-tract of HVS-II was reported in sample CS-2 with two insertions (309.2 C) in all eight reported Sanger data (Supplementary Table S4c, Fig. 3a), which was confirmed by independent data review. Seven Ion Torrent-based laboratories reported a single insertion (309.1 C) and six described two C-insertions (309.2 C). The review of the Ion Torrent data brought seven instances of 309.1 C and five times two C-insertions (Supplementary Table S4d, Fig. 3b). Six Illumina-based laboratories reported 309.2 C and one reported no insertions. The review of the Illumina data yielded 309.2 C in all instances.

The other sample with LHP in the HVS-II poly(C)-tract was CS-3. Six of eight Sanger data reported 309.1 C and two gave 309.2 C (Supplementary Table S4e, Fig. 3a). Data review by Lab 0 resulted in 309.1 C. All 13 Ion Torrent results reported 309.1 C and so did six of seven Illumina data; one Illumina-based laboratory reported no insertion. The review of the Ion Torrent data gave 309.1 C in all 12 results. The review of the Illumina data brought also eight times 309.1 C (Supplementary Table S4f, Fig. 3b).

The remaining samples did not show LHP in the HVS-II C-tract in Sanger and Illumina reported and reviewed data. However, Ion Torrent data reported four times 309del and eight times 309c in samples CS-1, CS-4, and CS-5, respectively. Data review resulted in 309del in the three samples in total (Supplementary Table S4a, b, g, h, i, j).

3.5.3. LHP in HVS-III (around ntp 573)

LHP in the poly(C)-tract of HVS-III around position 573 was observed in samples CS-1, CS-3 and CS-5. For CS-1 Sanger reports varied between zero and seven C-insertions, Ion Torrent data between zero and four C-insertions and Illumina data between zero and five C-insertions (Supplementary Table S4a, Fig. 4a). The review of the Sanger raw data following ISFG guidelines resulted in concordant LHP calls in all instances, indicating four insertions at 573. Ion Torrent data review resulted in nine instances of four C-insertions and one instance of five C-insertions, while the review of the Illumina data brought six times four C-insertions (Supplementary Table S4b, Fig. 4b).

In CS-3, Sanger data reported between two and five C-insertions, Ion Torrent data between zero and three C-insertions, while all six Illumina data were reported with three C-insertions (Supplementary Table S4e, Fig. 4a). Sanger data review resulted three times in 573.3 C and one time gave identical peak heights at 573.2 C and 573.3 C. Ion Torrent data review resulted in five times two and seven times three C-insertions, and Illumina data review resulted with six times three C-insertions and in a single instance in four C-insertions (Supplementary Table S4f, Fig. 4b).

For CS-5, Sanger data reported between zero and two C-insertions, Ion Torrent between zero and three C-insertions and Illumina data resulted in two reported C-insertions for all labs (Supplementary Table S4i, Fig. 4a). Sanger data review resulted in 573.2 C in all instances. Ion Torrent raw data review gave 11 instances of 573.2 C and one 573.3 C, whereas Illumina data review resulted in seven times 573.2 C (Supplementary Table S4j, Fig. 4b).

The remaining samples CS-2 and CS-4 were not reported with LHP in the HVS-III poly(C)-tract, which was confirmed by the review of the raw data (Supplementary Table S4c, d, g, h).

The observed discrepancies between reported and reviewed LHP results are summarized in Supplementary Table S5.

4. Discussion

This collaborative study on mtDNA heteroplasmy was carried out by the EDNAP group to compare the detection of PHP and LHP in standardized DNA extracts using the various platforms (combinations of instrument and software) in the participating laboratories. The most recent ISFG guidelines on mtDNA typing date back to the year 2014

[23], when Sanger represented the main sequencing method used in forensic laboratories. The guidelines recommend that "... laboratories must establish their own interpretation and reporting guidelines for observed length and point heteroplasmy" based on their internal validation results. Meanwhile, MPS-based technologies have been implemented in many forensic laboratories and industry is providing assays to generate mtDNA sequences for the CR and the entire mitogenome using various library generation methods. Earlier comparisons between some of these methods and sequencing technologies provided evidence for discrepant LHP data [19] and have alerted the field to exercise caution when reporting LHP in the forensic context. The EDNAP group thus set out to systematically evaluate the established platforms with respect to heteroplasmy detection.

