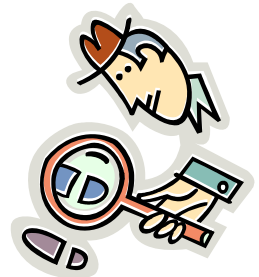




Mitchell M. Holland, Ph.D.  
Professor, Biochem & MolBio  
Former Director, Forensic Science Program  
Eberly College of Science  
Penn State University, University Park, PA

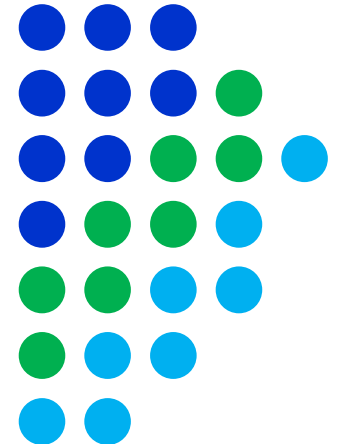


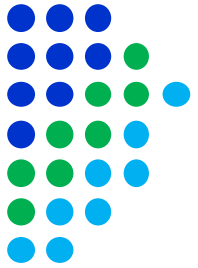
# *Routine mitoGenome MPS Analysis from 1 mm of Human Hair Shaft*

**17 November 2021**

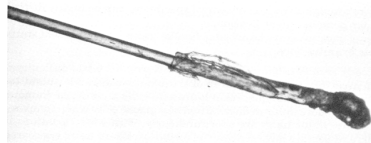
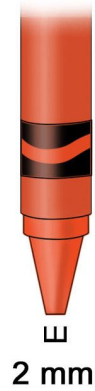


**TECHTOUR**  
2021





Millimeters (mm)  
10 mm  
0 1 2  
Centimeters (cm)



10 mm

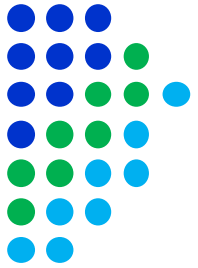


20 mm



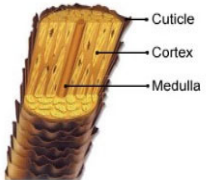
50 mm



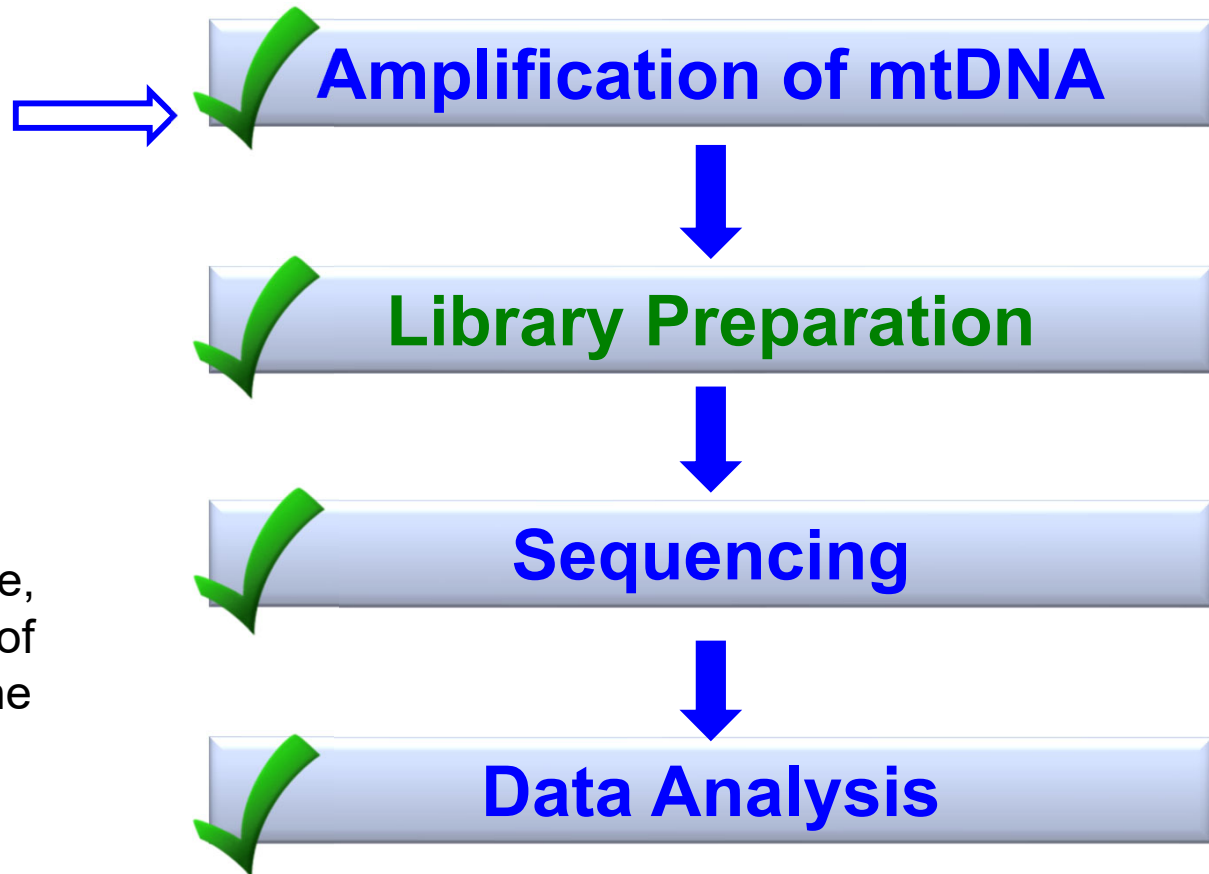


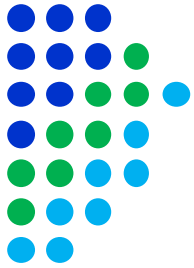
# MPS mtDNA Workflow

**DNA  
Extraction**



DNA extraction is the same, but with MPS the amount of information gained from the extract is significantly increased

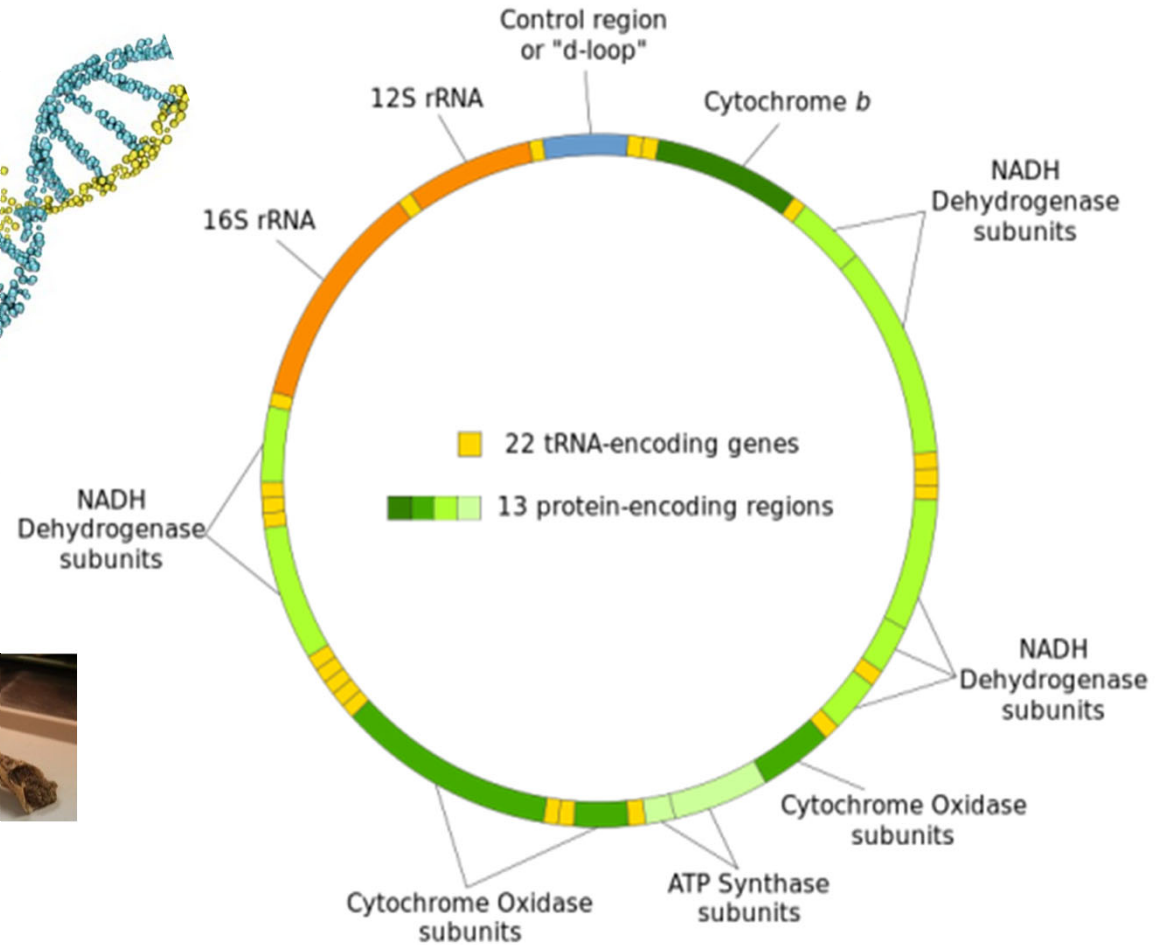
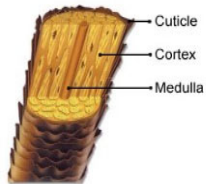
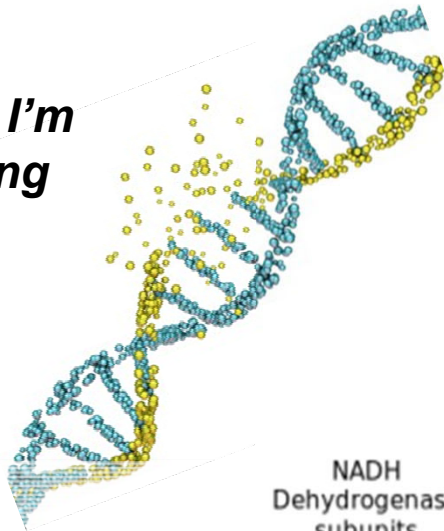




HV1/HV2 = 610 bps

CR = 1122 bps

Help me, I'm degrading



mitoGenome = ~16,569 bps



Forensic Science International: Genetics 12 (2014) 128–135

Contents lists available at ScienceDirect

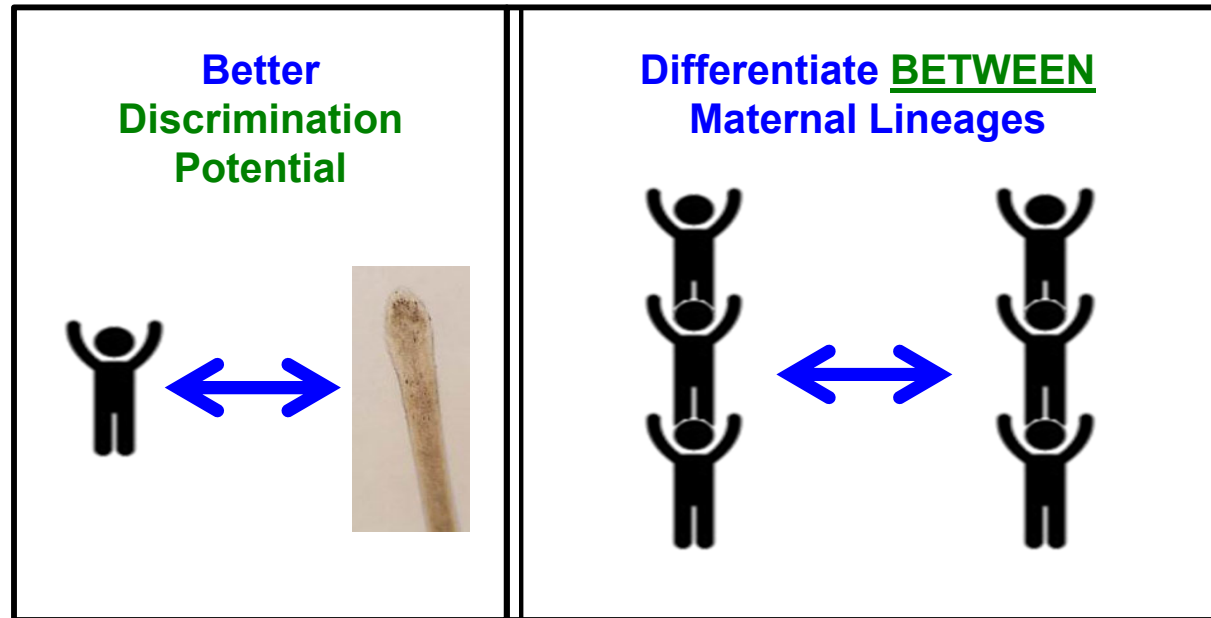
**Forensic Science International: Genetics**

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

CrossMark

High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq

Jonathan L. King<sup>a,1,\*</sup>, Bobby L. LaRue<sup>a,1</sup>, Nicole M. Novroski<sup>a</sup>, Monika Stoljarova<sup>a</sup>, Seung Bum Seo<sup>a</sup>, Xiangpei Zeng<sup>a</sup>, David H. Warshauer<sup>a</sup>, Carey P. Davis<sup>a</sup>, Walther Parson<sup>b,c</sup>, Antti Sajantila<sup>a,d</sup>, Bruce Budowle<sup>a,e</sup>



Populations	n	HVI/HVII		mtGenome	
		RMP	GD	RMP	GD
AFA	87	2.42%	98.72%	1.31%	99.84%
CAU	83	3.12%	98.06%	1.20%	100.00%
HIS	113	3.33%	97.53%	0.98%	99.91%
Mean ± SD		2.96 ± 0.48%	98.10 ± 0.59%	1.16 <sup>c</sup> ± 0.17%	99.91 <sup>d</sup> ± 0.08%



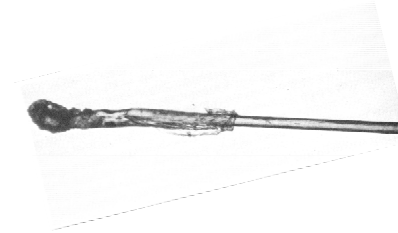
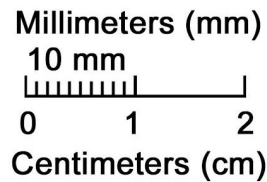
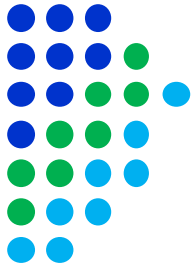
# Amplification Approaches & Kits Available



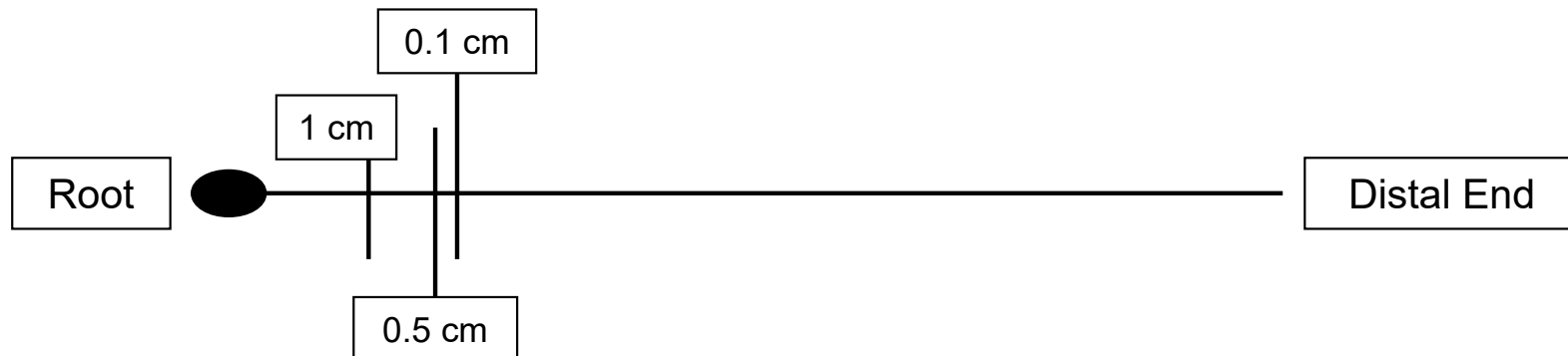
- Promega
  - PowerSeq CRM (control region, 1 multiplex, 144-237 bps)
  - **PowerSeq WGM (mitogenome, 1 multiplex of 161 amplicons averaging 167 bps, research product)**
- Verogen
  - ForenSeq mtDNA Control Region (2 multiplexes, 18 amplicons averaging 118 bps)
  - ForenSeq mtDNA Whole Genome (2 multiplexes, 245 amplicons averaging 131 bps)
- ThermoFisher
  - Precision ID mtDNA Control Region Panel (2 multiplexes)
  - Precision ID mtDNA Whole Genome Panel (2 multiplexes of 81 amplicons averaging 161 bps)

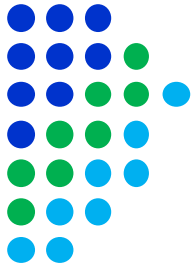


# Hair Study



Tested 1 mm and 5 mm cuttings from 60 hair shafts (120 samples).  
Approximately 1 cm of the root end was removed, followed by the 5 mm cutting, and finally the 1 mm cutting.