DNA extracts were centrally prepared at the organizing laboratory, aliquoted, and shipped to the participating laboratories. One participating partner received two batches of samples and returned one set immediately upon receipt to the organizing laboratory to assess the impact of transport conditions on the quantity and quality of mtDNA and nDNA in the extracts. Degradation-sensitive mtDNA quantification revealed substantial loss of detectable mtDNA upon return of the DNA extracts; however, no further degradation patterns were observed when compared to the DNA extracts sent out initially.

Six of the 24 participating laboratories performed DNA quantification experiments on the extracts, two of which quantified nDNA and four determined mtDNA quantities (one laboratory determined nDNA and estimated the mtDNA copy number based on a value of 150 copies mtDNA/cell). A previous study [29] systematically investigated nDNA-to-mtDNA ratios across various tissues and found that nDNA quantification may be a poor surrogate for mtDNA copy-number estimation. For this reason, we advocate the use of mtDNA-specific quantification assays. Nevertheless, some of the reported mtDNA quantification values differed substantially from the others. The reasons for this variability are likely multifactorial and may include shipment conditions, calibration of quantification methods, pipetting accuracy and other technical factors. Since this study did not focus on mtDNA quantification, no further investigations into the possible causes of the observed variation were conducted. However, these observations demonstrate that shipping DNA extracts under ambient temperature condition has a substantial effect on the measurable DNA quantity, but apparently not on the degradation state of the DNA, and should be considered in future collaborative exercises. Moreover, as the extracts were not shipped in low-bind tubes, DNA adsorption to the tube surface cannot be excluded as an additional factor contributing to the observed loss in measurable DNA quantity.

We observed some variation in the reported mitotypes caused by transcription errors, which constitute the most common form of errors in this context [30]. In total, 14 transcription errors were identified (Sanger: N = 3; Ion Torrent: N = 6; Illumina: N = 5) considering non-heteroplasmic variants. Regarding the submitted sequence raw data, we note a high consistency produced in this collaborative study across the different platforms. Aside from differences observed in the display of heteroplasmy, the raw data offered by the participating laboratories yielded matching mitotypes in the overlapping regions.

4.1. Reporting and review of point heteroplasmy

Point heteroplasmy is described as mixture of mostly two nucleotides at a given mtDNA position [3] that is present above the analytical threshold and has been excluded from representing other possible sources of mixtures, e.g., mtDNA mixtures from two or more individuals, NUMTs, damage patterns, elevated background or other artefacts (see [31] for a detailed review).

This study focused on PHP present at expected nucleotide positions based on the selected samples, i.e., 2413Y (CS-1), 7861Y (CS-2), 12483Y, 16093Y (CS-4) and 234 R (CS-5). Other instances of PHP, reported by some laboratories (typically at very low MAF) were