---

LMB

---

1x Terg-a-zyme Wash  
Qiagen ATL buffer  
Proteinase K  
DTT  
Qiagen AL Buffer  
Magnetic Beads  
Isopropanol  
PrepFiler Wash Buffer A  
PrepFiler Wash Buffer B  
PrepFiler Elution Buffer

---

**LMB** = **L**yse  
(dissolve) &  
**M**agnetic **B**eads

**40 uL extract**

Forensic Science International: Genetics 32 (2018) 7–17

Contents lists available at ScienceDirect

**Forensic Science International: Genetics**




journal homepage: [www.elsevier.com/locate/bscig](http://www.elsevier.com/locate/bscig)

Research Paper

Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach

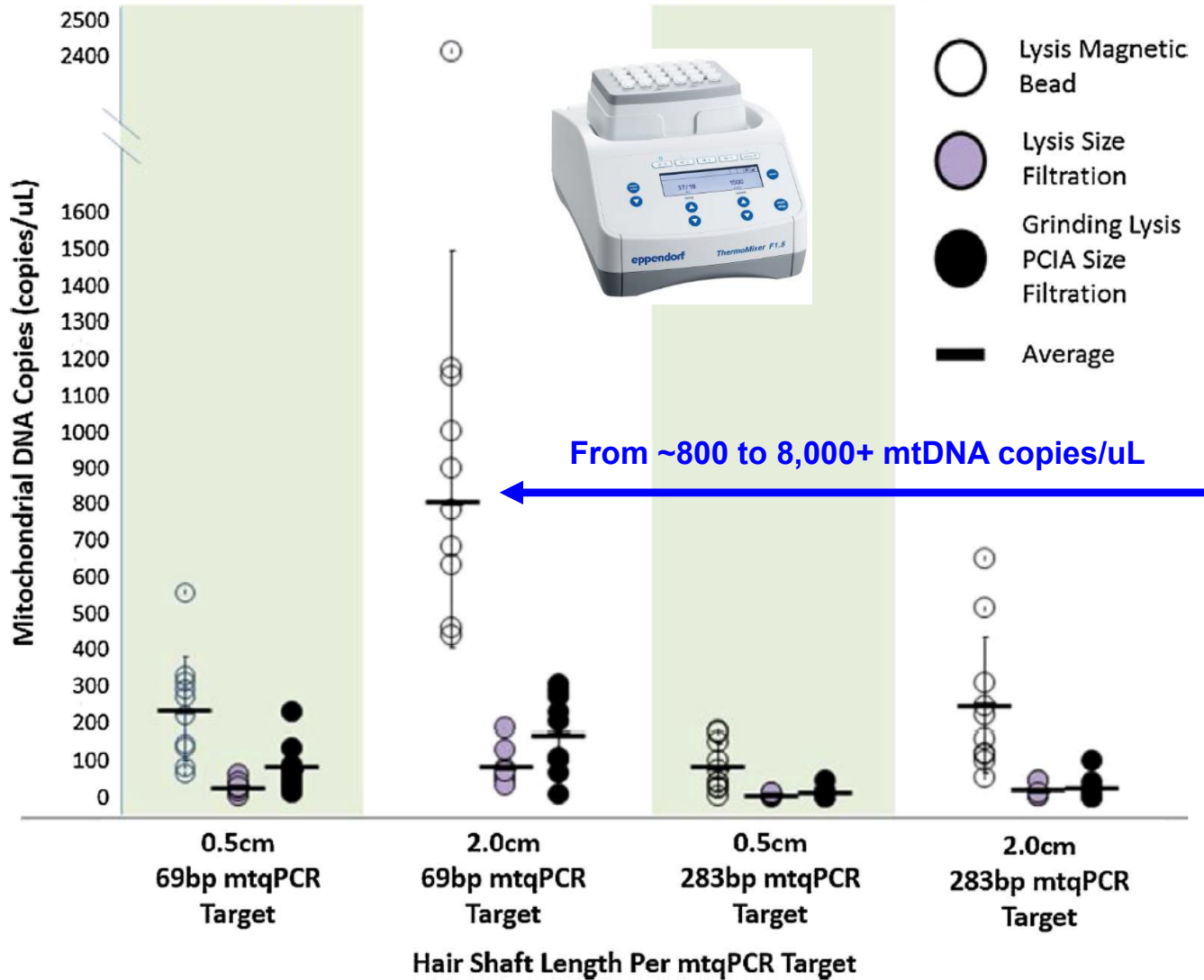
Jamie M. Gallimore, Jennifer A. McElhoe, Mitchell M. Holland\*

Forensic Science Program, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 014 Thomas Building, University Park, PA 16802, United States





# Mitochondrial DNA Extraction Method Comparison



**With the 2-TM approach yields increased by ~10X**

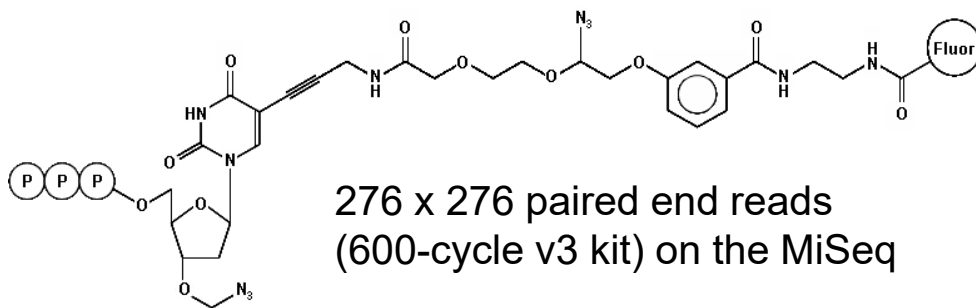
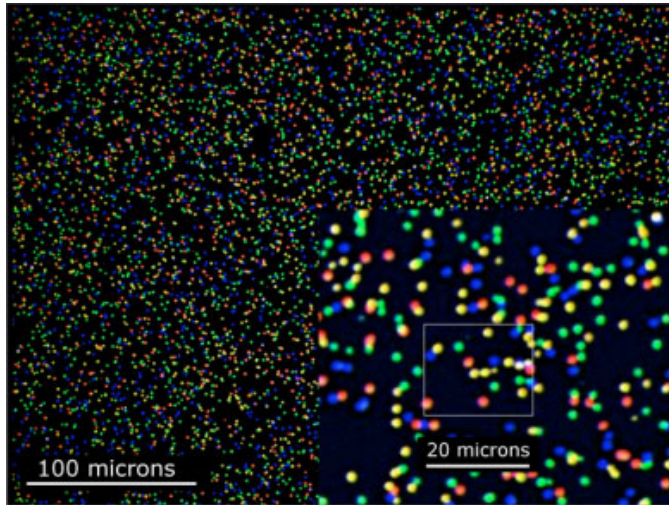
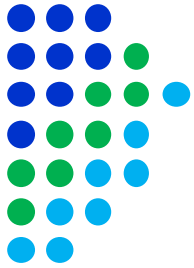


**Amplification (targeting 40,000 copies, but as low as 100 copies) and Library Prep with the Promega PowerSeq WGM kit**



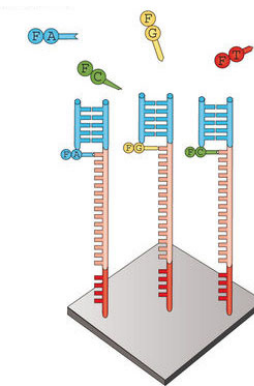


# Reverse Terminator Sequencing

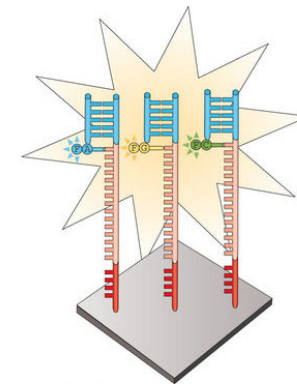


276 x 276 paired end reads  
(600-cycle v3 kit) on the MiSeq

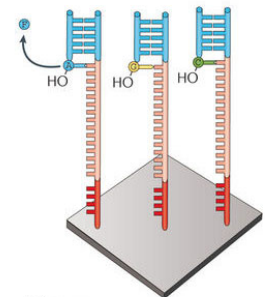
10's of millions of clusters  
generating Gbases of DNA  
sequence



**Nucleotide addition**  
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.



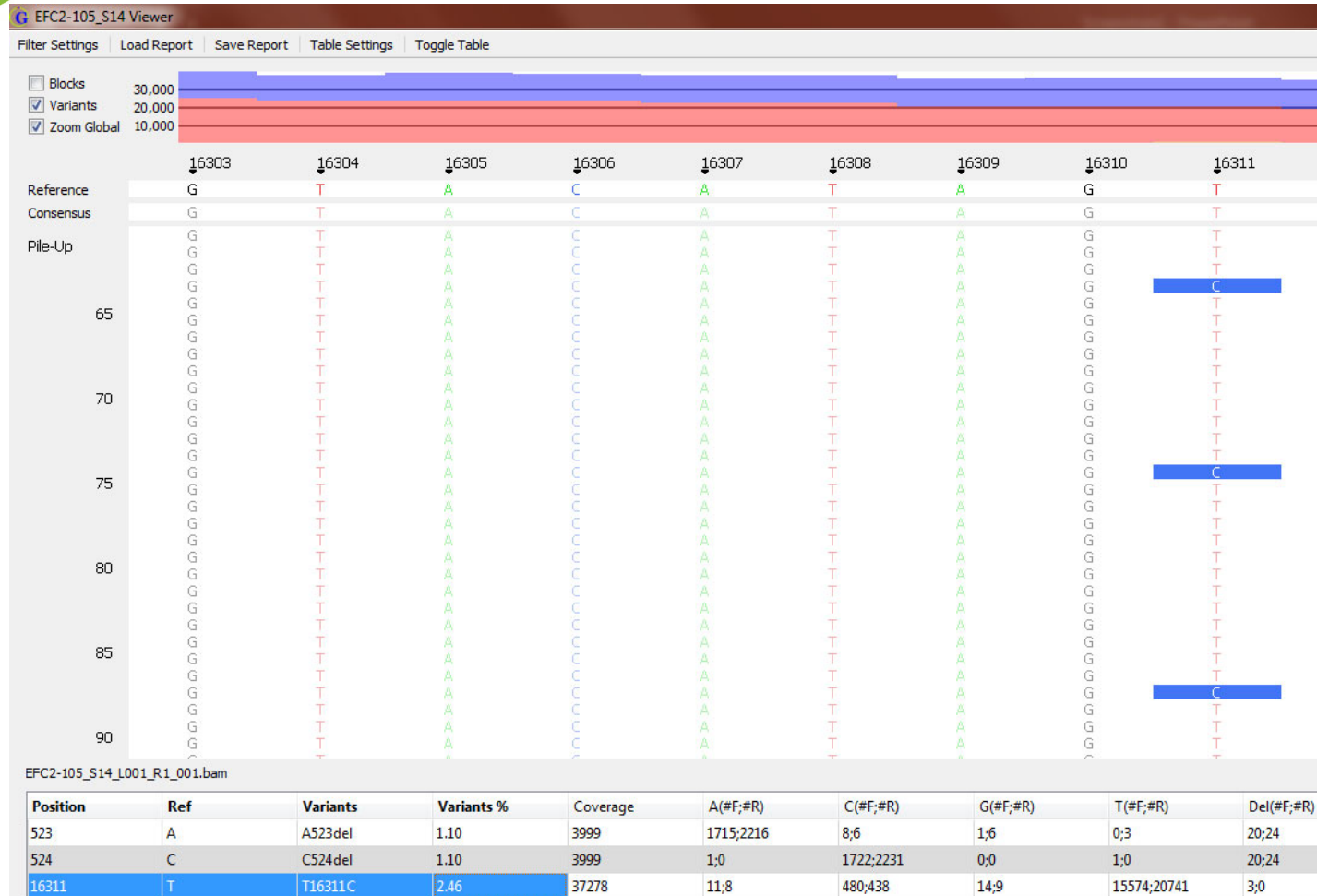
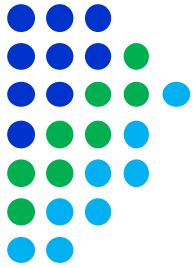
**Imaging**  
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.



**Cleavage**  
Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.



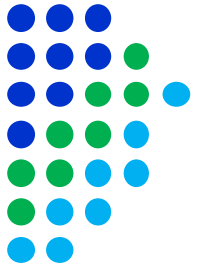
# Data Analysis



Minor Variant = a base call with <50% of the reads and  $\geq 2\%$  of the reads



# Hair Study



Head Hairs in Three Different Age Ranges:

Recent (**R**) = <5 years of age (13 hairs)

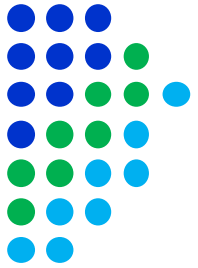
Old (**O**) = 5-27 years, avg of 13.6 (24 hairs)

Older (**VO**) = 41-46 years, avg of 43.4 (23 hairs)

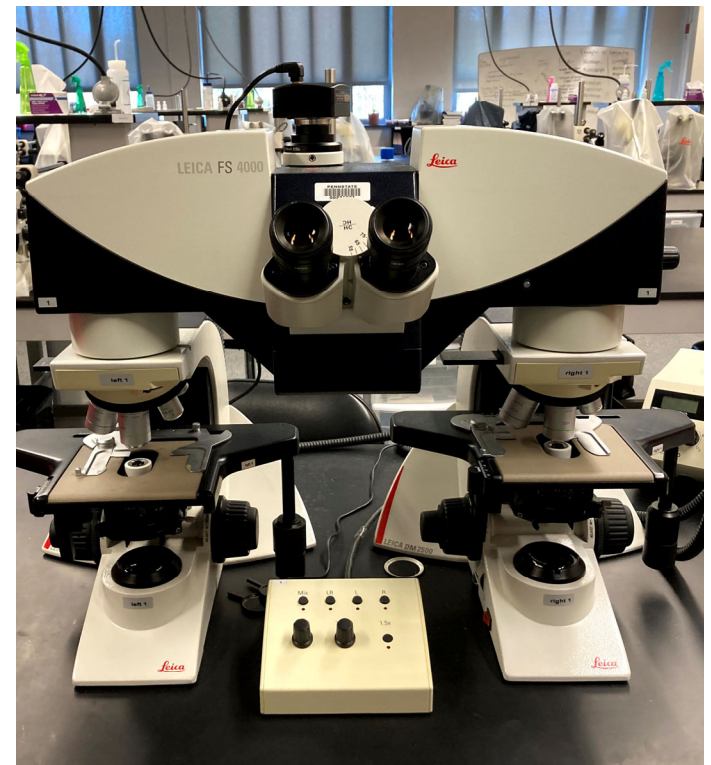
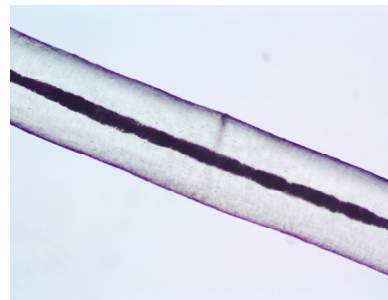
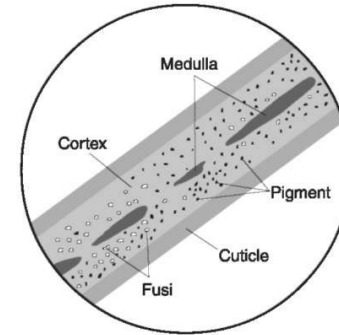




# Microscopic Characterization



- Microscopically characterized on a Leica FS 4000
- Medulla structure
- Diameter
- Other characteristics such as pigment, ovoid bodies, cortical fusi, cuticle structure, physical damage





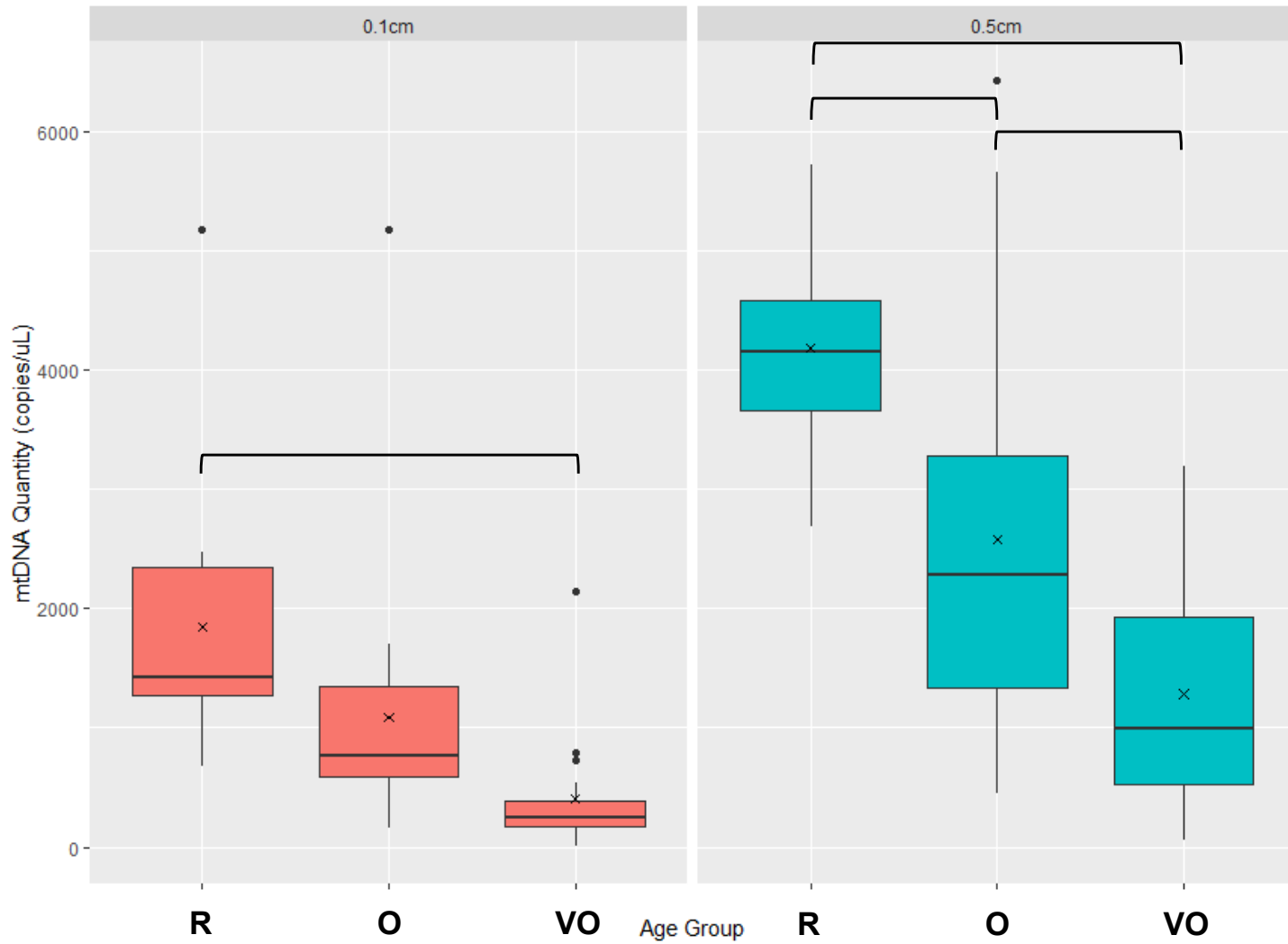
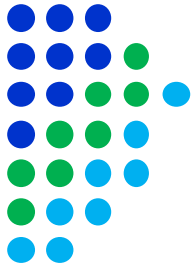
# Statistics



- Rosner test for **outliers**
- Shapiro-Wilkes test for **normality**
- **ANOVA** for datasets that exceed the 20 samples requirement for assumed normality
  - ***Tukey Post Hoc***
  - ***Kruskal-Wallis*** for datasets that don't meet 20 samples
    - ***Dunn's Post Hoc with Holm correction***
- **R Studio**



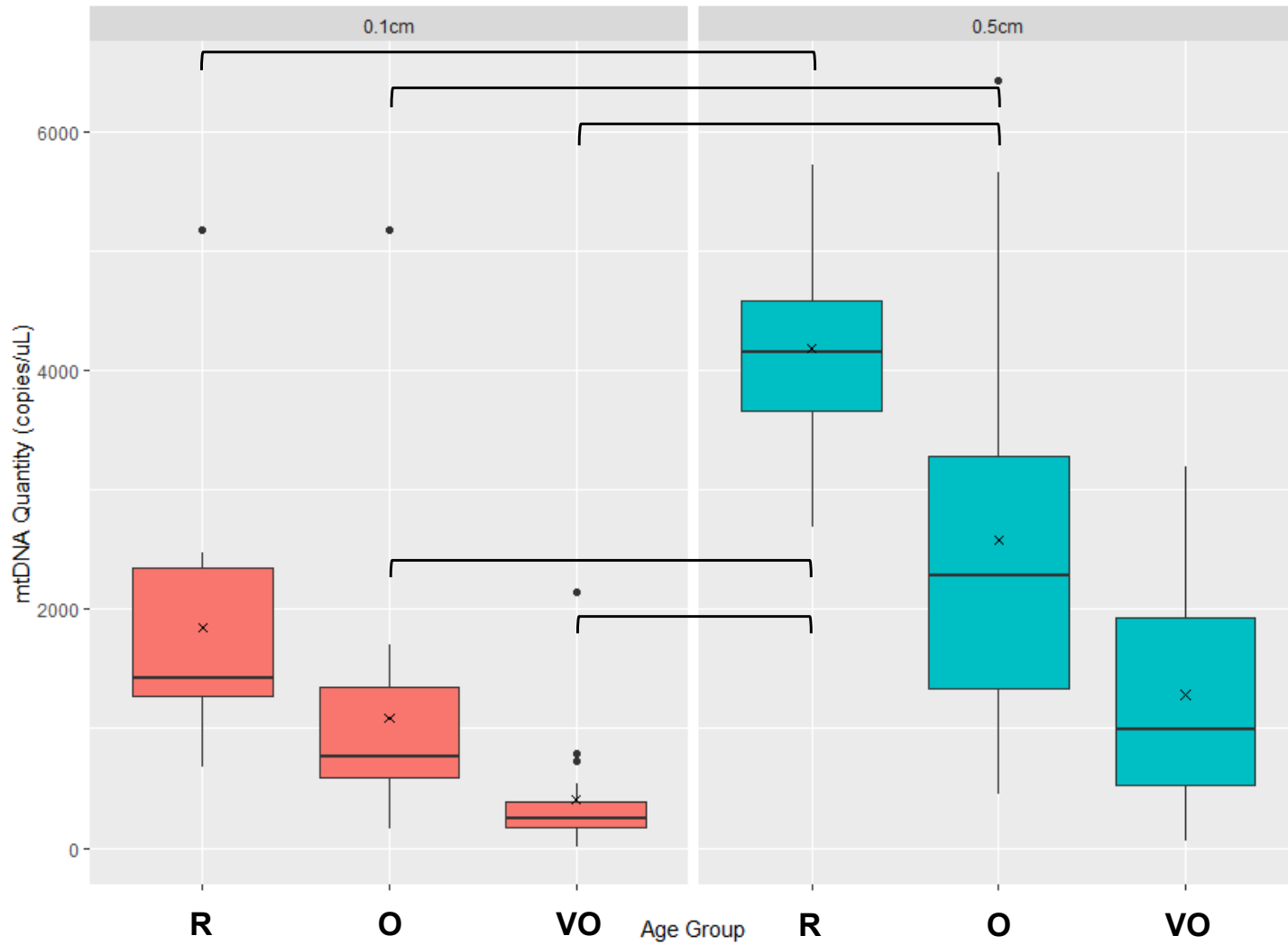
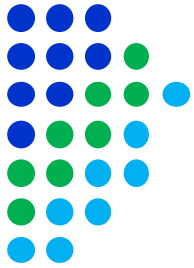
# mtDNA Yield v. Age of the Hair



┌───┐ p-value < 0.05 (within)

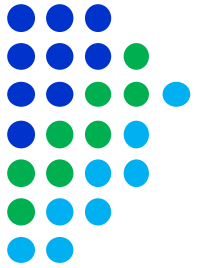


# mtDNA Yield v. Age of the Hair

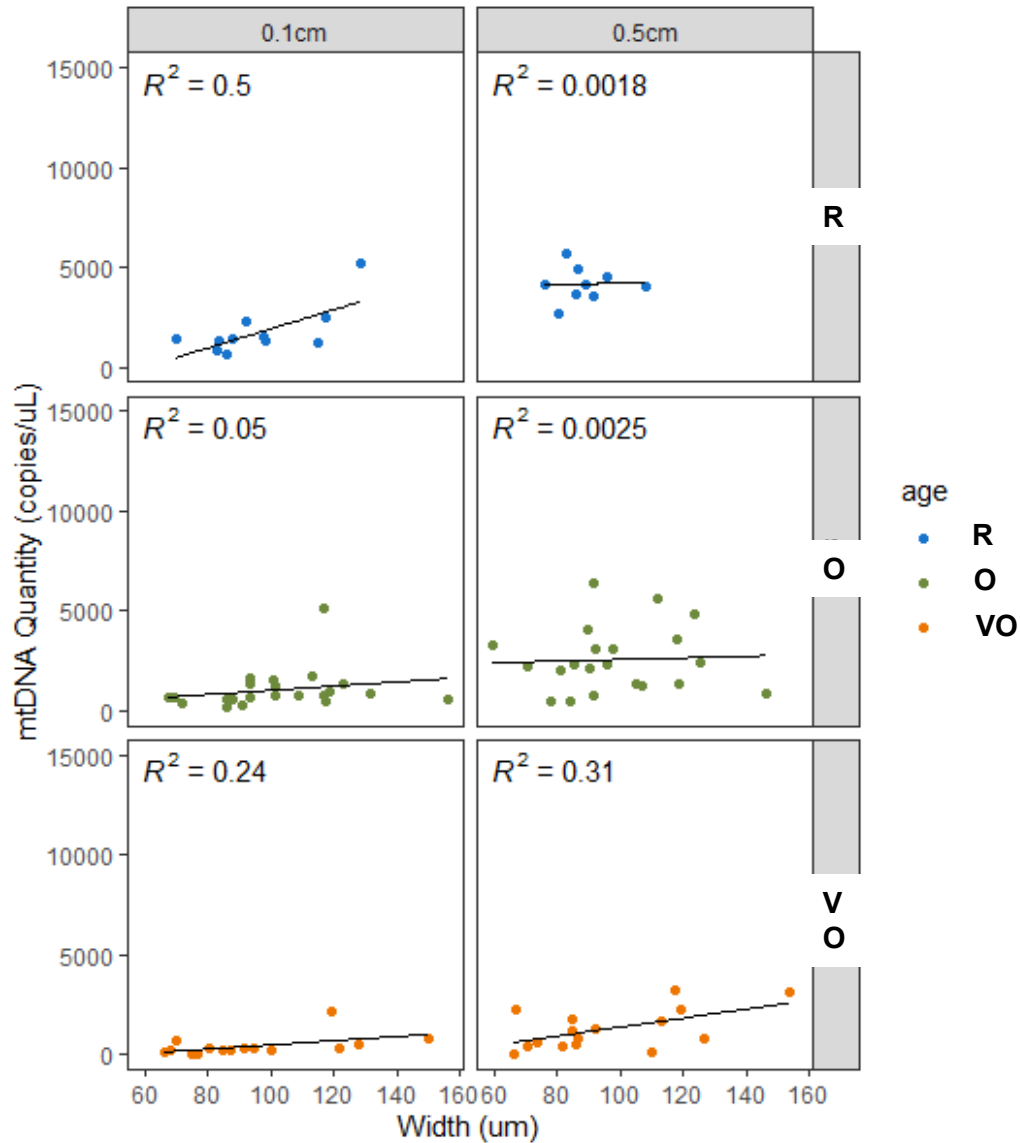


┌───┐ p-value < 0.05 (between)



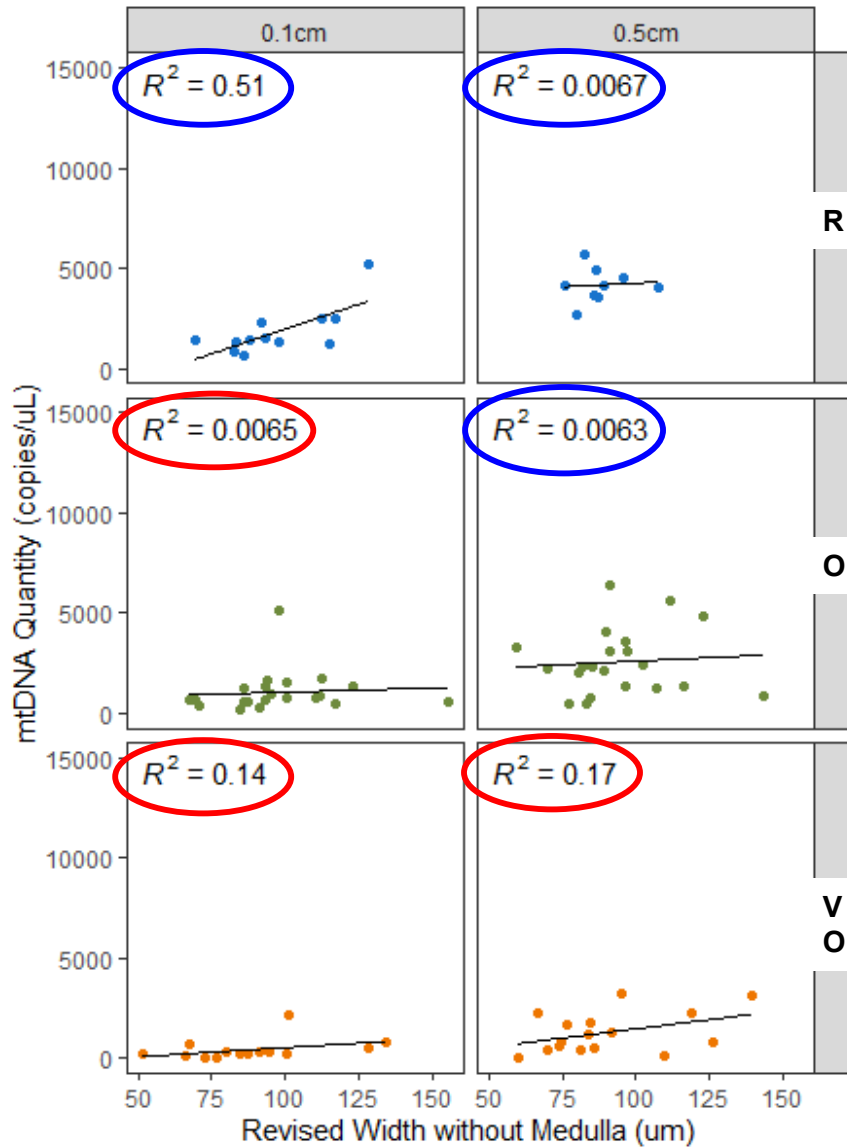
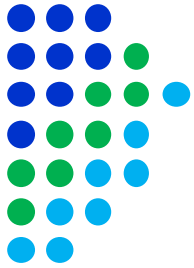