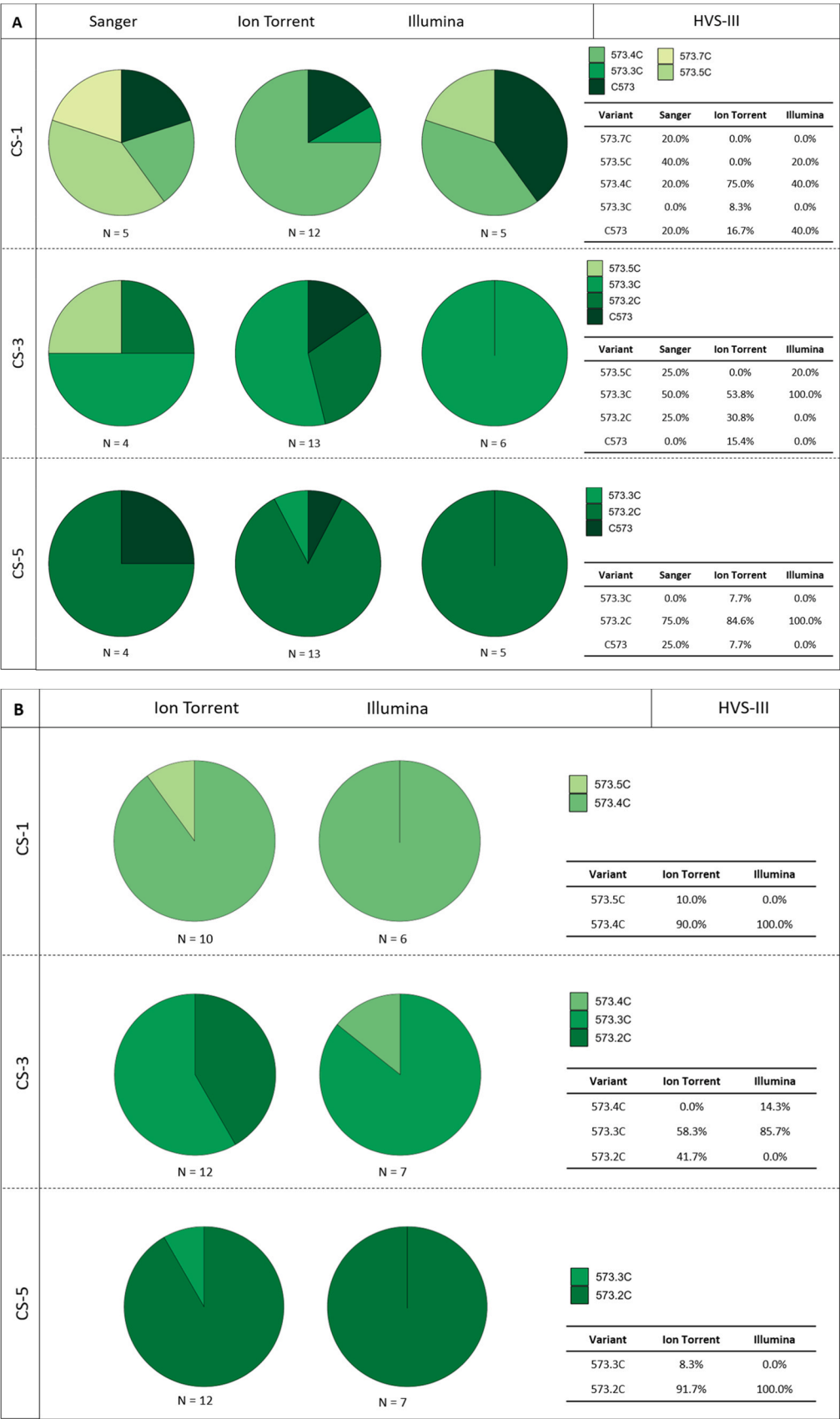


Fig. 4. Summary of a) reported and b) reviewed length heteroplasmy in the HVS-III poly(C)-tract of CS-1, CS-3 and CS-5.

documented here as well. However, they were not further investigated in detail. Some of these additional variants seemed to be related to the platforms used, whereas others represented isolated observations without a recurring pattern across laboratories. Depending on the available raw data support, these observations may reflect true low-level signals, sample-specific artefacts, or background noise, all of which can contribute to variation in mtDNA sequence interpretation.

The detection of PHP in Sanger-based data was reduced to the CR, and consistent results were reported for both PHPs 16093Y and 234 R, with two exceptions, where the laboratories reported the dominant type. However, the presence of the PHPs in these data was also confirmed by raw data review. All Sanger data were based on BigDye Terminator fluorescence technology that provides only semi-quantitative data, i.e. peak heights vary across the electropherogram depending on the sequence neighborhood, the nature of the incorporated ddNTPs and on the primers used. Therefore, PHP ratios typically cannot be reliably estimated from the peak height ratios of the related positions. In the two instances described above, the MAF were well above background and conventional interpretation guidelines (e.g., 20%) and therefore were reproducibly detected across all laboratories.

The same two PHPs were also consistently reported or detected by the laboratories that used Ion Torrent or Illumina technology. MPS-based sequencing is generally considered more quantitative in nature than Sanger-based sequencing, which was reflected by relatively consistent mixture ratios and small MAD values between laboratories and technologies.

During data review, we observed discrepant mixture ratios at position 234 R in the Illumina data compared with the reported results. This discrepancy appears to be attributable to one of the amplification/sequencing primers used by some laboratories, which covers ntp 234 and may therefore contribute primer-derived reads to the final alignment if primer sequences or concatemers were not completely removed during data processing. Because the exact primer sequence was not available to us, these reads could not be specifically accounted for during our review. This phenomenon should be considered when analyzing PCR-generated sequence data with independent software that does not incorporate assay-specific features such as primer sequences.

Two of the three coding region PHPs at 2413 and 7861 had low MAFs (below 10%). The majority of participants did not report the low PHPs as they fell under their defined MAF threshold, but the raw data review at Lab 0 brought consistent detection of these low-level PHPs. PHP at ntp 12483 yielded larger differences in medAF between Ion Torrent (~33%) and Illumina (~29%) data (Table 3, Table 4, Supplementary Table S2c-d). However, the latter was represented by a smaller sample size of only two and three contributing laboratories, respectively.

Overall, reported MPS-based PHP detection revealed similar mixture ratios across all positions and technologies, which has promising implications for practical application in forensic routine. The results indicate that PHP, including those with low MAF, were generally supported by the raw sequencing data across the different platforms used. This implies that PHP can be detected consistently in forensic mtDNA analysis when evaluated according to standardized interpretation criteria. It should be kept in mind that PHP-like mixtures can also arise due to artefacts caused by the sequencing platform used or the sample analyzed; but these typically occur at low signal level.

4.2. Reporting and review of length heteroplasmy

Length heteroplasmy is most commonly observed in the hypervariable poly(C)-tracts of the CR around ntps 16193 (HVS-I), 309 (HVS-II), and 573 (HVS-III). Four of the DNA samples evaluated in this study exhibited LHP in at least one of these regions. Sanger-based results yielded some variation in the reported LHP patterns; however, independent raw data reviews conducted according to standardized ISFG guidelines and based on consensus interpretation resulted in concordant pattern interpretations. The ISFG guidelines recommend reporting the

DV, i.e., the most abundant length variant present in the mixture determined by the peak height of a unique marker, e.g. G16196, T310 and G577. Strictly applying this method to the Sanger raw data provided by the participating laboratories resulted in the same DVs across laboratories.

This was different for MPS-based data. We observed differences between the reported and reviewed LHP patterns between both MPS and Sanger data, as well as between the two MPS technologies. Furthermore, and adding another layer of complexity, we detected variation in LHP patterns within the Ion Torrent and within the Illumina datasets. This highlights that the observed variation was not driven solely by human factors, such as individual analysis practices seen in the Sanger data, but also by differences in MPS workflows, including PCR-based versus capture-based enrichment and library preparation strategies, sequencing technologies themselves and the alignment software employed.

To investigate the impact of software, we standardized the analysis workflow for Ion Torrent and Illumina data by adapting the EMPOP-specific phylogenetic alignment software SAM2 [24] to enable the processing of BAM files. In its original version, SAM2 converts rCRS-coded mitotypes into unaligned nucleotide strings and aligns them according to phylogenetic principles. The adapted version uses BAM files as input and performs the alignment on the basis of the same phylogenetic framework. This software standardization improved concordance between resulting DVs to some extent, particularly in HVS-II and HVS-III poly(C)-tracts. However, some variation remained within and between MPS technologies, particularly in Ion Torrent HVS-I poly(C)-data, which has been described before [19]. As a general trend, laboratories tend to conservatively exclude LHP by reporting the rCRS rather than the observable LHP variants, resulting in an rCRS alignment bias whereby weak signals are preferentially assigned to rCRS. These findings suggest that the LHP differences observed between laboratories can be attributed to three independent sources: the library preparation and sequencing technologies itself, the analysis software, and human factors, such as data interpretation. In MPS data, the region chosen for LHP interpretation affects total read depth and thus the dominant variant. This highlights the need to update current mtDNA interpretation guidelines in the light of MPS-based data and their heterogeneity.