# mtDNA Yield v. Width of Hair





# mtDNA Yield v. Width of Hair



= Better

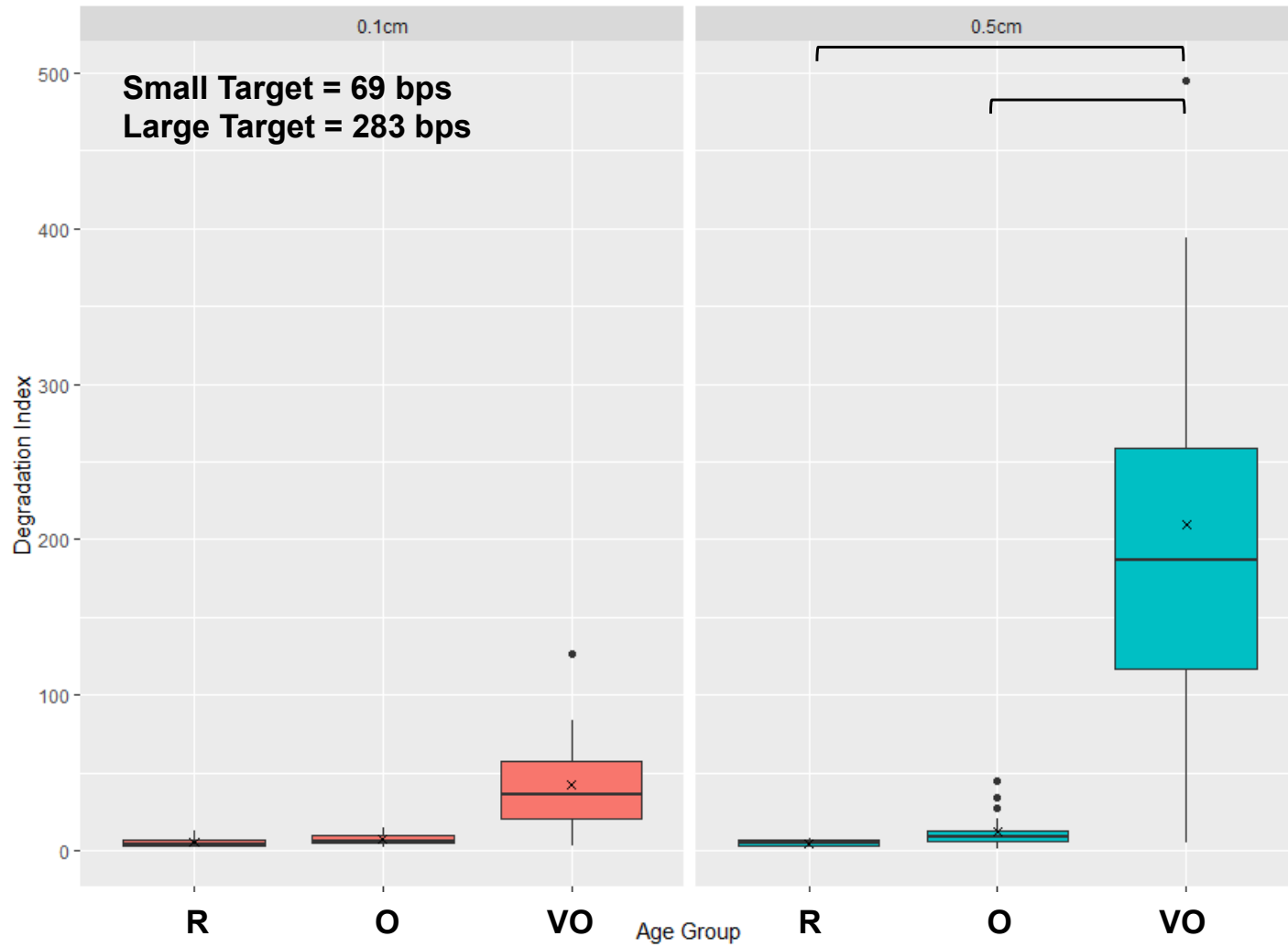
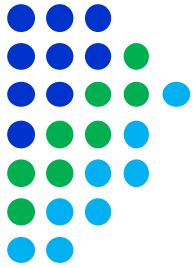
= Worse

age  
 • R  
 • O  
 • VO

**Width without a medulla**



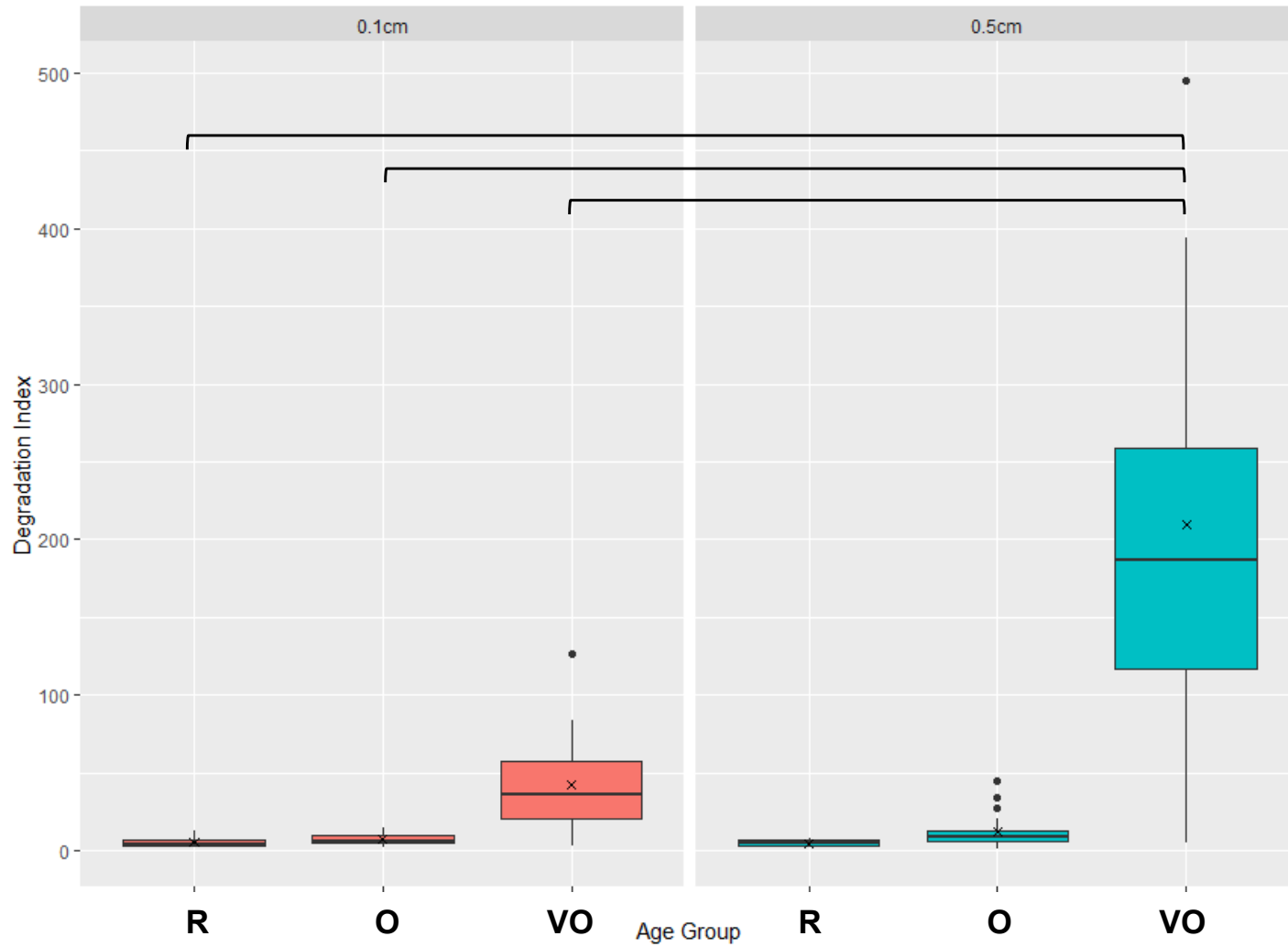
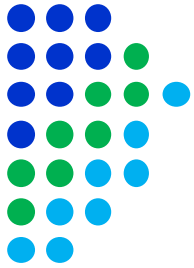
# Degradation Index (DI) v. Age



┌───┐ p-value < 0.05 (within)



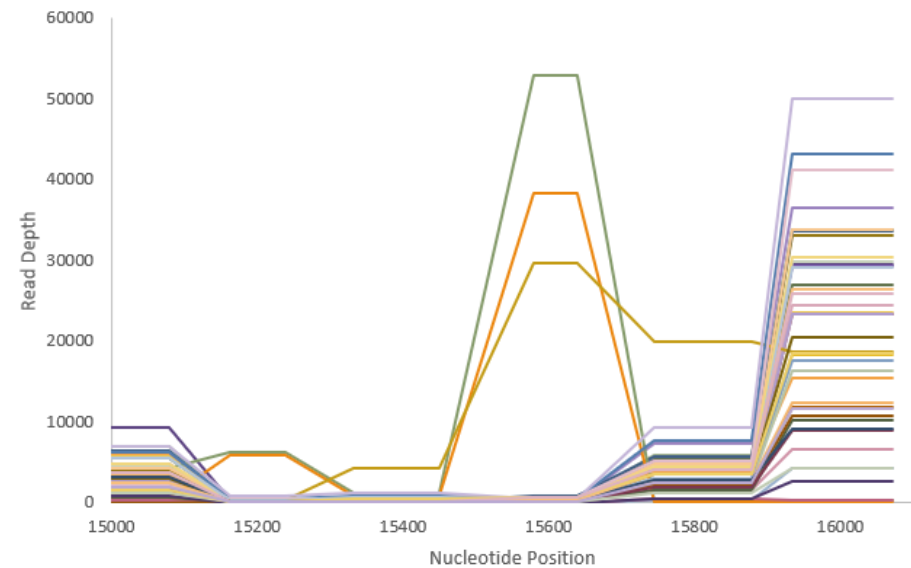
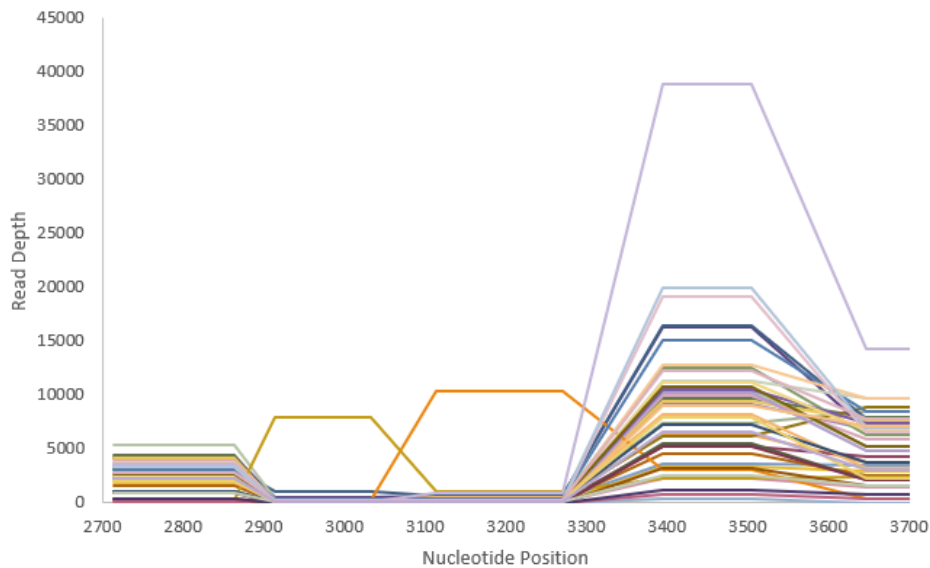
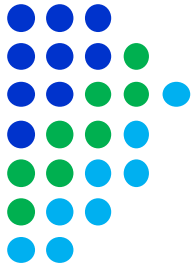
# Degradation Index (DI) v. Age



┌───┐ p-value < 0.05 (between)



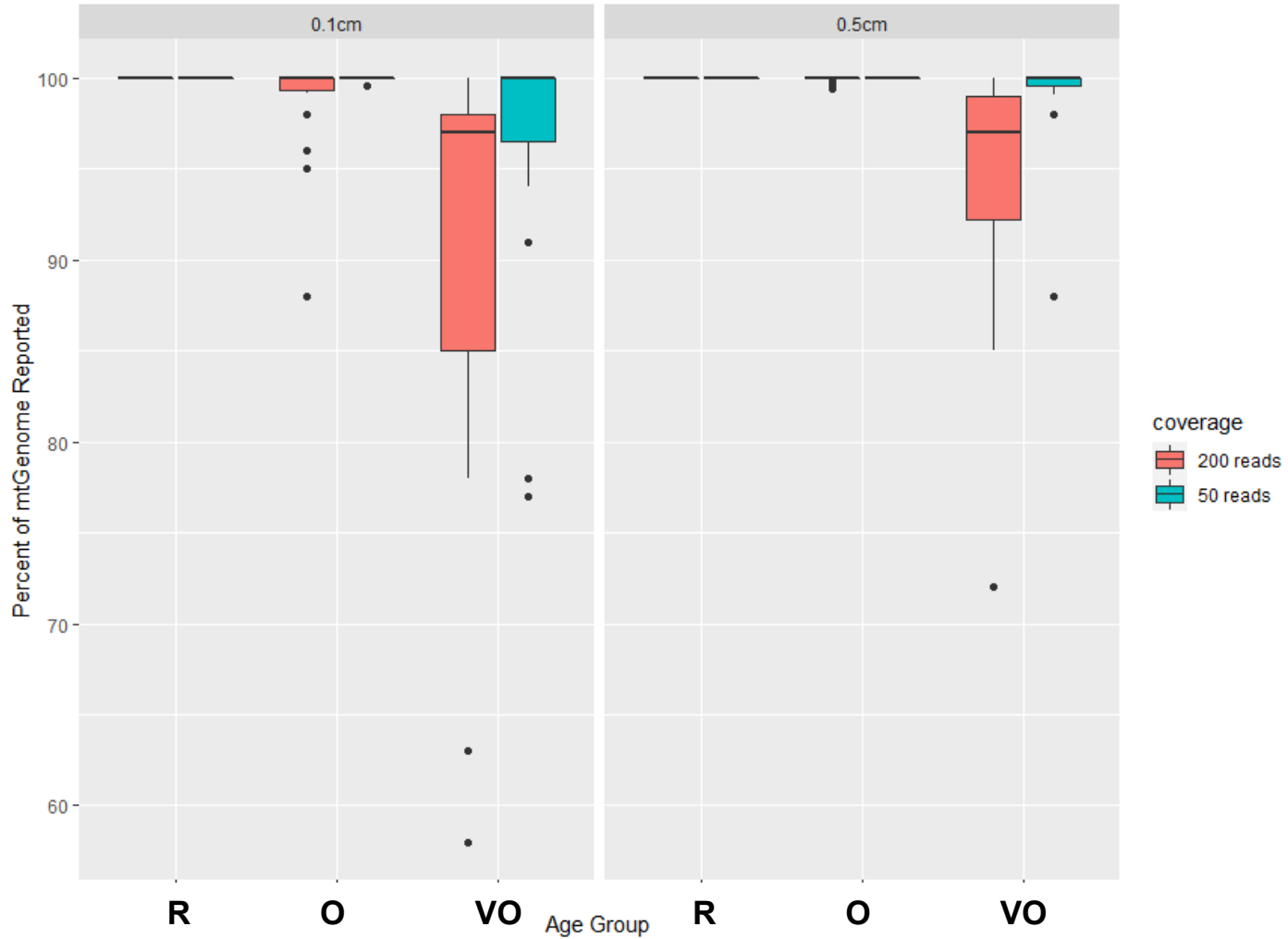
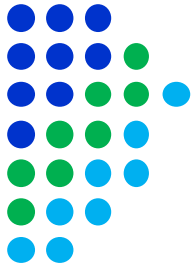
# Amplicon “Dropout”



Plot of MPS read depth across portions of the mitogenome for 41 samples with lower overall coverage. Certain amplicons dropout above 200 reads, however most haplotypes can be fully reported if read depth is lowered to 50. The dips in coverage may be due to regions with lower amplification or sequencing efficiencies.



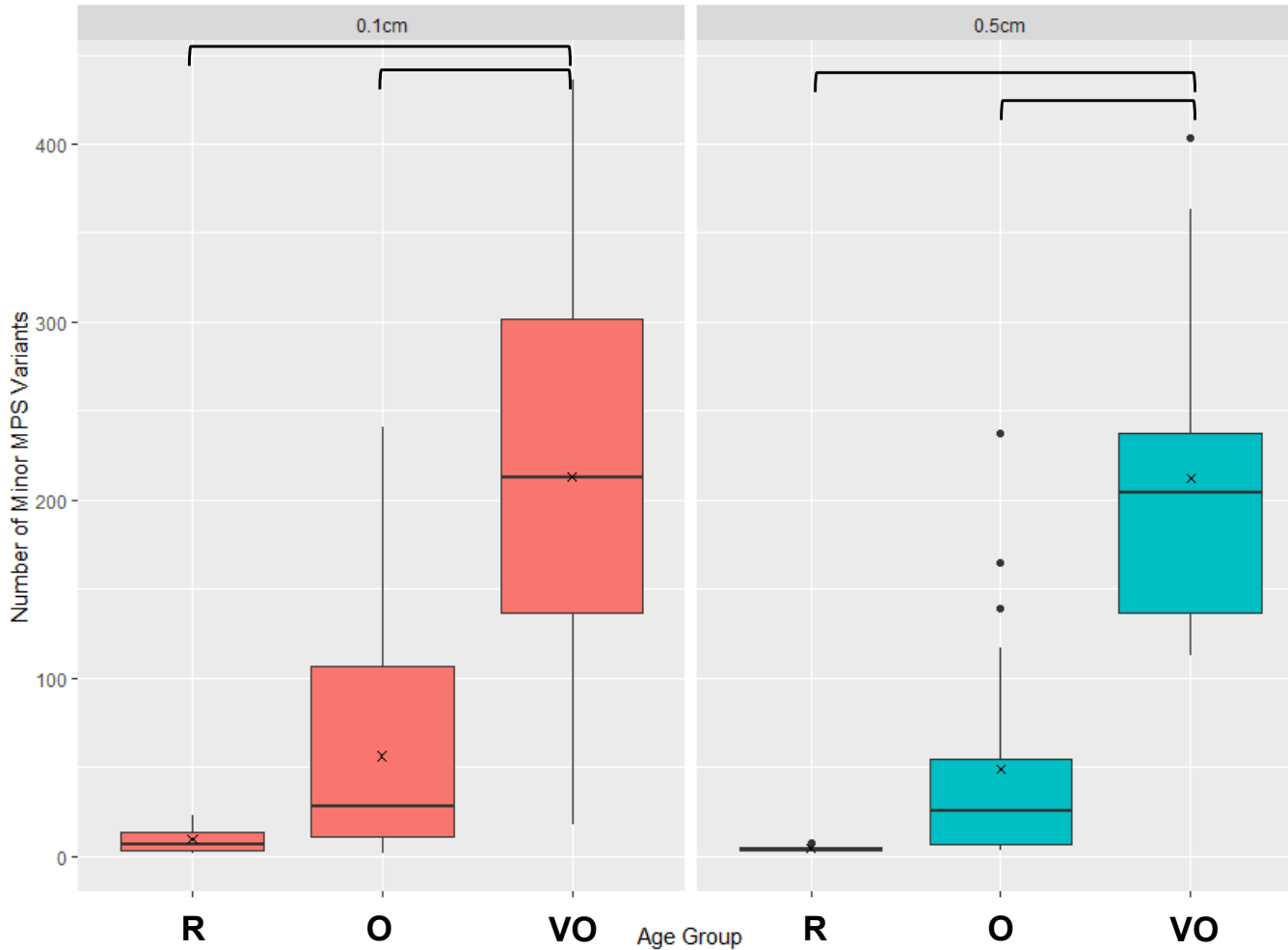
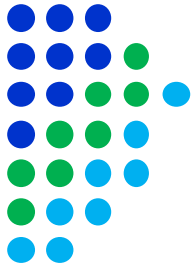
# Percent of the mitoGenome Reported v. Age



┌──────────┐ p-value < 0.05 (between)



# Number of Minor Variants v. Age

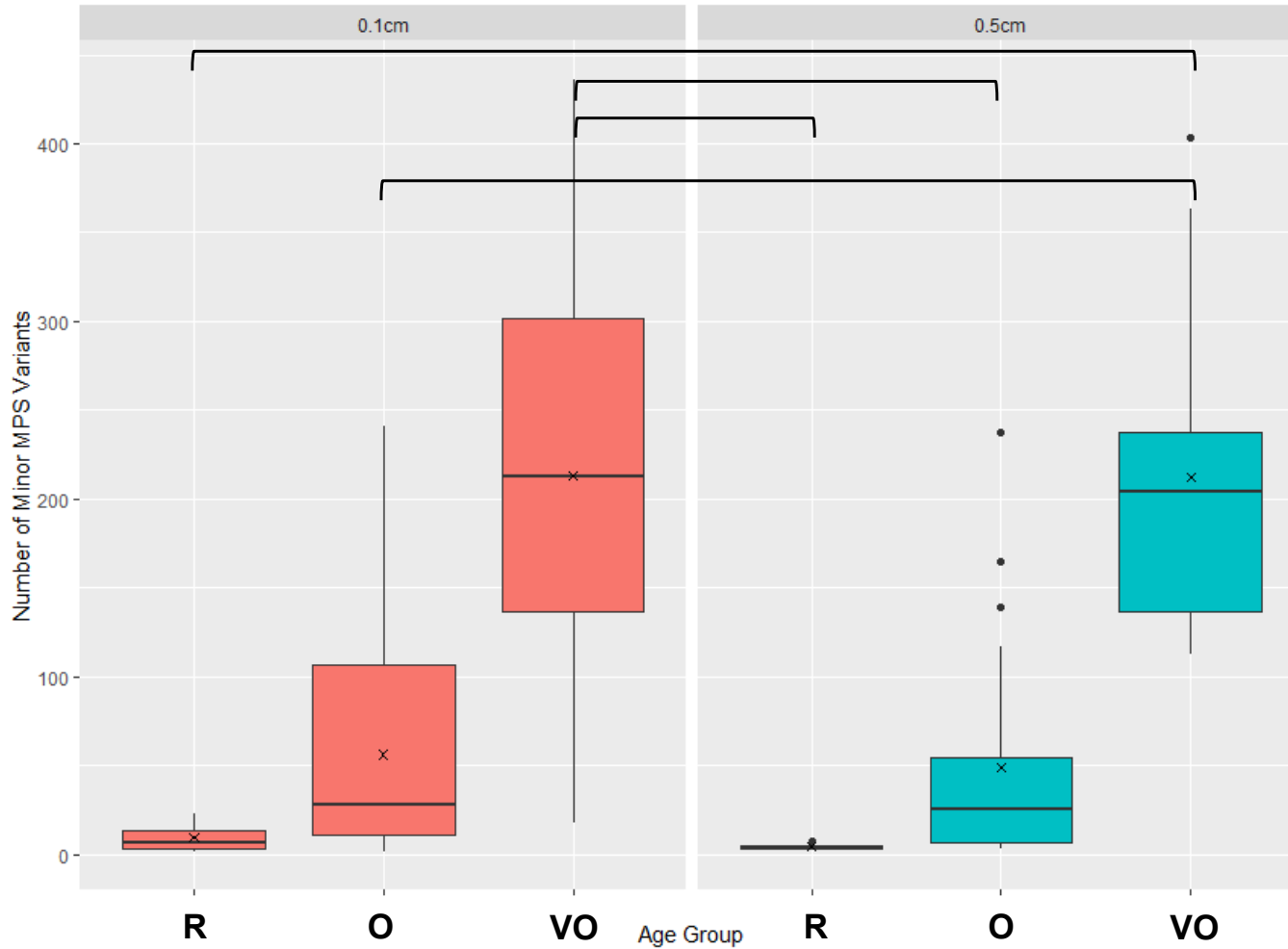
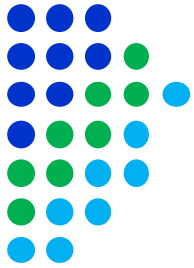


┌───┐ p-value < 0.05 (within)

Minor Variant = a base call with <50% of the reads and  $\geq$ 2% of the reads



# Number of Minor Variants v. Age

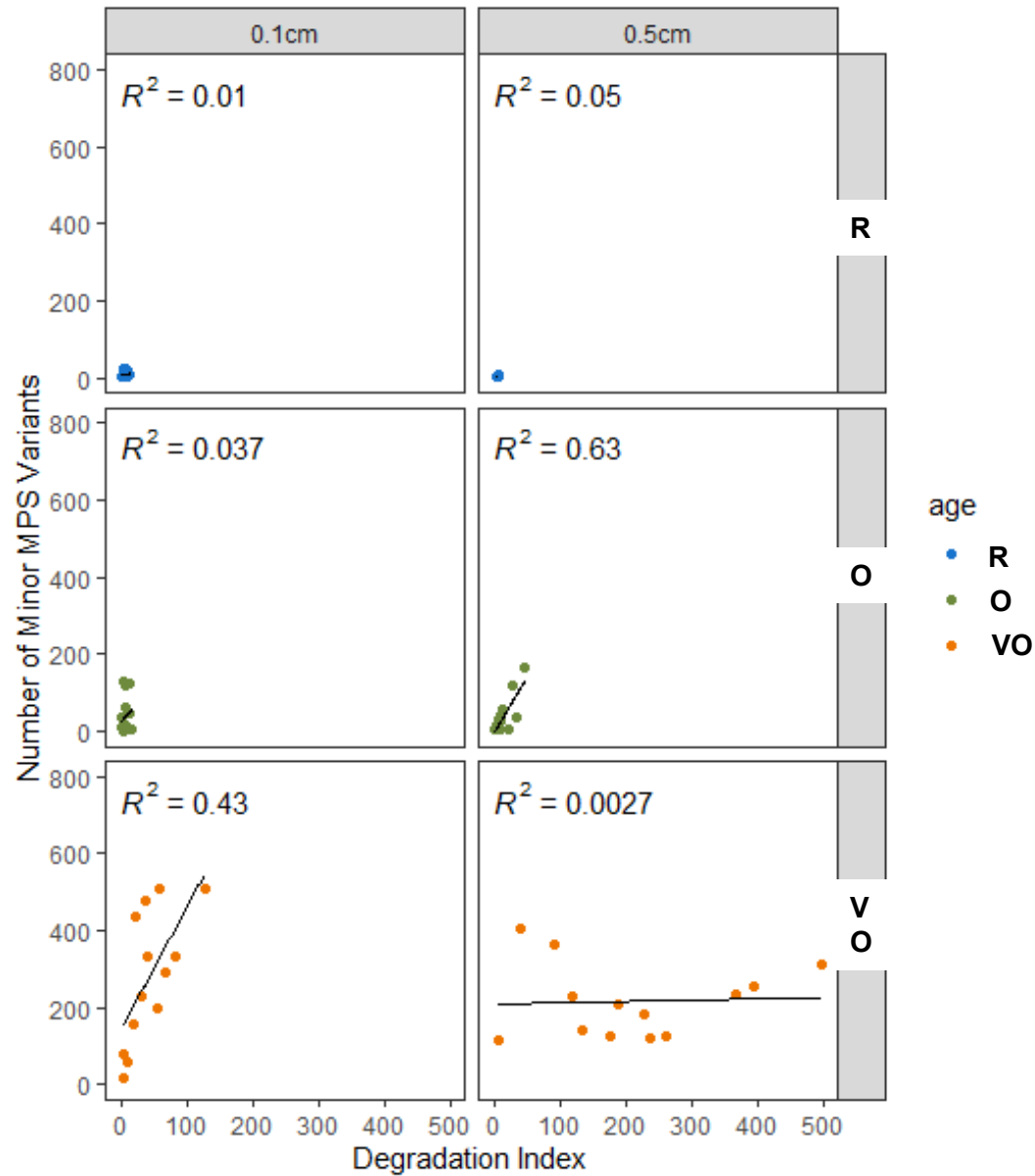
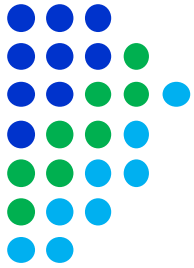


— p-value < 0.05 (between)



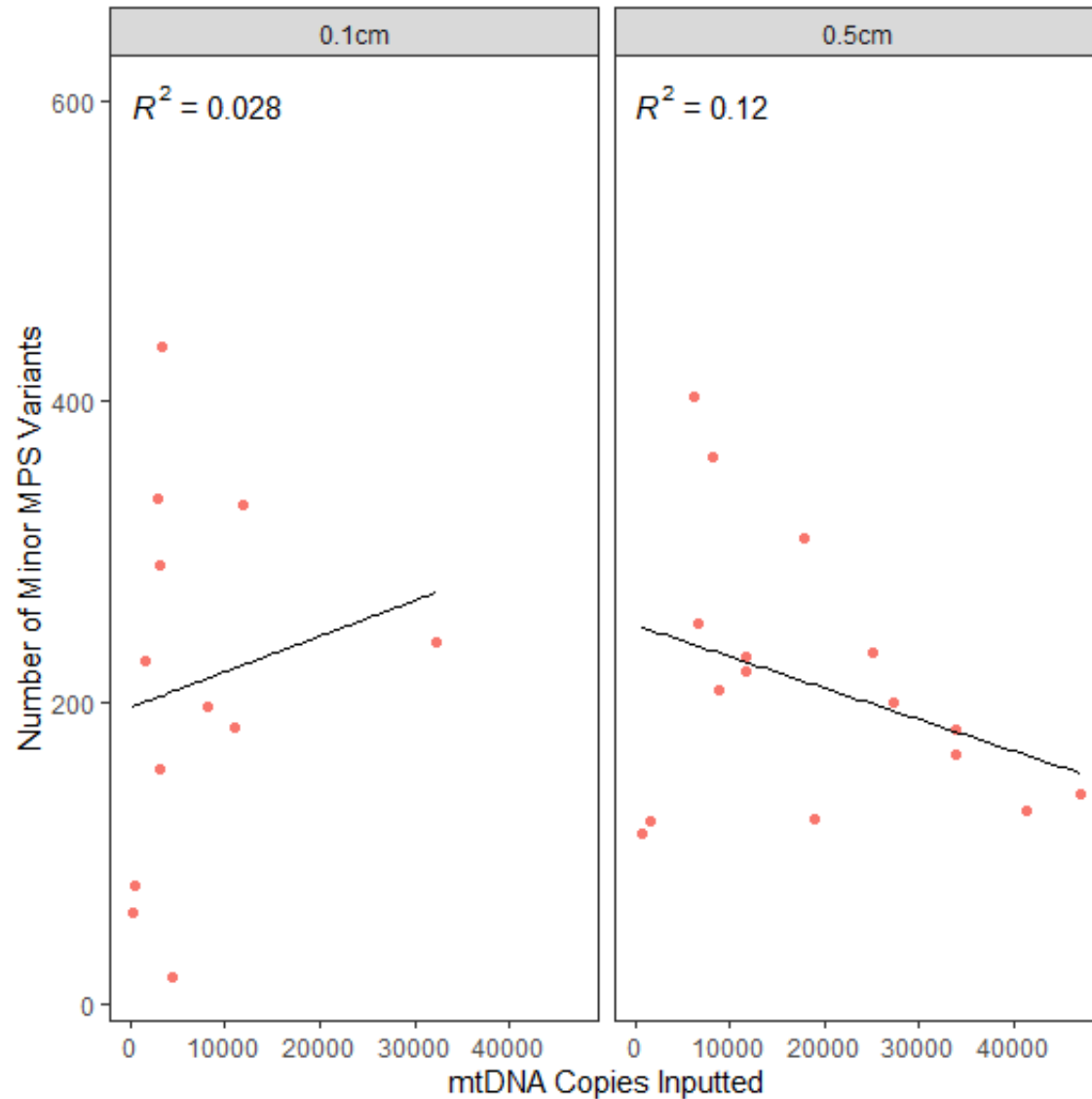


# Number of Minor Variants v. DI



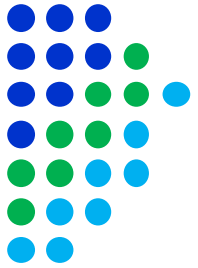


# Number of Minor Variants v. DNA Input

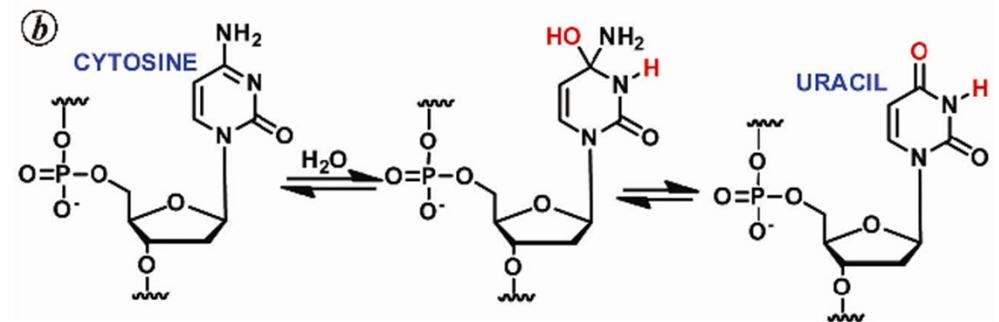




# DNA Damage

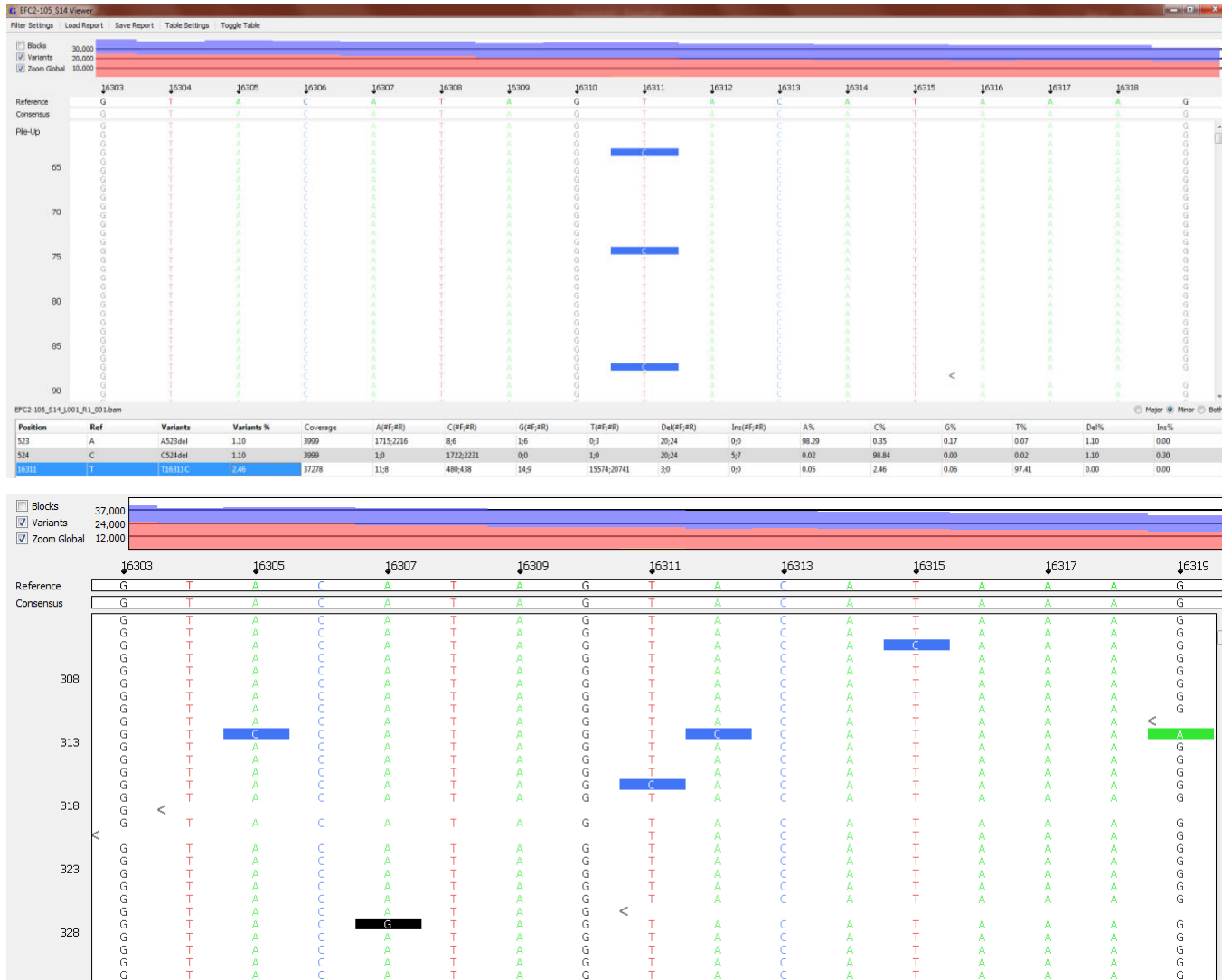
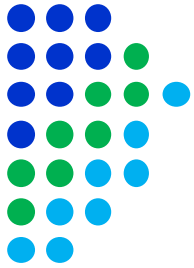


- Cytosine deamination has been identified in hair samples in the past
- Deamination appears as C to T or G to A transitions
- G to A: 20.1% of base changes
- C to T: 66.3% of base changes
- Occurred most often in older hairs



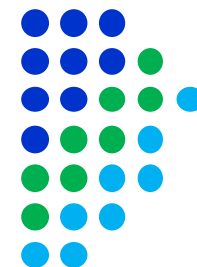


# “Noise” in the System





# “Noise” in the System



A.