5. Conclusions and outlook

This EDNAP collaborative study demonstrates that mtDNA point heteroplasmy can be detected and reported with high concordance across current forensic sequencing platforms, including Sanger, Ion Torrent, and Illumina. In contrast, length heteroplasmy remains considerably more variable and is influenced by sequencing technology, analysis software, and user interpretation. While standardized review improved agreement for some MPS datasets, important discrepancies persisted in some poly(C)-tract regions. The findings show that current mtDNA interpretation approaches, largely developed for Sanger sequencing, are no longer sufficient for routine MPS-based heteroplasmy analysis. Updated forensic guidelines are therefore needed, especially for the interpretation and reporting of length heteroplasmy, with greater emphasis on assay-specific validation, standardized bioinformatic workflows, and harmonized reporting criteria.

Disclaimer

For AFDIL authors: The assertions herein are those of the authors and do not necessarily represent the official position of the United States Department of Defense, the Defense Health Agency, or its entities including the Armed Forces Medical Examiner System. Any mention of commercial products was done for scientific transparency and should not be viewed as an endorsement of the product or manufacturer.

For FBI authors: Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply

endorsement of the manufacturer or its products or services by the FBI. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government. This is publication 26.12 of the FBI Laboratory Division.

For NIST authors: Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology (NIST), nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. All work performed at NIST has been reviewed and approved by the NIST Research Protections Office under determination #MML–2023–0466. This study was determined to be “not human subjects research” (often referred to as research not involving human subjects) as defined in U.S. Regulations, 15 CFR 27, also known as the Common Rule (45 CFR 46, Subpart A), for the Protection of Human Subjects by the NIST Human Research Protections Office and therefore not subject to oversight by the NIST Institutional Review Board.

CRediT authorship contribution statement

Cordula Haas: Writing – review & editing, Investigation, Data curation. **SallyAnn Harbison:** Writing – review & editing, Investigation, Data curation. **Johannes Hedman:** Writing – review & editing, Investigation, Data curation. **Mitchel M Holland:** Writing – review & editing, Investigation, Data curation. **Mario Gysi:** Writing – review & editing, Investigation, Data curation. **Alberto Marino:** Writing – review & editing, Investigation, Data curation. **Beatriz Heinrich Roson:** Writing – review & editing, Investigation, Data curation. **Kevin M. Kiesler:** Writing – review & editing, Investigation, Data curation. **Odile Loreille:** Writing – review & editing, Investigation, Data curation. **Sabine Lutz-Bonengel:** Writing – review & editing, Investigation, Data curation. **Vincent Castella:** Writing – review & editing, Investigation, Data curation. **Jennifer Daniels-Higginbotham:** Writing – review & editing, Investigation, Data curation. **Stijn Desmyter:** Writing – review & editing, Investigation, Data curation. **Marta Diepenbroek:** Writing – review & editing, Investigation, Data curation. **Claudia Wöss:** Writing – review & editing, Validation, Data curation. **Nicole Huber:** Writing – review & editing, Validation, Software. **Arne Dür:** Writing – review & editing, Validation, Software. **Walther Parson:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization. **Federica Giangasparo:** Writing – review & editing, Investigation, Data curation. **Nair Gouveia:** Writing – review & editing, Investigation, Data curation. **Sophie Dognaux:** Writing – review & editing, Investigation, Data curation. **Bethany Forsythe:** Writing – review & editing, Investigation, Data curation. **Kristiaan J. van der Gaag:** Writing – review & editing, Investigation, Data curation. **Oscar Garcia:** Writing – review & editing, Investigation, Data curation. **Peter Schneider:** Writing – review & editing, Investigation, Data curation. **Cibeles Serna-Menor:** Writing – review & editing, Investigation, Data curation. **Gabriela Huber:** Writing – review & editing, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Bing Hong Shue:** Writing – review & editing, Investigation, Data curation. **Katja Anslinger:** Writing – review & editing, Investigation, Data curation. **Maja Sidstedt:** Writing – review & editing, Investigation, Data curation. **David Ballard:** Writing – review & editing, Investigation, Data curation. **Maria de la Puente:** Writing – review & editing, Investigation, Data curation. **Erica L. Romsos:** Writing – review & editing, Investigation, Data curation. **Lisa Sampaio:** Writing – review & editing, Investigation, Data curation. **Francesca Brisighelli:** Writing – review & editing, Investigation, Data curation. **Lauren Canale:** Writing – review & editing, Investigation, Data curation. **Gianluca Casamassima:** Writing – review & editing,

Investigation, Data curation. **Titia Sijen:** Writing – review & editing, Investigation, Data curation. **Pedro A. Barrio Caballero:** Writing – review & editing, Investigation, Data curation. **Kimberly Sturk-Andreaggi:** Writing – review & editing, Investigation, Data curation. **Magdalena Bogus:** Writing – review & editing, Investigation, Data curation. **Christopher Kiu-Choong Syn:** Writing – review & editing, Investigation, Data curation. **Claus Børsting:** Writing – review & editing, Investigation, Data curation. **Michael D. Brandhagen:** Writing – review & editing, Investigation, Data curation. **Jennifer A. McElhoe:** Writing – review & editing, Investigation, Data curation. **Lena Ewers:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Formal analysis, Data curation. **Marta Montesino:** Writing – review & editing, Investigation, Data curation. **Marie-Laure Morerod:** Writing – review & editing, Investigation, Data curation. **Ana Mosquera-Miguel:** Writing – review & editing, Investigation, Data curation. **Charla Marshall:** Writing – review & editing, Investigation, Data curation. **Patrik Matejov:** Writing – review & editing, Investigation, Data curation. **Jana Naue:** Writing – review & editing, Investigation, Data curation. **Leire Palencia-Madrid:** Writing – review & editing, Investigation, Data curation. **Vincenzo Pascali:** Writing – review & editing, Investigation, Data curation. **Vania Pereira:** Writing – review & editing, Investigation, Data curation.

Editorial statement

Given his role as Editor-In-Chief of Forensic Science International Genetics, Walther Parson had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank the sample donors and the EDNAP group for their valuable discussions. We also gratefully acknowledge the following colleagues for their technical assistance and/or their supporting contributions to the manuscript through insightful discussions (in alphabetical order): Antonio Alonso, Birgit Bayer, Jodi A. Irwin, Anthony Onorato, Chris Phillips, Pablo Martín, Lourdes Prieto, Timo Sänger, Peter M. Vallone. The authors from the INTCF (National Institute of Toxicology and Forensic Science, Department of Madrid, Spain) would like to thank their Center for its institutional support and for many ongoing, meaningful discussions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.fsigen.2026.103593](https://doi.org/10.1016/j.fsigen.2026.103593).

References

- [1] B.D. Greenberg, J.E. Newbold, A. Sugino, Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA, *Gene* 21 (1–2) (1983) 33–49, [https://doi.org/10.1016/0378-1119\(83\)90145-2](https://doi.org/10.1016/0378-1119(83)90145-2).
- [2] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, et al., Identification of the remains of the Romanov family by DNA analysis, *Nat. Genet.* 6 (2) (1994) 130–135, <https://doi.org/10.1038/ng0294-130>.
- [3] P.L. Ivanov, M.J. Wadhams, R.K. Roby, M.M. Holland, V.W. Weedn, T.J. Parsons, Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nat. Genet.* 12 (4) (1996) 417–420, <https://doi.org/10.1038/ng0496-417>.
- [4] M.D. Coble, O.M. Loreille, M.J. Wadhams, S.M. Edson, K. Maynard, C.E. Meyer, et al., Mystery solved: the identification of the two missing Romanov children using DNA analysis, *PLoS. One* 4 (3) (2009) e4838, <https://doi.org/10.1371/journal.pone.0004838>.