	Metals-Power	Hair-Power	Blood/Buccal-Power	LowTemp-Power	LowTemp-Next	Buccal-Next
Total error	0.485±0.0049	0.325±0.1005	0.231±0.0663	0.297±0.0694	0.205±0.0021	0.182±0.0645
A error	0.118±0.0009	0.100±0.0417	0.052±0.0230	0.048±0.0035	0.070±0.0476	0.037±0.0121
C error	0.121±0.0006	0.085±0.0280	0.080±0.0191	0.069±0.0189	0.079±0.0297	0.060±0.0205
G error	0.131±0.0033	0.048±0.0227	0.044±0.0097	0.059±0.0013	0.061±0.0023	0.047±0.0204
T error	0.115±0.0038	0.091±0.0261	0.053±0.0145	0.049±0.0110	0.065±0.0243	0.039±0.0139

B.

	Buccal-Next CR	Buccal-Next mtgenome
Total error	0.158±0.0720	0.166±0.0745
A error	0.032±0.0148	0.036±0.0164
C error	0.057±0.0273	0.063±0.0292
G error	0.036±0.0180	0.036±0.0171
T error	0.032±0.0126	0.030±0.0124

Background noise or error ranged from 0.030% to 0.131% across the four nucleotides.

Samples with increasing DNA damage saw increases in error.

420+ million reads of data

Mitochondrion 52 (2020) 40–55

Contents lists available at ScienceDirect

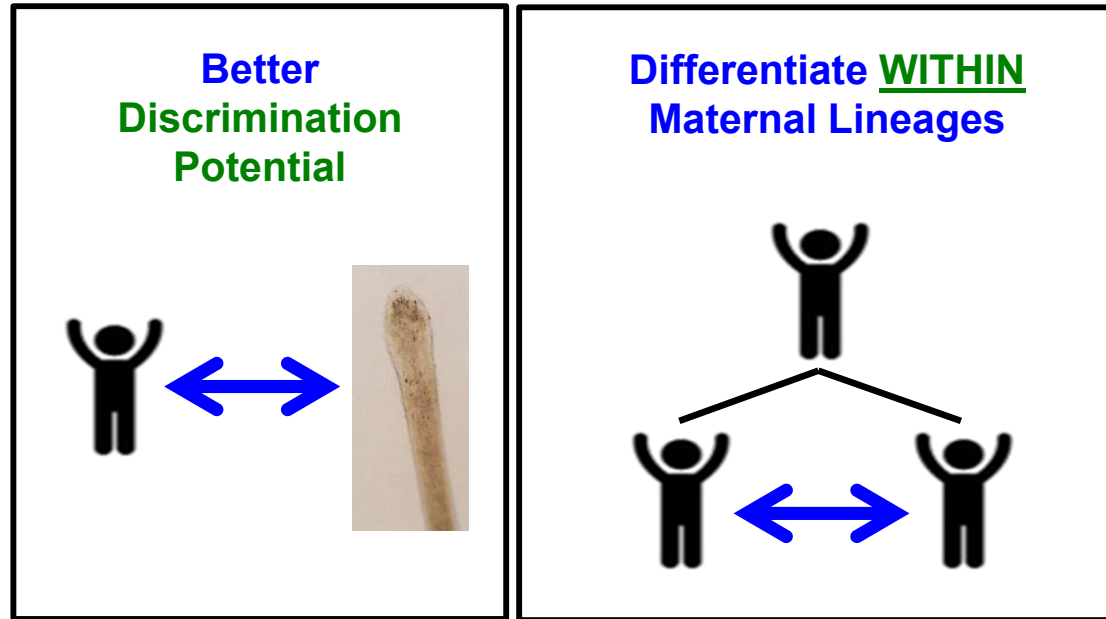
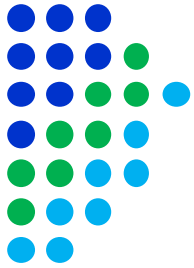
## Mitochondrion

journal homepage: [www.elsevier.com/locate/mito](http://www.elsevier.com/locate/mito)

Characterization of background noise in MiSeq MPS data when sequencing human mitochondrial DNA from various sample sources and library preparation methods

Jennifer A. McElhoe<sup>a</sup>, Mitchell M. Holland

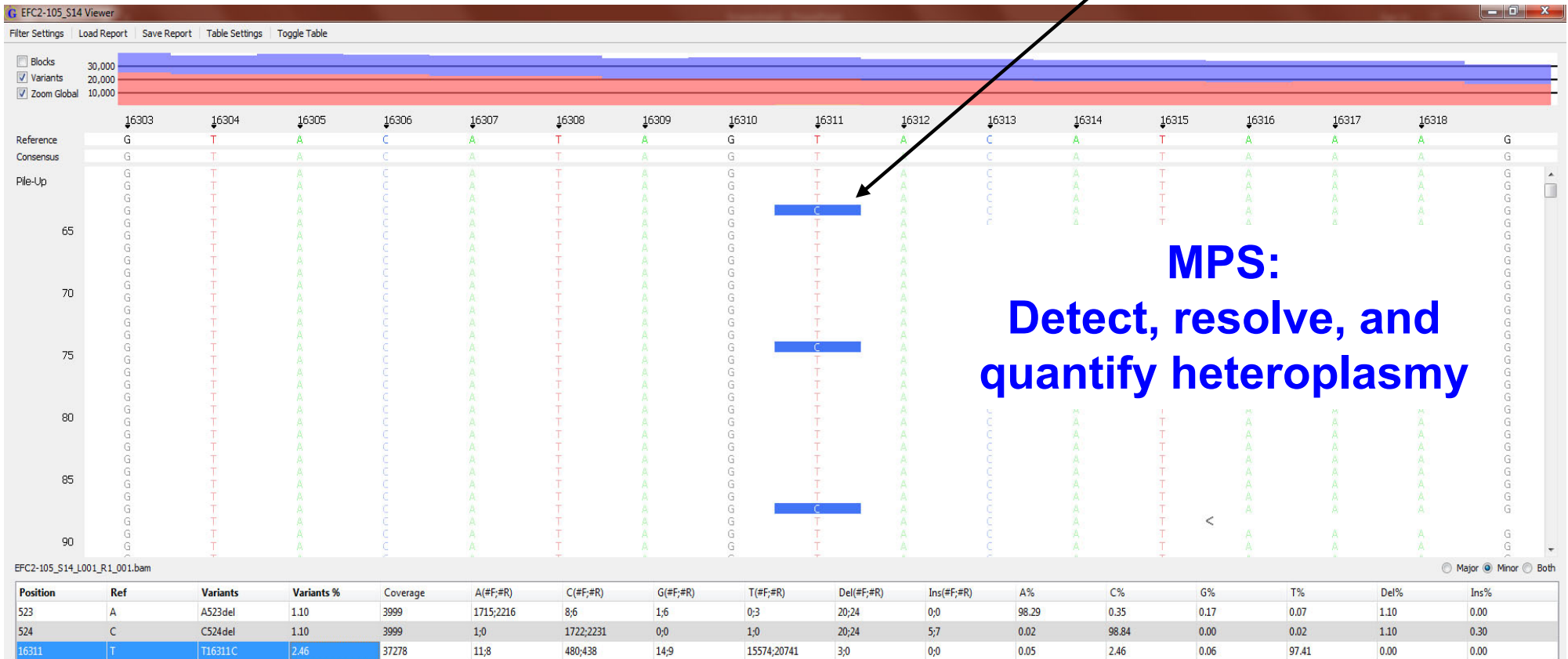
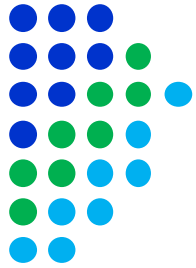
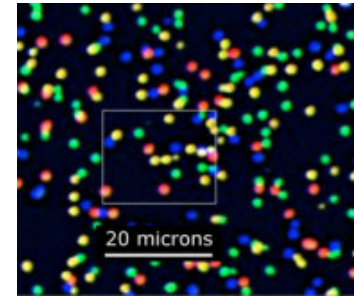
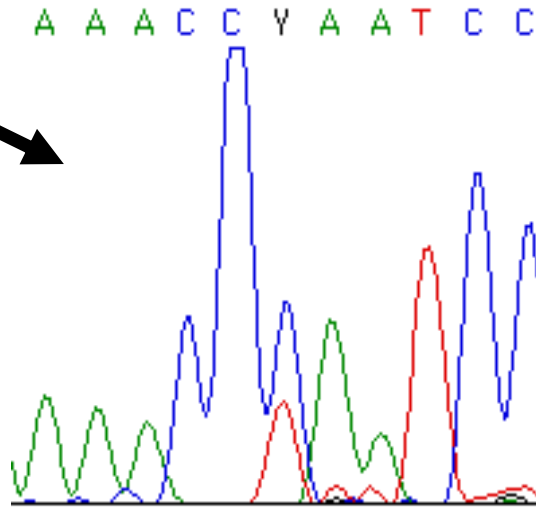
*Department of Biochemistry & Molecular Biology, Forensic Science Program, The Pennsylvania State University, University Park, PA 16802, USA*



**Deep Coverage MPS (DCMPS), i.e., deep read depth, allows for detection and resolution of heteroplasmy to ~2%**


# Sanger-Type Sequencing


**One Amplicon Target**  
**One Sample**  
**One Read/NP**





## 44% of Mother-Child Pairs were Differentiated





Article

### Deep-Coverage MPS Analysis of Heteroplasmic Variants within the mtGenome Allows for Frequent Differentiation of Maternal Relatives

Mitchell M. Holland <sup>1,\*</sup>, Kateryna D. Makova <sup>2</sup> and Jennifer A. McElhoe <sup>1</sup>

<sup>1</sup> Department of Biochemistry & Molecular Biology, Forensic Science Program, Eberly College of Science, Pennsylvania State University, University Park, PA 16802, USA; jam760@psu.edu

<sup>2</sup> Department of Biology, Eberly College of Science, Pennsylvania State University, University Park, PA 16802, USA; kmakova@bx.psu.edu

\* Correspondence: mmh20@psu.edu

Received: 1 January 2018; Accepted: 20 February 2018; Published: 26 February 2018



**Table 2. Differentiating heteroplasmy for 17 of the 39 mother-child pairs;** buccal cell (Bu) and blood (Bl) samples. Frequency of heteroplasmy at each np in percentage (%), with the minor variant annotated as the letter after the np; i.e., T2746C is heteroplasmy at np 2746, with 20.11% of the C variant. Coverage and read distribution (forward reads, #For, compared to reverse reads, #Rev) are provided for each np of differentiating heteroplasmy, along with the gene annotation and whether sites in protein coding genes result in a synonymous change (Y) or not (N). The gene annotations include: CR = control region, 12S & 16S = 12S & 16S rRNAs, ATP6 = ATP synthase 6, ND = NADH dehydrogenase, tRNA<sup>thr</sup> = tRNA for threonine, and CO = cytochrome oxidase. Metadata for samples without the heteroplasmy are provided to illustrate that read percentages are clearly below reporting threshold and that coverage was adequate for this assessment.

Mother-Child Pair	Nucleotide Position	Sample Number	Major Allele	Coverage (#For:#Rev Reads)	Major Frequency (%)	Minor Allele	Coverage (#For:#Rev Reads)	Minor Frequency (%)	Gene Annotation	Synonymous (Y or N)
T2746C		Mother - Bu (693)	T	2920:6014	79.67	C	655:1600	20.11	16S	
		Child - Bu (677)	T	4838:14038	99.64	C	1:9	0.053		
		Mother - Bl (M207)	T	14187:14328	80.3	C	3440:3528	19.62	16S	
		Child - Bl (M207-C)	T	24044:24176	99.88	C	6:12	0.037		
3	T9179C	Mother - Bu (1134)	T	3063:5076	85.02	C	538:892	14.93	ATP6	N (Val to Ala)
		Child - Bu (1099)	T	6651:8730	99.82	C	8:7	0.097		
		Mother - Bl (M502G)	T	16583:20269	87.14	C	2468:2934	12.77	ATP6	N (Val to Ala)
		Child - Bl (M501)	T	38769:44060	99.81	C	32:24	0.067		
4	G14040A	Mother - Bu (659)	G	5770:4227	92.01	A	474:381	7.86	ND5	Y (Gln)
		Child - Bu (722)	G	20789:16141	99.86	A	8:12	0.054		
		Mother - Bl (M242)	G	13200:12992	94.07	A	831:811	5.89	ND5	Y (Gln)
		Child - Bl (M242-C)	G	10355:10087	99.88	A	5:5	0.049		

... heteroplasmy must be observed in both tissues of one relative but not the other






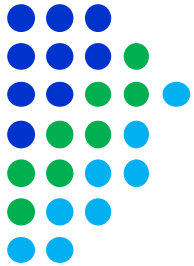
International Journal of Legal Medicine  
<https://doi.org/10.1007/s00414-020-02410-0>

ORIGINAL ARTICLE

## Damage patterns observed in mtDNA control region MPS data for a range of template concentrations and when using different amplification approaches

Charity A. Holland<sup>1</sup> · Jennifer A. McElhoe<sup>1</sup> · Sidney Gaston-Sanchez<sup>1,2</sup> · Mitchell M. Holland<sup>1</sup> 

Received: 30 June 2020 / Accepted: 21 August 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020



If we're going to consider low-level heteroplasmy (2-5% of the minor variant):

*What's the potential impact of DNA damage on the interpretation of MPS data when dealing with low-template samples?*