- [5] E.I. Rogaev, A.P. Grigorenko, Y.K. Moliaka, G. Faskhutdinova, A. Goltsov, A. Lahti, et al., Genomic identification in the historical case of the Nicholas II royal family, *Proc. Natl. Acad. Sci. USA* 106 (13) (2009) 5258–5263, <https://doi.org/10.1073/pnas.0811190106>.
- [6] C.L. Davis, *Mitochondrial DNA: State of tennessee v. paul ware*, *Profiles DNA* 1 (3) (1998) 6–7.
- [7] M.R. Wilson, D. Polanskey, J. Replogle, J.A. DiZinno, B. Budowle, A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes, *Hum. Genet.* 100 (2) (1997) 167–171, <https://doi.org/10.1007/s004390050485>.
- [8] T.J. Parsons, D.S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M. R. Wilson, et al., A high observed substitution rate in the human mitochondrial DNA control region, *Nat. Genet.* 15 (4) (1997) 363–368, <https://doi.org/10.1038/ng0497-363>.
- [9] J. Naue, S. Hörer, T. Sängler, C. Strobl, P. Hatzler-Grubwieser, W. Parson, et al., Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA, *Mitochondrion* 20 (2015) 82–94, <https://doi.org/10.1016/j.mito.2014.12.002>.
- [10] G. Tully, S.M. Barritt, K. Bender, E. Brignon, C. Capelli, N. Dima-Simonin, et al., Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts, *Forensic Sci. Int.* 140 (1) (2004) 1–11, [https://doi.org/10.1016/s0379-0738\(03\)00181-6](https://doi.org/10.1016/s0379-0738(03)00181-6).
- [11] S. Desmyter, M. Bodner, G. Huber, S. Dogniaux, C. Berger, F. Noël, et al., Hair matters: MtDNA quantity and sequence variation along and among human head hairs, *Forensic Sci. Int. Genet.* 25 (2016) 1–9, <https://doi.org/10.1016/j.fsigen.2016.07.012>.
- [12] J.M. Gallimore, J.A. McElhoe, M.M. Holland, Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach, *Forensic Sci. Int. Genet.* 32 (2018) 7–17, <https://doi.org/10.1016/j.fsigen.2017.09.013>.
- [13] K.J. van der Gaag, S. Desmyter, S. Smit, L. Prieto, T. Sijen, Reducing the Number of Mismatches between Hairs and Buccal References When Analysing mtDNA Heteroplasmic Variation by Massively Parallel Sequencing, *Genes (Basel)* 11 (11) (2020), <https://doi.org/10.3390/genes11111355>.
- [14] J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, et al., Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples, *J. Mol. Evol.* 68 (5) (2009) 516–527, <https://doi.org/10.1007/s00239-009-9227-4>.
- [15] M.M. Rathbun, J.A. McElhoe, W. Parson, M.M. Holland, Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data, *Forensic Sci. Int. Genet.* 26 (2017) 1–11, <https://doi.org/10.1016/j.fsigen.2016.09.008>.
- [16] M.D.M. González, A. Ramos, M.P. Aluja, C. Santos, Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing, *Mitochondrion* 50 (2020) 88–93, <https://doi.org/10.1016/j.mito.2019.10.006>.
- [17] R.S. Just, J.A. Irwin, W. Parson, Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing, *Forensic Sci. Int. Genet.* 18 (2015) 131–139, <https://doi.org/10.1016/j.fsigen.2015.05.003>.
- [18] M.D. Brandhagen, R.S. Just, J.A. Irwin, Validation of NGS for mitochondrial DNA casework at the FBI Laboratory, *Forensic Sci. Int. Genet.* 44 (2020) 102151, <https://doi.org/10.1016/j.fsigen.2019.102151>.