*Most importantly, what's the impact on reporting thresholds?*



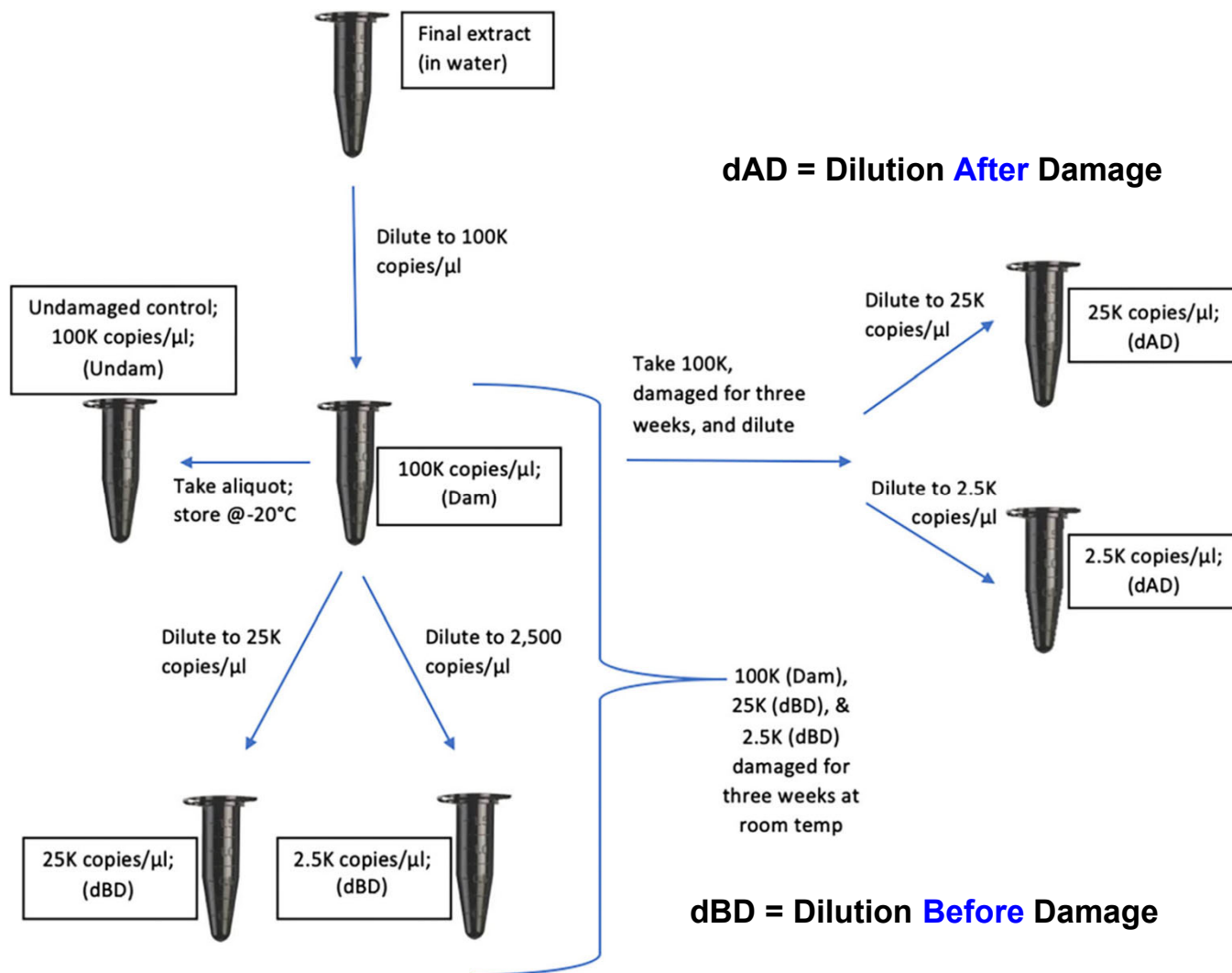
Buccal cells collected from donors, DNA extracted, and mtDNA quantified using the custom mtqPCR assay (Gallimore et al., 2018)



### Experimental Design:

Control sample at 100k copies/uL of mtDNA stored at -20C

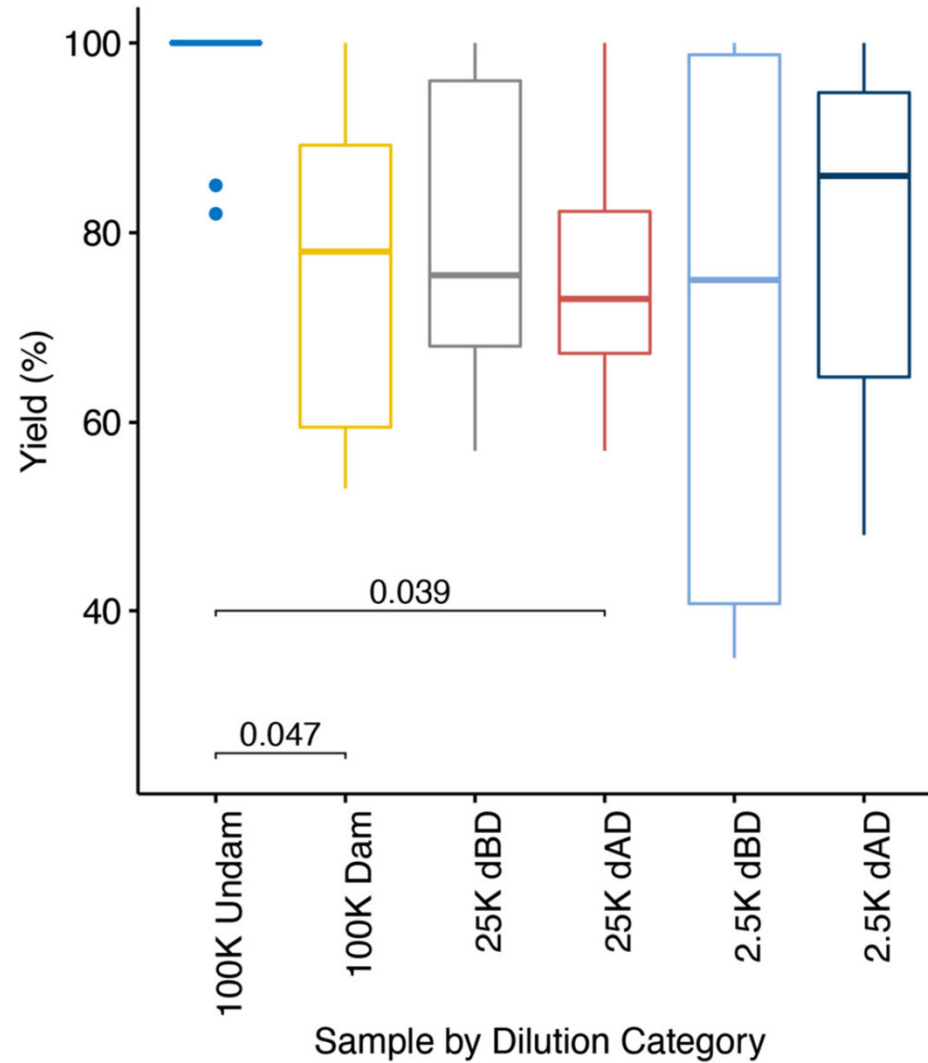
Dilutions made to 25k and 2.5k copies/uL before or after damage for three weeks at RT



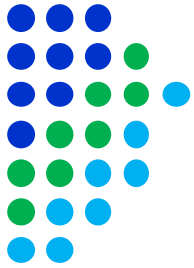


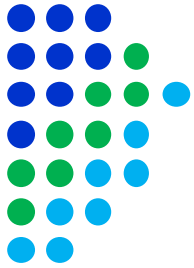
Control sample yield is high, reflecting that storage at -20C is effective, even when stored in water

Storage in water at RT results in lower yields across the data set



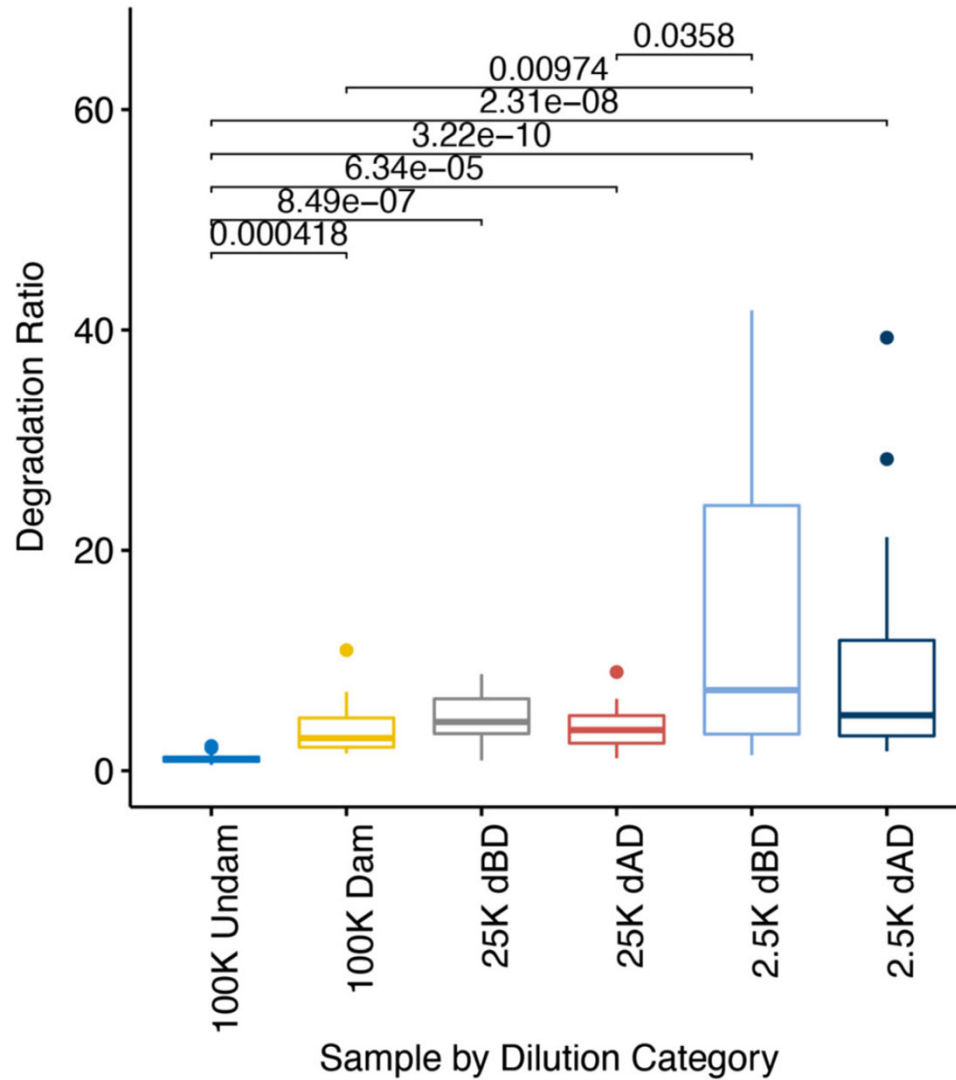
n = 20 per category





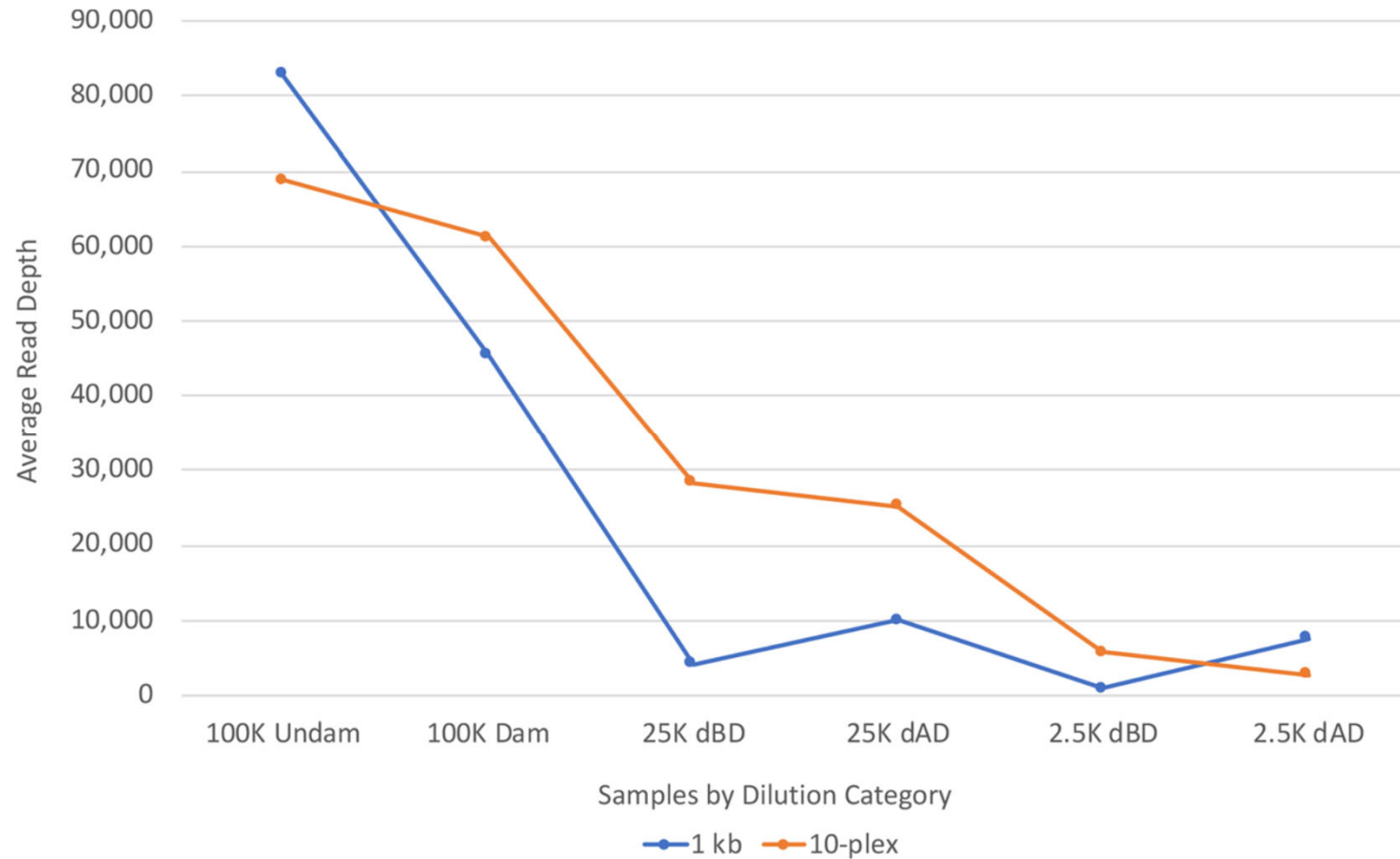
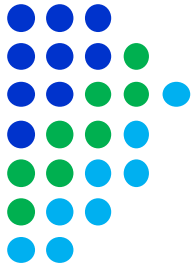
**Control sample degradation is low, again, reflecting that storage at -20C is effective**

**Dilutions before damage (dBD) result in higher degradation for both the 2.5k & 25k samples**





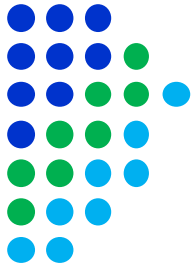
# MPS Results



Comparison of **normalized average** read depths between different DNA damage dilution categories and two amplification and library prep strategies; **1 kb v. PowerSeq<sup>®</sup> CRM (10-plex)**; n = 20 per category.

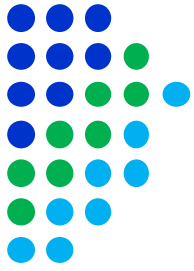
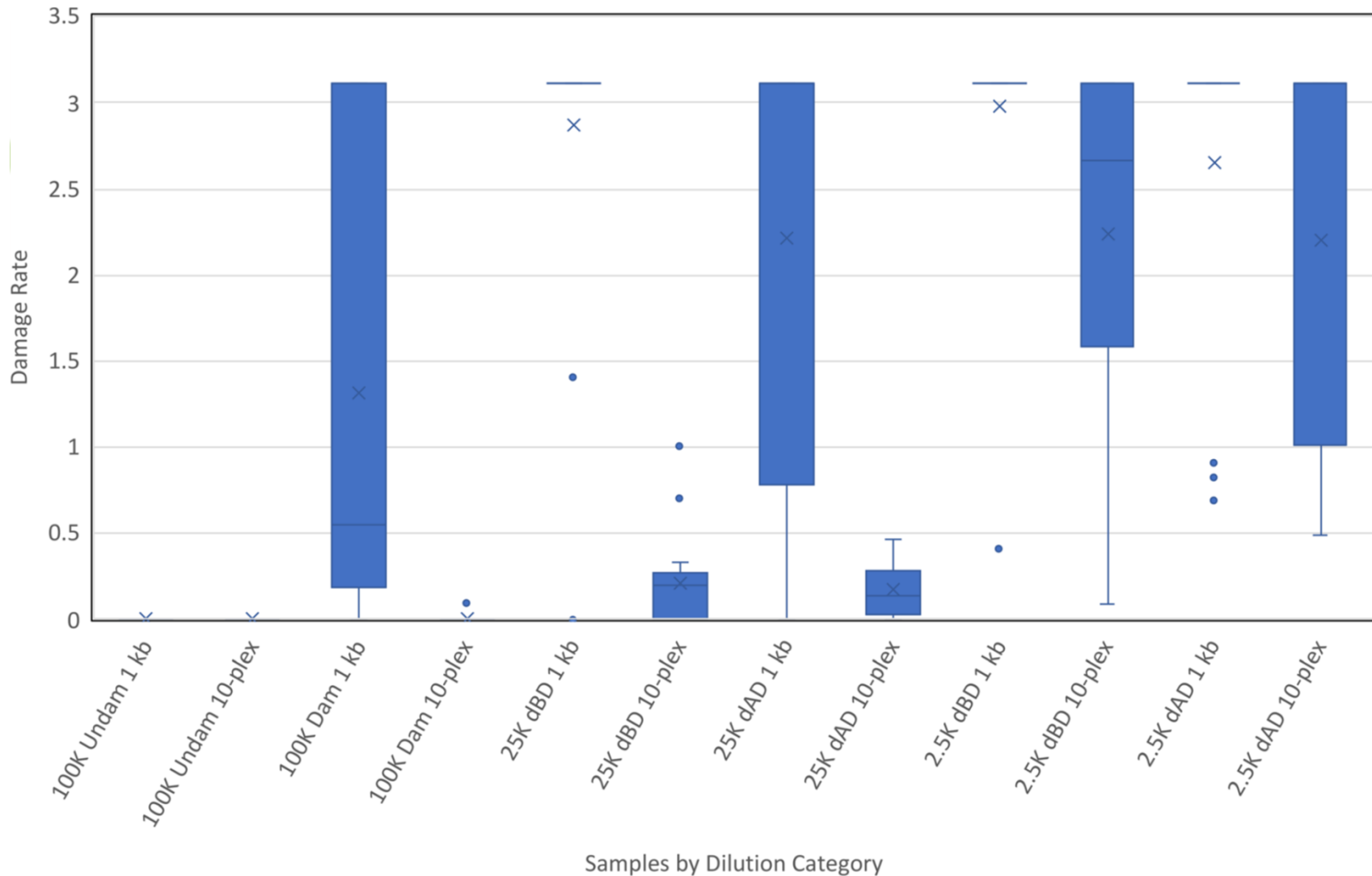


# MPS Results



	Full Profiles		Partial Profiles		No Results	
	1 kb	10-plex	1 kb	10-plex	1 kb	10-plex
100K Undam	100%	100%	0%	0%	0%	0%
100K Dam	70%	100%	0%	0%	30%	0%
25K dBD	10%	90%	10%	10%	80%	0%
25K dAD	45%	100%	0%	0%	55%	0%
2.5K dBD	10%	50%	0%	10%	90%	40%
2.5K dAD	20%	80%	0%	20%	80%	0%

Table 1: MPS results depicted as percentages of full profiles, partial profiles, and no results across each dilution category and comparing the two amplification and library prep strategies (**1 kb v. 10-plex**), with n = 20 per category.