- [19] K. Sturk-Andreaggi, W. Parson, M. Allen, C. Marshall, Impact of the sequencing method on the detection and interpretation of mitochondrial DNA length heteroplasmy, *Forensic Sci. Int. Genet.* 44 (2020) 102205, <https://doi.org/10.1016/j.fsigen.2019.102205>.
- [20] M. Eduardoff, G. Huber, B. Bayer, D. Schmid, K. Anslinger, T. Göbel, et al., Mass spectrometric base composition profiling: Implications for forensic mtDNA databasing, *Forensic Sci. Int. Genet.* 7 (6) (2013) 587–592, <https://doi.org/10.1016/j.fsigen.2013.05.007>.
- [21] C. Xavier, M. Eduardoff, C. Strobl, W. Parson, SD quantS-Sensitive detection tetraplex-system for nuclear and mitochondrial DNA quantification and degradation inference, *Forensic Sci. Int. Genet.* 42 (2019) 39–44, <https://doi.org/10.1016/j.fsigen.2019.06.004>.
- [22] M.F. Kavlick, H.S. Lawrence, R.T. Merritt, C. Fisher, A. Isenberg, J.M. Robertson, et al., Quantification of human mitochondrial DNA using synthesized DNA standards, *J. Forensic Sci.* 56 (6) (2011) 1457–1463, <https://doi.org/10.1111/j.1556-4029.2011.01871.x>.
- [23] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, et al., DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci. Int. Genet.* 13 (2014) 134–142, <https://doi.org/10.1016/j.fsigen.2014.07.010>.
- [24] A. Dür, N. Huber, W. Parson, Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences, *Int. J. Mol. Sci.* 22 (11) (2021), <https://doi.org/10.3390/ijms22115747>.
- [25] W. Parson, H.J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, *Forensic Sci. Int. Genet.* 1 (1) (2007) 13–19, <https://doi.org/10.1016/j.fsigen.2006.11.003>.
- [26] C. Strobl, M. Eduardoff, M.M. Bus, M. Allen, W. Parson, Evaluation of the precision ID whole MtDNA genome panel for forensic analyses, *Forensic Sci. Int. Genet.* 35 (2018) 21–25, <https://doi.org/10.1016/j.fsigen.2018.03.013>.
- [27] J.C. Cihlar, D. Peters, C. Strobl, W. Parson, B. Budowle, The lot-to-lot variability in the mitochondrial genome of controls, *Forensic Sci. Int. Genet.* 47 (2020) 102298, <https://doi.org/10.1016/j.fsigen.2020.102298>.
- [28] S. Desmyter, S. Dogniaux, F. Noel, L. Prieto, Base specific variation rates at mtDNA positions 16093 and 16183 in human hairs, *Forensic Sci. Int. Genet.* 43 (2019) 102142, <https://doi.org/10.1016/j.fsigen.2019.102142>.
- [29] J. Naue, C. Xavier, S. Hörer, W. Parson, S. Lutz-Bonengel, Assessment of mitochondrial DNA copy number variation relative to nuclear DNA quantity between different tissues, *Mitochondrion* 74 (2024) 101823, <https://doi.org/10.1016/j.mito.2023.11.006>.
- [30] W. Parson, A. Brandstätter, A. Alonso, N. Brandt, B. Brinkmann, A. Carracedo, et al., The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives, *Forensic Sci. Int.* 139 (2–3) (2004) 215–226, <https://doi.org/10.1016/j.forsci.2003.11.008>.
- [31] C. Marshall, W. Parson, Interpreting NUMTs in forensic genetics: Seeing the forest for the trees, *Forensic Sci. Int. Genet.* 53 (2021) 102497, <https://doi.org/10.1016/j.fsigen.2021.102497>.