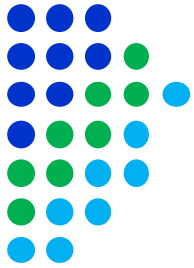


Box and whisker plot depicting damage rates (**number of damage sites/number of total sites reported × 100**) calculated across each dilution category using a 4000 read cutoff and a comparison of the 1 kb and 10-plex amplification and library prep strategies (n = 20 per category).

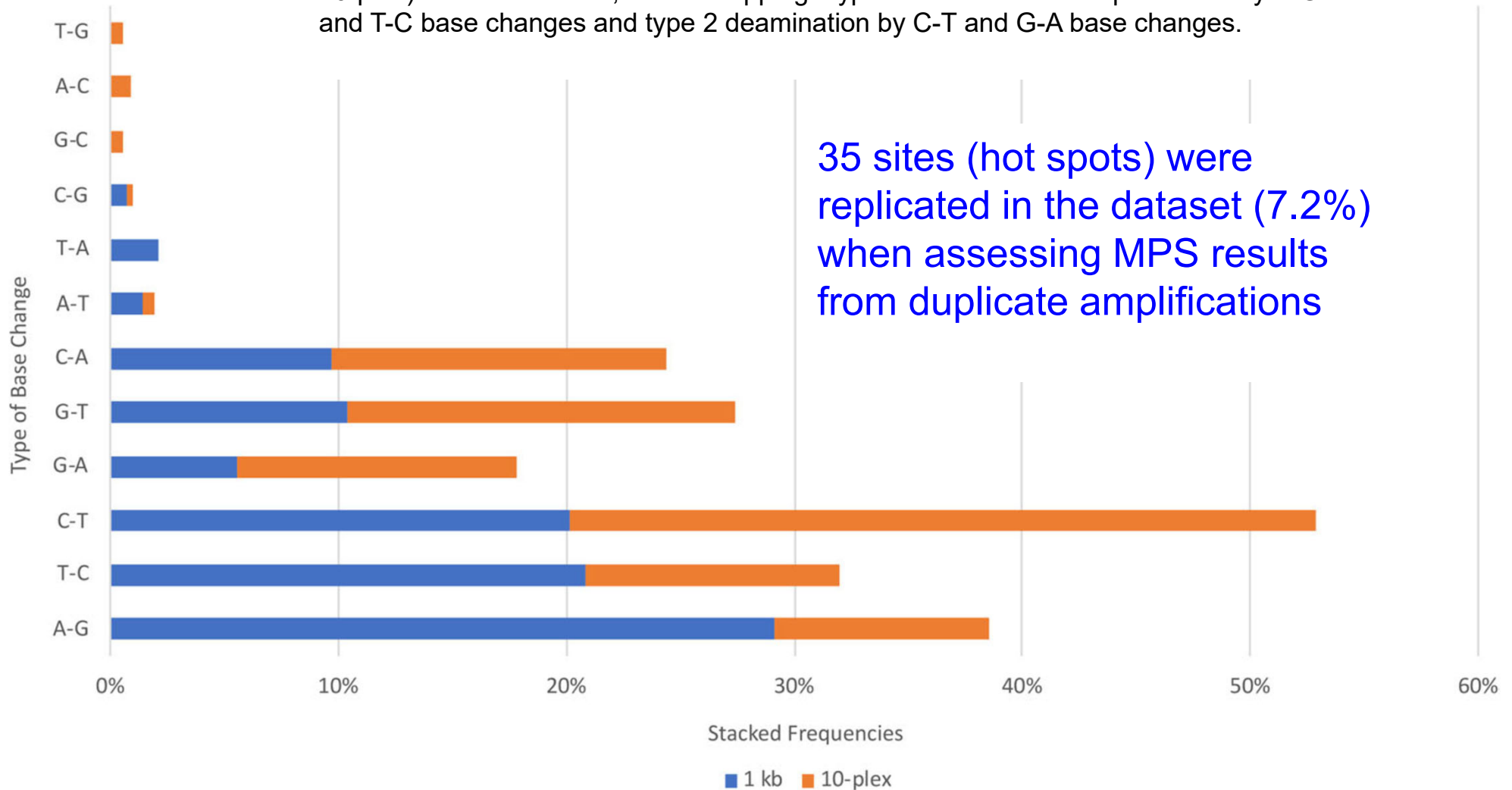
# Damage Rates



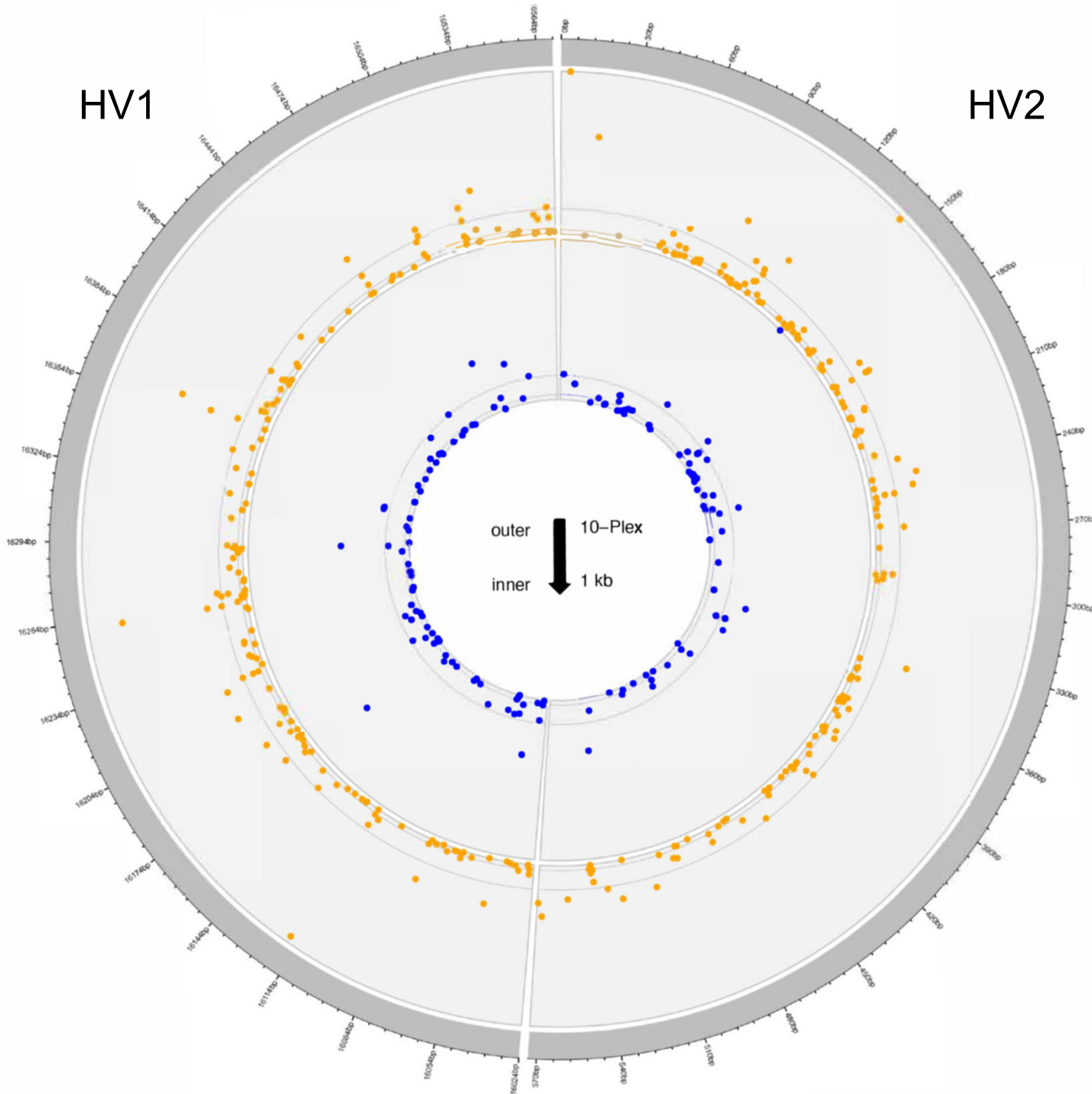
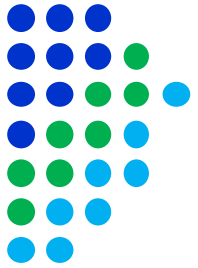
# Type of Base Change



Frequencies of each type of base change observed in **486 damage lesions** across all samples and comparison of two amplification and library prep strategies (1 kb v. 10-plex). Data is stacked, not overlapping. Type 1 deamination is represented by A-G and T-C base changes and type 2 deamination by C-T and G-A base changes.





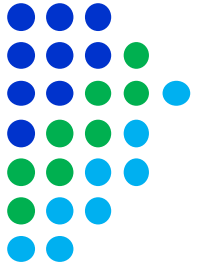


## Impact of Damage on Reporting Thresholds

Individual MVFs of 486 damage lesions observed across the control region for all samples (144 in the 1 kb samples and 342 in the 10-plex samples). **Proposed analytical (2%) and reporting (5%) thresholds** are marked by the first and second circles moving outward, respectively, for each dataset.



# Take Home Messages



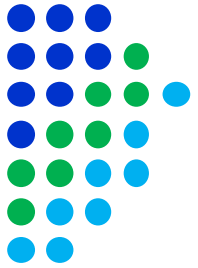
DNA damage impacts the quantity and quality of mitoMPS data when working with low-template samples.

DNA damage increases as template levels decrease, especially when the damage occurs after dilution.

Therefore, it's important to protect against further damage when working with forensic samples containing low amounts of extracted DNA.



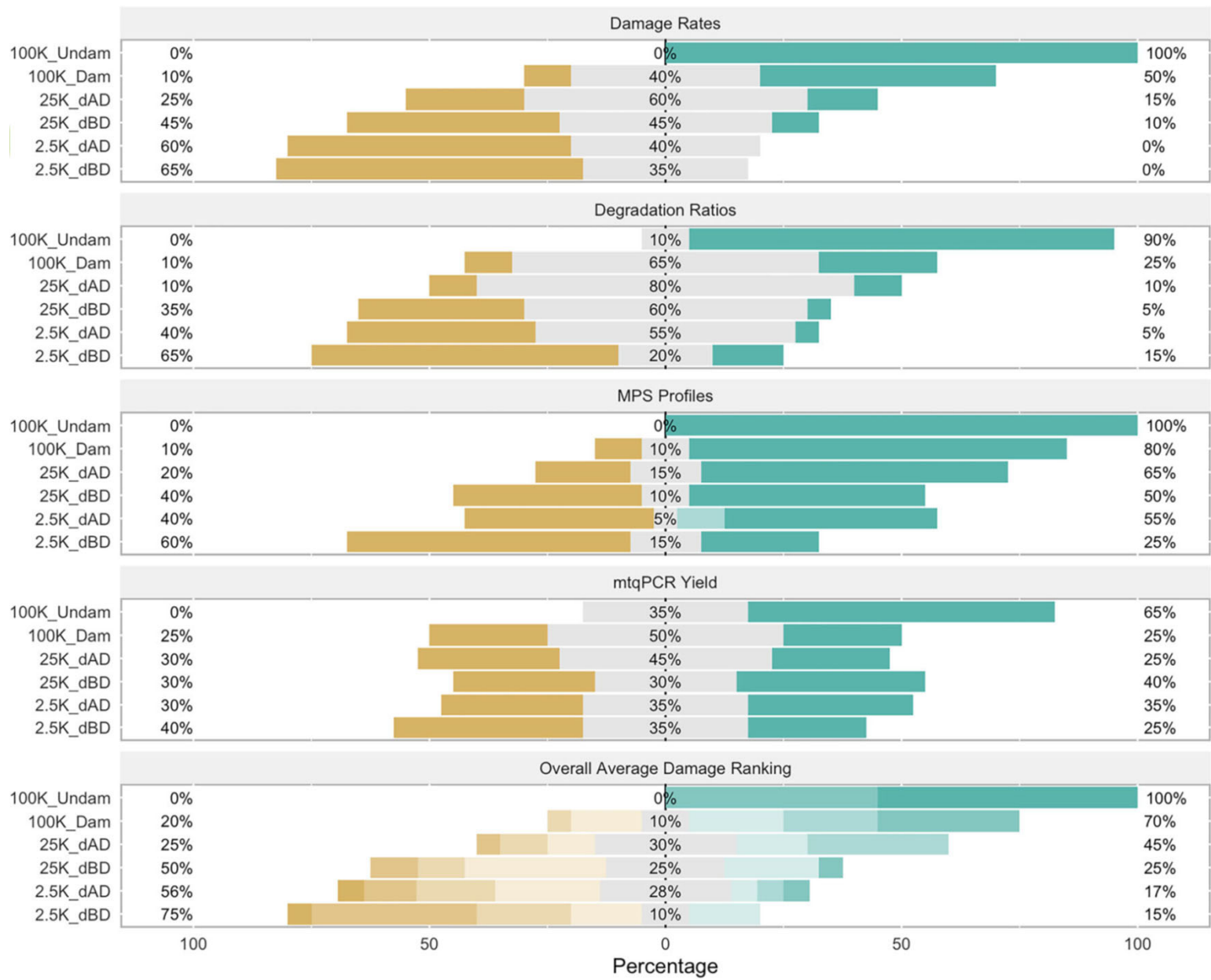
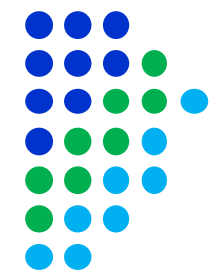
# Take Home Messages



Duplicate amplifications will mitigate the impact of the damage on interpretation of low-level heteroplasmy.

Reporting thresholds may be impacted by damage associated with low-level template samples.

The Promega 10-Plex helped to mitigate the impact of the damage, most likely due to the size of the amplicons being targeted.

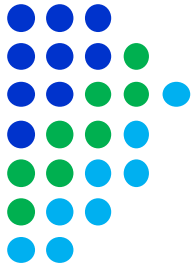


Overall damage assessment using a damage coefficient scale of 1–5 (1 = highest damage; 5 = little to no damage).

The last plot shows the overall damage assessment when averaging the damage coefficients for all four categories.



Thanks!!



**Promega**

**Doug Storts, Spencer Hermanson, Jeff Shaw, Margaret Ewing**



**National Institute of Justice (NIJ 2019-DU-BX-0045)  
National Institute of Justice (NIJ 2016-DN-BX-0171)  
National Institute of Justice (NIJ 2015-DN-BX-K025)**



**Lauren Canale, Charity Holland, Jennifer McElhoe, Shelby Bain,  
Kate DeHeer**



**Eberly College of Science, Department of Biochemistry & Molecular  
Biology, Forensic Science Program at Penn State**



**Daisy says hello!**

[mmh20@psu.edu](mailto:mmh20@psu.edu)

